გვანჯი მანია

ალბათობის თეორია და მათემატიკური სტატისტიკა

დამტკიცებულია სახელმძღვანელოდ
საქართველოს სსრ უმაღლესი და საშუალო
სპეციალური განათლების სამინისტროს მიერ
უნივერსიტეტის ეკონომიური ფაკულტეტების
სტუდენტთათვის

თბილისის უნივერსიტეტის გამომცემლობა
თბილისი 1976

517 . 8+517 . 8
530 . 16+519 . 2
მ 272

სახელმძღვანელო შედგენილია მოქმედი პროგრამის მიხედვით. იგი შედგება 2 ნაწილისაგან. პირველ ნაწილში განხილულია ალბათობის თეორიის, ხოლო მეორე ნაწილში კი—მათემატიკური სტატისტიკის ელემენტები.

წიგნი განკუთვნილია უნივერსიტეტის ეკონომიური ფაკულტეტების სტუდენტთათვის.

M $\frac{20203}{M—608(08)—76}$ 2—76

ავტორისაგან

წინამდებარე სახელმძღვანელო შედგენილია უნივერსიტეტის ეკონომიური ფაკულტეტების სტუდენტებისათვის. სახელმძღვანელო გამოადგებათ უნივერსიტეტის სხვა ფაკულტეტების (სახელდობრ, ქიმიის, ბიოლოგიის, გეოგრაფია-გეოლოგიის და სხვ.) სტუდენტებსაც, სადაც ისწავლება უმაღლესი მათემატიკა.

წიგნი სამსახურს გაუწევს საღამოს დასწრებული და დაუსწრებელი სწავლების სტუდენტებსაც.

სახელმძღვანელო ორი ნაწილისაგან შედგება. პირველ ნაწილში გადმოცემულია ალბათობის თეორიის ელემენტები, ხოლო მეორე ნაწილში კი—მათემატიკური სტატისტიკის ძირითადი მეთოდები. ყოველ თავს ახლავს სათანადო სავარჯიშო, რომელთა ამოხსნა სტუდენტებს ხელს შეუწყობს გადმოცემული მასალების შეგნებულად ათვისებაში.

სახელმძღვანელოს ბოლოში დართული აქვს ცხრილები, რომელთა დახმარებით სტუდენტებს შეეძლებათ რიგი პრაქტიკული ამოცანების ამოხსნა.

ეკონომიური ფაკულტეტების სტუდენტებისათვის ქართული სახელმძღვანელო ალბათობის თეორიასა და მათემატიკურ სტატისტიკაში დღემდე არ არსებობდა, თუ მხედველო-

ბაში არ მივიღებთ მათთვის შედგენილ უმაღლესი მათემატიკის კურსს (პ. ზერაგიასა და გ. მანიასი, 1967).

სახელმძღვანელოსთან დაკავშირებული ყოველგვარი საქმიანი შენიშვნა მადლიერების გრძნობით იქნება მიღებული.

შესავალი

ყოველგვარი მოვლენა სამყაროში მრავალ ფაქტორზეა დამოკიდებული, რომელთაგან ზოგიერთი მთავარია, ხოლო ზოგიერთი—მეორეხარისხოვანი. მთავარი ფაქტორები განსაზღვრავენ ამა თუ იმ პროცესის მიმდინარეობას ბუნებაში, ხოლო მეორეხარისხოვანი ფაქტორები იწვევენ უმნიშვნელო გადახრებს.

ხშირად პრაქტიკული ამოცანების ამოხსნისას მეორეხარისხოვანი ფაქტორების უგულებელყოფა სავსებით შესაძლებელია და მოვლენებს სწავლობენ მხოლოდ მთავარი, ძირითადი ფაქტორების საშუალებით. ასეთ შემთხვევებში ხდება ე. წ. დინამიური კანონზომიერების დადგენა, რომელიც პროცესის აღწერის შესაძლებლობას იძლევა მისი რომელიმე ერთ-ერთი რეალიზაციის დროს და უჩვენებს, თუ როგორი იქნება პროცესი გარკვეული პირობების შესრულებისას. ამგვარად, თუ ცნობილია რაიმე მოვლენის გამომწვევი მიზეზები და მისი შესაბამისი კანონი, უშეცდომოდ შეიძლება მისი წინასწარმეტყველება.

დინამიური კანონზომიერების დასადგენად ატარებენ ცდებს და აკვირდებიან ცალკეულ მოვლენებს, ხოლო შემდეგ ხდება მიღებული შედეგების განზოგადება. სამწუხაროდ ასეთი მეთოდის გამოყენება ყოველთვის როდია შესაძლებელი. შესასწავლ მოვლენაზე ხშირად მრავალი ფაქტორი მოქმედებს, რომელთაგან ძირითადის გამოყოფა ყოველთვის არ შეიძლება, ხოლო ყველა მათგანის ერთბაშად შესწავლა კი შეუძლებელია. სწორედ ასეთ შემთხვევაში ხდება ე. წ. სტატისტიკური მეთოდის გამოყენება. ამ მეთოდით სწავლობენ არა ცალკეულ მოვლენებს, არამედ მასობრივ მოვლენებს მთლიანად. ასეთ შემთხვევაში მეორეხარისხოვანი ფაქტორების მოქმედება შემთხვევით ხასიათს ატარებს და მათი მოქმედება საბოლოო ჯამში ერთობლიობისათვის სტატისტიკურ კანონზომიერებებს გვაძლევს. ასეთი კანონზომიერებანი არ იძლევიან ცალკეული მოვლენების წინასწარმეტყველების შესაძლებლობას, მაგრამ გვეხმარებიან მთელი ერთობლიობის ხასიათის საკმაოდ

ზუსტად დადგენაში, რის შემდეგაც ხდება ესა თუ ის მეცნიერული წინასწარმეტყველება.

მაგალითად, გაზების კინეტიკური თეორიიდან ცნობილია, რომ ჭურჭელში მოთავსებული გაზის ყოველი მოლეკულის მოძრაობის აღწერა შეუძლებელია, მაგრამ შესაძლებელია მოლეკულათა ჯგუფისათვის ისეთი კანონზომიერების დადგენა, რომლებიც განსაზღვრავენ გაზის წნევას ჭურჭლის კედლებზე, ამ წნევის დამოკიდებულებას ტემპერატურაზე და სხვ.

მასობრივი მოვლენების შესწავლას დიდი მნიშვნელობა აქვს ეკონომიკაში. ცნობილია, რომ დიდი სტატისტიკური მასალის გაანალიზების საფუძველზე კ. მარქსი მივიდა იმ დასკვნამდე, რომ კაპიტალიზმის დროს დაბადებისა და სიკვდილიანობის რაოდენობა ხელფასის დონის უკუპროპორციულია.

როდესაც ხდება სახალხო მეურნეობის გეგმის შესრულების გაანალიზება წარმოების ტემპის გადიდების ან შრომის ნაყოფიერების ზრდის, ან თვითღირებულების შემცირების მიხედვით, და ა. შ. ცხადია, მხედველობაში ღებულობენ არა ცალკეული მუშის ან დაწესებულების მაჩვენებლებს, არამედ ყველას ერთად, მასობრივ ერთობლიობაში.

თუ რომელიმე წარმოება გეგმას ასრულებს მთლიანად 108%-ით, ეს იმას ნიშნავს, რომ ზოგიერთ სახეობაში გეგმა სრულდება 108%-ზე მეტად და ზოგიერთში კი ნაკლებად. 108% გვიჩვენებს საშუალო მაჩვენებელს მთელი წარმოებისათვის.

მაგალითისათვის განვიხილოთ მყიდველთა მოთხოვნების დასაკმაყოფილებლად ფეხსაცმლის ასორტიმენტის განსაზღვრის საკითხი.

ეს საკითხი რომ კარგად გადაწყდეს, ამისათვის საჭიროა ვიცოდეთ ამა თუ იმ ნომრის ფეხსაცმელზე მოთხოვნილება. ვთქვათ, მამაკაცთა ფეხსაცმელზე მოთხოვნილება შემდეგნაირად არის განაწილებული

ცხრ. 1

ფეხსაცმლის №	38	39	40	41	42	43	44	45	ან მეტი
წყვილთა რაოდენობა პროცენტებში	8	9	18	24	22	13	5	1	

როგორც ცხრილიდან ჩანს, მოთხოვნილების 24% მოდის № 41 ფეხსაცმელზე. ეს იმას ნიშნავს, რომ ყოველი 1000 წყვილიდან 240 მოთხოვნილება მოდის № 41-ზე. თუ გვინდა დამზადებული პროდუქცია მოსახლეობის მოთხოვნილებას აკმაყოფილებდეს, მაშინ მისი გამოშვება უნდა მოხდეს სწორედ მოთხოვნილების მიხედვით.

ზემოთ მოყვანილი ცხრილი წარმოადგენს გარკვეულ სტატისტიკურ კანონზომიერებას. მრავალი დაკვირვება გვიჩვენებს, რომ ერთგვაროვანი მოვლენების მასობრივი ერთობლიობა სათანადოდ მდგრადია და სწორედ ეს უკანასკნელი წარმოადგენს სტატისტიკური მეთოდის საფუძველს.

მათემატიკური სტატისტიკის საგანს წარმოადგენს ერთგვაროვანი მოვლენების მასობრივ ერთობლიობათა ობიექტურად არსებული კანონზომიერებების შესწავლა.

შესაძლებელია ავაგოთ რეალური მასობრივი ერთობლიობის ამსახველი აბსტრაქტული მოდელი და შევისწავლოთ მისი კანონზომიერებანი. შედგენილი აბსტრაქტული მოდელის ძირითად ელემენტს წარმოადგენს შემთხვევითი ხდომილობა, რომელიც რაიმე მოვლენის მოხდენის ფაქტს წარმოადგენს და შესაძლებლობის სახითაა მოცემული. ყოველ ხდომილობას შესაძლებელია შევუსაბამოთ გარკვეული დადებითი რიცხვი, რომელსაც ამ ხდომილობის ალბათობა ეწოდება და წარმოადგენს მათემატიკურ აბსტრაქციას ჩვენთვის სასურველ ხდომილობათა მოხდენის რიცხვის ფარდობისა ყველა ჩატარებული ცდის საერთო რიცხვთან. კანონზომიერებანი, რომელთაც ადგილი აქვთ ასეთი მოდელების შემთხვევაში, წარმოადგენენ რეალური სტატისტიკური კანონზომიერების მათემატიკურ აბსტრაქციას და იწოდებიან ალბათურ ანუ სტოქასტურ კანონზომიერებად. სწორედ ალბათური კანონზომიერების შესწავლა მასობრივი შემთხვევითი ხდომილობებისათვის წარმოადგენს ალბათობის თეორიის საგანს.

ალბათობის თეორიის დებულებები ასახავენ რეალურ სტატისტიკურ კანონზომიერებებს, ამიტომ მათი დადასტურება ცდებითა და დაკვირვებებითაც შეიძლება. ამისათვის საჭიროა ექსპერიმენტული მონაცემების დამუშავება, მიღებული შედეგების მიხედვით აბსტრაქტული ალბათური მეთოდების აგება და ექსპერიმენტებით მიღებული შედეგების თეორიულ დასკვნებთან შეთანადება. სწორედ ამ საკითხების შესწავლა წარმოადგენს მათემატიკური სტატისტიკის ძირითად მიზანს.

ნაწილი პირველი

ალბათობის თეორია

I თავი

ალბათობის განსაზღვრა და მისგან გამომდინარე ზოგიერთი დებულება

§ 1. ძირითადი განმარტებანი

ალბათობის თეორია სწავლობს მასობრივ შემთხვევით მოვლენათა რაოდენობრივი ხასიათის კანონზომიერებებს.

ვთქვათ, ვაკვირდებით რაიმე მოვლენას. ცდის ჩატარებისას ამ მოვლენას შეიძლება ექნეს ან არ ექნეს ადგილი. რაიმე მოვლენის მოხდენის ფაქტს *ხდომილობა* ეწოდება. ის რაიმე რიცხვით სიდიდეს არ წარმოადგენს.

თუ ხდომილობა ისეთი ბუნებისაა, რომ ცდების ან დაკვირვებების განმეორებისას მას არასოდეს არ ექნება ადგილი, ასეთ ხდომილობას *შეუძლებელს* უწოდებენ. მაგალითად, თუ ყუთში, რომელშიც მხოლოდ წითელი ფერით შეღებილი ბურთებია მოთავსებული, გვინდა შემთხვევით ამოღებული ბურთი იყოს თეთრი ფერის, ცხადია ამ შემთხვევაში ცდა რამდენჯერაც არ უნდა გავიმეოროთ, ამოღებული ბურთი თეთრი ფერის არ იქნება და ამგვარად, განხილული ყუთიდან თეთრი ფერის ბურთის ამოღების შესაბამისი ხდომილობა *შეუძლებელი* იქნება.

თუ ხდომილობა ისეთია, რომ ყოველი ცდის ან დაკვირვების დროს მას ექნება ადგილი, მაშინ ასეთ ხდომილობას *აუცილებელს* უწოდებენ. მაგალითად, იმავე ყუთიდან, რომელშიც მხოლოდ წითელი ფერის ბურთებია მოთავსებული, თუ შემთხვევით ვიღებთ რომელიმე ბურთს, ცხადია, ის წითელი ფერის იქნება. ამგვარად, ასეთი

ყუთიდან წითელი ფერის ბურთის ამოღების შესაბამისი ხდომილობა აუცილებელი იქნება.

თუ ხდომილობა ისეთია, რომ ამა თუ იმ ცდის ან დაკვირვების დროს მას შეიძლება ექნეს ან არ ექნეს ადგილი, მაშინ მას შემთხვევითი ხდომილობა ეწოდება. მაგალითად, თუ ყუთში სხვადასხვა ფერის ბურთია მოთავსებული, მაშინ ერთი რომელიმე ფერის ბურთის ამოღების შესაბამისი ხდომილობა იქნება შემთხვევითი.

ხდომილობები საზოგადოდ აღინიშნება დიდი ასოებით: $A, B, C, \dots$

რამდენიმე ხდომილობიდან რაიმე ცდის ან დაკვირვების შედეგად თუ მხოლოდ ერთ-ერთს ექნება ადგილი, მაშინ ვიტყვით, რომ ასეთი ხდომილობანი არიან მხოლოდ შესაძლებელი.

თუ რამდენიმე ხდომილობიდან რაიმე ცდის ან დაკვირვების დროს ნებისმიერი ორი მათგანის ერთდროულად მოხდენა შეუძლებელია, მაშინ მათ უთავსებად ხდომილობებს უწოდებენ. მაგალითად, ყუთიდან, რომელშიც სხვადასხვა ფერის ბურთია მოთავსებული ორი დასახელებული ფერის ბურთის ამოღების შესაბამისი ხდომილობანი ერთი ცდის შემთხვევაში იქნებიან უთავსებადნი.

თუ ხდომილობათა ერთდროულად მოხდენა შესაძლებელია, მათ თავსებადი ეწოდება. მაგალითად, ვთქვათ ყუთში მოთავსებული ბურთები დანომრილია. ერთი ხდომილობა იყოს კენტნომრიანი ბურთის ამოღების აღმნიშვნელი, ხოლო მეორე—ამოღებული ბურთის ნომერია მარტივი რიცხვი. ცხადია ასეთი ხდომილობანი თავსებადნი არიან.

თუ რამდენიმე ხდომილობიდან არც ერთს არა აქვს არავითარი უპირატესობა მოხდენის აზრით, მაშინ იტყვიან, რომ გვაქვს ერთნაირად მოსალოდნელი ხდომილობანი. მაგალითად, თუ ყუთში თანაბარი რაოდენობითაა მოთავსებული თეთრი და შავი ბურთები და შემთხვევით ვიღებთ რომელიმე ერთს, მაშინ ამოღებული ბურთი თეთრი იქნება თუ შავი, მათი შესაბამისი ხდომილობანი ერთნაირად იქნება მოსალოდნელი.

თუ რომელიმე ხდომილობის მოხდენას ან არმოხდენას არავითარი გავლენა არა აქვს მეორის მოხდენაზე, მაშინ ვიტყვით, რომ ასეთი ხდომილობანი ურთიერთდამოუკიდებელია. წინააღმდეგ შემთხვევაში ხდომილობანი ურთიერთდამოკიდებული იქნებიან. მაგალითად, ვთქვათ ერთი ხდომილობა ნიშნავს ლითონის ფულის აგდებისას გერბის მოსვლას, ხოლო მეორე ხდომილობა კი—სხვადასხვა ფერის ბურთებიანი ყუთიდან წითელი ფერის ბურთის ამოღებას, მაშინ ასეთი ხდომილობანი ურთიერთდამოუკიდებელნი იქნებიან. თუ ერთი რომელიმე ხდომილობა ნიშნავს რაიმე სამიზნის ცენტრში გასრო-

ლილი ვაზნის მოხვედრას, ხოლო მეორე ხდომილობა — რომელიმე მსროლელის არჩევას, ასეთი ხდომილობანი ურთიერთდამოკიდებული იქნებიან, რადგანაც ამორჩეული მსროლელი რამდენად მაღალი კლასისაა, იმდენად უფრო ექნება ადგილი სამიზნეს ცენტრში მოხვედრას.

რომელიმე A ხდომილობის საწინააღმდეგო ხდომილობა ნიშნავს მის არმოხდენას და აღინიშნება $\bar{A}$-ით. მაგალითად, A ხდომილობა იყოს სხვადასხვა ფერის ბურთებიანი ყუთიდან თეთრი ფერის ბურთის ამოღება, მაშინ $\bar{A}$ იქნება იმავე ყუთიდან არათეთრი ფერის ბურთის ამოღება.

ცხადია, A და $\bar{A}$ ხდომილობათა ერთდროულად მოხდენა გამორიცხულია, ხოლო ერთ-ერთ მათგანს აუცილებლად ექნება ადგილი. ამგვარად, საწინააღმდეგო ხდომილობანი უთავსებადნი არიან.

თუ გვაქვს რამდენიმე ხდომილობა, რომელთაგან ერთ-ერთს აუცილებლად ექნება ადგილი, ხოლო ნებისმიერი ორის ერთდროულად მოხდენა შეუძლებელია, ვიტყვით, რომ ასეთი ხდომილობები ქმნიან ხდომილობათა სრულ სისტემას. მაგალითად, საწინააღმდეგო ხდომილობები ყოველთვის ქმნიან ხდომილობათა სრულ სისტემას. აგრეთვე სათამაშო კამათელზე თუ ერთიანის მოსვლას აღვნიშნავთ A_1-ით, ორიანის მოსვლას A_2-ით, და ა. შ. ექვსიანის მოსვლას A_6-ით, მაშინ A_1, A_2, A_3, A_4, A_5, A_6 ხდომილობები ქმნიან ხდომილობათა სრულ სისტემას.

განვიხილოთ შემდეგი მაგალითი. ვთქვათ ქარხანა ამზადებს რაიმე პროდუქციას. დავუშვათ, რომ დამზადებული პროდუქცია ორი ხარისხისაა—I და II, აგრეთვე შეიძლება იყოს წუნდებულიც. A იყოს ხდომილობა, რომელიც აღნიშნავს, რომ შემთხვევით აღებული მზა პროდუქცია I ხარისხისაა, B—II ხარისხისაა, ხოლო C—წუნდებულია.

ცხადია, როცა შემთხვევით ვიღებთ რომელიმე მზა პროდუქციას, ადგილი ექნება ან A, ან B, ან C ხდომილობას. არ შეიძლება ორ რომელიმე მათგანს ერთდროულად ჰქონდეს ადგილი. თუ ქარხანა კარგად მუშაობს, მაშინ უფრო უნდა ველოდოთ, რომ შემთხვევით შერჩეული პროდუქცია I ხარისხისაა, ე. ი. ადგილი ექნება A ხდომილობას და უფრო ნაკლებად უნდა ველოდოთ, რომ ადგილი ექნება B და C ხდომილობებს. როგორც ვხედავთ, A, B და C ხდომილობები არ არიან ერთნაირად მოსალოდნელი.

ყოველი ცალკეული ცდის დროს A, B და C ხდომილობათაგან ერთ-ერთს აუცილებლად ექნება ადგილი, ამიტომ ისინი მხოლოდ შესაძლებელი ხდომილობებია.

რადგანაც რომელიმე ორი მათგანის ერთდროულად მოხდენა შეუძლებელია, ამიტომ ისინი უთავსებადი ხდომილობებია.

რადგანაც A, B და C ხდომილობათაგან ერთ-ერთს აუცილებლად ექნება ადგილი, ხოლო ნებისმიერი ორი უთავსებადია, ამიტომ ისინი ქმნიან სრულ სისტემას.

ლითონის ფულის აგდებისას გერბის და საფასურის მოსვლის აღმნიშვნელი ხდომილობები იქნებიან საწინააღმდეგო, უთავსებადი, ერთნაირად შესაძლებელი და ქმნიან ხდომილობათა სრულ სისტემას.

ხდომილობები რიცხვით სიდიდეებს არ წარმოადგენენ, მაგრამ მათთვის შესაძლებელია ზოგიერთი არითმეტიკული ოპერაციის განმარტება. მაგალითად, ორი A და B ხდომილობის ჯამი $A+B$ მაშინ იქნება, როცა ერთ-ერთი შესაკრები მაინც არსებობს. თუ A აღნიშნავს სხვადასხვა ფერის ბურთებიანი ყუთიდან თეთრი ფერის ბურთის ამოღებას, ხოლო B—შავისას, მაშინ $A+B$ ჯამს მაშინ ექნება ადგილი, როცა შემთხვევით ამოღებული ბურთი იქნება ან თეთრი ან შავი.

ორი A და B ხდომილობის ნამრავლს (AB) მაშინ ექნება ადგილი, როდესაც ორივეს ერთდროულად ექნება ადგილი. მაგალითად, თუ A არის იგივე სხვადასხვა ფერის ბურთებიანი ყუთიდან თეთრი ფერის ბურთის ამოღების აღმნიშვნელი ხდომილობა ყუთიდან პირველად ერთ-ერთი ბურთის ამოღებისას, ხოლო B აღნიშნავს იმავე ყუთიდან მეორედ ამოღებისას წითელი ფერის ბურთის ამოღებას, მაშინ AB ნამრავლს ექნება ადგილი, როდესაც პირველად ამოღებული ბურთი იქნება თეთრი, ხოლო მეორედ ამოღებული—წითელი.

ორი A და B ხდომილობის სხვაობა $A-B$ მაშინ იქნება, როცა ადგილი ექნება A-ს და არ ექნება ადგილი B-ს. მაგალითად, უკანასკნელ მაგალითში, პირველად ამოღებული ბურთი თუ თეთრი იქნება, ხოლო მეორედ ამოღებული არაწითელი, მაშინ ადგილი ექნება $A-B$ სხვაობას.

ვთქვათ, გვაქვს ორი ხდომილობა—A და B. თუ A ხდომილობა იწვევს B-ს, მაშინ ვიტყვით, რომ A მოიცავს B-ს და ეს გარემოება სიმბოლურად ასე ჩაიწერება:

$$A \subset B$$

ამ ოპერაციას ჩართვა ეწოდება.

მაგალითად, ვთქვათ A არის კამათელზე ხუთიანის მოსვლის აღმნიშვნელი ხდომილობა, ხოლო B იმავე კამათელზე კენტი ნომრის მოსვლის აღმნიშვნელი ხდომილობა. მაშინ ცხადია, კამათელზე ხუთიანის მოსვლა იმავე დროს კამათელზე კენტი ნომრის მოსვლასაც ნიშნავს და გვექნება ჩართვა

$$A \subset B.$$

A და B ხდომილობანი ტოლძალოვანნი იქნებიან, თუ A ხდომილობა B-ს მოიცავს და ამავე დროს B მოიცავს A-ს, ე. ი. თუ $A \subset B$ და $B \subset A$, მაშინ დავწერთ:

$$A = B.$$

ამ შემთხვევაში ტოლობის ნიშანი ტოლძალოვნებას ნიშნავს.

ხდომილობათა ჯამის განმარტების თანახმად, $A + \bar{A}$ აუცილებელი ხდომილობა (U) იქნება, ხოლო $A\bar{A}$ კი შეუძლებელი (V), ე. ი. $A + \bar{A} = U$, $A\bar{A} = V$. საწინააღმდეგო ხდომილობის განმარტების თანახმად $\bar{A}$ ხდომილობის საწინააღმდეგო ხდომილობა იქნება თვითონ A ხდომილობა, ე. ი.

$$\bar{\bar{A}} = A.$$

ცხადია, რომ

$$\overline{A+B} = \bar{A} \cdot \bar{B}$$

და

$$\overline{AB} = \bar{A} + \bar{B}.$$

ვთქვათ, მოცემულია ორი ხდომილობა A და B. A ხდომილობა გვექნება, თუ ერთდროულად არის A-ც და B-ც, ანდა როდესაც ერთდროულად არის A და B-ს საწინააღმდეგო ხდომილობა $\bar{B}$. AB და $A\bar{B}$, რომლებიც უთავსებადი ხდომილობები არიან, იქნებიან A-ს შემადგენელი ხდომილობები. მაშასადამე სამართლიანია შემდეგი ტოლძალოვნება:

$$A = AB + A\bar{B}.$$

ანალოგიურად მივიღებთ:

$$B = AB + \bar{A}B.$$

$A + B$ ჯამის წარმოდგენა უთავსებად შემადგენელ ხდომილობათა ჯამის სახით შემდეგნაირად შეიძლება:

$$A + B = AB + A\bar{B} + \bar{A}B.$$

§ 2. ალბათობის კლასიკური განსაზღვრა

ვთქვათ, ყუთში 25 ერთნაირი სიდიდის ბურთია მოთავსებული, რომლებიც გადანომრილია 1-დან 25-მდე. გულდასმით ავურიოთ ყუთში მოთავსებული ბურთები და ამოვიღოთ შემთხვევით რომელიმე მათგანი. ცხადია, ერთნაირად მოსალოდნელი იქნება 1-დან 25-მდე ნებისმიერი ნომრიანი ბურთის ამოღება. ჩატარებული ცდის შედეგად შეიძლება გვქონდეს 25 სხვადასხვა შემთხვევა, რომელთა ცალ-ცალკე მოხ-

დენა შესაძლებელია. ამ შემთხვევაში ვიტყვით, რომ ყველა შესაძლებელ შემთხვევათა რიცხვია 25. დავუშვათ, გვაინტერესებს ისეთი ბურთის ამოღება, რომლის ნომერია 9. ცხადია, ამ ბურთის ამოღებას ყველა 25 შესაძლებელი შემთხვევიდან ხელს უწყობს მხოლოდ ერთი. თუ გვაინტერესებს ისეთი ბურთის ამოღება, რომლის ნომერია ლუწი რიცხვი, მაშინ მის ამოღებას ხელს შეუწყობს 25-დან 12 შემთხვევა. საზოგადოდ, ყველა შესაძლებელი შემთხვევიდან იმათ, რომლებიც ხელს უწყობენ ჩვენთვის სასურველი ხდომილობის მოხდენას, ხელშემწყობი შემთხვევები ეწოდება. ბურთის ამოღების ცდის დროს 25-დან რომელიმე ერთ-ერთის ამოღება აუცილებლობას წარმოადგენს.

მაშასადამე, რადგანაც 25 ბურთიდან რომელიმე ერთ-ერთის ამოღება აუცილებელია, ბუნებრივია, მას შევუსაბამოთ ისეთი რიცხვი, რომელიც აუცილებელი ხდომილობის ალბათობის საზომი იქნება. **აუცილებელი ხდომილობის ალბათობად მივიღოთ ერთი.** მაშინ, ცხადია, ჩვენ მიერ წინასწარ დასახელებული № 9 ბურთის ამოღების ალბათობა 25-ჯერ ნაკლები იქნება, ვიდრე ნებისმიერი ერთის ამოღების ალბათობა და ის იქნება $\frac{1}{25}$ სიდიდის ტოლი. თუ გვაინტერესებს ლუწნომრიანი ბურთის ამოღების ალბათობა, მაშინ, რადგანაც 25-დან 12 არის ასეთი, მათგან რომელიმე ერთ-ერთის ამოღების ალბათობა 12-ჯერ მეტი იქნება რომელიმე წინასწარ დასახელებული ნომრიანი ბურთის ამოღების ალბათობაზე, ე. ი. იქნება $\frac{12}{25}$. აქ მრიცხველში დგას ჩვენთვის სასურველი ხდომილობის ხელშემწყობ შემთხვევათა რიცხვი 12, ხოლო მნიშვნელში — ყველა შესაძლებელ შემთხვევათა რიცხვი 25. ყველა შესაძლო შემთხვევა ერთნაირად არის მოსალოდნელი. ერთი ცდის დროს მათგან მხოლოდ ერთ-ერთს ექნება ადგილი და რომელიმე ორის ერთდროულად მოხდენა გამორიცხულია. ამგვარად, ყველა შესაძლო შემთხვევა ქმნის ხდომილობათა სრულ სისტემას, რომელშიც შემავალი ხდომილობები ერთნაირად არის მოსალოდნელი.

განხილული მაგალითის დროს მიღებული ალბათობის განსაზღვრის განზოგადება საზოგადოდ შესაძლებელია, სახელდობრ:

თუ რომელიმე შემთხვევითი A ხდომილობისათვის ყველა შესაძლებელ შემთხვევათა რიცხვს აღვნიშნავთ n-ით, ხოლო მათგან ხელშემწყობ შემთხვევათა რიცხვს — m-ით, მაშინ **A ხდომილობის ალბათობა იქნება წილადი, რომლის მნიშვნელია ყველა შესაძლო შემთხვევათა n რიცხვი, ხოლო მრი-**

ცხვეული — ხელშემწყობ შემთხვევათა m რიცხვი. იგულისხმება, რომ ყველა შესაძლებელი შემთხვევა ერთნაირადაა მოსალოდნელი და ქმნიან ხდომილობათა სრულ სისტემას.

მიღებული განმარტება, რომელსაც ალბათობის კლასიკური განსაზღვრა ეწოდება, დაიწერება შემდეგნაირად:

$$P(A)=\frac{m}{n}\,. \tag{2.1}$$

ალბათობის კლასიკური განსაზღვრიდან შეიძლება შემდეგი დასკვნების გაკეთება:

1. რადგანაც ხელშემწყობ შემთხვევათა და ყველა შესაძლო შემთხვევათა რიცხვი არაუარყოფითი რიცხვებია, ამიტომ ალბათობაც არაუარყოფითია, ე. ი.

$$P(A)\geqslant 0.$$

2. რადგანაც ყველა შესაძლო შემთხვევათა რიცხვი ყოველთვის მეტი ან ტოლია ხელშემწყობ შემთხვევათა რიცხვზე, ამიტომ ალბათობა არ აღემატება ერთს, ე. ი.

$$P(A)\leqslant 1.$$

3. თუ ყველა შესაძლო შემთხვევა რაიმე ხდომილობისათვის ამავე დროს ხელშემწყობიცაა, ე. ი. თუ $n=m$, მაშინ გვექნება აუცილებელი ხდომილობა და მისი ალბათობა იქნება ერთის ტოლი.

ამგვარად, თუ აუცილებელ ხდომილობას აღვნიშნავთ U-თი, დავწერთ:

$$P(U)=\frac{n}{n}=1. \tag{2.2}$$

მაგალითად, ვთქვათ ყუთში 10 ერთნაირი ზომის წითელი ფერის ბურთია. ცდის შედეგად გვაინტერესებს წითელი ფერის ბურთის ამოღება. ცხადია, ყოველი ცდის დროს ამოვიღებთ წითელი ფერის ბურთს. მას ყველა შესაძლო 10 შემთხვევიდან ათივე უწყობს ხელს და ამგვარად, გვექნება აუცილებელი ხდომილობა.

4. თუ ყველა შესაძლო შემთხვევიდან რაიმე ხდომილობას არც ერთი არ უწყობს ხელს, ე. ი. $m=0$, მაშინ გვექნება შეუძლებელი ხდომილობა და მისი ალბათობა იქნება 0. თუ შეუძლებელ ხდომილობას აღვნიშნავთ V-თი, დავწერთ

$$P(V)=\frac{0}{n}=0. \tag{2.3}$$

მაგალითად, იმავე ყუთიდან, რომელშიც წითელი ფერის 10 ბურთია, შავი ფერის ბურთის ამოღება შეუძლებელია, რადგანაც მას არც ერთი შემთხვევა ხელს არ უწყობს.

5. რადგანაც შემთხვევითი ხდომილობისათვის ხელშემწყობ შემთხვევათა რიცხვი ყოველთვის ნაკლებია ყველა შესაძლო შემთხვევათა რიცხვზე, ამიტომ *შემთხვევითი ხდომილობის ალბათობა ყოველთვის წესიერ წილადს წარმოადგენს*, ე. ი.

$$0 < P(A) < 1.$$

მაგალითად, ვთქვათ, ყუთში 10 ბურთია, რომელთაგან 5 წითელი, 3 შავი და 2 თეთრია. მაშინ, თუ A, B და C-თი აღვნიშნავთ შესაბამისად წითელი, შავი და თეთრი ფერის ბურთების ამოღების აღმნიშვნელ ხდომილობებს, გვექნება:

$$P(A) = \frac{5}{10}, \quad P(B) = \frac{3}{10} \quad \text{და} \quad P(C) = \frac{2}{10}.$$

6. თუ $A = B$, ე. ი. თუ *ხდომილობანი ტოლძალოვანია*, მაშინ

$$P(A) = P(B).$$

მართლაც, ტოლძალოვან ხდომილობებს ერთი და იგივე ხელშემწყობ შემთხვევათა რიცხვი აქვს და ამიტომ მათი ალბათობანიც ერთმანეთის ტოლი იქნება.

§ 3. ალბათობის გამოთვლა კომბინატორიკის დახმარებით

სწავლებას წყობათა, გადანაცვლებათა და ჯუფთებათა შესახებ, როგორც ელემენტარული მათემატიკის კურსიდანაა ცნობილი, *კომბინატორიკა* ეწოდება.

თუ ალბათობის გამოსათვლელად საჭირო ყველა შესაძლო და ხელშემწყობ შემთხვევათა რიცხვის საპოვნელად კომბინატორიკის ფორმულებს გამოვიყენებთ, გამოთვლები საგრძნობლად გამარტივდება.

წყობები, გადანაცვლებები და ჯუფთებები n ელემენტისაგან ქმნიან გარკვეულ ჯგუფებს. ელემენტებად შეიძლება წარმოვიდგინოთ როგორც გარკვეული საგნები, ისე ხდომილობანიც.

ამ ახალი ცნებების შინაარსში რომ გავერკვეთ, განვიხილოთ შემდეგი მაგალითი.

ვთქვათ, პრაქტიკაზე გასაგზავნ სტუდენტთაგან შევარჩიეთ რომელიმე სამი: a, b და c, რომელთა გაგზავნა შეიძლება ორ სხვადასხვა A და B ქალაქში. დავუშვათ, რომ სამივე პრაქტიკანტის გაგზავნა

ნებისმიერ დასახელებულ ქალაქებში ერთნაირადაა შესაძლებელი. პრაქტიკანტების გაგზავნა შეიძლება მოხდეს შემდეგნაირად: a და b ან a და c, ან b და c გაიგზავნონ შესაბამისად A და B ქალაქებში; ანდა b და a ან c და a, ან c და b გაიგზავნონ იმავე მიმდევრობით იმავე A და B ქალაქში. ამგვარად, შესაძლებელია 6 სხვადასხვა ჯგუფის შედგენა, რომლებიც ერთმანეთისაგან განსხვავებულია ან ელემენტთა შემადგენლობით ან მიმდევრობით. მაგალითად, პირველი და მეოთხე ჯგუფი ერთი და იმავე ელემენტებისაგან შედგება, მაგრამ მათი მიმდევრობა განსხვავებულია. მიღებული 6 ჯგუფი წარმოადგენს წყობას 3 ელემენტისაგან ორ-ორად. იმ შემთხვევაში, როდესაც გვაქვს n ელემენტი და მათგან ვადგენთ m-ელემენტიან ($m \leqslant n$) ჯგუფებს, მაშინ შესაძლებელია წყობანი (გადანაცვლებანი) გვქონდეს, რომლებიც ერთმანეთისაგან განსხვავებული იქნება ელემენტთა შემადგენლობით ან მათი რიგით.

თუ სამი პრაქტიკანტი სტუდენტიდან ორს ვაგზავნით ერთსა და იმავე ქალაქში, მაშინ გაგზავნის რიგს არავითარი მნიშვნელობა არ ექნება და გვექნება მხოლოდ სამი ჯგუფი: a და b, a და c, b და c. ასეთ ჯგუფებს ჯ უ ფ თ ე ბ ა ეწოდება.

ამგვარად, სამი ელემენტისაგან ორ-ორელემენტიანი ჯგუფი, რომლებიც ერთმანეთისაგან მხოლოდ ელემენტთა შემადგენლობით განსხვავდება,. იქნება სამი და მას ჯუფთება ეწოდება.

ზოგადად, თუ გვაქვს n ელემენტი და მათგან შევადგენთ ისეთ m-ელემენტიან ჯგუფებს, რომლებიც ერთმანეთისაგან განსხვავდებიან მხოლოდ ელემენტთა შემადგენლობით, მივიღებთ n-ელემენტისაგან m-ელემენტიან ჯუფთებას.

წყობათა, გადანაცვლებათა და ჯუფთებათა რიცხვის გამოსაანგარიშებელი ფორმულები მარტივად მიიღება.

ვთქვათ, გვაქვს 4 ელემენტი a, b, c, d. მათგან თითოელემენტიანი წყობა გვექნება 4. სახელდობრ,

$$a, \ b, \ c, \ d.$$

თითოეულ მიღებულთაგანს მივუერთოთ თითოეული სხვა ელემენტი. კერძოდ a-ს მივუერთოთ დანარჩენი სამი, ასევე b-ს და ა. შ. გვექნება:

$$\begin{aligned}&ab, \ ac, \ ad,\\ &ba, \ bc, \ bd,\\ &ca, \ cb, \ cd,\\ &da, \ db, \ dc.\end{aligned}$$

წყობათა რიცხვი 4 ელემენტისაგან ორ-ორად ტოლია 4.3=12. თუ მიღებულ ორ-ორელემენტიან წყობებს მივუერთებთ დანარჩენებს (რომლებიც მათში არ შედის), მივიღებთ სამ-სამელემენტიან წყობებს:

$$abc,\ abd,$$

$$acb,\ acd,$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$dca,\ dcb.$$

ნებისმიერ ამ წყობათაგანში ერთდროულად არ შედის არც ერთი ორი ელემენტი თავიდან აღებული 4-დან. ამიტომ ყოველი ორელემენტიანი წყობიდან, რომელთა რაოდენობა 12-ია, მასში არშემავალი ელემენტის მიერთებით მივიღებთ სამ-სამელემენტიან წყობას, თითოეულისაგან—ორს. სულ გვექნება $12 \cdot 2 = 24$ ან $4 \cdot 3 \cdot 2 = 24$.

ახლა თუ სამ-სამელემენტიან წყობებს მივუერთებთ მეოთხე ელემენტს, რომელიც მასში არ შედის, მივიღებთ ოთხელემენტიან წყობას:

$$abcd,\ abdc, \ldots,\ dcba.$$

მათი რიცხვი იქნება $4 \cdot 3 \cdot 2 \cdot 1 = 24$.

თუ 4-ის ნაცვლად ავიღებთ n საგანს, ადვილად დავწერთ, თუ რამდენი იქნება მათგან—ერთი, ორი, სამი და ა. შ. ზოგადად $m (m \leqslant n)$ ელემენტიანი წყობა.

A_n^m-ით აღვნიშნოთ n-ელემენტისაგან m-ელემენტიანი წყობა. მაშინ ერთელემენტიან წყობათა რიცხვი იქნება თვით ელემენტთა რაოდენობა n:

$$A_n^1 = n.$$

ორელემენტიან წყობათა რიცხვი იქნება

$$A_n^2 = n(n-1),$$

რადგანაც ორელემენტიანი წყობა მიიღება, როდესაც n-თითოელემენტიან წყობას მივუერთებთ დანარჩენთაგან (რომელთა რაოდენობაა $n-1$) რომელიმე ერთს.

სამ-სამელემენტიანი წყობათა რიცხვი იქნება:

$$A_n^3 = n(n-1)(n-2),$$

რადგანაც ამ შემთხვევაში ორ-ორელემენტიან წყობებს, რომელთა რაოდენობაა $n(n-1)$, მივუწერთ დანარჩენთაგან რომელიმე ერთს (მათი რაოდენობა კი არის $n-2$) და ა. შ. შეიძლება ზოგადად დავწეროთ:

$$A_n^m = n(n-1)(n-2) \ldots [n-(m-1)]. \tag{3.1}$$

ე. ი. n ელემენტიდან m-ელემენტიან წყობათა რიცხვი ტოლია m ნატურალურ რიცხვთა ნამრავლისა, რომელთაგან ყველაზე დიდია n.

წყობათა რიცხვის გამოსაანგარიშებელი ფორმულიდან მიიღება აგრეთვე გადანაცვლებათა რიცხვის გამოსათვლელი ფორმულაც.

თუ $m=n$, ე. ი. წყობაში ყველა მოცემული n ელემენტი შედის, მაშინ ჯგუფები ერთმანეთისაგან განსხვავებული იქნება მხოლოდ ელემენტთა რიგით. ამიტომ ისეთ წყობას, რომლის ჯგუფები მხოლოდ ელემენტთა რიგითაა განსხვავებული, გადანაცვლება ეწოდება. გადანაცვლებათა რიცხვის გამოსათვლელი ფორმულა იქნება:

$$P_n = A_n^n = n(n-1)(n-2)\dots 3\cdot 2\cdot 1 = n!. \tag{3.2}$$

n ელემენტისაგან ყველა გადანაცვლების რიცხვი ტოლია ნატურალურ რიცხვთა მწკრივის პირველი n წევრის ნამრავლის, ე. ი. ტოლია n ფაქტორიალის.

ჯუფთებათა რიცხვის გამოსათვლელი ფორმულის მისაღებად ვიგულისხმოთ, რომ n ელემენტისაგან შევადგინეთ ყველა შესაძლო ჯგუფი თითოეული m ელემენტისაგან, რომლებიც განსხვავებულია ელემენტების შემადგენლობით და არავითარი მნიშვნელობა არა აქვს მათ რიგს. ეს ჯგუფები, განმარტების თანახმად, წარმოადგენს ყველა შესაძლო ჯუფთებებს n ელემენტიდან m-ად.

მაგალითად, თუ გვაქვს (4) ელემენტი a, b, c, d, მაშინ ასეთი ჯგუფები სამ-სამ ელემენტად შეიძლება გვქონდეს ოთხი: abc, abd, acd, bcd; თუ მათ რიცხვს აღვნიშნავთ ზოგადად C_n^m-ით, ოთხი ელემენტისათვის სამ-სამად ჯუფთებათა რიცხვი იქნება:

$$C_4^3 = 4.$$

ვთქვათ, n ელემენტიდან ავიღეთ რომელიმე ჯუფთება, რომელშიც შედის m ელემენტი. მისგან შევადგინოთ სხვადასხვა ჯგუფები, რომლებიც ერთმანეთისაგან განსხვავებული იქნება მხოლოდ ელემენტთა მიმდევრობით. სხვანაირად რომ ვთქვათ, ვაწარმოოთ ყოველგვარი გადანაცვლება m ელემენტისაგან. მაშინ აღებული ერთი ჯუფთებიდან მივიღებთ p_m გადანაცვლებას. მაგალითად, თუ ავიღებთ abc ჯუფთებას, მისგან მივიღებთ 6 გადანაცვლებას:

$$abc,\ acb,\ bac,\ bad,\ cab,\ cba.$$

ვთქვათ, იგივე გავაკეთეთ თითოეული ჯუფთებისათვის, მაშინ მივიღებთ $C_n^m P_m$ ჯგუფს, რომლებიც m ელემენტისაგან შედგება და ერთმანეთისაგან განსხვავდება რიგით ან ელემენტების შემადგენლობით.

მიღებული ჯგუფები წარმოადგენს n ელემენტისაგან m—m-ად წყობებს. მათი რიცხვია A_n^m. მაშასადამე,

$$C_n^m P_m = A_n^m,$$

საიდანაც

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n(n-1)(n-2)\ldots[n-(m-1)]}{1\,.\,2\ldots m}\,.$$

მიღებული ფორმულის მრიცხველი და მნიშვნელი გავამრავლოთ $1 \cdot 2 \ldots (n-m)$ ნამრავლზე, მივიღებთ:

$$C_n^m = \frac{1 \cdot 2 \ldots (n-m)\,[(n-m)+1]\,[(n-m)+2]\ldots[(n-m)+m]}{1 \cdot 2 \ldots m \cdot 1 \cdot 2 \ldots (n-m)},$$

ან

$$C_n^m = \frac{n!}{m!(n-m)!}\,. \tag{3.3}$$

გადანაცვლებათა რიცხვის გამოსათვლელი ფორმულა მიღებულია იმ პირობით, რომ ყველა ელემენტი ერთმანეთისაგან განსხვავებულია. ახლა დავუშვათ, რომ n ელემენტიდან m ერთნაირია. ასეთ შემთხვევაში მიღებული გადანაცვლებანი იწოდება განმეორებიან გადანაცვლებად.

ცხადია, მათი რიცხვი არ იქნება P_n. მათი რაოდენობა აღვნიშნოთ N-ით, ე. ი. N არის რიცხვი ყველა გადანაცვლებისა n ელემენტიდან, რომელთაგან m არის ერთნაირი.

წარმოვიდგინოთ, რომ რომელიმე ერთგანმეორებიან გადანაცვლებაში ერთნაირი ელემენტები შევცვალეთ ახალი განსხვავებული ელემენტებით და ვაწარმოოთ ყველა შესაძლო გადანაცვლება ამ ახალ ელემენტებს შორის. მაშინ ერთგანმეორებიანი გადანაცვლებისგან მიიღება $m!$ გადანაცვლება განმეორების გარეშე.

ვიგულისხმოთ, რომ ასეთი ოპერაცია ჩავატარეთ ყველა განმეორებიანი გადანაცვლებისათვის. მაშინ N განმეორებიან გადანაცვლებიდან მიიღება $N \cdot m!$ გადანაცვლება განმეორების გარეშე. ამასთან მიიღება ყოველგვარი გადანაცვლება n სხვადასხვა ელემენტისაგან. მათი რიცხვია $P_n = n!$ მაშასადამე,

$$N \cdot m! = n!,$$

საიდანაც

$$N = \frac{n!}{m!}\,.$$

გადანაცვლებათა რიცხვი n ელემენტისაგან მცირდება $m!$-ჯერ, თუ n ელემენტიდან m ერთნაირია.

თუ მოცემული n ელემენტიდან m ერთნაირია და დანარჩენი მათგან განსხვავებული $n-m$ აგრეთვე ერთნაირია, მაშინ გადანაცვლებათა

რიცხვი შემცირდება კიდევ $(n-m)!$-ჯერ და განმეორებიან გადანაცვლებათა რიცხვი იქნება:

$$\frac{n!}{m!\,(n-m)!}=C_n^m,$$

ე. ი. ტოლი იქნება n ელემენტიდან m ელემენტიან ჯუფთებათა რიცხვისა.

ახლა კომბინატორიკის ფორმულების დახმარებით გამოვთვალოთ ალბათობები.

1. ვთქვათ, ყუთში მოთავსებულია 10 ერთნაირი ზომის ბურთი, რომლებიც გადანომრილია 1-დან 10-მდე. ყუთიდან რიგრიგობით ვიღებთ 5 ბურთს ისე, რომ ერთხელ ამოღებულს უკან არ ვაბრუნებთ. რას უდრის ალბათობა იმისა, რომ ხუთივე ამოღებული ბურთი იქნება ლუწი ნომრის?

ამოხსნა. როცა შემთხვევით ამოვიღებთ 5 ბურთს, შეგვიძლია მივიღოთ სხვადასხვა ჯგუფი, რომლებიც ერთმანეთისაგან განსხვავებულია შემადგენლობით ან ელემენტთა რიგით, ე. ი. ვღებულობთ წყობას 10 ელემენტიდან 5-5-ად. მაშასადამე, ერთნაირად შესაძლებელ ყველა შემთხვევათა რიცხვი n იქნება წყობათა რიცხვი 10-დან 5-5-ად:

$$n=A_{10}^5=10\cdot 9\cdot 8\cdot 7\cdot 6.$$

ხელშემწყობ შემთხვევათა რიცხვს ეკუთვნის ის ჯგუფები, რომლებიც მხოლოდ ლუწ ნომრებს შეიცავს და ერთმანეთისაგან მხოლოდ ელემენტთა რიგით განსხვავდება. მათი რიცხვი m ტოლი იქნება 5 ელემენტისაგან გადანაცვლებათა რიცხვისა

$$m=P_5=1\cdot 2\cdot 3\cdot 4\cdot 5.$$

ალბათობა ამისა, რომ ყველა 5 ბურთი ლუწი ნომრისაა, იქნება:

$$P=\frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{10\cdot 9\cdot 8\cdot 7\cdot 6}=\frac{1}{252}.$$

2. იმავე ამოცანის პირობებში განსაზღვრეთ ალბათობა იმისა, რომ, როცა ზედიზედ ამოვიღებთ ათივე ბურთს (დაბრუნების გარეშე) და დავალაგებთ მიმდევრობით იმისდა მიხედვით, თუ რა მიმდევრობით იყო ამოღებული, მივიღებთ რიცხვებს 1, 2 და 3 ერთმანეთის გვერდით ზრდადი მიმდევრობით.

ამოხსნა. ათივე ბურთის ამოღებისას შეგვიძლია მივიღოთ ჯგუფები, რომლებიც ერთმანეთისაგან მხოლოდ მიმდევრობით განსხვავდება. აქედან ერთნაირად შესაძლებელი ყველა შესაძლო შემთხვევის რიცხვი იქნება გადანაცვლებათა რიცხვი 10 ელემენტიდან:

$$n=P_{10}=1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10.$$

ხელშემწყობ შემთხვევათა რიცხვს მიეკუთვნება ის ჯგუფები 10-10 ბურთიდან, რომლებშიც რიცხვები 1, 2 და 3 ერთმანეთის გვერდით ზრდადი მიმდევრობით იქნება. ასეთ ჯგუფთა რაოდენობას მივიღებთ, თუ 123-ს ჩავთვლით ერთ რიცხვად და ვაწარმოებთ ყველა შესაძლო გადანაცვლებას 8 ელემენტისაგან:

$$m=P_3=1\cdot2\cdot3\cdot4\cdot5\,.\,6\,.\,7\cdot8.$$

საძებნი ალბათობა იქნება:

$$P=\frac{m}{n}=\frac{1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7\cdot8}{1\cdot2\cdot3\cdot4\cdot5\cdot6\,.\,7\cdot8\cdot9\cdot10}=\frac{1}{90}.$$

3. ვთქვათ, ყუთში 25 ბურთია, რომელთაგან 10 შავია და 15 თეთრი. მათგან შემთხვევით ვიღებთ რომელიმე 5-ს. რას უდრის ალბათობა იმისა, რომ მათგან 2 იქნება შავი?

ამოხსნა. ყველა შესაძლო შემთხვევათა რიცხვი n ტოლია ყველა სხვადასხვა შემადგენლობის ჯგუფთა რაოდენობისა, რომელთა შექმნა შეიძლება 25 ელემენტიდან 5-5-ად. ე. ი.

$$n=C_{25}^5=\frac{25\cdot24\cdot23\cdot22\cdot21}{1\cdot2\cdot3\cdot4\cdot5}=5\cdot23\cdot22\cdot21.$$

ხელშემწყობი იქნება ის ჯგუფები 5 ელემენტისაგან, რომლებშიც 2 შავი და 3 თეთრი ბურთია. მათი რაოდენობის განსასაზღვრავად საჭიროა გაირკვეს, ერთი მხრივ, რამდენი შეიძლება გვქონდეს სხვადასხვა შემადგენლობის წყვილი შავი ბურთი და, მეორე მხრივ, რამდენი შეიძლება გვქონდეს სხვადასხვა შემადგენლობის სამ-სამი თეთრი ბურთი.

წყვილ შავ ბურთთა რიცხვი იქნება ჯუფთებათა რიცხვი 10-დან 2-2-ად

$$C_{10}^2=\frac{10\cdot9}{1\cdot2}=5\,.\,9.$$

ხოლო სამ-სამ თეთრ ბურთთა რიცხვი იქნება ჯუფთება 15 ელემენტიდან 3-3-ად

$$C_{15}^3=\frac{15\cdot14\,.\,13}{1\,.\,2\cdot3}=5\cdot7\cdot13.$$

ყოველი წყვილი შავი ბურთებისა შესაძლებელია მოხდეს ერთ ჯგუფში თითოეულ 3 თეთრ ბურთთან ერთად. ამიტომ ყველა იმ 5 ელემენტიან ჯგუფთა რიცხვი, რომლებშიც 2 შავია და 3 თეთრი, ტოლია

$$C_{10}^2\cdot C_{15}^3=5\cdot9\cdot5\,.\,7\cdot13.$$

ალბათობა იმისა, რომ 5 შერჩეული ბურთიდან 2 იქნება შავი, ტოლია

$$P=\frac{m}{n}=\frac{C_{10}^{2}\,C_{15}^{3}}{C_{25}^{5}}=\frac{5\cdot 9\cdot 5\cdot 7\cdot 13}{5\cdot 23\cdot 22\cdot 21}=\frac{195}{506}=0{,}385.$$

4. ვთქვათ, ყუთში მოთავსებულია ერთნაირი ზომის N ბურთი, რომლებიც გადანომრილია და რომელთაგან M შავი ფერისაა, ხოლო დანარჩენი $N-M$ არაშავია. ვთქვათ, ყუთიდან ზედიზედ ვიღებთ n ბურთს ისე, რომ ამოღებულს უკან არ ვაბრუნებთ. რას უდრის ალბათობა იმისა, რომ n ამოღებული ბურთიდან $m\ (m \leqslant n)$ იქნება შავი ფერის?

ამოხსნა. N ნომრიდან ვიღებთ n-ს და ამგვარად მივიღებთ ჯგუფებს, რომლებიც ერთმანეთისაგან ელემენტთა შემადგენლობით განსხვავდება (ელემენტების მიმდევრობას მნიშვნელობა არა აქვს), ასეთ ჯგუფთა რაოდენობა იქნება C_N^n. ამგვარად, ყველა შესაძლო შემთხვევის რიცხვი იქნება C_N^n.

ახლა ვიანგარიშოთ, თუ ამათგან რამდენი იქნება ხელშემწყობ შემთხვევათა რიცხვი, ე. ი. უნდა ვიპოვოთ n ელემენტიანი ჯგუფებიდან რამდენი შეიცავს m შავ ბურთს და $n-m$ არაშავ ბურთს.

ყუთში M შავი ბურთია; შევადგინოთ მათგან m ბურთიანი ჯგუფები, რომლებიც ერთმანეთისაგან ელემენტთა შემადგენლობით განსხვავდება. ასეთ ჯგუფთა რაოდენობა იქნება C_M^m. ყუთში $N-M$ არაშავი ბურთია, მათგან შევადგინოთ სხვადასხვა შემადგენლობის $n-m$ ბურთიანი ჯგუფები. მათი რიცხვი იქნება C_{N-M}^{n-m}. ყოველი m-ბურთიანი ჯგუფი, რომლებშიც მხოლოდ შავი ფერის ბურთებია, შეიძლება შევიდეს ისეთ ნებისმიერ ჯგუფში, რომელშიც $n-m$ არაშავი ბურთია. ამგვარად, ყველა ჯგუფის რაოდენობა, რომლებიც შეიცავს m შავ ბურთს, იქნება $C_M^m \cdot C_{N-M}^{n-m}$.

მაშასადამე, ალბათობა იმისა, რომ შემთხვევით შერჩეულ n ბურთში იქნება m შავი ბურთი, ტოლია

$$P=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

5. იგივე მაგალითის პირობებში რას უდრის ალბათობა იმისა, რომ, როცა ვარჩევთ n ბურთს, მათგან ორი იქნება წინასწარ დასახელებული ნომრიანი?

ამოხსნა. ამ შემთხვევაში ბურთები განსხვავდება მოცემული და არმოცემული ნომრებით. წინასწარდასახელებული ნომრიანი ბურთები

უუთში ორია, ე. ი. $M=2$ და აგრეთვე შერჩევაშიც ორია, ე. ი. $m=2$. თუ ამ მნიშვნელობებს მიღებულ ფორმულაში შევიტანთ, გვექნება:

$$P=\frac{C_2^2\, C_{N-2}^{n-2}}{C_N^n},$$

$$C_2^2=1,$$

$$C_N^n=C_{N-2}^{n-2}\,\frac{N(N-1)}{n(n-1)},$$

ამიტომ

$$P=\frac{n(n-1)}{N(N-1)}.$$

§ 4. ალბათობის სტატისტიკური განსაზღვრა

ალბათობის კლასიკური განსაზღვრა მხოლოდ იმ შემთხვევაში გამოდგება, როდესაც ყველა შესაძლებელი შემთხვევა ერთნაირად მოსალოდნელია. მაგრამ იმ შემთხვევაში, როდესაც ამ უკანასკნელი პირობის შესრულების შესახებ არაფერი ვიცით, შეიძლება ალბათობის სტატისტიკურ განსაზღვრას მივმართოთ, რომელიც შემდეგში მდგომარეობს.

ვატარებთ რაიმე ცდას, ვიმეორებთ მას სავსებით ერთსა და იმავე პირობებში და ვაკვირდებით ყოველი ცდისას ჩვენთვის სასურველი ხდომილობის მოხდენის ფაქტს. ვთქვათ, ცდა ჩავატარეთ n-ჯერ, ხოლო n ცდაში ჩვენთვის სასურველი ხდომილობა მოხდა k-ჯერ. $\frac{k}{n}$ შეფარდებას ჩვენთვის სასურველი ხდომილობის მოხდენის ფარდობითი სიხშირე ვუწოდოთ, k-ს კი სიხშირე. თუ რამდენჯერმე ჩავატარებთ განმეორებით იმავე პირობებში ცდათა იმავე n რაოდენობას და ერთმანეთს შევადარებთ მიღებულ ფარდობით სიხშირეებს, ვნახავთ, რომ მათ შორის სხვაობა უმნიშვნელო იქნება და, რაც მეტი იქნება ჩატარებულ ცდათა რიცხვი n, მით უფრო ნაკლები იქნება ეს სხვაობა.

თუ იმ შემთხვევისათვის, რომლისთვისაც ცდებს ვატარებთ, შესაძლებელია ალბათობის კლასიკური განსაზღვრის გამოყენება, ვნახავთ, რომ ფარდობით სიხშირეთა სიდიდე ცალკეული ხდომილობის ალბათობის p-ს მახლობლად იცვლება.

ამ ცდით დადგენილ ფაქტს, როგორც შემდგომ დავინახავთ, მეტად ღრმა საფუძველი აქვს ბერნულის თეორემაში. ის გარემოება, რომ ცდათა საკმაოდ დიდი რაოდენობით ჩატარებისას ფარდობითი სიხშირე

თითქმის მუდმივ მნიშვნელობას ინარჩუნებს, შესაძლებლობას მოგვცემს გავაფართოოთ მოვლენათა არე, რომელთათვის შეგვეძლება ალბათობაზე ვილაპარაკოთ და ისეთი ხდომილობათა ალბათობანი განვსაზღვროთ, რომელთაც კლასიკური განსაზღვრის დახმარებით ვერ გამოვთვლით. მართალია, ამას ცდის ჩატარებით მიღებული შედეგების მიხედვით ვასკვნით, მაგრამ ეს იმას როდი ნიშნავს, რომ ადრე გარკვეული კანონზომიერება ობიექტურად არ არსებობდა. რა თქმა უნდა, ასეთი კანონზომიერება არსებობდა, მაგრამ უცნობი იყო მეცნიერებისათვის.

თუ ცდას საკმაოდ დიდი რაოდენობით ჩავატარებთ და ასეთ პროცესს რამდენჯერმე გავიმეორებთ, აღმოჩნდება, რომ ჩვენთვის სასურველი ხდომილობის მოხდენის ფარდობითი სიხშირე რომელიმე მუდმივი რიცხვის გარშემო ირხევა; ამ შემთხვევაში ვიტყვით, რომ ამ ხდომილობას ალბათობა აქვს და შეიძლება ამ ხდომილობის ალბათობის მნიშვნელობა საკმაოდ დიდი მიახლოებით განვსაზღვროთ მიღებული ფარდობითი სიხშირის საშუალებით, ე. ი. საკმაო მიახლოებით შეგვიძლია დავწეროთ:

$$P(A) \simeq \frac{k}{n}. \tag{4.1}$$

რა თქმა უნდა, ეს სავსებით არ ნიშნავს ალბათობის განსაზღვრას; ცდის შედეგს აღვნიშნავთ და ვმსჯელობთ ალბათობის არსებობაზე და მის მიახლოებით გამოთვლაზე. აქ ჩანს, რომ ცდა შესაძლებლობას გვაძლევს შევნიშნოთ ბუნებაში არსებული კანონზომიერება, მიახლოებით ვიანგარიშოთ ჩვენთვის სასურველი ხდომილობის ალბათობა, რომელიც ჩვენთვის უცნობი იყო, და, რაც მთავარია, შევამოწმოთ წინასწარი თეორიული მოსაზრებანი. ეს უკანასკნელი იმას ნიშნავს, რომ წინასწარ გარკვეული მოსაზრებით შესაძლებელია რომელიმე p რიცხვი მივიღოთ რაიმე ხდომილობის ალბათობად. თუ ცდებმა ცხადყო, რომ ფარდობითი სიხშირე საგრძნობლად განსხვავდება p რიცხვისაგან, ეს იმის მაჩვენებელი იქნება, რომ ჩვენი ადრინდელი მოსაზრება არასწორი ყოფილა და საკითხი ხელმეორედ გარკვევას საჭიროებს.

მაგალითად, თუ ვიგულისხმებთ, რომ კამათელი სწორია, მაშინ კამათელზე რომელიმე ნომრის, ვთქვათ 4-ის, მოსვლის ალბათობა $\frac{1}{6}$-ის ტოლია. მაგრამ, თუ კამათელი რამდენჯერმე გავაგორეთ და გამოირკვა, რომ ფარდობითი სიხშირე საგრძნობლად განსხვავდება $\frac{1}{6}$-გან, ეს იმის მომასწავებელი იქნება, რომ ჩვენი დაშვება კამათლის

სისწორის შესახებ არ ყოფილა მართებული. ამ შემთხვევაში 4-ის მოსვლის აღმნიშვნელი ხდომილობის ალბათობა იქნება არა $\frac{1}{6}$, არამედ ის რიცხვი, რომლის მახლობლადაც ირყევა ფარდობითი სიხშირეები.

§ 5. უთავსებად ხდომილობათა ჯამის ალბათობის გამოთვლა

ვთქვათ, A და B უთავსებადი ხდომილობანია, რომელთა ალბათობებია $P(A)$ და $P(B)$. ვიანგარიშოთ $A+B$ ჯამის ალბათობა. ყველა შესაძლებელ შემთხვევათა რიცხვი აღვნიშნოთ n-ით, რომელთაგან A-ს ხელს უწყობს m, B-ს კი — k. მაშინ ალბათობის კლასიკური განმარტების თანახმად დავწერთ:

$$P(A)=\frac{m}{n} \quad \text{და} \quad P(B)=\frac{k}{n}.$$

$A+B$ ჯამის ხელშემწყობ შემთხვევათა რიცხვი იქნება $m+k$, მაშინ ალბათობის იმავე კლასიკური განსაზღვრის საფუძველზე დავწერთ

$$P(A+B)=\frac{m+k}{n},$$

საიდანაც

$$P(A+B)=\frac{m}{n}+\frac{k}{n}=P(A)+P(B). \tag{5.1}$$

ამგვარად, *უთავსებად ხდომილობათა ჯამის ალბათობა შესაკრებ ხდომილობათა ალბათობების ჯამის ტოლია.*

მაგალითად, ვთქვათ ყუთში 50 ბურთია, რომელთაგან 30 წითელი, 10 ყვითელი, 6 შავი და 4—თეთრია, რა იქნება ალბათობა იმისა, რომ შემთხვევით ამოღებული ბურთი წითელი ან შავი ფერისაა. ამ შემთხვევაში ყველა შესაძლებელი შემთხვევათა რიცხვი n ტოლია 50-ის; თუ სათანადოდ წითელი და შავი ფერის ბურთის ამოღების შესაბამის ხდომილობებს A და B-თი აღვნიშნავთ, A-ს ხელშემწყობ შემთხვევათა რიცხვი იქნება 30, ხოლო B-სი — 6. ორივეს ხელს უწყობს 36 შემთხვევა. ამგვარად,

$$P(A+B)=\frac{30+6}{50}=\frac{36}{50}=\frac{18}{25}.$$

§ 6. საწინააღმდეგო ხდომილობის ალბათობის გამოთვლა

ვთქვათ, A შემთხვევითი ხდომილობაა. მის საწინააღმდეგო ხდომილობას აღვნიშნავთ $\bar{A}$-ით. ცხადია, საწინააღმდეგო ხდომილობანი უთავსებადია და მათი ჯამი აუცილებელი ხდომილობის ტოლძალოვანია, ე. ი.

$$A+\bar{A}=U.$$

მაგრამ, რადგანაც აუცილებელი ხდომილობის ალბათობა ერთის ტოლია, ამიტომ დავწერთ:

$$P(A+\bar{A})=1.$$

ხოლო თუ გამოვიყენებთ ზემოთ მიღებულ უთავსებად ხდომილობათა ჯამის ალბათობის გამოსათვლელ ფორმულას, გვექნება:

$$P(A)+P(\bar{A})=1,$$

საიდანაც

$$P(\bar{A})=1-P(A). \tag{6.1}$$

ამგვარად, **საწინააღმდეგო ხდომილობის ალბათობა არის ერთისა და თვით აღებული ხდომილობის ალბათობის სხვაობა.**

მაგალითად, ვთქვათ, ყუთში 15 ბურთია, რომელთაგან 4 წითელი ფერისაა, ხოლო დანარჩენი—სხვადასხვა ფერის. რა იქნება ალბათობა იმისა, რომ შემთხვევით ამოღებული ბურთი იქნება არაწითელი ფერის?

A იყოს იმის აღმნიშვნელი ხდომილობა, რომ ამოღებული ბურთი იქნება წითელი ფერის, მაშინ

$$P(A)=\frac{4}{15}.$$

ალბათობა იმისა, რომ ამოღებული ბურთი არაწითელია (ე. ი. ადგილი ექნება $\bar{A}$-ს), იქნება:

$$P(\bar{A})=1-P(A)=1-\frac{4}{15}=\frac{11}{15}.$$

§ 7. ნებისმიერ ხდომილობათა ჯამის ალბათობის გამოთვლა

ვთქვათ, გვაქვს ორი ნებისმიერი A და B ხდომილობა. ყველა შესაძლო შემთხვევათა რიცხვი იყოს n, საიდანაც A-ს ხელშემწყობ შემთხვევათა რიცხვი აღვნიშნოთ m-ით, ხოლო B-ს ხელშემწყობ შემთხვევათა რიცხვი—k-თი. რადგანაც A და B ნებისმიერი ხდომილობებია, შესაძლებელია მათ საერთო ხელშემწყობი შემთხვევებიც ჰქონ-

დეთ. ასეთების რიცხვი აღვნიშნოთ r-ით. უნდა ვიანგარიშოთ $A+B$ ჯამის ალბათობა. ამ ჯამის ხელშემწყობ შემთხვევათა რიცხვი იქნება $m+k-r$; ამიტომ, ალბათობის კლასიკური განმარტების თანახმად, დავწერთ:

$$P(A+B)=\frac{m+k-r}{n}=\frac{m}{n}+\frac{k}{n}-\frac{r}{n},$$

მაგრამ

$$\frac{m}{n}=P(A), \quad \frac{k}{n}=P(B)$$

და, რადგანაც r არის შემთხვევათა რიცხვი, რომელიც ხელს უწყობს ორივე A და B ხდომილობას ერთდროულად, ე. ი. AB ნამრავლს, ამიტომ

$$\frac{r}{n}=P(AB).$$

მაშასადამე, საბოლოოდ მივიღებთ:

$$P(A+B)=P(A)+P(B)-P(AB). \tag{7.1}$$

ამგვარად, ნებისმიერ ხდომილობათა ჯამის ალბათობა უდრის შესაკრებ ხდომილობათა ალბათობების ჯამს მინუს ორივეს ერთდროულად მოხდენის ალბათობა.

§ 8. პირობითი ალბათობა და მისი გამოთვლა

ვთქვათ, მოცემული გვაქვს ორი A და B ხდომილობა, ვიპოვოთ B ხდომილობის ალბათობა, თუ ცნობილია, რომ A ხდომილობას ადგილი ჰქონდა. საძებნი ალბათობა აღვნიშნოთ $P_A(B)$-თი და მას ვუწოდოთ B ხდომილობის პირობითი ალბათობა. ვთქვათ, A და B ხდომილობათათვის ყველა შესაძლო შემთხვევათა რიცხვი არის n, ხოლო A ხდომილობის ხელშემწყობ შემთხვევათა რიცხვი — m, B ხდომილობის ხელშემწყობ შემთხვევათა რიცხვი იყოს k. A და B ხდომილობის ერთდროულად მოხდენის ხელშემწყობ შემთხვევათა რიცხვი აღვნიშნოთ r-ით. მაშინ, ალბათობის კლასიკური განსაზღვრის თანახმად, დავწერთ:

$$P(A)=\frac{m}{n}, \quad P(B)=\frac{k}{n}, \quad P(AB)=\frac{r}{n}. \tag{8.1}$$

ჩვენ გვინდა $P_A(B)$-ის გამოთვლა. რადგან ცნობილია, რომ A ხდომილობა მოხდა, ამიტომ ამის შემდეგ B ხდომილობისათვის ყველა შესაძლო შემთხვევათა რიცხვი იქნება არა n, არამედ m. ამ შესაძლე-

ბელ m შემთხვევიდან B-თვის ხელშემწყობი იქნება r, რომელიც ხელს უწყობს ერთდროულად A და B ხდომილობის მოხდენას, რადგან გარკვეული პირობის (A-ს ჰქონდა ადგილი) შემდეგ B-ს მოხდენა AB-ს მოხდენასაც ნიშნავს. ზემონათქვამის საფუძველზე ალბათობის კლასიკური განსაზღვრის თანახმად დავწერთ:

$$P_A(B)=\frac{r}{m}. \tag{8.2}$$

(8.2) ტოლობის მარჯვენა მხარის მრიცხველი და მნიშვნელი გავყოთ n-ზე, გვექნება:

$$P_A(B)=\frac{\frac{r}{n}}{\frac{m}{n}}.$$

აქედან კი (8.1) ფორმულის ძალით დავწერთ:

$$P_A(B)=\frac{P(AB)}{P(A)}. \tag{8.3}$$

მიღებული (8.3) ფორმულით ხდება პირობითი ალბათობის გამოთვლა და მისგან შეიძლება ნებისმიერ ხდომილობათა ნამრავლის ალბათობის გამოსათვლელი ფორმულის მიღება, სახელდობრ:

$$P(AB)=P(A)\,P_A(B). \tag{8.4}$$

სავსებით ანალოგიურად მივიღებთ:

$$P(AB)=P(B)\,P_B(A). \tag{8.5}$$

(8.4) და (8.5) ფორმულების ერთობლიობა გვაძლევს:

$$P(AB)=P(A)\,P_A(B)=P(B)\,P_B(A). \tag{8.6}$$

ამგვარად, ორი ნებისმიერი ხდომილობის ნამრავლის ალბათობა უდრის ერთ-ერთის ალბათობას გამრავლებულს მეორის პირობით ალბათობაზე.

თუ A და B ხდომილობანი ურთიერთდამოუკიდებელი არის, მაშინ განმარტების თანახმად, ცხადია, ერთ-ერთის მოხდენას არავითარი გავლენა არ ექნება მეორის მოხდენაზე და ეს ნიშნავს, რომ $P_A(B)$ იგივე იქნება, რაც $P(B)$; აგრეთვე $P_B(A)$ იგივე იქნება, რაც $P(A)$.

მაშასადამე, (8.6) ფორმულიდან დამოუკიდებელ ხდომილობათა შემთხვევაში გვექნება:

$$P(AB)=P(A)\,P(B). \tag{8.7}$$

ამგვარად, დამოუკიდებელ ხდომილობათა ნამრავლის ალბათობა თანამამრავლთა ალბათობების ნამრავლის ტოლია.

ცხადია, ეს დებულება თანამამრავლთა ნებისმიერი რიცხვისათვის სამართლიანია.

განვიხილოთ შემდეგი მაგალითი. ვთქვათ, ფეხსაცმელების რომელიმე ქარხანაში დამზადებული ყოველი ასი წყვილი ფეხსაცმელიდან საშუალოდ 97 არის ვარგისი და მათგან 68 პირველი ხარისხისაა. რას უდრის ალბათობა იმისა, რომ შემთხვევით აღებული ვარგისი წყვილი ფეხსაცმელი იქნება პირველი ხარისხის?

A-თი აღვნიშნოთ ხდომილობა, რომელიც ნიშნავს, რომ შემთხვევით აღებული წყვილი ფეხსაცმელი ვარგისია, ხოლო B-თი აღვნიშნოთ ხდომილობა, რომელიც ნიშნავს, რომ შემთხვევით არჩეული წყვილი ფეხსაცმელი პირველი ხარისხისაა. უნდა ვიპოვოთ B ხდომილობის პირობითი ალბათობა იმ პირობით, რომ A-ს უკვე ჰქონდა ადგილი, ე. ი. უნდა ვიპოვოთ $P_A(B)$.

პირობის თანახმად, ერთდროულად A და B-ს ხელს უწყობს ასიდან 68 შემთხვევა, მაშასადამე,

$$P(AB)=0{,}68.$$

A ხდომილობის ხელშემწყობ შემთხვევათა რიცხვია ასიდან 97, მაშასადამე,

$$P(A)=0{,}97.$$

(8.3) ფორმულის ძალით დავწერთ:

$$P_A(B)=\frac{0{,}68}{0{,}97}=\frac{68}{97}.$$

ანდა კიდევ განვიხილოთ ორი მაგალითი, ვთქვათ, ყუთში 16 ბურთია, რომელთაგან 9 თეთრია და 7 შავი. A იყოს ხდომილობა, რომელიც აღნიშნავს, რომ პირველად ამოღებული ბურთი თეთრია, ხოლო B—მეორედ ამოღებული ბურთია თეთრი. მაშინ ალბათობის კლასიკური განმარტების თანახმად $P(A)=\frac{9}{16}$. თუ ამოღებულ ბურთს უკანვე დავაბრუნებთ და ხელმეორედ შემთხვევით ამოვიღებთ ბურთს, მაშინაც $P(B)=\frac{9}{16}$. აქ A და B ხდომილობანი ურთიერთდამოუკიდებელი არიან. ახლა ცდის ჩატარების პირობები ცოტათი შევცვალოთ. სახელდობრ, პირველად ამოღებული ბურთი უკან ყუთში არ

დავაბრუნოთ. მაშინ, ცხადია $P(A)=\frac{9}{16}$, ხოლო $P(B)$-ს მნიშვნელობა დამოკიდებული იქნება იმაზე, თუ რა ფერის იყო პირველად ამოღებული ბურთი. თუ ის იყო თეთრი, მაშინ $P(B)=\frac{8}{15}$, ხოლო თუ ის იყო შავი, მაშინ $P(B)=\frac{9}{15}=\frac{3}{5}$. ამ შემთხვევაში A და B ხდომილობანი ურთიერთდამოკიდებული არიან. ამგვარად, როცა ბურთს უკან ვაბრუნებთ ყუთში, მაშინ $P_A(B)=P(B)$, ხოლო როცა არ ვაბრუნებთ, მაშინ $P_A(B)\neq P(B)$.

მაგალითი 2. ვთქვათ, მუშა ემსახურება ერთდროულად 5 დაზგას. ალბათობა იმისა, რომ დაზგები ერთი საათის განმავლობაში მუშების ჩარევას მოითხოვენ, არის შესაბამისად—0,1; 0,05; 0,09; 0,08; 0,07; რას უდრის ალბათობა იმისა, რომ 5-ვე დაზგა ერთი საათის განმავლობაში იმუშავებს შეუჩერებლივ?

პირობის თანახმად, პირველი დაზგა რომ მწყობრიდან გამოვა ერთი საათის განმავლობაში, ამის ალბათობა $P_1=0{,}1$. მაშასადამე, ის მწყობრიდან რომ არ გამოვა, ამის ალბათობა $q_1=0{,}9$. ასევე $P_2=0{,}05$; $q_2=0{,}95$; $P_3=0{,}09$, $q_3=0{,}91$; $P_4=0{,}08$, $q_4-0{,}92$; და $P_5=0{,}07$, $q_5=0{,}93$.

მაშასადამე, რადგანაც დაზგათა მუშაობა ურთიერთდამოუკიდებელია, ალბათობა იმისა, რომ ხუთივე დაზგა ერთდროულად შეუჩერებლივ იმუშავებენ, იქნება:

$$P=q_1q_2q_3q_4q_5=0{,}9\cdot 0{,}95\cdot 0{,}91\cdot 0{,}92\cdot 0{,}93=0{,}6657.$$

§ 9. ურთიერთდამოუკიდებელ ხდომილობათა ჯამის ალბათობის გამოთვლა

ვთქვათ, $A_1, A_2, \ldots, A_n$ ურთიერთდამოუკიდებელი ხდომილობანია. განმარტების თანახმად, $A_1+A_2+\cdots+A_n$ ჯამი მაშინ გვექნება, თუ გვაქვს ერთ-ერთი შესაკრები ხდომილობა მაინც. ამ ჯამის საწინააღმდეგო ხდომილობა ისეთი იქნება, რომლის დროსაც ამ ჯამში შემავალი არც ერთი ხდომილობა არ იქნება, ე. ი. როცა გვექნება $\bar{A}_1\bar{A}_2\cdots\bar{A}_n$ ნამრავლი. ამგვარად, $A_1+A_2+\cdots+A_n$ ჯამი და $\bar{A}_1\ \bar{A}_2\cdots\bar{A}_n$ ნამრავლი საწინააღმდეგო ხდომილობანია და მათი ჯამი აუცილებელი ხდომილობა იქნება. რადგანაც აუცილებელი ხდომილობის ალბათობა ერთის ტოლია, ამიტომ დავწერთ

$$P(A_1+A_2+\cdots+A_n+\bar{A}_1\ \bar{A}_2\cdots\bar{A}_n)=1.$$

როგორც ვიცით, საწინააღმდეგო ხდომილობანი უთავსებადნი არიან და ამიტომ გვექნება:

$$P(A_1+A_2+\cdots+A_n)+P(\bar{A}_1\ \bar{A}_2\cdots\bar{A}_n)=1,$$

საიდანაც,

$$P(A_1+A_2+\cdots+A_n)=1-P(\bar{A}_1\ \bar{A}_2\cdots\bar{A}_n).$$

რადგანაც $A_1, A_2,\ldots, A_n$ ხდომილობანი ურთიერთდამოუკიდებელია, ურთიერთდამოუკიდებელი იქნებიან მათი საწინააღმდეგო ხდომილობანი და, მაშასადამე, თუ დამოუკიდებელ ხდომილობათა ნამრავლის ალბათობის გამოსათვლელ ფორმულას გამოვიყენებთ, დავწერთ:

$$P(A_1+A_2+\cdots+A_n)=1-P(\bar{A}_1)\,P(\bar{A}_2)\cdots P(\bar{A}_n).$$

როგორც ვიცით,

$$P(\bar{A}_i)=1-P(A_i).$$

ამიტომ დავწერთ:

$$P(A_1+A_2+\cdots+A_n)=1-[1-P(A_1)]\,[1-P(A_2)]\cdots[1-P(A_n)].$$

ეს უკანასკნელი კი მოკლედ ჩაიწერება შემდეგნაირად:

$$P(A_1+A_2+\cdots+A_n)=1-\prod_{i=1}^{n}[1-P(A_i)]$$

როგორც ვხედავთ, მიღებული დამოკიდებულება საშუალებას იძლევა ურთიერთდამოუკიდებელ ხდომილობათა ჯამის ალბათობა შესაკრებ ხდომილობათა ალბათობების საშუალებით გამოვსახოთ.

§ 10. სრული ალბათობის ფორმულა

ვთქვათ,

$$A_1,\ A_2,\ldots,\ A_n \tag{10.1}$$

არის ხდომილობათა სრული სისტემა, ხოლო B იყოს ნებისმიერი ხდომილობა. ცხადია, თუ B ხდომილობას ადგილი ექნება, ის მოხდება (10.1) მიმდევრობის რომელიმე, ერთ-ერთ ხდომილობასთან ერთად [რადგან (10.1) მიმდევრობის ერთ-ერთ ხდომილობას აუცილებლად ექნება ადგილი]. მაშასადამე, B ხდომილობას მაშინ ექნება ადგილი, როდესაც ადგილი ექნება ან A_1B-ს, ან A_2B-ს, ან...ანდა A_nB-ს. ყველა ესენი წყვილ-წყვილად უთავსებადნი არიან და B ხდომილობის წარმოდგენა მათი ჯამის სახით შეიძლება. ამგვარად, დავწერთ:

$$B=A_1B+A_2B+\cdots+A_nB.$$

აქედან კი, რადგანაც ტოლძალოვან ხდომილობათა ალბათობა ტოლია, გვექნება:

$$P(B)=P(A_1B+A_2B+\cdots+A_nB).$$

ან, თუ გამოვიყენებთ უთავსებად ხდომილობათა ჯამის ალბათობის გამოსათვლელ ფორმულას, დავწერთ:

$$P(B)=P(A_1B)+P(A_2B)+\cdots+P(A_nB).$$

თუ თითოეული შესაკრებისათვის გამოვიყენებთ ნამრავლის ალბათობის გამოსათვლელ (8.6) ფორმულას, მაშინ გვექნება:

$$P(B)=P(A_1)\,P_{A_1}(B)+P(A_2)\,P_{A_2}(B)+\cdots+P(A_n)\,P_{A_n}(B),$$

ან საბოლოოდ დავწერთ:

$$P(B)=\sum_{i=1}^{n}P(A_i)\,P_{A_i}(B). \tag{10.2}$$

მიღებული ფორმულა სრული ალბათობის ფორმულას წარმოადგენს.

პრაქტიკაში ხშირია ისეთი შემთხვევები, როცა ჩვენთვის უცნობია რომელიმე ხდომილობის ალბათობა, მაგრამ ვიცით ამ ხდომილობის პირობითი ალბათობა, ე. ი. ვიცით B ხდომილობის ალბათობა, როდესაც ცნობილია, რომ მოხდა A. დავუშვათ გვესაჭიროება B ხდომილობისათვის უპირობო ალბათობის ცოდნა. სწორედ ასეთი შემთხვევების დროს გამოდგება სრული ალბათობის ფორმულა. ალბათობის თეორიაში ხშირად A_i ხდომილობებს (რომელთა მოხდენის ცოდნით B ხდომილობის ალბათობას ვეძებთ) ჰიპოთეზებს უწოდებენ.

მიუხედავად იმისა, რომ არ ვიცით B ხდომილობათა ალბათობა, შეგვიძლია დავუშვათ რაიმე ჰიპოთეზა, რომლის ალბათობა ჩვენთვის ცნობილია. მაშასადამე, თუ სხვადასხვა ჰიპოთეზის დაშვებით ცნობილია B ხდომილობის ალბათობა და ვიცით თვითონ ამ ჰიპოთეზების ალბათობანი, მაშინ ზემოთ მიღებული სრული ალბათობის ფორმულის გამოყენებით გამოვითვლით B ხდომილობის ალბათობას. სწორედ ალბათობის თეორიის ერთ-ერთ ძირითად ამოცანას წარმოადგენს ერთი რომელიმე შემთხვევითი ხდომილობის ალბათობის გამოთვლა, როცა ცნობილია სხვა ხდომილობის ალბათობა.

§ 11. ბაიესის ფორმულა

ვთქვათ, $A_1, A_2, \ldots, A_n$ არის ხდომილობათა სრული სისტემა. (8,6) ფორმულაში თუ A-ს შევცვლით A_i-თ, გვექნება:

$$P(A_i)\,P_{A_i}(B)=P(B)\,P_B(A_i),$$

საიდანაც

$$P_B(A_i)=\frac{P(A_i)\,P_{A_i}(B)}{P(B)}.$$

ან, თუ $P(B)$-ს შევცვლით მისი მნიშვნელობით (10.2), მივიღებთ:

$$P_B(A_i)=\frac{P(A_i)\,P_{A_i}(B)}{\sum_{s=1}^{n}P(A_s)\,P_{A_s}(B)}. \tag{11.1}$$

ეს უკანასკნელი წარმოადგენს ბაიესის ფორმულას. (11.1) ფორმულა სამართლიანია $i=1, 2\ldots n$ მნიშვნელობებისათვის. ყველა $P_B(A_i)$ ალბათობათა ჯამი 1-ის ტოლია, რადგან მრიცხველი ერთ-ერთი შესაკრებია მნიშვნელისა, ხოლო მნიშვნელი ყველა ამ $P_B(A_i)$ ალბათობას ერთი და იგივე ექნება. მართლაც,

$$\sum_{i=1}^{n}P_B(A_i)=\frac{\sum_{i=1}^{n}P(A_i)\,P_{A_i}(B)}{\sum_{s=1}^{n}P(A_s)\,P_{A_s}(B)}=1.$$

ეს უკანასკნელი გარემოება ხელს გვიწყობს, რომ გამოთვლების სისწორე შევამოწმოთ.

სავარჯიშო

1. ყუთში ერთნაირი ზომის 15 ბურთია, რომელთაგან 6 თეთრია და 9 შავი. რას უდრის ალბათობა იმისა, რომ შემთხვევით ამოღებული ბურთი იქნება შავი?

პას. 3/5.

2. ყუთში 11 წითელი, 12 ყვითელი, 10 თეთრი და 17 შავი ბურთია. რას უდრის ალბათობა იმისა, რომ შემთხვევით ამოღებული ბურთი იქნება წითელი ან შავი?

პას. 14/25.

3. ყუთში 5 ბურთია, რომლებიც გადანომრილია 1-დან 5-მდე. რიგრიგობით ვიღებთ ბურთებს. რას უდრის ალბათობა იმისა, რომ ამოღებული ბურთების ნომრები ზრდადი მიმდევრობით დალაგდებიან?

პას. $\frac{1}{120}$.

4. ყუთში 15 საპერფორაციო ბარათია, რომელთა ნომრებია რიგ-რიგობით 51, 52,..., 65. ყუთიდან შემთხვევით ვიღებთ რომელიმე ორ მათგანს. რას უდრის ალბათობა იმისა, რომ მათი ნომრები იქნება 55 და 61?

პას. $\frac{1}{105}$.

5. ყუთში 15 ბურთია, რომელთაგან 10 შეღებილია. შემთხვევით ვიღებთ 3-ს. რა იქნება ალბათობა იმისა, რომ ამოღებული ბურთები შეღებილი იქნება?

პას. $P=\frac{C_{10}^3}{C_{15}^3}=\frac{24}{91}$.

6. კონვერტში 100 ფოტოსურათია, რომელთაგან ერთზე ძებნაა გამოცხადებული. კონვერტიდან შემთხვევით ვიღებთ 10 ფოტოსურათს. რას უდრის ალბათობა იმისა, რომ მათ შორის მოხვდება საძებნი ფოტოსურათი?

პას. $P=\frac{C_{99}^9}{C_{100}^{10}}=0,1$.

7. აპარატი 5 ელემენტისაგან შედგება, რომელთაგან 2 გაცვეთილია. აპარატის ჩართვისას რთავენ შემთხვევით 2 ელემენტს. რას უდრის ალბათობა იმისა, რომ ჩართული იქნება ორივე გაუცვეთელი ელემენტი?

პას. $P=\frac{C_3^2}{C_5^3}=0,3$.

8. ტელეფონის ნომრის აკრეფისას აბონენტს დაავიწყდა ბოლო 3 ციფრი. მან მხოლოდ ის იცის, რომ სამივე ერთმანეთისაგან განსხვავებულია. შემთხვევით იღებს მათ. რა იქნება ალბათობა იმისა, რომ ის სწორად აკრეფს მისთვის საჭირო ნომრებს?

პას. $P=\frac{1}{C_{10}^3}=\frac{1}{720}$.

9. საამქროში მუშაობს 12 მამაკაცი და 8 ქალი. მათგან პირად საქმეთა ნომრების მიხედვით შემთხვევით ვირჩევთ 10-ს. რას უდრის ალბათობა იმისა, რომ არჩეულთაგან 6 იქნება მამაკაცი, ხოლო 4 ქალი?

პას. $P=\frac{C_8^4\,C_{12}^6}{C_{20}^{10}}=\frac{16170}{31189}$.

10. ჯგუფში 20 სტუდენტია. მათ შორის 4 ფრიადოსანია. სიის მიხედვით შემთხვევით ვარჩევთ 6 ს. რას უდრის ალბათობა იმისა, რომ მათ შორის აღმოჩნდება 2 ფრიადოსანი?

პას. $P=\frac{C_4^2 C_{16}^4}{C_{20}^6}=\frac{91}{323}$.

11. ორი მსროლელი ესვრის სამიზნეს, ერთი გასროლის დროს სამიზნეში მოხვედრის ალბათობა პირველი მსროლელისათვის არის 0,6, მეორისათვის კი 0,9. რას უდრის ალბათობა იმისა, რომ სამიზნეში მოხვედრას ექნება ადგილი, თუკი ორივე მსროლელი ერთდროულად ისვრის და სამიზნეში მოხვედრილად ჩაითვლება, თუ ერთ-ერთი მაინც მოახვედრებს?

მ ი თ ი თ ე ბ ა: აქ ისარგებლეთ ნებისმიერ ხდომილობათა ჯამის ალბათობის გამოსათვლელი ფორმულით.

პას. $P=0{,}96$.

12. წინა ამოცანის პირობებში, რას უდრის ალბათობა იმისა, რომ სამიზნეში, როცა ორივე მსროლელი ერთდროულად ისვრის, მხოლოდ რომელიმე ერთი მოახვედრებს?

პას. $P=0{,}42$.

13. ალბათობა იმისა, რომ დამზადებული დეტალი სტანდარტულია, უდრის 0,8. რა იქნება ალბათობა იმისა, რომ ორი შემოწმებული დეტალიდან მხოლოდ ერთი იქნება სტანდარტული.

პას. $P=0{,}18$.

14. საკითხველო დარბაზში ალბათობის თეორიის სახელმძღვანელო სულ 10 ცალია, მათგან 4 ყდაშია ჩასმული. ბიბლიოთეკარი შემთხვევით იღებს სამ მათგანს. რა იქნება ალბათობა იმისა, რომ სამივე იქნება ყდიანი?

მ ი თ ი თ ე ბ ა: ისარგებლეთ ნებისმიერი ხდომილობებისათვის ნამრავლის ალბათობის გამოსათვლელი ფორმულით:

$$P(ABC)=P(A)\,P_A(B)\,P_{AB}(C).$$

პას. $P=\frac{1}{120}$. ? $\frac{1}{30}$

15. საამქროში 7 მამაკაცი და 3 ქალია. მათგან შემთხვევით ვარჩევთ 3-ს. რას უდრის ალბათობა იმისა, რომ ყველანი მამაკაცები იქნებიან?

პას. $P=7/24$.

16. ყუთში 10 ბურთია, რომელთაგან 6 შეღებილია. შემთხვევით ვიღებთ 4-ს. რას უდრის ალბათობა იმისა, რომ ოთხივე შეღებილი იქნება?

პას. $P = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} = 1/14.$

17. სტუდენტმა პროგრამის 25 საკითხიდან იცის 20. რა იქნება ალბათობა იმისა, რომ გამომცდელის მიერ დასმულ 3 საკითხზე ის უპასუხებს სწორად?

პას. $P = \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} = \frac{57}{115}.$

18. ორი მსროლელი ისვრის სამიზნეში. თვითეულისათვის სამიზნეში მოხვედრის ალბათობა უდრის 0,3. მსროლელები ისვრიან რიგრიგობით. თითოეული ორჯერ ესვრის სამიზნეს. ვინც პირველი მოახვედრებს სამიზნეს, ღებულობს ჯილდოს. რას უდრის ალბათობა იმისა, რომ მსროლელები მიიღებენ ჯილდოს?

პას. $P \simeq 0{,}76.$

19. მსროლელი სამი გასროლიდან მიზანში რომ ერთხელ მაინც მოახვედრებს, არის 0,875. რა იქნება ალბათობა ერთი გასროლით სამიზნეში მოხვედრისა?

პას. $P = 0{,}5.$

20. ყუთში, რომელშიც 2 ბურთია, ჩავუმატეთ ერთი თეთრი ბურთი, რის შემდეგ ყუთიდან ამოვიღეთ შემთხვევით ერთი ბურთი.

რას უდრის ალბათობა იმისა, რომ ამოღებული ბურთი იქნება თეთრი?

მითითება: ისარგებლეთ სრული ალბათობის ფორმულით.

პას. $P = 2/3.$

21. გამოთვლით ლაბორატორიაში აქვთ 6 ავტომატი და 4 ნახევარავტომატი. გამოთვლების დროს ავტომატი რომ არ გაფუჭდება, ამის ალბათობაა 0,95, ხოლო ნახევარავტომატი რომ არ გაფუჭდება— 0,8. გამომთვლელი ანგარიშობს შემთხვევით აღებულ მანქანაზე. რას უდრის ალბათობა იმისა, რომ მუშაობის დამთავრებამდე მანქანა არ გამოვა მწყობრიდან?

პას. $P = 0{,}89.$

22. გამომთვლელი მოწყობილობის 3 ურთიერთდამოუკიდებლად მომუშავე ელემენტიდან ორი გაფუჭდა. მონახეთ ალბათობა იმისა, რომ გაფუჭდა პირველი და მეორე ელემენტი, თუ პირველი, მეორე და მესამე ელემენტის გაფუჭების ალბათობებია შესაბამისად: 0,2; 0,4; 0,3.

მ ი თ ი თ ე ბ ა: ისარგებლეთ ბაიესის ფორმულით.

პას. $P \simeq 0,3$.

23. ხელსაწყოს 4 ურთიერთდამოუკიდებლად მომუშავე ნათურიდან ორი გაფუჭდა. იპოვეთ ალბათობა იმისა, რომ გაფუჭდა პირველი და მეორე ნათურა, თუკი პირველი, მეორე, მესამე და მეოთხე ნათურათა გაფუჭების ალბათობებია შესაბამისად: $P_1 = 0,1$; $P_2 = 0,2$; $P_3 = 0,3$ და $P_4 = 0,4$.

პას. $P \simeq 0,039$.

24. გვაქვს დეტალების ორი პარტია. ცნობილია, რომ ერთ პარტიაში ყველა დეტალი ვარგისია, ხოლო მეორეში კი 1/4 გაფუჭებულია. დეტალი, რომელიც შემთხვევით არჩეული პარტიიდანაა აღებული, აღმოჩნდა ვარგისი. განსაზღვრეთ ალბათობა იმისა, რომ მეორე დეტალი იმავე პარტიიდან იქნება უვარგისი, თუკი პირველი დეტალი შემოწმების შემდეგ დავაბრუნეთ პარტიაში.

პას. $P = 3/28$.

25. 20 ნათურიდან 5-ია უმაღლესი ხარისხის. შემთხვევით ვიღებთ 7-ს. რას უდრის ალბათობა იმისა, რომ მათ შორის აღმოჩნდება 3 უმაღლესი ხარისხის?

პას. $P = 0,176$.

26. ვთქვათ, ლითონის ფულს ვაგდებთ ზედიზედ n-ჯერ. რას უდრის ალბათობა იმისა, რომ გერბი მოვა ლუწ რიცხვჯერ?

პას. $P = 1/2$.

27. ერთდროულად ვყარეთ 6 კამათელი. რას უდრის ალბათობა იმისა, რომ ექვსივეგან სხვადასხვა ნომერი დავარდება?

$$\text{პას.} \quad P = \frac{6!}{6^6} = \frac{5}{324}.$$

28. საჭადრაკო ტურნირში მონაწილეობს ორ ტოლ ჯგუფად განაწილებული 30 მოჭადრაკე.

ა) რას უდრის ალბათობა იმისა, რომ ორი უძლიერესი მოჭადრაკე სხვადასხვა ჯგუფში მოხვდება?

ბ) რას უდრის ალბათობა იმისა, რომ ოთხი უძლიერესი მოჭადრაკეთაგან ორი ერთ ჯგუფში მოხვდება, მეორე ორი კი—მეორე ჯგუფში?

პას. ა) $P = 15/29 > 1/2$

$$\text{ბ)} \quad P = \frac{C_{26}^{13} C_4^2}{C_{30}^{15}} = \frac{35}{87} < 1/2.$$

29. ტეტრაედრის ერთ-ერთი გვერდი შეღებილია სამ ფერად. დანარჩენი გვერდები კი — ამ ფერთა შორის ერთ-ერთით (ორი ერთნაირი ფერით შეღებილი გვერდი არ არის). A, B, C-თი აღვნიშნოთ ხდომილობანი, რომ ტეტრაედრის შემთხვევით აღებული გვერდი შეღებილი იქნება სათანადოდ ერთი, მეორე ან მესამე ფერით. ცხადყავით, დამოუკიდებელნი იქნებიან თუ არა A, B და C ხდომილობანი?

პას. წყვილ-წყვილად დამოუკიდებელნი იქნებიან, ხოლო სამივე მთლიანობაში ურთიერთდამოკიდებულია.

30. ვთქვათ, გვაქვს 3 სავსებით ერთნაირი მაგიდა, A, B და C; რომელთაც ორ-ორი უჯრა აქვთ. A მაგიდის თითოეულ უჯრაში თითო ოქროს ფულია, C მაგიდის უჯრებში თითო ვერცხლის ფულია, ხოლო B მაგიდის ერთ უჯრაში ოქროს ფულია, მეორეში კი ვერცხლისა. შემთხვევით მივდივართ რომელიმე მაგიდასთან, ვაღებთ ერთ-ერთ უჯრას და ამოგვაქვს ოქროს ფული, რას უდრის ალბათობა იმისა, რომ ეს მაგიდა არის B?

მ ი თ ი თ ე ბ ა : ისარგებლეთ ბაიესის ფორმულით.

პას. $P = 1/3$.

II თ ა ვ ი

განმეორებით ცდათა სქემა

§ 1. ალბათობათა ბინომიალური განაწილება და მისი თვისებები

ვთქვათ, ვატარებთ რაიმე ცდას და ვაკვირდებით ჩვენთვის სასურველი ხდომილობის მოხდენის თუ არმოხდენის ფაქტს. იგივე ცდას სავსებით იმავე პირობებში ვიმეორებთ საკმაოდ დიდი რაოდენობით. ვიგულისხმოთ, რომ ცალკეულ ცდათა შედეგები ერთმანეთისაგან სავსებით დამოუკიდებელია. ცალკეული ცდის დროს ჩვენთვის სასურველი A შემთხვევითი ხდომილობის ალბათობა აღვნიშნოთ p-თი, ე. ი. $P(A) = p$, მაშინ ჩვენთვის სასურველი A ხდომილობის არმოხდენის ალბათობა, რომელსაც აღვნიშნავთ q-თი, იქნება $1 - p$. ასე რომ, ყოველთვის $p + q = 1$.

დავუშვათ, ცდებს ვატარებთ n-ჯერ. ვთქვათ, პირველი ცდის დროს მოხდა A ხდომილობა, მეორე ცდის დროს—არ მოხდა, მესამე ცდის დროს—მოხდა, შემდეგი ცდის დროს—მოხდა და ა. შ. ბოლოს წინა

ცდის დროს მოხდა და ბოლო ცდის დროს არ მოხდა. მიღებული შედეგები დავაფიქსიროთ. მივიღებთ შემდეგ სქემას:

$$A\bar{A}AA\cdots A\bar{A}.$$

ახლა დავსვათ ასეთი კითხვა: რას უდრის ალბათობა იმისა, რომ n ჩატარებული ცდის დროს ჩვენთვის სასურველ A ხდომილობას ადგილი ექნება ზუსტად m-ჯერ და, მაშასადამე, ადგილი არ ექნება $n-m$-ჯერ. ეს ალბათობა აღვნიშნოთ $P_n(m)$-ით. ცხადია, m-ს შეუძლია მიიღოს ნებისმიერი მნიშვნელობა 0-დან n-მდე უკანასკნელის ჩათვლით.

ვიგულისხმოთ, რომ ჩვენ მიერ ზემოთ შედგენილი სქემა სწორედ ისეთია, სადაც ჩვენთვის სასურველ A ხდომილობას ზუსტად m-ჯერ აქვს ადგილი. ვიპოვოთ ამ სქემის ალბათობა.

რადგანაც პირობის თანახმად ცალკეულ ცდათა შედეგები ერთმანეთისაგან დამოუკიდებელია, ამიტომ მიღებული სქემის ალბათობა უდრის ცალკეული ცდის შედეგთა ალბათობების ნამრავლს, ე. ი.

$$P(A\bar{A}AA\cdots A\bar{A})=P(A)\,P(\bar{A})\,P(A)\,P(A)\cdots P(A)\,P(\bar{A}).$$

რადგანაც დაშვების თანახმად A ხდომილობას m-ჯერ ჰქონდა ადგილი, ამიტომ $\bar{A}$-ს $n-m$-ჯერ ექნება ადგილი და, მაშასადამე, მივიღებთ:

$$P(A\bar{A}AA\cdots A\bar{A})=p^m q^{n-m}.$$

მაგრამ ისეთი სქემა, სადაც n ცდაში A ხდომილობა m-ჯერ ხდება, ერთადერთი არ იქნება, რადგანაც მიღებულ სქემაში თუ შევცვლით, მაგალითად, მეორე და მესამე ცდის შედეგს, გვექნება აგრეთვე ისეთი სქემა, სადაც A ხდომილობა m-ჯერ ხდება და $n-m$-ჯერ არ ხდება. ამგვარად, დასმულ კითხვაზე რომ ვუპასუხოთ, ამისათვის საჭიროა ვიცოდეთ ყველა ასეთი სქემის რაოდენობა. ალბათობათა შეკრების შესახებ თეორემის დახმარებით ყველა ასეთ სქემათა ალბათობების ჯამი იქნება სწორედ n ცდაში A ხდომილობის m-ჯერ მოხდენის ალბათობა. ცალკეული სქემის ალბათობა $p^m q^{n-m}$-ის ტოლია. ვთქვათ, ასეთი სქემა არის სულ N, მაშინ მივიღებთ:

$$P_n(m)=N\,p^m q^{n-m},$$

სადაც N არის შემთხვევათა რიცხვი, რომელიც გამოსახავს იმას, თუ n ელემენტიდან რამდენნაირად შეგვიძლია შევარჩიოთ m ელემენტი, ე. ი.

$$N=C_n^m=\frac{n!}{m!(n-m)!}.$$

ამგვარად, $P_n(m)$-თვის საბოლოოდ გვექნება:

$$P_n(m)=\frac{n!}{m!\,(n-m)!}\,p^m q^{n-m}. \tag{1.1}$$

მიღებულ ბინომიალურ ფორმულას ბ ე რ ნ უ ლ ი ს ფორმულა ეწოდება. ის შესაძლებლობას გვაძლევს n ჩატარებული ცდის დროს განვსაზღვროთ ჩვენთვის სასურველი A ხდომილობის მოხდენათა რიცხვის ალბათობა.

მაგალითად, ვთქვათ ლითონის ფულს ვაგდებთ 5-ჯერ და გვაინტერესებს, თუ რას უდრის ალბათობა იმისა, რომ გერბს მივიღებთ ზუსტად 3-ჯერ.

ამ შემთხვევაში, როგორც ვიცით, $p=q=1/2$ და, მაშასადამე, ბერნულის ფორმულის თანახმად, დავწერთ

$$P_5(3)=\frac{5!}{3!\,2!}\,\frac{1}{2^5}=\frac{4\cdot 5}{2}\cdot\frac{1}{2^5}=\frac{5}{16}.$$

ამგვარად, მივიღეთ, რომ თუ ლითონის ფულს 5-ჯერ ავაგდებთ, მაშინ, იმის ალბათობა, რომ გერბს ზუსტად 3-ჯერ მივიღებთ, ყოფილა 5/16-ის ტოლი.

ნიუტონის ბინომის ფორმულის გამოყენებით დავწერთ:

$$(q+p)^n=C_n^0\,q^n+C_n^1\,q^{n-1}\,p+C_n^2\,q^{n-2}\,p^2+\dots+C_n^n\,P^n.$$

აქ მარჯვენა მხარის პირველი შესაკრები არის $P_n(0)$, მეორე — $P_n(1)$ და ა. შ. ბოლო შესაკრები $P_n(n)$-ს წარმოადგენს. ამგვარად,

$$(q+p)^n=P_n(0)+P_n(1)+\dots+P_n(n)=\sum_{m=0}^{n}P_n(m).$$

რადგანაც $p+q=1$, ამიტომ გვექნება:

$$\sum_{m=0}^{n}P_n(m)=1. \tag{1.2}$$

§ 2. $P_n(m)$-ის, როგორც m-ის ფუნქციის შესწავლა

როგორც ვხედავთ, $P_n(m)$-ის გამოსახულებაში ცვლადია მხოლოდ m. განვიხილოთ $P_n(m)$-ის მნიშვნელობა $m+1$-თვის, დავწერთ:

$$P_n(m+1)=\frac{n!}{(m+1)!\,(n-m-1)!}\,p^{m+1}\,q^{n-m-1}.$$

$P_n(m+1)$-ის მიღებული მნიშვნელობა გავყოთ $P_n(m)$-ზე. გვექნება:

$$\frac{P_n(m+1)}{P_n(m)}=\frac{(n-m)p}{(m+1)q}.$$

შევისწავლოთ, თუ როდის არის აღებული შეფარდება ერთზე მეტი, ტოლი ან ნაკლები. განვიხილოთ სამივე შემთხვევა ცალ-ცალკე:

1. $$\frac{P_n(m+1)}{P_n(m)}=\frac{(n-m)p}{(m+1)\,q}>1.$$

ამისათვის საჭიროა ადგილი ჰქონდეს უტოლობას:

$$(n-m)\,p>(m+1)\,q,$$

ან, რაც იგივეა,

$$m<np-q.$$

მაგრამ $P_n(m+1)>P_n(m)$ იმას ნიშნავს, რომ m-ის გაზრდით $P_n(m)$-ის მნიშვნელობა იზრდება, მაშასადამე, სანამ $m<np-q$ სიდიდეზე, მანამ $P_n(m)$-ის მნიშვნელობა იზრდება.

2. $$\frac{P_n(m+1)}{P_n(m)}=\frac{(n-m)p}{(m+1)\,q}=1.$$

ამისათვის საჭიროა ადგილი ჰქონდეს ტოლობას:

$$(n-m)\,p=(m+1)\,q$$

ან, რაც იგივეა,

$$m=np-q.$$

ამგვარად, $np-q$ გამოსახულება მთელი რიცხვი უნდა იყოს, რადგან m მხოლოდ მთელ რიცხვით მნიშვნელობებს ღებულობს.

3. $$\frac{P_n(m+1)}{P_n(m)}=\frac{(n-m)p}{(m+1)q}<1.$$

ამისათვის საჭიროა ადგილი ჰქონდეს უტოლობას:

$$(n-m)\,p<(m+1)\,q,$$

ან, რაც იგივეა,

$$m>np-q,$$

ე. ი. $P_n(m)$ მაშინ დაიწყებს კლებას, როდესაც m გახდება $np-q$ სიდიდეზე მეტი.

ამგვარად მივიღეთ, რომ:

$$\frac{P_n(m+1)}{P_n(m)}\begin{cases}>1, & \text{როცა } m<np-q\\ =1, & \text{როცა } m=np-q\\ <1, & \text{როცა } m>np-q.\end{cases}$$

m-ის იმ მნიშვნელობას, რომლისთვისაც $P_n(m)$ ყველაზე დიდ მნიშვნელობას იღებს, უ ა ლ ბ ა თ ე ს ი მ ნ ი შ ვ ნ ე ლ ო ბ ა ეწოდება.

m-ის უალბათესი მნიშვნელობა m_0-ით აღვნიშნოთ. ცხადია, რადგანაც $m<np-q$ უტოლობა სრულდება მანამ, სანამ $P_n(m)$-ის მნიშვნელობა იზრდება და ამიტომ m_0-ის მნიშვნელობა $np-q$ სიდიდეზე ნაკლები არ იქნება, ე. ი.

$$m_0 \geqslant np-q.$$

რადგანაც $m>np-q$ მნიშვნელობისათვის $P_n(m)$ იწყებს კლებას, ამიტომ m-ის უალბათესი m_0 მნიშვნელობა არ აღემატება $np-q+1$-ს, ე. ი.

$$m_0 \leqslant np-q+1.$$

ამგვარად, m_0-თვის გვაქვს სავსებით გარკვეული საზღვრები

$$np-q \leqslant m_0 \leqslant np-q+1,$$

ან, რაც იგივეა,

$$np-q \leqslant m_0 \leqslant np+p. \tag{2.1}$$

როგორც ვხედავთ, m_0-ის საზღვრებს შორის სხვაობა უდრის ერთს.

თუკი $np-q$ არ არის მთელი რიცხვი, მაშინ ცხადია, $np+p$-ც არ არის მთელი. ამიტომ m_0 იქნება ის ერთადერთი მთელი რიცხვი, რომელიც $np-q$ და $np+p$-ს შორისაა მოთავსებული.

ამგვარად, როდესაც $np-q$ არამთელი რიცხვია, m_0-თვის გვექნება ერთადერთი უალბათესი მნიშვნელობა და ის იქნება $np-q$ სიდიდის შემდეგ მარჯვნივ მდებარე პირველი მთელი რიცხვი.

თუ $np-q$ მთელი რიცხვია, მაშინ ცხადია, $np+p$-ც იქნება მთელი და როგორც $np-q$-ზე, ისე $np+p$-ზე m-ს ექნება უალბათესი მნიშვნელობანი. მართლაც, როდესაც $m=np-q$, მაშინ, როგორც ვნახეთ, $P_n(m+1)=P_n(m)$, ე. ი. როდესაც $np-q$ მთელი რიცხვია, მაშინ m-ს ექნება ორი უალბათესი მნიშვნელობა: ერთი იქნება თვით $np-q$, ხოლო მეორე $np+p$.

განვიხილოთ შემდეგი მაგალითი. ვთქვათ, ლითონის ფულს ვაგდებთ 9-ჯერ ზედიზედ. ვიპოვოთ რას უდრის ალბათობა იმისა, რომ გერბი მოვა 0-ჯერ, 1-ჯერ, 2-ჯერ და ა. შ. 9-ჯერ. ვიპოვოთ გერბის მოსვლათა უალბათესი რიცხვი და ავაგოთ სათანადო დიაგრამაც.

რადგანაც განხილულ მაგალითში $n=9$, ხოლო $p=q=\dfrac{1}{2}$, დავწერთ;

$$P_9(0)=\frac{9!}{0!\,9!}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^9.$$

0!-ს ღებულობენ ერთის ტოლად, ამიტომ გვექნება:

$$P_9(0)=\frac{1}{2^9}\simeq 0{,}00195,$$

ასევე:

$$P_9(1)=9\cdot\frac{1}{2^9}\simeq 0{,}01755,$$

$$P_9(2)=36\cdot\frac{1}{2^9}\simeq 0{,}07020,$$

$$P_9(3)=84\cdot\frac{1}{2^9}\simeq 0{,}16380,$$

$$P_9(4)=126\cdot\frac{1}{2^9}\simeq 0{,}24570,$$

$$P_9(5)=126\cdot\frac{1}{2^9}\simeq 0{,}24570,$$

$$P_9(6)=84\cdot\frac{1}{2^9}\simeq 0{,}16380,$$

$$P_9(7)=36\cdot\frac{1}{2^9}\simeq 0{,}07020,$$

$$P_9(8)=9\cdot\frac{1}{2^9}\simeq 0{,}01755,$$

$$P_9(9)=\frac{1}{2^9}\simeq 0{,}00195.$$

რომ ვიპოვოთ m-ის უალბათესი მნიშვნელობა, ამისათვის საჭიროა განვიხილოთ $np-q$ გამოსახულება. ჩვენი მაგალითისათვის

$$np-q=9\cdot\frac{1}{2}-\frac{1}{2}=4.$$

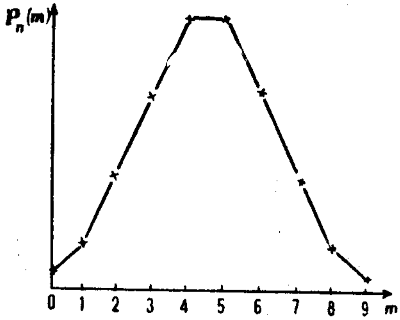
რადგანაც $np-q$ მთელი რიცხვია, გვექნება ორი უალბათესი მნიშვნელობა: ერთი იქნება თვით $np-q=4$-ს, ხოლო მეორე კი მას მიმატებული 1, ე. ი. 5.

მართლაც, როგორც გამოანგარიშებიდან ჩანს,

$$P_9(4)=P_9(5)=126\cdot\frac{1}{2^9}\simeq 0{,}24570.$$

თუ ჩატარებული გამოთვლების დახმარებით $P_n(m)$ ფუნქციის დიაგრამას ავაგებთ, მაშინ გვექნება შემდეგი სურათი (ნახ. 1).

როგორც № 1 ნახაზიდან ჩანს, დიაგრამას სიმეტრიული ხასიათი აქვს უდიდესი ორდინატების მიმართ. $P_9(0)$ და $P_9(9)$ მცირე სიდიდეებია.



ნახ. 1.

როდესაც n საკმაოდ დიდია, (2.1) ტოლობის ძალით შეგვიძლია შემდეგი მიახლოებითი ტოლობის დაწერა:

$$m_0 \simeq np.$$

მაგალითად, ვთქვათ $n=1000$, $p=q=1/2$, მაშინ

$$np-q=499{,}5$$

და შეგვიძლია დავწეროთ

$$m_0 \simeq 500.$$

m-ს ვუწოდებთ A ხდომილობის მოხდენის სიხშირეს, ხოლო $\frac{m}{n}$ შეფარდებას—ფარდობით სიხშირეს. n მუდმივი რიცხვია, m კი შემთხვევაზეა დამოკიდებული.

ცხადია, უალბათესი ფარდობითი სიხშირე იქნება $\frac{m_0}{n}$, რომელიც მოთავსებული იქნება შემდეგ საზღვრებში:

$$p-\frac{q}{n} \leqslant \frac{m_0}{n} \leqslant p+\frac{p}{n}.$$

თუ $n \to \infty$, მაშინ როგორც მარცხენა, ისე მარჯვენა მხარე p-კენ მიისწრაფვის და მივიღებთ:

$$\frac{m_0}{n} \to p.$$

ამგვარად, ადგილი აქვს შემდეგ დებულებას:

n-ის უსაზღვროდ ზრდის დროს ფარდობითი სიხშირის უალბათესი მნიშვნელობა p-კენ მიისწრაფვის.

ამ დებულებაში არც ერთი სიდიდე არ არის შემთხვევითი, ყველა სიდიდე მუდმივია, n ცდათა რიცხვი ჩვენზეა დამოკიდებული, p მოცემულია და m_0 არის m-ის უალბათესი მნიშვნელობა, რომელიც თეო-

რიულად გამოითვლება. ეს დებულება საინტერესოა მხოლოდ იმით, რომ იგი გვიჩვენებს

$$\frac{m_0}{n} \to p.$$

ყველა შესაძლებელი შემთხვევიდან უნდა ველოდოთ, სახელდობრ, ამას, მაგრამ ეს არ ნიშნავს, რომ აუცილებლად m_0 გვექნება. უბრალოდ, ამ შემთხვევის ფარდობითი ალბათობა ყველა დანარჩენზე მეტია, ხოლო აბსოლუტური მნიშვნელობით შესაძლებელია ძალიან პატარაც იყოს.

სავარჯიშო

1. ვატარებთ 8-ჯერ ურთიერთდამოუკიდებელ ცდებს. ცალკეული ცდის დროს ჩვენთვის სასურველი A ხდომილობის ალბათობა არის 0,5. რას უდრის ალბათობა იმისა, რომ 8 ჩატარებულ ცდაში A-ს ექნება ადგილი არა ნაკლებ 5-ჯერ?

პას. $P = P_8(5) + P_8(6) + P_8(7) + P_8(8) = \frac{93}{256}$.

2. ორი თანაბარი ძალის მოჭადრაკე ეთამაშება ერთმანეთს. რა უფრო შესაძლებელია: ერთი მეორეს 4 პარტიიდან მოუგებს 3-ს, თუ 8-დან—5-ს?

პას. $P_4(3) = \frac{1}{4}$, $P_8(5) = \frac{7}{32}$; $P_4(3) > P_8(5)$.

3. ორი თანაბარი ძალის მოჭადრაკე ეთამაშება ერთმანეთს. რა უფრო შესაძლებელია: ერთი მეორეს 4 პარტიიდან მოუგებს არანაკლებ 3-ს, თუ 8-დან—არა ნაკლებ 5-ს?

პას. $P_4(3) + P_4(4) = \frac{5}{16}$; $P_8(5) + P_8(6) + P_8(7) + P_8(8) = \frac{93}{256}$

$$5/16 < 93/256.$$

4. ოჯახში 5 ბავშვია. იპოვეთ ალბათობა იმისა, რომ მათ შორის ა) ორი ბიჭია; ბ) არა უმეტეს ორი ბიჭია, გ) ბიჭები ორზე მეტია, დ) ბიჭებია არანაკლებ 2-ზე და არა მეტი—3-ზე. ვიგულისხმოთ, რომ ბიჭის დაბადების ალბათობა არის 0,51.

პას. მიმდევრობით, მიახლოებით გვექნება: 0,31; 0,48; 0,52; 0,62.

5. AB მონაკვეთი, რომლის სიგრძეა 15 სმ, C წერტილით იყოფა ფარდობით $2:1$. ამ მონაკვეთზე შემთხვევით ვაგდებთ 4 წერტილს.

იპოვეთ ალბათობა იმისა, რომ ორი მათგანი მოხვდება C წერტილის მარცხნივ და 2 კი — მარჯვნივ. იგულისხმება, რომ წერტილის მოხვედრა მონაკვეთში მონაკვეთის სიგრძის პროპორციულია და მისი მდებარეობისაგან დამოუკიდებელია.

პას. $P_4(2)=C_4^2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2=8/27.$

6. საბითუმო ბაზა ამარაგებს 20 მაღაზიას, ყოველი მათგანისაგან სხვებისაგან დამოუკიდებლად მოსალოდნელია შემდეგი დღისათვის განაცხადის მოცემა 0,5-ის ტოლი ალბათობით. იპოვეთ დღის განმავლობაში შეკვეთათა რიცხვის უალბათესი მნიშვნელობა და ალბათობა ამ უალბათესი მნიშვნელობისა.

პას. $m_0=10$; $P_{20}(10)=\frac{4125}{1048576}$.

7. B ხდომილობას მაშინ ექნება ადგილი, თუ A ხდომილობა მოხდება სამჯერ მაინც. იპოვეთ B-ს მოხდენის ალბათობა, თუ A-ს მოხდენის ალბათობა ცალკეული ცდის დროს არის 0,3 და ჩატარებულია 5 ურთიერთდამოუკიდებელი ცდა.

პას. $P=0{,}163$.

8. ოთხი ურთიერთდამოუკიდებელი ცდის დროს A ხდომილობის ერთხელ მაინც მოხდენის ალბათობა არის 0,59. რას უდრის ალბათობა A ხდომილობის მოხდენის ცალკეული ცდის დროს, თუ ყოველი ცდის დროს ეს ალბათობა ერთი და იგივეა?

პას. $P\simeq 0{,}2$.

9. პირველი, მეორე და მესამე ნათურის გადაწვის ალბათობებია სათანადოდ: 0,1; 0,2; 0,3. ხელსაწყოს მწყობრიდან გამოსვლა ერთი, ორი ან სამი ნათურის გადაწვის გამო არის შესაბამისად 0,25; 0,6 და 0,9. იპოვეთ ხელსაწყოს მწყობრიდან გამოსვლის ალბათობა.

პას. $P=0{,}159$.

10. კალათბურთელისათვის კალათში ბურთის მოთავსების ალბათობა ცალკეული გასროლისას არის 0,4. მან 10-ჯერ ისროლა. იპოვეთ კალათში ჩაგდებულ ბურთთა უალბათესი რიცხვი და მონახეთ ამ რიცხვის შესაბამისი ალბათობა.

პას. $m_0=4$; $P_{10}(4)=0{,}251$.

11. იანგარიშეთ ალბათობა იმისა, რომ 10-ბავშვიან ოჯახში ქალ-ვაჟთა რაოდენობა თანაბარი იქნება, თუ ვაჟიშვილისა და ქალიშვილის დაბადება ტოლალბათია.

პას. $P_{10}(5)\simeq 0{,}25$.

12. ვთქვათ, ქვემეხით 14-ჯერ ვისროლეთ სამიზნეში. ცალკეული გასროლისას მოხვედრის ალბათობა 0,2-ის ტოლია.

განსაზღვრეთ მოხვედრათა უალბათესი რიცხვი და იპოვეთ მისი ალბათობა.

პას. $m_0=2$ და $m_0=3$.

$P_{14}(2)=P_{14}(3)=0,25$.

13. ზოგიერთ მრავალშვილიან ოჯახში 7 ბავშვია. რას უდრის ალბათობა იმისა, რომ ასეთ ოჯახში ქალიშვილი იქნება არა უმეტეს სამისა? იგულისხმეთ, რომ ვაჟისა და ქალიშვილის დაბადების ალბათობა ტოლია.

პას. $P=1/2$.

14. რაიმე გენერალურ ერთობლიობაში წუნდებული ნაწარმის რაოდენობა 3%-ის ტოლია. გენერალური ერთობლიობის მოცულობაა 800.

რა იქნება წუნდებულთა უალბათესი რიცხვი და რას უდრის მისი ალბათობა?

პას. $m_0=24$, $P_{800}(24)\simeq 0,083$.

III თავი

ლაპლასისა და პუასონის ფორმულა

§ 1. ალბათობის მრუდი

წინა თავში განვიხილეთ ალბათობათა ბინომიალური განაწილების ფორმულა

$$P_n(m)=\frac{n!}{m!\,(n-m)!}\,p^m\,q^{n-m} \tag{1.1}$$

და ავაგეთ სათანადო დიაგრამაც.

ამოცანა იმაში მდგომარეობს, რომ მიღებული ტეხილი წირები ისეთი მრუდი წირებით შევცვალოთ, რომლებიც მიახლოებით ასახავს მათ. ასეთი ტიპის მრუდებს ალბათობის მრუდებს უწოდებენ. გამოვიყვანოთ ალბათობის მრუდის განტოლება. ამისათვის მოვიქცეთ შემდეგნაირად: ავიღოთ დეკარტის მართკუთხა კოორდინატთა სისტემა. აბსცისათა ღერძზე დავნიშნოთ რამდენიმე წერტილი, რომელთა აბსცისები განვსაზღვროთ ტოლობიდან

$$x=m-np. \tag{1.2}$$

დანიშნული წერტილებიდან აბსცისათა ღერძისადმი აღვმართოთ პერპენდიკულარები, რომლებზედაც მოვზომოთ

$$y = P_n(m)$$

სიდიდეები, მოზომვის წერტილები შევაერთოთ ტეხილი ხაზით, მივიღებთ ალბათობის განაწილების ტეხილ მრუდს, რომელსაც პოლიგონს უწოდებენ.

შევადგინოთ ისეთი მრუდის განტოლება, რომელიც რაც შეიძლება ახლოს გაივლის ჩვენს პოლიგონურ ტეხილთან. ამ მიზნით განვიხილოთ აღებული ტეხილის ორი მეზობელი წვეროს კოორდინატები. ვთქვათ, ერთ-ერთი წვეროს კოორდინატებია (x, y); (1.1) და (1.2) ტოლობათა ძალით გვექნება:

$$x = m - np,$$

$$y = P_n(m) = \frac{n!}{m!\,(n-m)!}\, p^m q^{n-m} =$$

$$= \frac{n(n-1)\dots(n-m+1)}{1 \cdot 2 \dots m}\, p^m\, q^{n-m}. \tag{1.3}$$

მეორე მეზობელი წვეროს კოორდინატების განსასაზღვრავად (1.1) და (1.2) ტოლობაში m შევცვალოთ $m+1$-ით, მივიღებთ:

$$x + \Delta x = m + 1 - np, \tag{1.4}$$

$$y + \Delta y = P_n(m+1) = \frac{n!}{(m+1)!\,(n-m-1)!}\, p^{m+1}\, q^{n-m-1} =$$

$$= \frac{n(n-1)\dots(n-m+1)(n-m)}{1 \cdot 2 \dots m(m+1)}\, p^{m+1}\, q^{n-m-1}. \tag{1.5}$$

(1.4) ტოლობიდან მოვნახავთ Δx-ს:

$$\Delta x = m + 1 - np - x = m + 1 - np - m + np = 1.$$

ახლა გამოვთვალოთ ფარდობა $\dfrac{y+\Delta y}{y}$ ან, რაც იგივეა, გამოვთვალოთ ფარდობა

$$\frac{P_n(m+1)}{P_n(m)}.$$

თუ ამ ფარდობის მრიცხველსა და მნიშვნელში შევიტანთ სათანადო მნიშვნელობებს, მაშინ გამარტივების შემდეგ მივიღებთ:

$$\frac{y+\Delta y}{y} = \frac{P_n(m+1)}{P_n(m)} = \frac{(n-m)p}{(m+1)q},$$

საიდანაც

$$\frac{\Delta y}{y}=\frac{(n-m)p}{(m+1)q}-1=\frac{np-mp-mq-q}{(m+1)q}=\frac{np-q-m(p+q)}{(m+1)q}.$$

რადგანაც $p+q=1$, გვექნება:

$$\frac{\Delta y}{y}=\frac{np-q-m}{(m+1)q}.$$

რადგან $\Delta x=1$, შეგვიძლია დავწეროთ:

$$\frac{\Delta y}{y\Delta x}=\frac{np-q-m}{(m+1)q}=\frac{-q-(m-np)}{(m+1)q}. \tag{1.6}$$

(1.2) ტოლობიდან შეგვიძლია m-ის განსაზღვრა:

$$m=x+np.$$

(1.6) ტოლობის მარჯვენა მხარის მრიცხველში $m-np$ შევცვალოთ x-ით, ხოლო მნიშვნელში m-ის ნაცვლად შევიტანოთ $x+np$. გვექნება:

$$\frac{\Delta y}{y\Delta x}=\frac{-q-x}{(np+x+1)q}.$$

გარდავქმნათ ეს ტოლობა. ამისათვის ტოლობის მარჯვენა მხარის მრიცხველი და მნიშვნელი გავყოთ n-ზე, მივიღებთ:

$$\frac{1}{y}\frac{\Delta y}{\Delta x}=\frac{-\frac{q}{n}-\frac{x}{n}}{pq+\frac{xq}{n}+\frac{q}{n}}. \tag{1.7}$$

როდესაც n საკმაოდ დიდია, მაშინ, ცხადია, $\frac{xq}{n}$ და $\frac{q}{n}$ საკმაოდ მცირე სიდიდეები იქნება და მიახლოებით შეგვიძლია დავწეროთ:

$$\frac{1}{y}\frac{\Delta y}{\Delta x}\simeq\frac{-\frac{x}{n}}{pq}=-\frac{x}{npq}, \tag{1.8}$$

ვიგულისხმოთ, რომ საძიებელი მრუდი უწყვეტია:

$\frac{\Delta y}{\Delta x}$ შევცვალოთ $\frac{dy}{dx}$-ით,

გვექნება:

$$\frac{1}{y}\frac{dy}{dx}=-\frac{x}{npq}. \tag{1.9}$$

მიღებული დიფერენციალური განტოლება წარმოვადგინოთ შემდეგნაირად:

$$\frac{dy}{y}=-\frac{xdx}{npq},$$

საიდანაც ინტეგრების შემდეგ გვექნება:

$$\ln y=-\frac{x^2}{2npq}+\ln c.$$

ეს უკანასკნელი შეგვიძლია წარმოვადგინოთ შემდეგნაირად:

$$\ln y=\ln e^{-\frac{x^2}{2npq}}+\ln c,$$

საიდანაც

$$y=ce^{-\frac{x^2}{2npq}}. \tag{1.10}$$

შემოვიღოთ აღნიშვნა

$$h=\frac{1}{\sqrt{2npq}}, \tag{1.11}$$

გვექნება:

$$y=ce^{-h^2x^2}. \tag{1.12}$$

მიღებული განტოლება საძიებელი ალბათობის მრუდის განტოლებაა. საჭიროა მხოლოდ c მუდმივის განსაზღვრა.

როგორც უკვე აღვნიშნეთ, პოლიგონურ ტეხილსა და აბსცისათა ღერძს შორის მოთავსებული ფართობი ერთის ტოლია. ამის საფუძველზე შეგვიძლია მივიღოთ, რომ ალბათობის მრუდსა და აბსცისათა ღერძს შორის მოთავსებული ფართობიც ერთის ტოლია.

ჩვენს შემთხვევაში, რადგან ტეხილს ორივე მხრივ უსასრულოდ ვაგრძელებთ, ამიტომ გვექნება:

$$c\int_{-\infty}^{\infty} e^{-h^2x^2}\,dx=1.$$

შემოვიღოთ აღნიშვნა

$$hx=z,$$

საიდანაც

$$dx=\frac{dz}{h}.$$

მაშასადამე,

$$c\int_{-\infty}^{\infty} e^{-h^2x^2}\,dx=\frac{c}{h}\int_{-\infty}^{\infty} e^{-z^2}dz=1.$$

ცნობილია, რომ

$$\int_{-\infty}^{\infty} e^{-z^2}dz=\sqrt{\pi}.$$

ამიტომ

$$\frac{c}{h}\sqrt{\pi}=1,$$

საიდანაც

$$c=\frac{h}{\sqrt{\pi}}.$$

თუ ამას შევიტანთ y-ის გამოსახულებაში, გვექნება:

$$y=\frac{h}{\sqrt{\pi}}e^{-h^2x^2} \tag{1.13}$$

(1.13) განტოლებით გამოსახულ მრუდს ეწოდება ალბათობათა განაწილების მრუდი. მას აგრეთვე ნორმალურ მრუდსაც უწოდებენ.

(1.13) გამოსახულება ლაპლასის მიერ იყო მიღებული და მას ლაპლასის ფორმულასაც უწოდებენ. თუ ამ ფორმულაში h-ს შევცვლით თავისი მნიშვნელობით, მივიღებთ:

$$y=\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x^2}{2npq}}. \tag{1.14}$$

ეს არის ალბათობის განაწილების მიახლოებითი ფორმულა. ამგვარად გვაქვს:

$$y=P_n(m)\simeq\frac{h}{\sqrt{\pi}}e^{-h^2x^2}. \tag{1.15}$$

უნდა შევნიშნოთ, რომ ლაპლასის ფორმულა მით უფრო ზუსტი იქნება, რაც მეტი იქნება n-ის მნიშვნელობა, რადგანაც ლაპლასის ფორმულა იმ დაშვებით იყო მიღებული, რომ n დიდი რიცხვია.

ლაპლასის ფორმულას შეიძლება უფრო მარტივი სახე მიეცეს. (1.14) ფორმულიდან ჩანს, რომ ალბათობის ყოველგვარი განაწილება დამოკიდებულია $\sqrt{npq}$ გამოსახულებაზე. ამის საფუძველზე შემოვიყვანოთ ახალი ცვლადი t, რომელიც x-თან დაკავშირებულია შემდეგნაირად:

$$x=t\sqrt{npq}.$$

მაშინ ლაპლასის ფორმულა მიიღებს შემდეგ სახეს:

$$y=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{npq}}e^{-t^2/2}, \tag{1.16}$$

სადაც

$$t=\frac{m-np}{\sqrt{npq}}.$$

ლაპლასის ფორმულის დახმარებით შეიძლება m-ის უალბათესი მნიშვნელობის $\tilde{m}_0$-ის მონახვა. ამისათვის საჭიროა ვიპოვოთ y-ის უდიდესი მნიშვნელობა.

უდიდეს y-ს მაშინ მივიღებთ, თუ (1.14) ფორმულაში ჩავსვამთ $x=0$-ს.

ამგვარად, დავწერთ:

$$m_0=\frac{1}{\sqrt{2\pi npq}}.$$

გავითვალისწინოთ, რომ

$$\frac{1}{\sqrt{\pi}}\simeq 0{,}56419,$$

მაშინ m_0-თვის გვექნება:

$$m_0=\frac{0{,}56419}{\sqrt{2npq}}. \tag{1.17}$$

ლაპლასის ფორმულის გამოყენებით შეიძლება აგრეთვე ალბათობის გამოთვლა. ასეთ შემთხვევებში სარგებლობენ ცხრილით, რომელიც შედგენილია

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-t^2/2} \tag{1.18}$$

ფუნქციისათვის.

$\varphi(t)$ ფუნქციის საშუალებით ლაპლასის ფორმულა დაიწერება შემდეგნაირად:

$$y=\frac{1}{\sqrt{npq}}\varphi(t).$$

ამრიგად, რომ გამოვთვალოთ y, საჭიროა ცხრილში ვიპოვოთ $\varphi(t)$

$$t=\frac{m-np}{\sqrt{npq}}$$

მნიშვნელობისათვის და შემდეგ ის გავამრავლოთ

$$\frac{1}{\sqrt{npq}}\text{-ზე.}$$

ამით ვიპოვით საძიებელ ალბათობას.

როგორც ცნობილია, ბურთის ამოღებასთან დაკავშირებით არსებობს ორი ტიპის ამოცანა: 1. როდესაც ცდათა სერიის ჩატარებისას

ამოღებული ბურთი უკანვე ბრუნდება და 2. როდესაც ამოღებული ბურთი უკან არ ბრუნდება. პირველ შემთხვევაში საქმე გვაქვს „ბურთის დაბრუნების სქემასთან“, ხოლო მეორე შემთხვევაში — „ბურთის არდაბრუნების სქემასთან“. ეს უკანასკნელი სქემა, რასაკვირველია, უფრო რთულია, რადგანაც ამ შემთხვევაში ბურთის ამოღება დამოკიდებულია იმაზეც, თუ ამოღებული ბურთებიდან რამდენია ჩვენთვის სასურველი.

ჯერ სხვა გზით ამოვხსნათ მაგალითი, რომელიც „ბურთების არდაბრუნების სქემას“ ეკუთვნის; ამ მაგალითის გადასაწყვეტად გამოვიყენოთ ლაპლასის ფორმულა.

მ ა გ ა ლ ი თ ი. ვთქვათ, ყუთში N ბურთია, რომელთაგან M თეთრია, ხოლო დანარჩენი შავი. ყუთიდან ამოვიღოთ n ბურთი ისე, რომ უკანვე არ დავაბრუნოთ. რას უდრის ალბათობა იმისა, რომ n ამოღებული ბურთიდან m იქნება თეთრი?

ამ ამოცანის ამოსახსნელად ვიპოვოთ როგორც ყველა შესაძლებელ შემთხვევათა რიცხვი, ისე ყველა ხელშემწყობ შემთხვევათა რიცხვი. ცხადია, ყველა შესაძლებელ შემთხვევათა რიცხვი იქნება N ელემენტიდან n-ად ჯუფთებათა რიცხვი, ე. ი. C_N^n.

ახლა ვიპოვოთ ხელშემწყობ შემთხვევათა რიცხვიც.

თეთრი ბურთის ამოღების ხელშემწყობი შემთხვევები იქნება ისეთი n ბურთიანი ჯგუფთა რიცხვები, რომლებშიც m თეთრი და $n-m$ შავი ბურთია; როგორც ვიცით, სულ M თეთრი ბურთია, ამიტომ m თეთრი ბურთი, ცალკე აღებული, წარმოადგენს ჯუფთებათა რიცხვს M ელემენტიდან m-ით, ე. ი. C_M^m; ასეთნაირად ვიპოვით იმ ჯგუფთა რიცხვს, რომლებიც შედგენილი იქნება შავი ბურთებისაგან. შავი ბურთების რაოდენობა მთლიანად არის $N-M$, შერჩეულ n ჯგუფში შავი ბურთი უნდა იყოს $n-m$, ამიტომ $n-m$ შავი ბურთი ცალკე აღებული წარმოადგენს ჯუფთებათა რიცხვს $N-M$ ელემენტიდან $n-m$-ად, ე. ი.

$$C_{N-M}^{n-m}.$$

ჯგუფთა რიცხვი, რომლებიც m თეთრ და $n-m$ შავ ბურთებს შეიცავს, ცხადია, იქნება შემდეგი ნამრავლი:

$$C_M^m\, C_{N-M}^{n-m};$$

თუ საძიებელ ალბათობას აღვნიშნავთ $P_{N,M,n,m}$-ით, დავწერთ:

$$P_{N,M,n,m}=\frac{C_M^m\, C_{N-M}^{n-m}}{C_N^n}.$$

ამ შემთხვევაში ლაპლასის ფორმულა ისეთნაირად მიიღება, როგორც უკვე იყო მითითებული. მართლაც, თუ აქ განვიხილავთ ფარდობას

$$\frac{P_{N,M,n,m+1}}{P_{N,M,n,m}}$$

და შემოვიღებთ აღნიშვნებს:

$$\frac{M}{N}=p, \quad 1-\frac{M}{N}=1-p=q,$$

ანალოგიური გამოთვლებით $P_{N,M,n,m}$-თვის მივიღებთ:

$$P_{N,M,n,m}=\frac{1}{\sqrt{2\pi npq\,\frac{N-n}{N}}}e^{-\frac{Nx^2}{2npq(N-n)}}. \qquad (1.19)$$

მაგალითი. ყუთში 30 ბურთია, რომელთაგან 10 თეთრია და დანარჩენი შავი. რას უდრის ალბათობა იმისა, რომ ბურთების 20-ჯერ ამოღების შემთხვევაში თეთრი ბურთი ამოღებული იქნება 3-ჯერ. ყოველი ცალკეული ამოღების შემდეგ ბურთი უკანვე ბრუნდება ყუთში.

ამოხსნა: აქ $p=\frac{1}{3}$, $q=\frac{2}{3}$, $n=20$ და $m=3$.

ვისარგებლოთ (1.17) ფორმულით. ამისათვის ჯერ ვიპოვოთ t-ს მნიშვნელობა:

$$t=\frac{m-np}{\sqrt{npq}}=\frac{3-20\cdot\frac{1}{3}}{\sqrt{20\cdot\frac{1}{3}\cdot\frac{2}{3}}}=-\sqrt{3{,}025}=-1{,}7392.$$

ახლა უნდა ვიპოვოთ φ (1,7392) ან, რაც იგივეა, φ (—1,7392). რადგანაც, როგორც ფორმულიდან ჩანს, $\varphi(t)=\varphi(-t)$.

წიგნის ბოლოში დართული I ცხრილის დახმარებით ვნახავთ

$$\varphi(1{,}73)=0{,}0893.$$

გვაინტერესებს φ (1,7392). ცხრილში 0,0893-ის შემდეგი მნიშვნელობაა 0,0878. მაშასადამე, სხვაობა

$$d=0{,}0878-0{,}0893=-0{,}0015.$$

გვაქვს

$$0{,}01 \;.\;.\;.\;.\;.\;.\;.\;.\;.\; 0{,}0015,$$

$$0{,}0092 \;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\; x,$$

საიდანაც

$$x=\frac{-0,0092\cdot 0,0015}{0,01}=-0,0014.$$

მაშასადამე,

$$\varphi(1,7392)=0,0893-0,0014=0,0879.$$

ახლა გამოვთვალოთ

$$\frac{1}{\sqrt{npq}}.$$

გვაქვს

$$\frac{1}{\sqrt{npq}}=\frac{1}{\sqrt{20\cdot\frac{1}{3}\cdot\frac{2}{3}}}\simeq 0,4743.$$

ბოლოს, (1.14) ფორმულის გამოყენებით დავწერთ:

$$y=0,4743\cdot 0,0879=0,0417.$$

თუ გამოვთვლით იგივე ალბათობას ზუსტი ბინომიალური ფორმულის დახმარებით, მივიღებთ:

$$P_{20}(3)=\frac{20!}{3!\ 17!}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{17}\simeq 0,0429.$$

როგორც ვხედავთ, ლაპლასის მიახლოებითი ფორმულით მიღებული შედეგი მცირედ განსხვავდება ზუსტი ფორმულით მიღებული შედეგისაგან. რა თქმა უნდა, n-ის უფრო დიდი მნიშვნელობისათვის ლაპლასის ფორმულით მიღებული შედეგი უფრო ზუსტი იქნებოდა. ამიტომ პრაქტიკაში ალბათობის განსასაზღვრავად ხშირად ლაპლასის ფორმულით სარგებლობენ.

მაგალითი. ხდომილობის ალბათობა არის 0,3, მონახეთ 20 ცდის შემთხვევაში ჩვენთვის სასურველი ხდომილობის მოხდენათა უალბათესი რიცხვი m_0 და აგრეთვე იპოვეთ ამ უალბათესი რიცხვის ალბათობა.

ა მ ო ხ ს ნ ა: აქ $n=20$; $p=0,3$; $q=0,7$.

რადგან m_0 მოთავსებულია შემდეგ საზღვრებში:

$$np-q\leqslant m_0\leqslant np+p,$$

ამიტომ გვაქვს:

$$5,3\leqslant m_0\leqslant 6,3.$$

რადგანაც $np-q$ არამთელი რიცხვია, ამიტომ ერთადერთი უალბათესი რიცხვი გვექნება და m_0-ის ეს მნიშვნელობა იქნება 6. გამოვთვალოთ $P_{20}(6)$ ზუსტი ფორმულით:

$$P_{20}(6)=\frac{20!}{6!\ 14!}\left(\frac{3}{10}\right)^6\left(\frac{7}{10}\right)^{14}\approx 0,192.$$

ეს უკანასკნელი ალბათობა ვიპოვოთ (1.17) ფორმულით. გვექნება:

$$m_0=\frac{0,56419}{\sqrt{2npq}}=\frac{0,56419}{\sqrt{\frac{2\cdot 20\cdot 3\cdot 7}{10\cdot 10}}}\approx 0,194.$$

როგორც ვხედავთ, მიღებული შედეგი მცირედ განსხვავდება ზუსტი ფორმულით მიღებული შედეგისაგან.

§ 2. ლაპლასის თეორემა

ლაპლასის ფორმულა გვაძლევს ჩვენთვის სასურველი ხდომილობის m-ჯერ მოხდენის ალბათობის მიახლოებით მნიშვნელობას, როდესაც ცდას ვატარებთ n-ჯერ.

ახლა ჩვენი მიზანია, ვიპოვოთ ალბათობა იმისა, რომ n ჩატარებული ცდის დროს ჩვენთვის სასურველი ხდომილობა მოხდება არა ნაკლებ a-ჯერ და არა უმეტეს b-ჯერ; ეს იმას ნიშნავს, რომ n ჩატარებულ ცდაში ჩვენთვის სასურველი ხდომილობა მოხდება a-ჯერ ან $a+1$-ჯერ და ა. შ., ან b-ჯერ.

თუ საძიებელ ალბათობას p-თი აღვნიშნავთ, მაშინ ალბათობათა შეკრების თეორემის თანახმად დავწერთ:

$$p=P_n(a)+P_n(a+1)+\dots+P_n(b)=\sum_{m=a}^{b}P_n(m). \qquad (2.1)$$

რადგან

$$x=m-np,$$

ამიტომ m-ის ცვლილებას (a, b) შუალედში ეთანადება x-ის ცვლილება $(-t\sqrt{npq},\ t\sqrt{npq})$ შუალედში, ე. ი. m-ის ყოველ მნიშვნელობას (a, b) შუალედში შეესაბამება x-ის გარკვეული მნიშვნელობა $(-t\sqrt{npq},\ t\sqrt{npq})$ შუალედში.

აქედან ცხადია, რომ x-ის და m-ის მნიშვნელობათა ალბათობანი ერთი და იგივეა და არის ალბათობა იმისა, რომ n ცდაში x მიიღებს რომელიმე მნიშვნელობას $(-t\sqrt{npq},\ t\sqrt{npq})$ შუალედში. (2.1) ფორმულაში შევიტანოთ $P_n(m)$-ის მნიშვნელობა და a და b შევცვალოთ $-t\sqrt{npq}$, $t\sqrt{npq}$-თი, მაშინ (2.1) ფორმულა გადაიწერება შემდეგნაირად:

$$p=\sum_{x=-t\sqrt{npq}}^{t\sqrt{npq}}\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x^2}{2npq}}.$$

უკანასკნელი გამოსახულება ზღვარში მოგვცემს:

$$p=\frac{1}{\sqrt{2\pi npq}}\int_{-t\sqrt{npq}}^{t\sqrt{npq}} e^{-\frac{x^2}{2npq}}\,dx.$$

ეს ტოლობა სამართლიანია n-ის საკმაოდ დიდი მნიშვნელობისათვის. რადგანაც მიღებულ გამოსახულებაში ინტეგრალქვეშა ფუნქცია ლუწია, ამიტომ ის შეგვიძლია გადავწეროთ ასე:

$$p=\frac{2}{\sqrt{2\pi npq}}\int_{0}^{t\sqrt{npq}} e^{-\frac{x^2}{2npq}}\,dx.$$

შემოვიღოთ ჩასმა:

$$x=z\sqrt{npq}.$$

მაშინ

$$dx=\sqrt{npq}\,dz,$$

$$\frac{x^2}{2npq}=\frac{z^2}{2}.$$

აქ ინტეგრების ახალი საზღვრები იქნება 0 და t და, ამგვარად, p-თვის მივიღებთ შემდეგ გამოსახულებას:

$$p=\frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{z^2}{2}}\,dz. \tag{2.2}$$

ამრიგად, (2.2) ფორმულა არის ალბათობა იმისა, რომ n ჩატარებულ ცდაში x მიიღებს ერთ-ერთ რომელიმე მნიშვნელობას, რომელიც მოთავსებული იქნება

$\left(-t\sqrt{npq},\ t\sqrt{npq}\right)$ შუალედში.

მაგრამ x-თვის რომელიმე მნიშვნელობის მიღება ზემოგანხილულ შუალედში იმის ტოლფასია, რომ $\frac{x}{n}$ მიიღებს რომელიმე მნიშვნელობას

$$\left(-t\sqrt{\frac{pq}{n}},\ t\sqrt{\frac{pq}{n}}\right)$$

შუალედში.

მაგრამ

$$\frac{x}{n}=\frac{m-np}{n}=\frac{m}{n}-p,$$

სადაც

$$\frac{m}{n} - p$$

ფარდობით სიხშირესა და ცალკეული ხდომილობის ალბათობას შორის სხვაობას წარმოადგენს.

ამ უკანასკნელის საფუძველზე შეგვიძლია ვთქვათ, რომ (2.2) ინტეგრალი გამოსახავს ალბათობას იმისა, რომ

$$\frac{m}{n} - p$$

სხვაობას ექნება ერთ-ერთი ისეთი მნიშვნელობა, რომელიც მოთავსებული იქნება

$$\left(-t\sqrt{\frac{pq}{n}},\ t\sqrt{\frac{pq}{n}}\right)$$

შუალედში.

ამგვარად, დავამტკიცეთ ლაპლასის თეორემა, რომელიც მდგომარეობს შემდეგში: ალბათობა იმისა, რომ საკმაოდ დიდი რაოდენობით ჩატარებული ცდის დროს ფარდობით სიხშირესა და ცალკეული ხდომილობის ალბათობას შორის სხვაობა მოთავსებული იქნება $\left(-t\sqrt{\frac{pq}{n}},\ t\sqrt{\frac{pq}{n}}\right)$ შუალედში, გამოისახება ინტეგრალით:

$$\frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz.$$

სადაც t გარკვეული მოცემული რიცხვია. ამგვარად

$$P\left\{-t\sqrt{\frac{pq}{n}} < \frac{m}{n} - p < t\sqrt{\frac{pq}{n}}\right\} = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz. \qquad (2.3)$$

შემოვიღოთ აღნიშვნა:

$$\varphi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz.$$

„ბურთების არდაბრუნების სქემის“ შემთხვევაში ლაპლასის თეორემა წინა პარაგრაფის (1.18) ფორმულის გამოყენებით დაიწერება შემდეგნაირად:

$$P\left\{-t\cdot\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}<\frac{m}{n}-p<t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}\right\}=$$

$$=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{z^2}{2}}dz.$$

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx \qquad (2.4)$$

გამოსახულებას ლაპლასის ფუნქციას უწოდებენ. $\Phi(t)$-ს გამოსახულებაში ზედა საზღვარი t შეიძლება იყოს ნებისმიერი ნამდვილი რიცხვი.

ცხადია, რომ

$$\Phi(-t)=\frac{1}{\sqrt{2\pi}}\int_0^{-t} e^{-\frac{x^2}{2}}dx=-\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx=-\Phi(t),$$

ე. ი.

$$\Phi(-t)=-\Phi(t).$$

ეს იმას ნიშნავს, რომ, თუ ცხრილში არგუმენტის დადებით ნიშანს შევცვლით მინუსით, მივიღებთ ლაპლასის ფუნქციის მნიშვნელობებს უარყოფითი t-თვის.

როცა $t=0$,

$$\Phi(0)=0.$$

$\Phi(t)$ ზრდადი ფუნქციაა. მართლაც ვთქვათ, $t_1<t_2$, მაშინ

$$\Phi(t_2)=\frac{1}{\sqrt{2\pi}}\int_0^{t_2} e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_0^{t_1} e^{-\frac{x^2}{2}}dx+$$

$$+\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{x^2}{2}}dx=\Phi(t_1)+\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{x^2}{2}}dx.$$

რადგანაც $\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{x^2}{2}}dx>0$ (თუ ინტეგრალქვეშა ფუნქცია დადებითია და ინტეგრება ხდება დადებითი მიმართულებით, მაშინ ინტეგრალი დადებითია). ამიტომ

$$\Phi(t_2)>\Phi(t_1).$$

$\Phi(t)$ ფუნქციის ზღვარი, როცა $t \to \infty$, იქნება $^1/_2$, ე. ი.

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx = {}^1/_2.$$

მართლაც, როგორც ცნობილია,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1,$$

ხოლო

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx.$$

მაგრამ, რადგანაც

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

მრუდი სიმეტრიულია, დავწერთ

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx = {}^1/_2.$$

$\Phi(t)$ ფუნქცია t-ს არადიდი დადებითი მნიშვნელობებისათვის ძალიან ახლოსაა 0, 5-თან (ცხრილი 2).

ც ხ რ ი ლ ი 2

t	4	5	6	10
$\Phi(t)$	0,49997	0,49999	0,49999	0,49999

ამიტომ $|t|>6$ მნიშვნელობა შეიძლება პრაქტიკულად შეუძლებლად ჩაითვალოს.

ლაპლასის თეორემიდან დავწერთ:

$$P\left(t_1 \sqrt{\frac{pq}{n}} < \frac{m}{n} - p < t_2 \sqrt{\frac{pq}{n}}\right) =$$

$$= P\left(t_1 \sqrt{\frac{pq}{n}} < \frac{m - np}{n} < t_2 \sqrt{\frac{pq}{n}}\right) =$$

$$=P\left(t_1<\frac{m-np}{n\sqrt{\frac{pq}{n}}}<t_2\right)=P\left(t_1<\frac{m-np}{\sqrt{npq}}<t_2\right)=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{0}e^{-\frac{x^2}{2}}dx+\frac{1}{\sqrt{2\pi}}\int_{0}^{t_2}e^{-\frac{x^2}{2}}dx=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{0}^{t_2}e^{-\frac{x^2}{2}}dx-\frac{1}{\sqrt{2\pi}}\int_{0}^{t_1}e^{-\frac{x^2}{2}}dx=\Phi(t_2)-\Phi(t_1).$$

ამგვარად,

$$\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{x^2}{2}}dx=\Phi(t_2)-\Phi(t_1).$$

დავუშვათ, რომ $t_2=t=-t_1$, მაშინ გვექნება:

$$\frac{1}{\sqrt{2\pi}}\int_{-t}^{t}e^{-\frac{x^2}{2}}dx=\Phi(t)-\Phi(-t)=\Phi(t)+\Phi(t)=2\Phi(t).$$

ლაპლასის ფორმულა პრაქტიკული ამოცანების ამოხსნისას საკმაოდ კარგ მიახლოებას იძლევა, როდესაც npq საკმაოდ დიდია, დაახლოებით რამდენიმე ასეულის ტოლია. როცა npq არადიდია, მაშინაც შეიძლება ლაპლასის ფორმულის გამოყენება, თუკი არ არის საჭირო მაინცდამაინც დიდი სიზუსტე.

განვიხილოთ რამდენიმე მაგალითი.

1. ყუთში 20 ბურთია, რომელთაგანაც ნახევარი თეთრია და ნახევარი შავი. ბურთის ამოღებაზე ვატარებთ 10 000 ცდას. ამოღებულ ბურთს ყოველი ცდის შემდეგ უკან ვაბრუნებთ. მოვნახოთ ფარდობითი სიხშირის ცალკეული ხდომილობის ალბათობისაგან ის უდიდესი გადახრა, რომელიც პრაქტიკულად შესაძლებელია.

ა მ ო ხ ს ნ ა: (2.3) ფორმულიდან ჩანს, რომ ფარდობითი სიხშირის უდიდესი გადახრა p ალბათობისაგან $t\sqrt{\frac{pq}{n}}$-ის ტოლია.

ჩვენს შემთხვევაში გვაქვს:

$$p=q=1/2, \quad n=10\,000.$$

ახლა უნდა მოვნახოთ ისეთი t, რომლისთვისაც (2.3) ალბათობა ახლოს იქნება ერთთან.

II ცხრილით ვიპოვით, რომ

$$\Phi(3,88)=0.9999. \quad \text{აქ} \quad t=3,88.$$

მაშასადამე, საძიებელი უდიდესი გადახრა იქნება:

$$t\sqrt{\frac{pq}{n}}=3,88\sqrt{\frac{1}{4\cdot 10000}}=0,0194.$$

2. ალბათობა იმისა, რომ ქარხანაში დამზადებული პროდუქცია პირველი ხარისხისაა, უდრის 0,75-ს. ვამოწმებთ 1000 ერთეულს. უნდა გაირკვეს:

ა) მათგან პირველი ხარისხის პროდუქციათა რა რიცხვია უალბათესი და რას უდრის ალბათობა იმისა, რომ პირველი ხარისხის პროდუქციათა ეს რაოდენობა მასში სინამდვილეში იქნება? ბ) რას უდრის ალბათობა იმისა, რომ პირველი ხარისხის ფაქტიურ რაოდენობასა და უალბათეს რიცხვს შორის სხვაობის აბსოლუტური მნიშვნელობა 4%-ს არ აღემატება?

ამოხსნა: $np=1000\cdot 0,75=750$. უალბათესი რიცხვი იქნება $m_0=750$. სწორედ ეს იქნება უალბათესი რიცხვი პირველი ხარისხის პროდუქციისა 1000 ერთეულიდან. მისი ალბათობა მოინახება (1.16) ფორმულით:

$$P_n(m)=\frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{npq}}\,e^{-\frac{t^2}{2}},$$

$$m_0-np=750-750=0,$$

$$\sqrt{npq}=\sqrt{1000\cdot\frac{3}{4}\cdot\frac{1}{4}}=13,7.$$

$$t=\frac{m_0-np}{\sqrt{npq}}=\frac{0}{13,7}=0.$$

ვიპოვით, რომ

$$\varphi(0)=\frac{1}{\sqrt{2\pi}}\quad e^0=\frac{1}{\sqrt{2\pi}}=0,399.$$

ხოლო

$$P_{1000}(750)=\frac{1}{13,7}\cdot 0,399=0,029.$$

როგორც ვხედავთ, ალბათობა იმისა, რომ 1000 ერთეულიდან 750 იქნება ზუსტად პირველი ხარისხის, ძალიან მცირეა, ის არ აღემატება 0,03-ს.

ბ) პირობის თანახმად, გადახრის აბსოლუტური მნიშვნელობა:

$$|m-np| \leqslant 750 \cdot 0{,}04 = 30.$$

ალბათობა იმისა, რომ პირველი ხარისხის პროდუქციათა რაოდენობის გადახრა უალბათესი რიცხვისაგან არ აღემატება 4%-ს, არის აგრეთვე ალბათობა იმისა, რომ პირველი ხარისხის პროდუქცია არანაკლებია, ვიდრე $750-30=720$ და არაუმეტესი, ვიდრე $750+30=780$. ეს ალბათობა გამოითვლება ფორმულით:

$$2\Phi(t)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx,$$

სადაც

$$t=\frac{30}{13{,}7}=2{,}19,$$

ხოლო

$$2\Phi(t)=2\Phi(2{,}19)=0{,}972.$$

მაშასადამე, საძებნი ალბათობა მეტია, ვიდრე 0,97.

ადრე განხილული მაგალითის შემთხვევაში ალბათობა იმისა, რომ 1000-დან ზუსტად 750 ცალი იქნებოდა პირველი ხარისხის, ახლოს იყო ნულთან, ნაკლები იყო 0,03-ზე, ხოლო ალბათობა იმისა, რომ გადახრა 750-დან იქნება არაუმეტეს 4%-სა, ახლოს იყო ერთთან, მეტი იყო, ვიდრე 0,97.

საზოგადოდ, თუ რაიმე ხდომილობის ალბათობა ძალიან პატარაა, მაშინ პრაქტიკულად შეიძლება დარწმუნებული ვიყოთ, რომ გარკვეული პირობების შესრულებისას ამ ხდომილობას არ ექნება ადგილი; თუკი რაიმე ხდომილობის ალბათობა გარკვეულ პირობებში ახლოსაა ერთთან, მაშინ პრაქტიკულად შეიძლება დარწმუნებული ვიყოთ, რომ მას აუცილებლად ექნება ადგილი. სწორედ ამაში მდგომარეობს **პ რ ა ქ ტ ი კ უ ლ ა დ დ ა რ წ მ უ ნ ე ბ ი ს პ რ ი ნ ც ი პ ი**.

ალბათობა როდის შეიძლება ჩაითვალოს ძალიან მცირედ ან ძალიან დიდად, ეს კონკრეტული ამოცანის შინაარსზეა დამოკიდებული. ხშირ შემთხვევაში, თუ ალბათობა არ აღემატება 0,03-ს, ხდომილობას თვლიან პრაქტიკულად შეუძლებლად, ხოლო თუ ალბათობა 0,97-ზე მეტია, მაშინ მის შესაბამის ხდომილობას პრაქტიკულად აუცილებლად თვლიან.

უკანასკნელ მაგალითში გამოვიყენეთ ლაპლასის ფორმულა. ჩვენთვის სასურველი ხდომილობის მოხდენათა რიცხვი m-ის უალბათესი მნიშვნელობიდან გადახრა შევაფასეთ. შეფასება გამოისახება გადახრის ალბათობით (0,972).

ასევე შეიძლება ფარდობითი სიხშირის შეფასება ცალკეული ხდომილობის ალბათობით, ე. ი. საჭიროა

$$\frac{m}{n}-p$$

გამოსახულების შეფასება.

ამისათვის საჭიროა მოინახოს

$$P\left(\left|\frac{m}{n}-p\right|\leqslant\varepsilon\right),$$

სადაც ε წინასწარ მოცემული მცირე დადებითი რიცხვია. თანაფარდობიდან

$$\left|\frac{m}{n}-p\right|\leqslant\varepsilon$$

გვაქვს, რომ

$$-\varepsilon n\leqslant m-np\leqslant\varepsilon n.$$

აქედან გამომდინარეობს, რომ $\frac{m}{n}$ ფარდობითი სიხშირის ცალკეული ხდომილობის ალბათობიდან გადახრა იგივეა, რაც m სიხშირის გადახრა მოხდენათა უალბათესი np რიცხვისაგან. ეს უკანასკნელი კი გამოითვლება ფორმულით:

$$2\Phi(t)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx.$$

მაშასადამე, საძებნი ალბათობა

$$P\left\{\left|\frac{m}{n}-p\right|\leqslant\varepsilon\right\}\simeq2\Phi(t).$$

ჩვენს შემთხვევაში $|m-np|\leqslant\varepsilon n$ და $t=\frac{\varepsilon n}{\sqrt{npq}}$, საიდანაც

$$t=\varepsilon\sqrt{\frac{n}{pq}}.$$

t-ს საკმაოდ დიდი მნიშვნელობისათვის, მაგალითად, როცა $t=5$-ს, $2\Phi(t)$ ახლოსაა ერთთან. მაშასადამე, t-ს საკმაოდ დიდი მნიშვნელობისათვის პრაქტიკულად უზრუნველყოფილია $\left|\frac{m}{n}-p\right|\leqslant\varepsilon$ თანაფარდობის შესრულება.

t და p-ს მივცეთ გარკვეული მნიშვნელობანი და შემოვიღოთ აღნიშვნა

$$t\sqrt{pq}=k.$$

მაშინ

$$\varepsilon=\frac{k}{\sqrt{n}},$$

$$\left|\frac{m}{n}-p\right|\leqslant\frac{k}{\sqrt{n}}.$$

ხოლო

$$P\left\{\left|\frac{m}{n}-p\right|\leqslant\frac{k}{\sqrt{n}}\right\}\simeq 1.$$

ამგვარად, ერთის ტოლი ალბათობით შეიძლება ჩაითვალოს, რომ $\frac{m}{n}$ ფარდობითი სიხშირის გადახრა ცალკეული ხდომილობის p ალბათობიდან ცდათა n რიცხვიდან კვადრატული ფესვის უკუპროპორციულია.

იმ შემთხვევაში, როდესაც ჩვენთვის სასურველი ხდომილობის ალბათობა უცნობია, მაშინ საჭიროა შეფასდეს ყველაზე არახელსაყრელი შემთხვევა, როდესაც pq ნამრავლი უდიდესია, ე. ი., როცა $p=q=\frac{1}{2}$ და $pq=\frac{1}{4}$, მაშინ

$$t=\frac{k}{\sqrt{pq}}=2k=2\cdot\varepsilon\sqrt{n}.$$

მაგალითები

1) რას უდრის ალბათობა იმისა, რომ, როცა ვატარებთ 10000 ურთიერთდამოუკიდებელ ცდას, ჩვენთვის სასურველი ხდომილობის ფარდობითი სიხშირე $\frac{m}{n}$ თავისი $p=0,36$ ალბათობიდან გადაიხრება არა უმეტეს 1%-ით.

ამოხსნა. პირობის თანახმად, $n=10\,000$, $p=0,36$, $q=0,64$

$$\left|\frac{m}{n}-p\right|\leqslant 0,01 \quad \text{ან} \quad \varepsilon=0,01.$$

გვქონდა

$$t=\varepsilon\sqrt{\frac{n}{pq}},$$

საიდანაც

$$t=0{,}01\sqrt{\frac{10\,000}{0{,}36\cdot 0{,}64}}=\frac{1}{0{,}48}=2{,}08,$$

ხოლო

$$P\left\{\left|\frac{m}{n}-0{,}36\right|\leqslant 0{,}01\right\}=2\Phi(2{,}08).$$

II ცხრილის მიხედვით ვიპოვით, რომ

$$2\Phi(2{,}08)=0{,}962,$$

მაშასადამე, ალბათობა $P_t=0{,}962$.

2) ვთქვათ, ისევ $n=10000$ და $\varepsilon=0{,}01$, ხოლო p უცნობია, მაშინ გვექნება

$$t=2\varepsilon\sqrt{n}=2\cdot 0{,}01\cdot\sqrt{10\,000}=2,$$

ხოლო

$$P_t=2\Phi(2)=0{,}954.$$

3) რამდენჯერ უნდა ჩავატაროთ ცდა, რომ p_t-ზე არანაკლები ალბათობით ვამტკიცოთ, რომ $\frac{m}{n}$-ის გადახრა p-საგან არ იქნება მოცემულ ε რიცხვზე მეტი?

ა მ ო ხ ს ნ ა. მოცემული p_t ალბათობით, ცხრილის დახმარებით, ვიპოვით t-ს; გარდა ამისა, ცნობილია p ალბათობა და ფარდობითი სიხშირის ცალკეული ხდომილობის ალბათობისაგან გადახრა ε.

როგორც გვქონდა

$$t=\varepsilon\sqrt{\frac{n}{pq}},$$

საიდანაც

$$n=\frac{t^2}{\varepsilon^2}\,pq.$$

თუ p უცნობი ან განუზღვრელია, მაშინ, როგორც ვიცით, უნდა მივიღოთ $p=q=\frac{1}{2}$ და, მაშასადამე, გვექნება:

$$n=\frac{t^2}{4\varepsilon^2}.$$

4) რამდენი ცდა უნდა ჩავატაროთ, რომ $\frac{m}{n}$ ფარდობითი სიხშირის გადახრა $p=0{,}36$ ალბათობისაგან იყოს არაუმეტეს 0,01-ისა. მოითხოვება შედეგის უზრუნველყოფა 0,99-ზე არანაკლები ალბათობით.

ა მ ო ხ ს ნ ა. მოცემულია $p_t=0{,}99$, $\varepsilon=0{,}01$, $p=0{,}36$, $q=1-0{,}36=0{,}64$.

ცხრილით ვიპოვით, რომ $p_t=0{,}99$-ს, $t=2{,}58$. თუ ამათ შევიტანთ n-ის გამოსახულებაში, მივიღებთ:

$$n=\frac{(2{,}58)^2}{(0{,}01)^2}\cdot 0{,}36\cdot 0{,}64=15\,335.$$

5) ვთქვათ, ცდათა რიცხვი $n=2000$, დავუშვათ, რომ აქედან ჩვენთვის სასურველ A ხდომილობას ადგილი ჰქონდა 150-ჯერ. მონახეთ ის საზღვრები, რომელშიც მოთავსებულია A ხდომილობის ალბათობა. შედეგი უზრუნველყავით $p_t=0{,}9973$-ის ტოლი ალბათობით.

ა მ ო ხ ს ნ ა. საზღვრები, რომელშიც იქნება მოთავსებული

$$p=\frac{m}{n}\pm\varepsilon,$$

მოინახება

$$\varepsilon=t\sqrt{\frac{pq}{n}}$$

ტოლობიდან.

p უცნობია, ამიტომ მას მივიღებთ დაახლოებით $\frac{m}{n}=0{,}075$ ფარდობითი სიხშირის ტოლად

$$p\simeq 0{,}10,\quad q=0{,}90,$$

t-ს მნიშვნელობას I ცხრილით ვიპოვით, $p_t=0{,}9973$-დან. ვნახავთ, რომ $t=3$. მაშასადამე,

$$\varepsilon=3\sqrt{\frac{0{,}1\cdot 0{,}9}{2000}}=\frac{0{,}9}{\sqrt{2000}}\simeq 0{,}02,$$

ხოლო

$$p=0{,}075\pm 0{,}02.$$

ცდა რომ ჩაგვეტარებინა 200-ჯერ და ხდომილობას 15-ჯერ ჰქონოდა ადგილი, მაშინ ფარდობითი სიხშირე იქნებოდა იგივე 0,075, მაგრამ ის საზღვრები, რომელშიც უნდა მოთავსებულიყო p ალბათობა, შეიცვლებოდა. მართლაც,

$$\varepsilon=3\sqrt{\frac{0{,}1\cdot 0{,}9}{200}}=\frac{0{,}9}{\sqrt{200}}\simeq 0{,}065$$

და

$$p=0{,}075\pm 0{,}065.$$

§ 3. პუასონის ასიმპტოტური ფორმულა

პ უ ა ს ო ნ ი ს ფორმულას მათემატიკურ სტატისტიკაში იშვიათი ხდომილობების შემთხვევაში მიმართავენ. მას დიდი გამოყენება აქვს.

საკითხი ეხება იმავე $P_n(m)$ ალბათობის გამოთვლას. ზემოთ გვქონდა

$$P_n(m)=\frac{n!}{m!\,(n-m)!}\;p^m q^{n-m}. \tag{3.1}$$

გარკვეული პირობების შემთხვევაში ამ ფუნქციის მიახლოებით მნიშვნელობას ლაპლასის ფორმულით ვპოულობთ.

ახლა ვიპოვოთ სავსებით განსხვავებული მიახლოებითი ფორმულა, რომელიც წინა ფორმულისაგან განსხვავდება არა მარტო ფორმით, არამედ შინაარსითაც. ლაპლასის თეორემის დროს m-ს ვგულისხმობდით ძალიან დიდს, დაახლოებით np-ს რიგისას, ახლა განვიხილავთ იშვიათ ხდომილობებს, ე. ი. ისეთ ხდომილობებს, რომელთა მოხდენის ალბათობანი ძალიან მცირე სიდიდეებია. ასე რომ, n-ის საკმაოდ დიდი მნიშვნელობისათვის ნამრავლი np არ იქნება დიდი.

ტექნიკაში ასეთი ხდომილობანი ძალიან ბევრია, საიდანაც ჩანს, რომ პუასონის ფორმულას პრაქტიკაში მეტად დიდი გამოყენება აქვს.

ვთქვათ, $np=a$, სადაც a არც ისე დიდი მუდმივია. m-ის განხილვას ისეთ სიდიდედ, რომელსაც შეუძლია მიიღოს ნებისმიერად დიდი მნიშვნელობა, როგორც ამას წინათ ვგულისხმობდით, არავითარი აზრი არა აქვს, რადგან, თუ მაგალითად დავუშვებთ, რომ $a=10$, ხოლო m-ს ავიღებთ 1000-ს, მაშინ $m-np$ გადახრის ალბათობაზე საკითხის დასმას აზრი არ ექნება.

$P_n(m)$ ფუნქციის გამოსახულებაში p შევცვალოთ $\frac{a}{n}$-ით, მაშინ გვექნება:

$$P_n(m)=\frac{n(n-1)(n-2)\ldots(n-m+1)}{m!}\left(\frac{a}{n}\right)^m\left(1-\frac{a}{n}\right)^{n-m}$$

ან

$$P_n(m)=\frac{a^m}{m!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)\frac{\left(1-\frac{a}{n}\right)^n}{\left(1-\frac{a}{n}\right)^m}.$$

ახლა, თუ გადავალთ ზღვარზე, როდესაც $n\to\infty$, მაშინ, რადგან

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)=1,$$

$$\lim_{n\to\infty}\left(1-\frac{a}{n}\right)^m=1,$$

ხოლო

$$\lim_{n\to\infty}\left(1-\frac{a}{n}\right)^n=e^{-a},$$

გვექნება:

$$\lim_{n\to\infty}P_n(m)=\frac{a^m}{m!}e^{-a}.$$

ამგვარად,

$$P_n(m)\simeq\frac{a^m}{m!}e^{-a}. \tag{3.2}$$

ეს არის პუასონის მიახლოებითი ფორმულა.

სავარჯიშო

1. იპოვეთ ალბათობა იმისა, რომ 243 ჩატარებული დამოუკიდებელი ცდიდან ჩვენთვის სასურველ A ხდომილობას ექნება ადგილი ზუსტად 70-ჯერ, თუ ცალკეული ცდის დროს $P(A)=0{,}25$.

პას. $P_{243}(70)=0{,}0231$.

2. იპოვეთ ალბათობა იმისა, რომ A ხდომილობას 2400 ცდის დროს 1400-ჯერ ექნება ადგილი, თუ ცალკეული ცდის დროს $P(A)=0{,}6$.

პას. $P_{2400}(1400)=0{,}0041$.

3. ცალკეული გასროლისას სამიზნეში მოხვედრის ალბათობა არის 0,8. რა იქნება ალბათობა იმისა, რომ 100 გასროლიდან სამიზნეში მოხვედრას ექნება ადგილი 75-ჯერ.

პას. $P_{100}(75)=0{,}04565$.

4. A ხდომილობის ალბათობა ცალკეული ცდის დროს არის 0,7. ვატარებთ ურთიერთდამოუკიდებელ ცდას 2100-ჯერ. მონახეთ ალბათობა იმისა, რომ A მოხდება არა ნაკლებ 1470-ისა და არაუმეტეს 1500-ისა.

პას. $P_{2100}(1470;\ 1500)=0{,}4236$.

5. ყუთში 5 თეთრი და 50 შავი ბურთია. ვთქვათ 10-ჯერ ვიღებთ ბურთს ყუთიდან და ამოღების შემდეგ ისევ უკან ვაბრუნებთ. რას უდრის ალბათობა იმისა, რომ 3-ჯერ ამოვიღებთ თეთრ ბურთს?

მითითება: თუ ბინომიალური ფორმულით ამოხსნით მიიღებთ: $P_{10}(3)\simeq 0{,}013$. გამოიყენეთ პუასონის ასიმპტოტური ფორმულა და ნახავთ, რომ პასუხი მცირედ იქნება განსხვავებული მიღებულისაგან. პუასონის ფორმულის გამოყენებით მიღებული პასუხი იქნება მიახლოებით 0,011.

6. ქარხანამ ბაზაში გაგზავნა 5000 ვარგისი ნაწარმი. ალბათობა იმისა, რომ გზაში ნაწარმი გაფუჭდება, არის 0,0002. მონახეთ ალბათობა იმისა, რომ ბაზაში მოიტანენ 3 ცალ უვარგის ნაწარმს.

მ ი თ ი თ ე ბ ა: ისარგებლეთ პუასონის ასიმპტოტური ფორმულით.

პას. $P_{5000}(3)=0,06$.

IV თ ა ვ ი

შემთხვევითი სიდიდეები და განაწილების ფუნქცია

§ 1. შემთხვევითი სიდიდე და მისი განაწილების კანონი

ჩვენს ყოველდღიურ ცხოვრებაში საქმე გვაქვს შემთხვევით სიდიდეებთან. მათი მნიშვნელობა შემთხვევაზეა დამოკიდებული. წინასწარ შეუძლებელია იმის განსაზღვრა, თუ რა მნიშვნელობას მიიღებს ესა თუ ის შემთხვევითი სიდიდე. ის რიცხვითი მნიშვნელობა, რომლის მიღება შეუძლია შემთხვევით სიდიდეს, იქნება მისი შესაძლო მნიშვნელობა. შემთხვევითი სიდიდის შესასწავლად პირველ რიგში უნდა ვიცოდეთ ყველა ის რიცხვითი მნიშვნელობანი, რომელთა მიღება მას შეუძლია. თუ შემთხვევითი სიდიდე ღებულობს რიცხვით მნიშვნელობათა სასრულ რაოდენობას ან ისეთ უსასრულო რაოდენობას, რომელთა დანომვრა შეიძლება, იტყვიან, რომ შემთხვევითი სიდიდე არის დისკრეტული ტიპის, ხოლო, თუ მას შეუძლია მიიღოს რიცხვით მნიშვნელობათა უსასრულო რაოდენობა, რომელთა დანომვრა არ შეიძლება, მაშინ ის იქნება უწყვეტი ტიპისა.

დისკრეტული ტიპის შემთხვევით სიდიდეთა მაგალითებია: წლის განმავლობაში წვიმიან დღეთა რიცხვი, რომელიმე ქალაქში ერთი დღეღამის განმავლობაში ახალშობილთა რაოდენობა, ბერნულის სქემაში ჩვენთვის სასურველი ხდომილობის მოხდენათა რიცხვი m, ერთი თვის განმავლობაში რომელიმე სასურსათო მაღაზიის ნავაჭრი თანხა, საკოლმეურნეო ბაზარში ერთ-ერთ კვირა დღეს გაყიდულ კარტოფილთა რაოდენობა კილოგრამობით და სხვ.

უწყვეტი ტიპის შემთხვევით სიდიდეთა მაგალითებია: შემთხვევით აღებული ადამიანის სიგრძე, რომელიმე ვაშლის წონა, რომელიმე ქალაქის მიერ დღე-ღამის განმავლობაში დახარჯული გაზის რაოდენობა და სხვ.

მარტო შესაძლო რიცხვით მნიშვნელობათა ცოდნა შემთხვევითი სიდიდის შესასწავლად არ არის საკმარისი. სრული წარმოდგენა რომ გვქონდეს შემთხვევით სიდიდეზე, ამისათვის საჭიროა აგრეთვე ვიცოდეთ, თუ რა ხშირად ღებულობს ის ამა თუ იმ შესაძლო რიცხვით მნიშვნელობას, ე. ი. უნდა ვიცოდეთ, თუ რა ალბათობით ღებულობს შემთხვევითი სიდიდე ამა თუ იმ შესაძლო მნიშვნელობას.

შემთხვევითი სიდიდეები აღვნიშნოთ ლათინური ანბანის დიდი ასოებით X, Y, $Z, \ldots U$, V, W, ხოლო შესაძლო რიცხვითი მნიშვნელობანი კი შესაბამისად—პატარა ასოებით და ინდექსებით.

ვთქვათ, X დისკრეტული ტიპის რაიმე შემთხვევითი სიდიდეა, მაშინ ვიტყვით, რომ ის სავსებით განსაზღვრულია, თუკი მოცემულია შემდეგი ცხრილი:

$$X\begin{cases} x_1, & x_2, \ldots, & x_n \\ p_1, & p_2, \ldots, & p_n, \end{cases} \tag{1.1}$$

სადაც პირველ სტრიქონში მოცემულია შესაძლო რიცხვითი მნიშვნელობანი, რომელთაც ვარიანტებს უწოდებენ, ხოლო მეორე სტრიქონში კი—ამ რიცხვით მნიშვნელობათა შესაბამისი ალბათობანი. სახელდობრ, p_i არის ალბათობა იმისა, რომ X შემთხვევითი სიდიდე მიიღებს x_i მნიშვნელობას

$$p_i = P\{X = x_i\}. \tag{1.2}$$

რადგანაც X შემთხვევითი სიდიდე თავის შესაძლო რიცხვითი მნიშვნელობებიდან ერთ-ერთს აუცილებლად მიიღებს, ხოლო ორი მათგანის მნიშვნელობის მიღება ერთდროულად არ შეუძლია, ამიტომ ცხადია, p_i ალბათობები წარმოადგენს ხდომილობათა სრულ სისტემაში შემავალ ხდომილობათა შესაბამის ალბათობებს. მაგრამ, როგორც ვიცით, სრული სისტემის ხდომილობათა ალბათობების ჯამი უდრის ერთს, ამიტომ:

$$\sum_{i=1}^{n} p_i = 1.$$

(1.1) ცხრილს, რომელშიც მოცემულია შემთხვევითი სიდიდის შესაძლო რიცხვითი მნიშვნელობანი და სათანადო ალბათობები, ეწოდება შემთხვევითი სიდიდის განაწილების კანონი.

ამგვარად, შემთხვევითი სიდიდის განსასაზღვრავად საჭიროა მისი განაწილების კანონის მოცემა.

(1.1) არის დისკრეტული ტიპის შემთხვევითი სიდიდის განაწილების კანონი. უწყვეტი ტიპის შემთხვევითი სიდიდისათვის ამგვარი განაწილების კანონის მოცემა არ შეიძლება, ამიტომ მთელ შუალედს,

რომლის შიგნითაც ღებულობს უწყვეტი ტიპის შემთხვევითი სიდიდე თავის შესაძლო რიცხვით მნიშვნელობებს, ყოფენ რამდენიმე პატარ-პატარა შუალედებად და ადგენენ (1.1)-ის ანალოგიურ ცხრილს, რომლის პირველ სტრიქონში მოცემულია პატარ-პატარა შუალედები, ხოლო მეორე სტრიქონში კი ამ შუალედებში შემთხვევითი სიდიდის მოხვედრის ალბათობანი. ამგვარად, უწყვეტი ტიპის შემთხვევითი სიდიდის განაწილების კანონი შეიძლება ჩაიწეროს შემდეგნაირად:

$$X \begin{cases} (x_0, x_1), & (x_1, x_2); & \ldots; & (x_{n-1}, x_n) \\ p_1, & p_2, & \ldots, & p_n. \end{cases} \tag{1.3}$$

ცხადია, ასეთი ცხრილით მოცემული განაწილების კანონი X შემთხვევით სიდიდეს ახასიათებს მხოლოდ მიახლოებით, რადგანაც ის არ იძლევა ცალკეული შესაძლო რიცხვითი მნიშვნელობის ალბათობის განსაზღვრის შესაძლებლობას.

ყველა ზემონათქვამიდან შეგვიძლია დავასკვნათ, რომ **სიდიდეს, რომელსაც მოცემულ პირობებში შეუძლია მიიღოს სხვადასხვა შესაძლო რიცხვითი მნიშვნელობები სათანადო ალბათობებით, ეწოდება შემთხვევითი სიდიდე.**

§ 2. განაწილების ფუნქცია

ნებისმიერი ტიპის შემთხვევითი სიდიდის განაწილების მოცემა შეიძლება განაწილების ფუნქციის საშუალებით. ვთქვათ X შემთხვევითი სიდიდეა და x კი რაიმე ფიქსირებული რიცხვია, მაშინ ალბათობას იმისა, რომ X შემთხვევითი სიდიდე მიიღებს უფრო ნაკლებ მნიშვნელობას, ვიდრე x, ეწოდება X შემთხვევითი სიდიდის განაწილების ფუნქცია და აღინიშნება $F(x)$-ით, ე. ი.:

$$P(X<x)=F(x). \tag{2.1}$$

ცხადია, რადგანაც განაწილების ფუნქცია ალბათობითაა განსაზღვრული, ამიტომ იგი არაუარყოფითი იქნება და მოთავსდება 0-სა და 1 შორის:

$$0 \leqslant F(x) \leqslant 1.$$

განაწილების ფუნქცია მონოტონურად არაკლებადი ფუნქციაა, რაც იმას ნიშნავს, რომ არგუმენტის მეტ მნიშვნელობას ფუნქციის არანაკლები მნიშვნელობა შეესაბამება, ე. ი.

$$F(x_1) \leqslant F(x_2),$$

თუკი

$$x_1 < x_2.$$

მართლაც,

$$X<x_2$$

უტოლობას მაშინ ექნება ადგილი, როდესაც

$$X<x_1$$

ან, როდესაც

$$x_1 \leqslant X<x_2.$$

ეს იმას ნიშნავს, რომ $X<x_2$ უტოლობის შესაბამისი ხდომილობა რთული ხდომილობაა და მისი წარმოდგენა შეიძლება ორი უთავსებადი ხდომილობათა ჯამის სახით:

$$X<x_2=(X<x_1)+(x_1 \leqslant X<x_2).$$

ამიტომ, რადგანაც ტოლძალოვან ხდომილობათა ალბათობანიც ტოლია, დავწერთ:

$$P(X<x_2)=P[(X<x_1)+(x_1 \leqslant X<x_2)].$$

მაგრამ, თუ გამოვიყენებთ ჯამის ფორმულას უთავსებადი ხდომილობების შემთხვევაში, გვექნება:

$$P(X<x_2)=P(X<x_1)+P(x_1 \leqslant X<x_2), \tag{2.2}$$

საიდანაც

$$P(x_1 \leqslant X<x_2)=P(X<x_2)-P(X<x_1).$$

განაწილების ფუნქციის განმარტების თანახმად

$$F(x_1)=P(X<x_1)$$

და

$$F(x_2)=P(X<x_2).$$

მაშასადამე, შეგვიძლია დავწეროთ

$$P(x_1 \leqslant X<x_2)=F(x_2)-F(x_1). \tag{2.3}$$

რადგან ალბათობა არაუარყოფითი სიდიდეა, ამიტომ ადგილი ექნება უტოლობას:

$$F(x_2)-F(x_1) \geqslant 0$$

ან, რაც იგივეა,

$$F(x_1) \leqslant F(x_2).$$

თუ ცნობილია X-ის განაწილების ფუნქციის მნიშვნელობა, როგორც (2.3)-დან ჩანს, შესაძლებელია X შემთხვევითი სიდიდის ნებისმიერ შუალედში მოხვედრის ალბათობის განსაზღვრა.

განაწილების ფუნქცია საზოგადოდ ყოველთვის უწყვეტი ფუნქცია არ იქნება, მაგრამ ის ყოველთვის იქნება მარცხნიდან უწყვეტი, ე. ი.

$$F(x-0)=F(x).$$

აგრეთვე ცხადია, რომ

$$F(+\infty)=P(X<+\infty)=1$$

და

$$F(-\infty)=P(X<-\infty)=0.$$

სხვაობას

$$F(x_0+0)-F(x_0-0)=c\geqslant 0$$

ეწოდება განაწილების ფუნქციის ნახტომი $x=x_0$ წერტილზე.

ამგვარად, ვხედავთ, რომ ყოველი განაწილების ფუნქცია არის ისეთი ფუნქცია, რომელიც არაკლებადია, მარცხნივ უწყვეტია და აკმაყოფილებს პირობას $F(+\infty)=1$, $F(-\infty)=0$. მართებული იქნება შებრუნებითაც: თუ რომელიმე ფუნქცია არაკლებადია, უწყვეტია მარცხნივ და აკმაყოფილებს პირობებს $F(+\infty)=1$, $(F-\infty)=0$, მაშინ ის აუცილებლად რომელიმე შემთხვევითი სიდიდის განაწილების ფუნქცია იქნება.

თუ ცნობილია რაიმე შემთხვევითი სიდიდის განაწილების კანონი, მაშინ მისი დახმარებით შეიძლება განაწილების ფუნქციის შედგენა.

მაგალითად, ვთქვათ, მოცემული x შემთხვევითი სიდიდის განაწილების კანონია:

$$X\begin{cases} 0, & 1, & 2, & 3, & 4, & 5 \\ 0{,}1; & 0{,}1; & 0{,}2; & 0{,}3; & 0{,}2; & 0{,}1. \end{cases}$$

ვთქვათ, $x=0$, რადგან X სიდიდე არაუარყოფითია, ამიტომ არ შეიძლება X იყოს 0-ზე ნაკლები და, რადგანაც შეუძლებელი ხდომილობის ალბათობა ნულია, გვექნება $F(x)=P(X<x)=0$, როცა $x\leqslant 0$.

ვთქვათ $x=1$, მაშინ X მიიღებს მხოლოდ ერთ მნიშვნელობას — ნულის ტოლს, რომელიც ნაკლებია 1-ზე და შესაბამისი ალბათობა ცხადია, იქნება 0,1:

$$F(x)=F(1)=P(X<x)=P(X<1)=0{,}1.$$

ამგვარად,

$$F(x)=0{,}1, \quad \text{როცა} \quad 0<x\leqslant 1.$$

ვთქვათ, $x=2$, მაშინ

$$F(2)=P(X<2)=P(X<1)+P(1\leqslant X<2).$$

მაგრამ, როგორც ვნახეთ $P(X<1)=0{,}1$, ხოლო

$$P(1\leqslant X<2)=0{,}1.$$

მაშასადამე,

$$F(2)=P(X<2)=0{,}1+0{,}1=0{,}2,$$

თუ $1<x\leqslant 2$.

ახლა, ვთქვათ, $x=3$; მაშინ ალბათობა იმისა, რომ $X<3$, (2.2) ფორმულის ძალით იქნება $P(X<2)$ და $P(2\leqslant X<3)$ ალბათობათა ჯამი, ე. ი.

$$F(3)=P(X<3)=0{,}2+0{,}2=0{,}4$$

თუკი

$$2\leqslant x<3.$$

ანალოგიურად გამოვთვლით, რომ

$$F(4)=P(X<4)=0{,}4+0{,}3=0{,}7,$$

თუკი

$$3\leqslant x<4.$$

ასევე

$$F(5)=P(X<5)=0{,}7+0{,}2=0{,}9,$$

როცა

$$4\leqslant x<5,$$

აგრეთვე,

$$F(6)=P(X<6)=0{,}9+0{,}1=1{,}0,$$

როცა

$$5\leqslant x<6.$$

II თავში ავაგეთ ბინომიალური განაწილების კანონის შესაბამისი დიაგრამა, სადაც $n=9$ და $p=q=\frac{1}{2}$ (ნახ. 1). ამის შესაბამისი განაწილების კანონი ჩაიწერება შემდეგნაირად:

$$X\begin{cases}0, & 1, & 2, & 3, & 4, & 5,\\ 0{,}00195, & 0{,}01755, & 0{,}07020, & 0{,}16380, & 0{,}24570, & 0{,}24570,\\ 6, & 7, & 8, & 9 & & \\ 0{,}16380, & 0{,}07020, & 0{,}01755, & 0{,}00195 & & \end{cases}$$

ამის შესაბამისი განაწილების ფუნქცია იქნება:

$$F(x)=0,\quad \text{როცა}\quad x\leqslant 0,$$

$$F(x)=\frac{1}{2^9}\simeq 0{,}00195,\quad \text{როცა}\quad 0\leqslant x<1,$$

$$F(x)=10\cdot\frac{1}{2^9}\simeq 0{,}01950,\quad \text{როცა}\quad 1\leqslant x<2,$$

$$F(x)=46\cdot\frac{1}{2^9}\simeq 0{,}08970,\quad \text{როცა}\quad 2\leqslant x<3,$$

$$F(x)=130\cdot\frac{1}{2^9}\simeq 0{,}25350.\quad \text{როცა}\quad 3\leqslant x<4$$

$$F(x)=256\cdot\frac{1}{2^9}\simeq 0{,}49920,\quad \text{როცა}\quad 4\leqslant x<5,$$

$$F(x)=282\cdot\frac{1}{2^9}\simeq 0{,}74490,\quad \text{როცა}\quad 5\leqslant x<6,$$

$$F(x)=466\cdot\frac{1}{2^9}\simeq 0{,}90870,\quad \text{როცა}\quad 6\leqslant x<7,$$

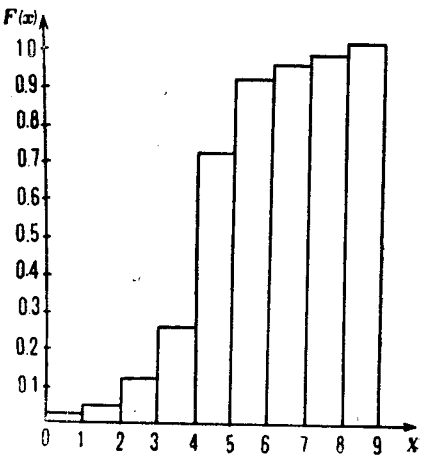
$$F(x)=502\cdot\frac{1}{2^9}\simeq 0{,}97890,\quad \text{როცა}\quad 7\leqslant x<8,$$

$$F(x)=511\cdot\frac{1}{2^9}\simeq 0{,}99745,\quad \text{როცა}\quad 8\leqslant x<9,$$

$$F(x)=512\cdot\frac{1}{2^9}=1,\quad \text{როცა}\quad 9\leqslant x<10.$$

თუ ამის მიხედვით ავაგებთ განაწილების ფუნქციის შესაბამის გრაფიკს, გვექნება შემდეგი სურათი (ნახ. 2).

რა თქმა უნდა, მე-2 ნახაზი ზუსტად მასშტაბის დაცვით არ არის აგებული. როგორც ვხედავთ, განაწილების ფუნქცია წარმოადგენს კიბისებრ ტეხილ მრუდს.



ნახ. 2.

დისკრეტული ტიპის შემთხვევითი სიდიდის ყოველი შესაძლო მნიშვნელობა წარმოადგენს $F(x)$-ის წყვეტის წერტილს; $F(x)$ აკეთებს ნახტომს, რომელიც X შემთხვევითი სიდიდის შესაბამისი შესაძლო მნიშვნელობის მიღების ალბათობის ტოლია.

გრაფიკულად ეს იმით გამოიხატება, რომ, როგორც კი შემთხვევითი სიდიდე მიიღებს ახალ მნიშვნელობას, კიბე მაშინვე აიწევა ზევით იმ სიდიდით, რომელიც ამ მნიშვნელობის შესაბამისი ალბათობის ტოლია.

საზოგადოდ, ბერნულის სქემის შემთხვევაში, როდესაც ურთიერთდამოუკიდებელ ცდას ვატარებთ, ჩვენთვის სასურველი ხდომილობა

შეიძლება ნებისმიერ რიცხვჯერ მოხდეს 0-დან n-მდე ჩათვლით. იმის ალბათობის გამოსათვლელად, რომ ხდომილობა n დამოუკიდებელ ცდაში ზუსტად m-ჯერ მოხდეს, მივიღეთ შემდეგი ბინომიალური ფორმულა:

$$P_n(m) = \frac{n!}{m!\,(n-m)!}\, p^m q^{n-m}.$$

თუ ხდომილობის მოხდენათა რიცხვის განაწილების ფუნქციას აღვნიშნავთ $F(x)$-ით, მაშინ მას ექნება ასეთი სახე:

$$F(x) = \begin{cases} 0, & \text{როცა} \quad x \leqslant 0 \\ \sum\limits_{k<x} P_n(k), & \text{როცა} \quad 0 < x \leqslant n \\ 1, & \text{როცა} \quad x > n. \end{cases}$$

განაწილების ფუნქცია წარმოადგენს კიბისებურ წირს, რომელსაც ნახტომები აქვს წერტილებზე: $x=0, 1, 2, \ldots, n$; ხოლო $x=k$ წერტილზე მისი ნახტომი იქნება $P_n(k)$.

იტყვიან, რომ ყოველი შემთხვევითი სიდიდე, რომელიც ზემოთ დაწერილ ბერნულის ფორმულას აკმაყოფილებს, ემორჩილება ბინომიალური განაწილების კანონს.

§ 3. განაწილების სიმკვრივე

განაწილების ფუნქციის წარმოებულს **განაწილების სიმკვრივე** ეწოდება

$$\frac{dF(x)}{dx} = f(x),$$

სადაც $f(x)$ არაუარყოფითი სიდიდეა და აკმაყოფილებს პირობას

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{3.1}$$

მართლაც,

$$\int_{-\infty}^{\infty} f(x)\,dx = F(+\infty) - F(-\infty) = 1 - 0 = 1.$$

საზოგადოდ კი

$$\int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1). \tag{3.2}$$

პრაქტიკულად, განაწილების სიმკვრივე გვიჩვენებს ერთეულ შუალედში მოხვედრილი შემთხვევითი სიდიდის მიერ მიღებულ რიცხვით მნიშვნელობათა რაოდენობას.

თუ X შემთხვევითი სიდიდის მიერ მიღებული შესაძლო რიცხვითი მნიშვნელობანი განაწილებულია რაიმე (x_1, x_2) შუალედის შიგნით, მაშინ, ცხადია, $f(x)$ ნულის ტოლი იქნება, როდესაც $x<x_1$ ან $x>x_2$. ამ შემთხვევაში შეიძლება დავწეროთ:

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{x_1}^{x_2} f(x)\,dx=1.$$

$f(x)$ განაწილების სიმკვრივეს უწყვეტი ტიპის X შემთხვევითი სიდიდის განაწილების დიფერენციალურ კანონს უწოდებენ, ხოლო $F(x)$-ს კი—X შემთხვევითი სიდიდის ინტეგრალური განაწილების ფუნქციას ან ინტეგრალური განაწილების კანონს.

როგორც ვიცით, ალბათობა იმისა, რომ $X<x$, არის განაწილების ფუნქცია

$$F(x)=P(X<x)=\int_{-\infty}^{x} f(x)\,dx. \tag{3.3}$$

შემდეგ, გვაქვს

$$P(x_1\leqslant X<x_2)=F(x_2)-F(x_1).$$

ამიტომ (3.2)-ის ძალით დავწერთ

$$P(x_1\leqslant X<x_2)=\int_{x_1}^{x_2} f(x)\,dx.$$

ეს უკანასკნელი გეომეტრიულად გამოსახავს ფართობს, რომელიც მოთავსებულია $y=f(x)$ მრუდსა, OX ღერძსა და იმ ორდინატებს შორის, რომელთა შესაბამისი აბსცისებია $x=x_1$ და $x=x_2$ (ნახ. 3).

ალბათობა იმისა, რომ x_1 ემთხვევა x_2-ს, (3.2)-ის ძალით ნულის ტოლი იქნება. ეს იმას ნიშნავს, რომ უწყვეტი ტიპის შემთხვევითი სიდიდეებისათვის ალბათობა იმისა, რომ ის რომელიმე თავის შესაძლო მნიშვნელობას მიიღებს, ნულის ტოლია.

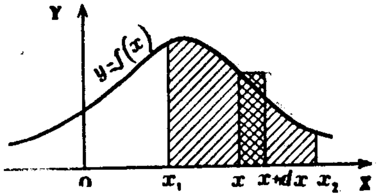
ალბათობა იმისა, რომ X შემთხვევითი სიდიდე მოთავსებულია რაიმე მცირე შუალედში $(x, x+dx)$, დაახლოებით $f(x)\,dx$ გამოსახულების ტოლია; მას ალბათობის ელემენტი ეწოდება.

$$P(x\leqslant X<x+dx)\simeq f(x)\,dx.$$

მართლაც,

$$P\{x \leqslant X < x+dx\} = F(x+dx) - F(x) \simeq dF(x) = f(x)\,dx.$$

ორ X და Y შემთხვევით სიდიდეს ურთიერთდამოუკიდებელი ეწოდება, თუკი $X=x$ ნებისმიერი ტოლობის ან $x_1 < X < x_2$ ნებისმიერი უტოლობის ალბათობა არ შეიცვლება, რა მნიშვნელობაც არ უნდა მიიღოს Y შემთხვევითმა სიდიდემ და, პირიქით, ნებისმიერი $Y=y$ ტოლობის და ნებისმიერი $y_1 < Y < y_2$ უტოლობის ალბათობა არ არის იმაზე დამოკიდებული, თუ რა მნიშვნელობა მიიღო X შემთხვევითმა სიდიდემ.



ნახ. 3.

ვთქვათ, X შემთხვევითი სიდიდე თანაბრად განაწილებულია რომელიმე (x_1, x_2) შუალედში. ეს იმას ნიშნავს, რომ მისი განაწილების სიმკვრივე

$$f(x) = \begin{cases} 0, & \text{როცა} \quad x < x_1, \\ k, & \text{როცა} \quad x_1 < x < x_2; \\ 0, & \text{როცა} \quad x > x_2. \end{cases}$$

აქ k მუდმივი განისაზღვრება ტოლობიდან:

$$\int_{x_1}^{x_2} f(x)\,dx = \int_{x_1}^{x_2} k\,dx = 1,$$

ამგვარად,

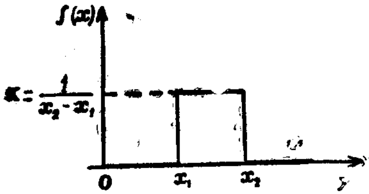
$$k = \frac{1}{x_2 - x_1}.$$

მიღებული განაწილება, რომლის სიმკვრივეა $f(x) = \dfrac{1}{x_2 - x_1}$ განხილული (x_1, x_2) შუალედის შიგნით, ხოლო მის გარეთ ნულის ტოლია, თანაბარი იქნება ან, ამ შემთხვევაში იტყვიან, რომ შემთხვევითი სიდიდე განაწილებულია ტოლი ალბათობიანი კანონით.

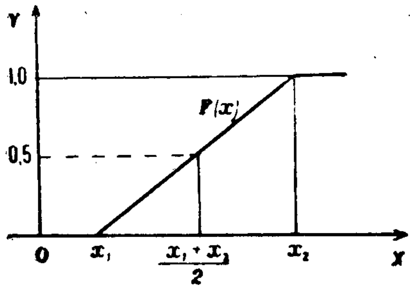
$f(x) = \dfrac{1}{x_2 - x_1}$ განაწილების სიმკვრივის შესაბამისი განაწილების ფუნქცია იქნება წრფივი ფუნქცია, რომელიც იზრდება 0-დან 1-მდე, როცა x იცვლება x_1-დან x_2-მდე.

თანაბრად განაწილების სიმკვრივე გრაფიკულად წარმოიდგინება შემდეგნაირად (ნახ. 4).

თანაბრად განაწილების ფუნქცია კი გამოისახება შემდეგნაირად (ნახ. 5).



ნახ. 4.



ნახ. 5.

§ 4. შემთხვევით სიდიდეთა სისტემა

რამდენიმე შემთხვევით სიდიდეთა ერთობლიობას შემთხვევით სიდიდეთა სისტემა ეწოდება.

ხშირ შემთხვევაში, რომელიმე ცდის ან დაკვირვების შედეგის აღწერა ხდება არა ერთი, არამედ რამდენიმე შემთხვევითი სიდიდის საშუალებით. მაგალითად, შემთხვევითი წერტილის დახასიათება სიბრტყეზე ხდება ორი შემთხვევითი სიდიდით: აბსცისით და ორდინატით. შემთხვევითი წერტილის დახასიათება სივრცეში კი ხდება სამი შემთხვევითი სიდიდით.

შემთხვევითი წერტილების მაგივრად ხშირად შემთხვევით ვექტორებს ხმარობენ. მაგალითად, სიბრტყეზე შემთხვევით (X, Y) წერტილს შეესაბამება შემთხვევითი ვექტორი, რომლის სათავეა კოორდინატთა სისტემის სათავე, ხოლო ბოლო კი არის წერტილი კოორდინატებით X და Y.

ჯერ შევისწავლოთ სისტემა, რომელიც შედგება ორი დისკრეტული ტიპის შემთხვევითი სიდიდისაგან.

ვთქვათ (X, Y) სისტემის კომპონენტები X და Y არიან დისკრეტული შემთხვევითი სიდიდეები, რომელთა განაწილების კანონებს აქვთ შემდეგი სახე

$$X \begin{cases} x_1, x_2, \dots, x_n \\ p_1, p_2, \dots, p_n \end{cases}$$

და

$$Y \begin{cases} y_1, y_2, \dots, y_m \\ q_1, q_2, \dots q_m. \end{cases}$$

ორი დისკრეტული შემთხვევითი სიდიდისაგან შემდგარი (X, Y) სისტემის განაწილების კანონი მოიცემა შემდეგი ცხრილის სახით:

$Y \backslash X$	x_1	$\dots$	x_i	$\dots$	x_n
y_1	p_{11}	$\dots$	p_{i1}	$\dots$	p_{n1}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_j	p_{1j}	$\dots$	p_{ij}	$\dots$	p_{nj}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	p_{1m}	$\dots$	p_{im}	$\dots$	p_{nm}

სადაც ჩამოთვლილია (x_i, y_j) წყვილების ყველა შესაძლო მნიშვნელობები და მათი შესაბამისი ალბათობები

$$p_{ij}=P\{X=x_i;\ Y=y_j\},\qquad \begin{array}{l} i=1,\dots\ n, \\ j=1,\dots, m. \end{array}$$

შევნიშნოთ, რომ (X, Y) სისტემის განაწილების კანონიდან შეიძლება აღვადგინოთ X და Y შემთხვევითი სიდიდეების განაწილების კანონები. მაგალითად

$$p_i=P(X=x_i)=\sum_{j=1}^{m} p_{ij},$$

$$q_j=P(Y=y_j)=\sum_{i=1}^{n} p_{ij}.$$

ამრიგად, X და Y შემთხვევითი სიდიდეების განაწილების კანონებს ექნებათ შემდეგი სახე

$$X\left\{\begin{array}{l} x_1, \ \cdots\ , \ x_n \\ \sum\limits_{j=1}^{m} p_{1j}\cdots\sum\limits_{j=1}^{m} p_{nj}, \end{array}\right.$$

$$Y\left\{\begin{array}{l} y_1, \ \cdots\ , \ y_m \\ \sum\limits_{i=1}^{n} p_{i1}\cdots\sum\limits_{i=1}^{n} p_{im}. \end{array}\right.$$

6 გვანჯი მანია

$$\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m}P\{X=x_i;\ Y=y_j\}=1.$$

ვთქვათ $(X,\ Y)$ სისტემის განაწილების კანონი მოცემულია ცხრილით, აღვადგინოთ ცალ-ცალკე X და Y-ის განაწილების კანონი

$Y \backslash X$	0,01	0,02	0,03	0,04
0,002	0,01	0,02	0,04	0,04
0,004	0,03	0,24	0,15	0,06
0,006	0,04	0,10	0,08	0,08
0,008	0,02	0,04	0,03	0,02

გვაქვს $p_1=P(X=0{,}01)=0{,}01+0{,}03+0{,}04+0{,}02=0{,}10$, $p_2=P(X=0{,}02)=0{,}02+0{,}24+0{,}10+0{,}04=0{,}40$, $p_3=P(X=0{,}03)=0{,}04+0{,}15+0{,}08+0{,}03=0{,}30$, $p_4=P(X=0{,}04)=0{,}04+0{,}06+0{,}08+0{,}02=0{,}20$, $q_1=P(Y=0{,}002)=0{,}01+0{,}02+0{,}04+0{,}04=0{,}11$, $q_2=P(Y=0{,}004)=0{,}03+0{,}24+0{,}15+0{,}06=0{,}48$, $q_3=P(Y=0{,}006)=0{,}04+0{,}10+0{,}08+0{,}08=0{,}30$, $q_4=P(Y=0{,}008)=0{,}02+0{,}04+0{,}03+0{,}02=0{,}11$.

ამრიგად, X შემთხვევითი სიდიდის განაწილების კანონს ექნება შემდეგი სახე

$$X\begin{cases}0{,}01; & 0{,}02; & 0{,}03; & 0{,}04\\ 0{,}1 & 0{,}4 & 0{,}3 & 0{,}2,\end{cases}$$

ხოლო Y შემთხვევითი სიდიდის განაწილების კანონს კი ექნება სახე

$$Y\begin{cases}0{,}002; & 0{,}004; & 0{,}006; & 0{,}008\\ 0{,}11 & 0{,}48 & 0{,}3 & 0{,}11.\end{cases}$$

ორი უწყვეტი ტიპის X და Y შემთხვევითი სიდიდის განაწილების ფუნქცია ეწოდება ორი $X<x$ და $Y<y$ უტოლობის ერთდროულად შესრულების ალბათობას

$$F(x,\ y)=P\{X<x;\ Y<y\},$$

სადაც x და y ფიქსირებული რიცხვითი სიდიდეებია.

გეომეტრიულად, $F(x,\ y)$ განაწილების ფუნქცია არის ალბათობა იმისა, რომ $(X,\ Y)$ შემთხვევითი წერტილი მოხვდება იმ უსასრულო

კვადრატის შიგნით, რომლის წვეროა (x, y) და მდებარეობს ამ წერტილის მარცხნივ და ქვემოთ (ნახ. 6).

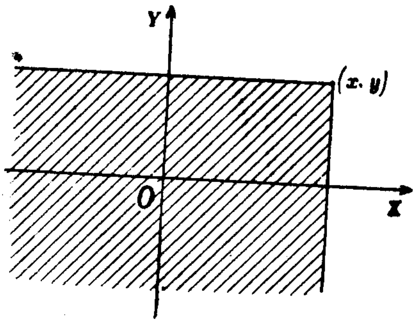
განვიხილოთ $F(x, y)$ განაწილების ფუნქციის თვისებები:

1. $F(x, y)$ ფუნქცია არის თავისი ორივე არგუმენტის მიმართ არაკლებადი ფუნქცია, ე. ი.

$$\text{როცა} \quad x_2 > x_1, \quad \text{მაშინ} \quad F(x_2, y) \geq F(x_1, y)$$

$$\text{და როცა} \quad y_2 > y_1, \quad \text{მაშინ} \quad F(x, y_2) \geq F(x, y_1).$$

ამ თვისების სამართლიანობაში უბრალოდ დავრწმუნდებით, თუკი მივმართავთ $F(x, y)$ განაწილების ფუნქციის გეომეტრიულ ინტერპრეტაციას; როგორც ნახ. 6-დან ჩანს, თუ x-ს გავადიდებთ, ე. ი. მას მარჯვნივ გადავაადგილებთ, ან y-ს გადავაადგილებთ ზემოთ, ამით, რა თქმა უნდა, მიღებული უსასრულო კვადრატის შიგნით მოხვედრის ალბათობას ვერ შევამცირებთ.



ნახ. 6.

2. ყველგან $-\infty$-ზე განაწილების ფუნქციის მნიშვნელობა ნულია.

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

ამ თვისების სამართლიანობაშიაც დავრწმუნდებით გეომეტრიული ინტერპრეტაციის მომარჯვებით. მართლაც, თუ უსასრულოდ გადავაადგილებთ მარცხნივ კვადრატის მარჯვენა საზღვარს, ანდა თუ უსასრულოდ ქვევით დავწევთ მის ზედა საზღვარს, ან თუ ორივე ამ პროცესს ერთდროულად ჩავატარებთ, ამით ცხადია, კვადრატში მოხვედრის ალბათობა ნულისაკენ მიისწრაფვის.

3. როდესაც ერთ-ერთი არგუმენტია $+\infty$, მაშინ სისტემის განაწილების ფუნქცია ხდება მეორე არგუმენტის შესაბამისი განაწილების ფუნქცია.

$$F(x, +\infty) = F_1(x),$$
$$F(+\infty, y) = F_2(y),$$

სადაც $F_1(x)$ და $F_2(y)$ არიან X და Y შემთხვევით სიდიდეთა განაწილების ფუნქციები შესაბამისად.

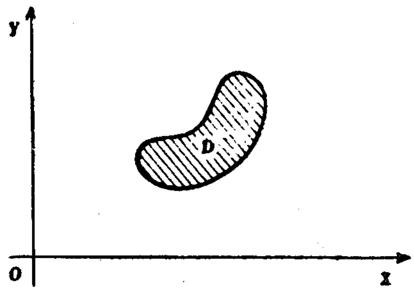
თუ ერთ ან მეორე მხარეს კვადრატის საზღვარს გადავაადგილებთ $+\infty$-მდე, მაშინ ზღვარში კვადრატი ნახევარსიბრტყედ იქცევა. რო-

მელშიც მოხვედრის ალბათობა არის სისტემაში შემავალი ერთ-ერთი შემთხვევითი სიდიდის განაწილების ფუნქცია.

4. თუ ორივე არგუმენტი ტოლია $+\infty$-ის, მაშინ სისტემის განაწილების ფუნქცია ერთის ტოლია

$$F(+\infty, +\infty)=1.$$

სინამდვილეში, როცა $x\to\infty$ და $y\to\infty$, მაშინ კვადრატი, რომლის წვეროა (x, y), იქცევა მთელ XOY სიბრტყედ, რომელშიც მოხვედრის ალბათობა არის აუცილებელი ხდომილობის ალბათობა.



ნახ. 7.

ახლა ვიანგარიშოთ ალბათობა იმისა, რომ შემთხვევითი (X, Y) წერტილი მოხვდება XOY სიბრტყის რაიმე D არეში (ნახ. 7).

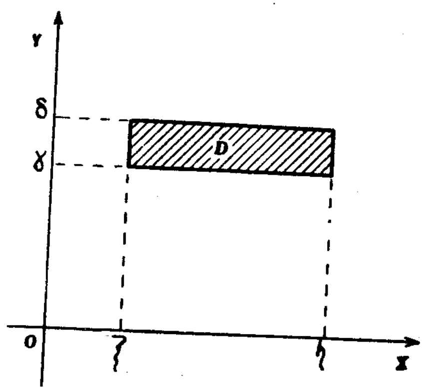
პირველ რიგში განვიხილოთ D არედ ისეთი მართკუთხედი, რომლის გვერდები საკოორდინატო ღერძების პარალელურია (ნახ. 8) და მიღებული შედეგი განვაზოგადოთ.

სისტემის განაწილების ფუნქციის საშუალებით გამოვსახოთ (X, Y) შემთხვევითი წერტილის მოხვედრის ალბათობა D მართკუთხედში, რომელიც შემოსაზღვრულია ξ და η აბსცისებით და γ და δ ორდინატებით (ნახ. 8). მაშინ (X, Y) წერტილის D არეში მოხვედრის ალბათობა ტოლძალოვანი იქნება ორი ხდომილობის:

$$\xi\leqslant x<\eta$$

და

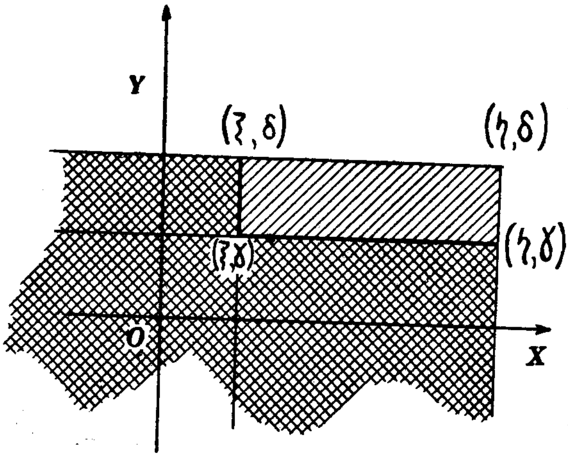
$$\gamma\leqslant y<\delta.$$



ნახ. 8.

ამ ხდომილობათა ალბათობები გამოვსახოთ სისტემის განაწილების ფუნქციის საშუალებით. ამისათვის XOY სიბრტყეზე განვიხილოთ ოთხი უსასრულო კვადრატი, რომელთა წვეროებია წერტილები: (η, δ), (ξ, δ), (η, γ) და (ξ, γ) (ნახ. 8).

ცხადია, D მართკუთხედში მოხვედრის ალბათობა ტოლია (η, δ) წვეროიან კვადრატში მოხვედრის ალბათობას მინუს (ξ, δ) წვეროიან კვადრატში მოხვედრის ალბათობა, მინუს აგრეთვე (η, γ) წვეროიან კვადრატში მოხვედრის ალბათობა და პლუს (ξ, γ) წვეროიან კვად-



ნახ. 9.

რატში მოხვედრის ალბათობა, რადგანაც უკანასკნელი ორჯერ ზედიზედ გამოვაკელით. ამგვარად გვექნება

$$P\{(X, Y)\subset D\}=F(\eta, \delta)-F(\xi, \delta)-F(\eta, \gamma)+F(\xi, \gamma).$$

$(X, Y)\subset D$ ნიშნავს იმას, რომ (X, Y) წერტილი შედის D მართკუთხედში.

ვთქვათ, მოცემულია უწყვეტი ტიპის ორი X და Y შემთხვევითი სიდიდე.

XOY სიბრტყეზე განვიხილოთ (x, y) წერტილზე მიდგმული პატარა D_Δ მართკუთხედი, რომლის გვერდებია Δx და Δy (ნახ. 10).

ამ მართკუთხედში მოხვედრის ალბათობა, როგორც ვნახეთ, ტოლია

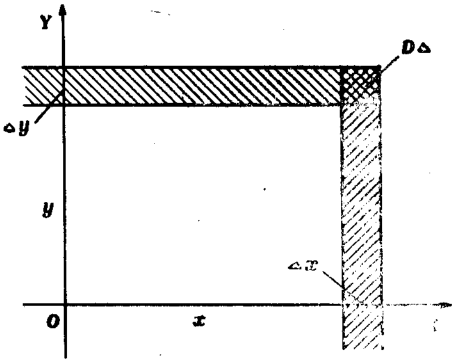
$$P\{(X, Y)\subset D_\Delta\}=F(x+\Delta x, y+\Delta y)-F(x+\Delta x, y)-$$
$$-F(x, y+\Delta y)+F(x, y).$$

მართკუთხედში მოხვედრის ალბათობა გავყოთ თვით D_Δ მართკუთხედის ფართზე და გადავიდეთ ზღვარზე, როდესაც ერთდროულად $\Delta x\to 0$ და $\Delta y\to 0$, გვექნება

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P\{X, Y) \subset D_\Delta\}}{\Delta x \Delta y} =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F(x+\Delta x, y+\Delta y) - F(x+\Delta x, y) - F(x, y+\Delta y) + F(x, y)}{\Delta x \Delta y}$$

ვიგულისხმოთ, რომ $F(x, y)$ უწყვეტია და დიფერენცირებადი, მაშინ უკანასკნელი ფორმულის მარჯვენა მხარე წარმოადგენს $F(x, y)$-ის მეორე რიგის შერეულ კერძო წარმოებულს x და y-ის მიმართ. თუ ამ წარმოებულს $f(x, y)$-ით აღვნიშნავთ, გვექნება



ნახ. 10.

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} =$$

$$= F''_{xy}(x, y).$$

$f(x, y)$ ფუნქციას სისტემის განაწილების სიმკვრივე ეწოდება.

ამგვარად, სისტემის განაწილების სიმკვრივე წარმოადგენს პატარა მართკუთხედში მოხვედრის ალბათობისა და თვით ამ პატარა მართკუთხედის ფართობის ფარდობის ზღვარს, როდესაც მართკუთხედის ორივე გვერდი ნულისკენ მიისწრაფვის, განაწილების სიმკვრივე გამოისახება როგორც სისტემის განაწილების ფუნქციის ორივე არგუმენტის მიმართ მეორე რიგის შერეული წარმოებული.

შემთხვევითი (X, Y) წერტილის ნებისმიერ D არეში მოხვედრის ალბათობა იქნება:

$$P\{(X, Y) \subset D\} = \iint_D f(x, y)\, dxdy.$$

$F(x, y)$ განაწილების ფუნქცია $f(x, y)$ განაწილების სიმკვრივის საშუალებით შემდეგნაირად გამოისახება:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\, dxdy.$$

განაწილების სიმკვრივეს აქვს შემდეგი თვისებები:

1. სისტემის განაწილების სიმკვრივე არაუარყოფითია. ეს იქიდან ჩანს, რომ სისტემის განაწილების სიმკვრივე წარმოადგენს ორ არაუარყოფით სიდიდეთა ფარდობის ზღვარს და, მაშასადამე, ის არ შეიძლება იყოს უარყოფითი.

2. უსასრულო საზღვრებში ორჯერადი ინტეგრალი სისტემის განაწილების სიმკვრივიდან ერთის ტოლია

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx dy = 1.$$

აღებული ინტეგრალი წარმოადგენს მთელ XOY სიბრტყეზე მოხვედრის ალბათობას, რაც აუცილებელი ხდომილობის ალბათობის ტოლფასია.

ჩვენ ზემოთ გვქონდა, რომ

$$F_1(x) = F(x, \infty), \quad F_2(y) = F(\infty, y)$$

და

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x, y)\, dx dy.$$

საიდანაც დავწერთ

$$F_1(x) = F(x, \infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x, y)\, dx dy.$$

თუ ამ გამოსახულებას გავადიფერენციალებთ x-ით, მივიღებთ X შემთხვევითი სიდიდის განაწილების სიმკვრივის გამოსახულებას

$$f_1(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x, y)\, dy.$$

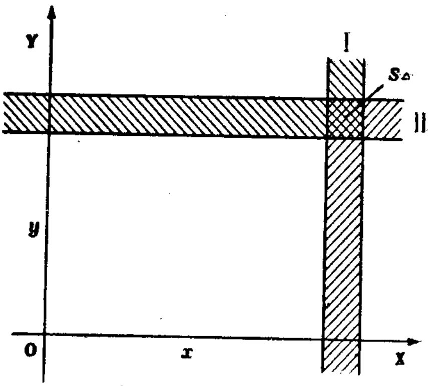
ანალოგიურად გვექნება

$$f_2(y) = F_2'(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

ამგვარად, რომ მივიღოთ სისტემაში შემავალი ერთი რომელიმე შემთხვევითი სიდიდის განაწილების სიმკვრივე, ამისათვის საჭიროა სისტემის განაწილების სიმკვრივე უსასრულო საზღვრებში ვაინტეგროთ იმ არგუმენტის მიმართ, რომელიც მეორე შემთხვევით სიდიდეს შეესაბამება.

(X, Y) სისტემაში შემავალი X შემთხვევითი სიდიდის პირობითი განაწილების კანონი ეწოდება ამ სისტემის განაწილების კანონს, რომელიც გამოთვლილია იმ პირობით, რომ სისტემაში შემავალმა მეორე შემთხვევითმა სიდიდემ Y-მა მიიღო განსაზღვრული მნიშვნელობა.

პირობითი განაწილების ფუნქცია აღვნიშნოთ $F_y(x)$-ით, ხოლო პირობითი განაწილების სიმკვრივე კი $f_y(x)$-ით. რადგანაც პრაქტიკულად მეტად მნიშვნელოვანია უწყვეტი ტიპის შემთხვევითი სიდიდეები. ამიტომ მათთვის განვიხილოთ პირობითი განაწილების კანონიც.



ნახ. 11.

ამრიგად, თუ ცნობილია სისტემაში შემავალი ერთი შემთხვევითი სიდიდის განაწილების კანონი და მეორის პირობითი განაწილების კანონი, მაშინ ჩვენ შეგვიძლია აღვადგინოთ სისტემის განაწილების კანონი.

მართლაც, განვიხილოთ (x, y) წერტილზე მიდგმული s_Δ მართკუთხედი, რომლის გვერდებია Δx და Δy (ნახ. 11). ამ მართკუთხედში მოხვედრის ალბათობა იქნება

$$f(x, y)\,dxdy = P\{(X, Y) \subset S_\Delta\} = P\{x < X < x + \Delta x;\ y < Y < y + dy\}.$$

ამ ტოლობის მარჯვენა მხარეში გვაქვს ორი ხდომილობის ნამრავლის ალბათობა, რომელიც, როგორც ვიცით, უდრის ერთ-ერთის ალბათობას, გამრავლებულს მეორის პირობით ალბათობაზე, ე. ი. ტოლი იქნება I ზონაში მოხვედრის ალბათობა, გამრავლებული II ზონაში მოხვედრის ალბათობაზე იმ პირობით, რომ I ზონაში მოხვედრას უკვე ჰქონდა ადგილი.

ეს პირობა ზღვარში $\mathbf{X} = x$ პირობის ტოლძალოვანია; მაშასადამე,

$$f(x, y)\,dxdy = f_1(x)\,dx\,f_x(y)\,dy$$

საიდანაც

$$f(x, y) = f_1(x)\,f_x(y).$$

ამგვარად მივიღეთ, რომ სისტემაში შემავალი ორი შემთხვევითი სიდიდის განაწილების სიმკვრივე უდრის ერთ-ერთის განაწილების სიმკვრივეს გამრავლებულს მეორის განაწილების სიმკვრივეზე იმ პირობით, რომ პირველმა სიდიდემ მიიღო გარკვეული მნიშვნელობა.

$f(x, y)$ სიმკვრივე ცხადია შემდეგნაირად წარმოიდგინება

$$f(x, y)=f_2(y)\, f_y(x).$$

მიღებული ტოლობებიდან დავწერთ:

$$f_x(y)=\frac{f(x, y)}{f_1(x)}$$

და

$$f_y(x)=\frac{f(x, y)}{f_2(y)}$$

თუ აქ შევიტანთ $f_1(x)$ და $f_2(y)$-ის მნიშვნელობებს, მივიღებთ:

$$f_x(y)=\frac{f(x, y)}{\int\limits_{-\infty}^{\infty} f(x, y)\, dy}$$

და

$$f_y(x)=\frac{f(x, y)}{\int\limits_{-\infty}^{\infty} f(x, y)\, dx}.$$

ვიტყვით, რომ Y შემთხვევითი სიდიდე X შემთხვევითი სიდიდისაგან დამოუკიდებელია, თუ Y შემთხვევითი სიდიდის განაწილების კანონი არ არის დამოკიდებული იმაზე, რა მნიშვნელობა მიიღო X შემთხვევითმა სიდიდემ. ეს დამოუკიდებლობა შემდეგნაირად ჩაიწერება:

$$f_x(y)=f_2(y)$$

და

$$f_y(x)=f_1(x).$$

ამგვარად, X და Y შემთხვევით სიდიდეებს ურთიერთდამოუკიდებელნი ეწოდებათ, თუ თითოეული მათგანის განაწილების კანონი არ არის დამოკიდებული იმაზე, რა მნიშვნელობას ღებულობს მეორე. წინააღმდეგ შემთხვევაში X და Y იქნებიან ურთიერთდამოკიდებულნი.

თუ X და Y შემთხვევითი სიდიდეები ურთიერთდამოუკიდებელნი არიან, მაშინ

$$f(x, y)=f_1(x)\, f_2(y), \tag{4.1}$$

ე. ი. ურთიერთდამოუკიდებელ შემთხვევით სიდიდეთა განაწილების სიმკვრივე უდრის სისტემაში შემავალ ცალკეულ შემთხვევით სიდიდეთა განაწილების სიმკვრივეების ნამრავლს.

(4.1) პირობა წარმოადგენს შემთხვევით სიდიდეთა ურთიერთდამოუკიდებლობის აუცილებელ და საკმარის პირობას.

1. შემთხვევითი სიდიდის შესაძლო რიცხვითი მნიშვნელობებია $x_1=1$; $x_2=5$; $x_3=7$, ხოლო პირველი ორის ალბათობებია შესაბამისად 0,35 და 0,3. იპოვეთ მესამე მნიშვნელობის შესაბამისი ალბათობა.

პას. $P_3=0{,}35$.

2. მოცემულია X და Y შემთხვევით სიდიდეთა განაწილების კანონი

$$X \begin{cases} 1, & -1, & 2, \\ 0{,}1; & 0{,}1; & 0{,}8 \end{cases} \qquad Y \begin{cases} 2, & 3, & -1 \\ 0{,}2; & 0{,}3: & 0{,}5 \end{cases}$$

იპოვეთ $X+Y$ და XY შემთხვევით სიდიდეთა განაწილების კანონი.

3. შეადგინეთ A ხდომილობის მოხდენათა რიცხვის განაწილების კანონი სამი დამოუკიდებელი ცდის დროს, თუ ცალკეული ცდის დროს მისი მოხდენის ალბათობაა 0,3.

$$\text{პას.}\quad X \begin{cases} 0, & 1, & 2, & 3, \\ 0{,}343; & 0{,}441; & 0{,}189; & 0{,}027 \end{cases}$$

4. დაწესებულების კომუტატორი ემსახურება 100 აბონენტს. ალბათობა იმისა, რომ წუთის განმავლობაში აბონენტი დაურეკავს კომუტატორს, არის 0,02. რომელი ხდომილობაა უფრო ალბათი: წუთის განმავლობაში დარეკავს 3 აბონენტი თუ დარეკავს 4 აბონენტი?

პას. $P_{100}(3)=0{,}18$; $P_{100}(4)=0{,}09$.

5. მანქანაზე დაბეჭდილი 1000-გვერდიანი ხელნაწერის ტექსტი შეიცავს 100 კორექტურულ შეცდომას. იპოვეთ ალბათობა იმისა, რომ შემთხვევით აღებული გვერდი შეიცავს: ა) ერთს მაინც კორექტურულ შეცდომას, ბ) ზუსტად ორ კორექტურულ შეცდომას; გ) არა ნაკლებ 2 კორექტურულ შეცდომას. იგულისხმეთ, რომ კორექტურული შეცდომები განაწილებულია პუასონის კანონით.

პას. ა) $P=1-e^{-1}=0{,}6321$,
ბ) $P_{100}(2)=0{,}18395$,
გ) $P=0{,}2642$.

6. სატელეფონო სადგურში წუთში გამოძახებათა საშუალო რიცხვია 5. მონახეთ ალბათობა იმისა, რომ 2 წუთის განმავლობაში ადგილი ექნება: ა) ორ გამოძახებას, ბ) ორზე ნაკლებ გამოძახებას, გ) არა ნაკლებ ორ გამოძახებას.

მ ი თ ი თ ე ბ ა: $e^{-10}=0{,}000045$.

პას. ა) $P=000025$,
ბ) $P=000495$,
გ) $P=0{,}999505$.

7. იპოვეთ ორი შემთხვევითი სიდიდის სისტემას მდგენელთა განაწილების კანონი, თუ სისტემის განაწილების კანონია:

$Y \backslash X$	x_1	x_2	x_3
y_1	0,14	0,25	0,19
y_2	0,11	0,13	0,18

პას. $X\begin{cases} x_1, & x_2, & x_3 \\ 0{,}25, & 0{,}38, & 0{,}37 \end{cases}$

$Y\begin{cases} y_1, & y_2 \\ 0{,}58; & 0{,}42. \end{cases}$

8. იპოვეთ ალბათობა იმისა, რომ ორი შემთხვევითი სიდიდის სისტემის მდგენელი $X<1/2$ და $Y<1/3$, თუკი

$$F(x, y)=\left(\frac{1}{\pi}\operatorname{arctg} 2x+\frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg} 3y+1/2\right).$$

პას. $P(X<1/2;\ Y<1/3)=9/16.$

9. იპოვეთ (X, Y) შემთხვევითი წერტილის ისეთ მართკუთხედში მოხვედრის ალბათობა, რომელიც შემოსაზღვრულია წრფეებით: $x=\frac{\pi}{4}$, $x=\frac{\pi}{2}$, $y=\frac{\pi}{6}$, $y=\frac{\pi}{3}$ და სისტემის განაწილების ფუნქციაა

$$F(x, y)=\sin x\sin y\left(0\leqslant x\leqslant\frac{\pi}{2};\ 0\leqslant y\leqslant\frac{\pi}{2}\right).$$

პას. $P\left\{\frac{\pi}{4}<X<\frac{\pi}{2};\ \frac{\pi}{6}<Y<\frac{\pi}{3}\right\}=0{,}11.$

10. ლატარიაში ყოველ ას ბილეთზე თამაშდება ერთი 5000-იანი მოგება, 2—1000-იანი, 10 ას-ასმანეთიანი მოგება. განსაზღვრეთ ერთი ბილეთის მოგების განაწილების კანონი, თუ მისი ღირებულებაა 80 მანეთი.

პას: $X\begin{cases} -80, & 20, & 920, & 4920 \\ 0{,}87, & 0{,}10, & 0{,}02; & 0{,}01. \end{cases}$

11. მსროლელი ისვრის სამიზნეში. ცალკეული გასროლისას სამიზნეში მოხვედრის ალბათობა არის 0,2. სამიზნეში მოხვედრა მსროლელს

აძლევს 5 ქულას. იპოვეთ განაწილების კანონი ქულათა რიცხვისა, თუ მსროლელი სამიზნეში სამჯერ ისვრის.

$$\text{პას. } X \begin{cases} 0; & 5; & 10; & 15: \\ 0{,}512; & 0{,}384; & 0{,}096: & 0{,}008. \end{cases}$$

12. შეადგინეთ მსროლელის მიერ 4 გასროლისას დაგროვილ ქულათა რაოდენობის განაწილების კანონი, თუ მიზანში მოხვედრისას მას 5 ქულა ეწერება, აცდენისას აკლდება 2 ქულა და ცალკეული გასროლის მოხვედრის ალბათობა არის 0,3.

$$\text{პას. } X \begin{cases} -8; & -1; & 6; & 13; & 20; \\ 0{,}24; & 0{,}41; & 0{,}26; & 0{,}08; & 0{,}010. \end{cases}$$

13. x შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases} a\cos x, & \text{როცა } |x|\leqslant \dfrac{\pi}{2} \\ 0, & \text{როცა } |x|>\dfrac{\pi}{2}. \end{cases}$$

იპოვეთ a მუდმივი და $F(x)$ განაწილების ფუნქცია.

პას. $a=1/2;$

$$F(x)=\begin{cases} 0, & \text{როცა } x<-\dfrac{\pi}{2} \\ 1/2\,(\sin x+1), & \text{როცა } -\dfrac{\pi}{2}\leqslant x<\dfrac{\pi}{2} \\ 1, & \text{როცა } x>\dfrac{\pi}{2}. \end{cases}$$

14. პარტიაში 10 დეტალია, რომელშიაც 8 სტანდარტულია. შემთხვევით ვარჩევთ 2-ს. შეადგინეთ შერჩევაში მოხვედრილი სტანდარტული დეტალებისათვის განაწილების კანონი.

$$\text{პას. } X \begin{cases} 0, & 1, & 2 \\ \dfrac{1}{45}, & \dfrac{16}{45}, & \dfrac{28}{45}. \end{cases}$$

15. პარტიაში 6 დეტალია, მათგან 4 სტანდარტულია. შემთხვევით ვარჩევთ 3-ს. შეადგინეთ შერჩევაში მოხვედრილი სტანდარტული დეტალებისათვის განაწილების კანონი.

$$\text{პას. } X \begin{cases} 0, & 1, & 2, & 3, \\ 0, & 1/5, & 3/5, & 1/5, \end{cases}$$

16. ორი ყუმბარმშენი რიგრიგობით ესვრის სამიზნეს პირველ მოხვედრამდე. პირველისათვის მოხვედრის ალბათობა არის 0,7, ხოლო მეორისათვის—0,8. სროლას იწყებს პირველი ყუმბარმშენი. შეადგინეთ ორივე ყუმბარმშენის მიერ ჩამოგდებულ ყუმბარათა რიცხვის განაწილების კანონის პირველი 4 წევრი.

პას. $X\begin{cases} 1, & 2, & 3, & 4 \\ 0{,}7; & 0{,}24; & 0.04; & 0{,}02. \end{cases}$

17. პარტიაში 10% არასტანდარტული დეტალია. შემთხვევით ვიღებთ 4 დეტალს. შეადგინეთ შერჩევაში მოხვედრილ არასტანდარტულ დეტალთა რიცხვის ბინომიალური განაწილების კანონი.

პას. $X\begin{cases} 0, & 1, & 2, & 3, & 4, \\ 0{,}6561; & 0{,}2916; & 0{,}0486; & 0{,}0036; & 0{,}0001. \end{cases}$

18. ორ კამათელს ერთდროულად ვყრით ორჯერ. შეადგინეთ ორთავე კამათელზე ჯამში ლუწ რიცხვთა ბინომიალური განაწილების კანონი.

პას. $X\begin{cases} 0, & 1, & 2, \\ 9/16, & 6/16, & 1/16. \end{cases}$

19. ხელსაწყო შედგება 3 ურთიერთდამოუკიდებლად მომუშავე ელემენტისაგან. ცალკეული ცდისას თითოეული ელემენტის მწყობრიდან გამოსვლის ალბათობაა 0,1. შეადგინეთ ერთი ცდის დროს მწყობრიდან გამოსულ ელემენტთა რიცხვის განაწილების კანონი.

პას. $X\begin{cases} 0, & 1, & 2, & 3, \\ 0{,}729; & 0{,}243; & 0{,}027; & 0{,}001. \end{cases}$

V თავი

შემთხვევითი სიდიდის მახასიათებლები

§ 1. მათემატიკური ლოდინი

შემთხვევითი სიდიდის მთლიანად დახასიათების შესაძლებლობას განაწილების კანონი ან განაწილების ფუნქცია იძლევა, მაგრამ ხშირად მათ მაგივრად შემთხვევითი სიდიდის დასახასიათებლად ზოგიერთ რიცხვით მახასიათებლებს მიმართავენ, რომლებიც შესაძლებლობას იძლევა შემთხვევით სიდიდეებზე საორიენტაციო წარმოდგენა შევიქმნათ.

შემთხვევითი სიდიდის ერთ-ერთი ძირითადი მახასიათებელია მათემატიკური ლოდინი ან, რაც იგივეა, საშუალო მნიშვნელობა, რომლის არსი შემდეგში მდგომარეობს.

ვთქვათ, X დისკრეტული ტიპის შემთხვევითი სიდიდეა, რომლის განაწილების კანონია:

$$X \begin{cases} x_1, & x_2, \dots, x_n \\ p_1, & p_2, \dots, p_n. \end{cases} \tag{1.1}$$

მაშინ X შემთხვევითი სიდიდის მათემატიკური ლოდინი ვუწოდოთ ამ შემთხვევითი სიდიდის შესაძლო რიცხვით მნიშვნელობათა შესაბამის ალბათობებზე ნამრავლთა ჯამს.

თუ X შემთხვევითი სიდიდის მათემატიკურ ლოდინს $E(X)$-ით აღვნიშნავთ, დავწერთ:

$$E(x) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^{n} x_i p_i. \tag{1.2}$$

თუ X უწყვეტი ტიპის შემთხვევითი სიდიდეა, მაშინ მისი მათემატიკური ლოდინი გამოისახება შემდეგნაირად:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx, \tag{1.3}$$

თუკი ასეთ ინტეგრალს აზრი აქვს.

მათემატიკური ლოდინის ეს უკანასკნელი გამოსახულება ანალოგიურია დისკრეტული ტიპის შემთხვევითი სიდიდისათვის დაწერილი გამოსახულებისა. მართლაც, ვთქვათ, $F(x)$ არის X შემთხვევითი სიდიდის ინტეგრალური განაწილების ფუნქცია. დავუშვათ, რომ აბსცისათა ღერძი დაფარულია ძალიან მცირე Δx მონაკვეთებით, რომლებიც ერთმანეთის მიმდევრობითაა განლაგებული, როგორც უკვე ვნახეთ,

$$P\{x < X < x + \Delta x\} = F(x + \Delta x) - F(x).$$

მაგრამ, რადგან Δx მცირეა, შეგვიძლია დავწეროთ:

$$F(x + \Delta x) - F(x) \simeq dF(x) = f(x)\, dx.$$

ვთქვათ, რომ X-ის მნიშვნელობა იცვლება ყოველი Δx შუალედის დასაწყისში, ხოლო $(x, x+\Delta x)$ შუალედის შიგნით არ იცვლება და ტოლია x-ის. მაშინ $(x, x+\Delta x)$ უსასრულოდ მცირე შუალედიდან x-ის ნებისმიერი მნიშვნელობის მის $f(x)\Delta x$ ალბათობაზე ნამრავლი იქნება $xf(x)\Delta x$. თუ მთელ აბსცისათა ღერძზე ავჯამავთ ასეთ ნამრავლებს

მივიღებთ $\sum\limits_{-\infty}^{\infty}{}' x f(x)\,\Delta x$; ამ გამოსახულების ზღვარი, როცა $\Delta x \to 0$, იქნება

$$\int\limits_{-\infty}^{\infty} x f(x)\,dx.$$

ცხადია, თუ რომელიმე X შემთხვევითი სიდიდის $f(x)$ განაწილების სიმკვრივე ლუწი ფუნქციაა, მაშინ ასეთი შემთხვევითი სიდიდის მათემატიკური ლოდინი ნულის ტოლი იქნება.

§ 2. მათემატიკური ლოდინის თვისებები

1. მუდმივი სიდიდის მათემატიკური ლოდინი თვით ამ მუდმივის ტოლია.

მართლაც, ვთქვათ, a არის რომელიმე მუდმივი რიცხვი, მაშინ განაწილების კანონის პირველ სტრიქონში ყველგან იქნება a, ხოლო მეორე სტრიქონში—ერთმანეთის ტოლი სათანადო ალბათობანი (სინამდვილეში პირველ სტრიქონში გვექნება მხოლოდ a, ხოლო მეორე სტრიქონში—1). მაშინ, საშუალო მნიშვნელობის განმარტების თანახმად,

$$E(a) = a \cdot 1 = a.$$

2. მუდმივი თანამამრავლი მათემატიკური ლოდინის ნიშნის გარეთ გამოდის.

ვთქვათ, k მუდმივი რიცხვია, ხოლო X — შემთხვევითი სიდიდე. მაშინ kX შემთხვევითი სიდიდის განაწილების კანონი იქნება:

$$kX \left\{ \begin{array}{llll} kx_1, & kx_2, & \ldots, & kx_n, \\ p_1, & p_2, & \ldots, & p_n. \end{array} \right.$$

მივიღებთ:

$$E(kX) = \sum_{i=1}^{n} kx_i p_i = k \sum_{i=1}^{n} x_i p_i = kE(X).$$

3. თუ X და Y ისეთი შემთხვევითი სიდიდეებია, რომელთაც აქვთ მათემატიკური ლოდინი, მაშინ მათ ჯამსაც ექნება მათემატიკური ლოდინი და ჯამის მათემატიკური ლოდინი შესაკრებთა მათემატიკური ლოდინების ჯამის ტოლი იქნება.

ვთქვათ, X შემთხვევითი სიდიდის განაწილების კანონია:

$$X \begin{cases} x_1, & x_2, \dots, & x_n \\ p_1, & p_2, \dots, & p_n, \end{cases}$$

ხოლო Y შემთხვევითი სიდიდის განაწილების კანონი არის

$$Y \begin{cases} y_1, & y_2, \dots, & y_m \\ q_1, & q_2, \dots, & q_m. \end{cases}$$

მაშინ $X+Y$ შემთხვევითი სიდიდის განაწილების კანონი იქნება:

$$X+Y \begin{cases} x_i+y_j \\ p_{ij} \end{cases}$$

სადაც

$$i=1, 2, \dots, n; \quad j=1, 2, \dots, m,$$

ხოლო

$$p_{ij}=p(X=x_i; \quad Y=y_j).$$

საზოგადოდ, $p_{ij} \neq p_i q_j$, მაგრამ, როდესაც X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ

$$p_{ij}=p_i q_j.$$

მათემატიკური ლოდინის განმარტების თანახმად,

$$E(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)\,p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i\,p_{ij}+\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}.$$

ცალ-ცალკე განვიხილოთ ჯამები:

$$\sum_{j=1}^{m} p_{ij} \quad \text{და} \quad \sum_{i=1}^{n} p_{ij}.$$

პირველი $\sum_{j=1}^{m} p_{ij}$ არის $P(X=x_i;\ Y=y_j)$ ალბათობათა ჯამი, სადაც i არ იცვლება, j კი იცვლება 1-დან m-მდე. რადგანაც $Y=y_j$ ხდომილობანი სხვადასხვა j-თვის უთავსებადია, ამიტომ, შეკრების აქსიომის თანახმად, $\sum_{j=1}^{m} p_{ij}$ იქნება

$$(X+x_i;\ Y=y_j), \quad j=1, 2, \dots, m$$

ხდომილობათაგან რომელიმე ერთ-ერთის მოხდენის ალბათობა. ცხადია, $(Y=y_j)$, $j=1, 2, \dots m$ ხდომილობათაგან ერთ-ერთი აუცილებლად

მოხდება, ამიტომ ($X=x_i$; $Y=y_j$), $j=1, 2\ldots m$ ხდომილობის მოხდენა ნიშნავს უბრალოდ ($X=x_i$) ხდომილობის მოხდენას.

ამგვარად,

$$\sum_{j=1}^{m} p_{ij}=P(X=x_i)=p_i.$$

ასევე მივიღებთ, რომ

$$\sum_{i=1}^{n} p_{ij}=q_j$$

და ჯამის მათემატიკური ლოდინისათვის გვექნება

$$E(X+Y)=\sum_{i=1}^{n} x_i\, p_i+\sum_{j=1}^{m} y_j q_j=E(X)+E(Y).$$

დამტკიცებული თვისების განზოგადება შეიძლება შესაკრებთა ნებისმიერი რიცხვისათვის. ამის შესრულებას ვანდობთ მკითხველს.

4. *თუ X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ მათი ნამრავლის მათემატიკური ლოდინი თანამამრავლთა მათემატიკური ლოდინების ნამრავლის ტოლია.*

მართლაც, ვთქვათ, X და Y შემთხვევით სიდიდეებს აქვს იგივე განაწილების კანონი, რაც წინა თვისების დამტკიცებისას ჰქონდა. შევადგინოთ განაწილების კანონი მათი ნამრავლისათვის:

$$XY \begin{cases} x_i\, y_j \\ p_{ij} \end{cases}$$

მაშინ, მათემატიკური ლოდინის განმარტების თანახმად,

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i\, y_j\, p_{ij}.$$

მაგრამ, რადგანაც X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, ამიტომ

$$p_{ij}=p_i\, q_j$$

და მივიღებთ:

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i\, y_j\, p_i\, q_j.$$

უკანასკნელი გადავწეროთ შემდეგნაირად:

$$E(XY)=\sum_{i=1}^{n} x_i\, p_i \sum_{j=1}^{m} y_j\, q_j = E(X)\, E(y).$$

თანამამრავლთა ნებისმიერი რიცხვისათვის ამ თვისების განზოგადება შეიძლება. ამის შესრულებას მკითხველს ვანდობთ.

განვიხილოთ ჯამის მათემატიკური ლოდინის შესახებ თეორემის გამოყენების მაგალითი.

დავუბრუნდეთ ბერნულის სქემას. გვქონდა ფორმულა:

$$P_n(m)=\frac{n!}{m!\,(n-m)!}\, p^m\, q^{n-m}.$$

როდესაც საქმე გვაქვს ბერნულის სქემასთან, მაშინ ჩვენთვის სასურველი ხდომილობის მოხდენისას, შემთხვევით სიდიდეს ეძლევა მნიშვნელობა 1, ხოლო თუ არ მოხდა, მაშინ—ნული. ეს სიდიდეები ერთმანეთზე დამოუკიდებელია. ალბათობა იმისა, რომ შემთხვევითი სიდიდე 1-ის ტოლ მნიშვნელობას მიიღებს, არის p, ხოლო რომ 0-ის ტოლ მნიშვნელობას მიიღებს, არის $1-p$.

ამგვარად, განხილული შემთხვევათი სიდიდის რომელიმე შესაძლო მნიშვნელობის განაწილების კანონი იქნება:

$$X_k \begin{cases} 1, & 0 \\ p, & 1-p, \end{cases}$$

მაშასადამე, მათემატიკური ლოდინის განმარტების თანახმად,

$$E(X_k)=p.$$

ყველა სხვა შესაძლო მნიშვნელობას იგივე მათემატიკური ლოდინი ექნება. ყველა შესაძლო მნიშვნელობათა ჯამი იქნება m:

$$\sum_{k=1}^{n} X_k = m.$$

რადგანაც ეს ჯამი 1-ებისა და 0-ებისაგან შედგება და 1 იმდენჯერ იქნება, რამდენჯერაც ჩვენთვის სასურველი ხდომილობა მოხდა,

$m=\sum_{k=1}^{n} X_k$ ჯამის განაწილების კანონი იქნება:

$$m=\sum_{k=1}^{n} X_k \begin{cases} 0, & 1, & 2, & \ldots, & m \\ P_n(0), & P_n(1), & P_n(2), & \ldots, & P_n(m). \end{cases}$$

აქედან კი, განმარტების თანახმად დავწერთ:

$$E(m)=E\left(\sum_{k=1}^{n}X_k\right)=\sum_{k=1}^{n}E(X_k).$$

მაგრამ, როგორც უკვე ვნახეთ, ყოველი X_k-ს მათემატიკური ლოდინი p-ს ტოლია. ამიტომ

$$\sum_{k=1}^{n}E(X_k)=np.$$

ამგვარად,

$$E(m)=np. \tag{2.1}$$

ამ თეორემას აქვს აგრეთვე ძალიან დიდი პრაქტიკული ღირებულება.

5. შემთხვევითი სიდიდის კვადრატის საშუალო მნიშვნელობა ყოველთვის მეტია ან ტოლია იმავე შემთხვევითი სიდიდის საშუალო მნიშვნელობის კვადრატისა.

მართლაც, ვთქვათ მოცემულია შემთხვევითი სიდიდე X, რომლის განაწილების კანონია:

$$X\begin{cases} x_1, & x_2, \dots, & x_n \\ p_1, & p_2, \dots, & p_n, \end{cases}$$

მაშინ X^2-იც შემთხვევითი სიდიდე იქნება შემდეგი განაწილების კანონით:

$$X^2\begin{cases} x_1^2, & x_2^2, \dots, & x_n^2 \\ p_1, & p_2, \dots, & p_n. \end{cases}$$

საშუალო მნიშვნელობის განმარტების თანახმად

$$E(X^2)=\sum_{i=1}^{n}x_i^2\,p_i \quad \text{და} \quad E(X)=\sum_{i=1}^{n}x_i\,p_i.$$

სხვაობა

$$E(X^2)-[E(X)]^2$$

წარმოვადგინოთ შემდეგნაირად:

$$E(X^2)-[E(X)]^2=E(X^2)-2\,[E(X)]^2+[E(X)]^2=$$
$$=E(X^2)-2E(X)\,E(X)+[E(X)]^2.$$

როგორც ვიცით,

$$\sum_{i=1}^{n}p_i=1.$$

განხილული სხვაობა გადავწეროთ შემდეგნაირად:

$$E(X^2)-[E(X)]^2=\sum_{i=1}^{n} x_i^2\, p_i-2E(X)\sum_{i=1}^{n} x_i\, p_i+$$

$$+[E(X)]^2\sum_{i=1}^{n} p_i=\sum_{i=1}^{n} x_i^2\, p_i-\sum_{i=1}^{n} 2E(X)\, x_i\, p_i+\sum_{i=1}^{n}[E(X)]^2\, p_i=$$

$$=\sum_{i=1}^{n}\{x_i^2-2x_i\,[E(X)]+[E(X)]^2\}\,p_i=\sum_{i=1}^{n}[x_i-E(X)]^2\, p_i\geqslant 0.$$

აქედან დავასკვნით, რომ

$$E(X^2)\geqslant [E(X)]^2. \tag{2.2}$$

სხვაობას

$$X-E(X)$$

X შემთხვევითი სიდიდის თავისი საშუალო მნიშვნელობიდან გადახრა ან უბრალოდ გადახრა ეწოდება.

6. გადახრის მათემატიკური ლოდინი ნულის ტოლია. მართლაც,

$$E[X-E(X)]=E(X)-E[E(X)]=E(X)-E(X)=0. \tag{2.3}$$

რადგანაც $E(X)$ თვითონ მუდმივი სიდიდეა, ამიტომ მისი მათემატიკური ლოდინიც იგივე იქნება.

მათემატიკური ლოდინი წარმოადგენს იმ „ფასს", რაზედაც უნდა „გაიყიდოს" შემთხვევითი სიდიდე, სანამ გავიგებდეთ, თუ რა რაცხვით მნიშვნელობას მიიღებს იგი.

შემთხვევითი სიდიდის მათემატიკური ლოდინის სწორად შეფასება ამ სიდიდის პირველი და ძირითადი დახასიათებაა. მაგრამ მარტო მათემატიკური ლოდინი მთლიანად ვერ განსაზღვრავს შემთხვევით სიდიდეს. მაგალითად, ვთქვათ თამაშის დროს მათემატიკური ლოდინი ნულის ტოლია, ე. ი. თამაში ტოლიღბლიანია; მოთამაშისათვის მთავარია სანაძლეოს სიდიდის განსაზღვრა, მოგების ან წაგების გადახრის ცოდნა მათემატიკური ლოდინიდან.

განვიხილოთ მსროლელის მაგალითი. ვთქვათ, მოხვედრის მათემატიკური ლოდინი ნულის ტოლია; ეს არაფერს არ გვეუბნება მსროლელის ოსტატობაზე; ამ შემთხვევაში მთავარია ვიცოდეთ, თუ რამდენად იქნება დაშორებული მიზნიდან რომელიმე მხარის მოხვედრა.

§ 3. შემთხვევითი სიდიდის დისპერსია

ზოგიერთ შემთხვევაში შესაძლებელია შემთხვევით სიდიდეზე გარკვეული წარმოდგენა შევიქმნათ მხოლოდ ერთი რიცხვის მოცემით. ასეთი რიცხვია აღებული შემთხვევითი სიდიდის მათემატიკური ლოდინი. მაგალითად, თუ გვინდა ერთმანეთს შევადაროთ ორი მსროლელი, რომელთა მათემატიკური ლოდინის მნიშვნელობა ცნობილია, მაშინვე განვსაზღვრავთ, თუ რომელი მსროლელი სჯობია. მაგრამ შემთხვევითი სიდიდე ყოველთვის ვერ განისაზღვრება მისი მარტო ერთი მახასიათებლით.

მაგალითისათვის განვიხილოთ ქვემეხიდან გასროლილი ყუმბარის დაცემის მანძილის განსაზღვრა. მანძილი ქვემეხიდან ყუმბარის დაცემის ადგილამდე აღვნიშნოთ X-ით. ეს მანძილი, რა თქმა უნდა, შემთხვევით სიდიდეს წარმოადგენს. თუ დამიზნება აბსოლუტურად ზუსტია, მაშინ საშუალო მნიშვნელობა ქვემეხიდან ყუმბარის დაცემის ადგილამდე მანძილის ტოლი იქნება.

ვთქვათ, ამ შემთხვევაში ცნობილია საშუალო მნიშვნელობა $E(X)$. შეგვიძლია თუ არა განვსაზღვროთ, როგორ განლაგდება მოხვედრები მიზნის გარშემო? მიზნიდან შორს იქნება თუ ახლოს ცალკეული მოხვედრები? ცხადია, ამ კითხვებზე მარტო საშუალო მნიშვნელობის ცოდნით ვერ ვუპასუხებთ. ამ შემთხვევაში შეგვიძლია ვთქვათ, რომ მოხვედრა იქნება მიზნის იქითაც და მიზნის აქეთაც. ამაზე მეტს ვერაფერს ვიტყვით.

ორი ქვემეხიდან რომ გავისროლოთ, შესაძლებელია მათთვის ერთი და იგივე იყოს საშუალო მნიშვნელობა. მაგრამ ერთი მიზანთან ახლოს მოხვედრებს იძლეოდეს, მეორე კი პირიქით. მიზნიდან შორს, ე. ი. პირველის მოხვედრები ნაკლებად გაფანტული იყოს მიზნის გარშემო, ვიდრე მეორე ქვემეხის მოხვედრები.

როგორც ვხედავთ, შემთხვევითი სიდიდის დასახასიათებლად მარტო მისი მათემატიკური ლოდინის ცოდნა არ არის საკმარისი. საჭიროა აგრეთვე ვიცოდეთ, თუ როგორია ამა თუ იმ შემთხვევითი სიდიდის გადახრა თავისი საშუალო მნიშვნელობიდან; ეს წარმოდგენას მოგვცემს იმაზე, თუ რამდენად გაფანტულია აღებული შემთხვევითი სიდიდე თავისი მათემატიკური ლოდინის გარშემო.

ვთქვათ, მოცემულია X და Y შემთხვევითი სიდიდეები, რომელთა განაწილების კანონებია:

$$X \begin{cases} 0{,}5 & -0{,}5 \\ 0{,}5 & 0{,}5 \end{cases}$$

$$Y \begin{cases} 50, & -50 \\ 0{,}5, & 0{,}5 \end{cases} \qquad (3.1)$$

როგორც ვხედავთ,

$$E(X)=E(Y)=0.$$

იმის მიუხედავად, რომ ორივე ნულის ტოლ საშუალო მნიშვნელობებს ღებულობენ, X შემთხვევითი სიდიდე შედარებით ნაკლებად გაფანტულია თავისი საშუალო მნიშვნელობის გარშემო, Y კი—უფრო მეტად და თავისი საშუალო მნიშვნელობისაგან დიდად განსხვავებულ მნიშვნელობებს იღებს.

ამგვარად, საჭიროა გვქონდეს X შემთხვევითი სიდიდის გაფანტულობის საზომი სიდიდე; ასეთი სიდიდე დაკავშირებული იქნება $X-E(X)$ გადახრასთან, რომელიც აგრეთვე შემთხვევით სიდიდეს წარმოადგენს. შემთხვევითი სიდიდე იქნება გადახრის აბსოლუტური მნიშვნელობაც $|X-E(X)|$.

ამგვარად, შემთხვევითი სიდიდის დასახასიათებლად, მის პირველ მახასიათებელ მათემატიკურ ლოდინთან ერთად, გამოგვადგება მეორე მახასიათებელი, რომელიც დაკავშირებულია მათემატიკური ლოდინიდან გადახრასთან.

თუ შემთხვევითი სიდიდის მიერ მიღებული ყველა რიცხვითი მნიშვნელობა მცირედ განსხვავდება ამ შემთხვევითი სიდიდის მათემატიკური ლოდინისაგან, მაშინ ვიტყვით, რომ შემთხვევითი სიდიდე ნაკლებად გაფანტულია.

როგორც მოცემული ცხრილებიდან ჩანს, ორი სხვადასხვა შემთხვევის დროს მათემატიკური ლოდინი ერთი და იგივე იყო, მაგალითად, მიზანთან ახლოს მოხვედრების შემთხვევაში და, აგრეთვე, მიზნიდან საკმაოდ შორს მოხვედრების შემთხვევაშიც. ორივე შემთხვევაში შეიძლება მათემატიკური ლოდინი ნულის ტოლიც იყოს, მაგრამ პირველ შემთხვევაში მცირე გაფანტულობა გვექნება, მეორე შემთხვევაში კი—საკმაოდ დიდი გაფანტულობა. პირველი შემთხვევა მსროლელის კარგ ოსტატობას გვიჩვენებს, მეორე კი — პირიქით, მსროლელის უვარგისობას.

ამგვარად, საჭიროა მოინახოს ამ გაფანტულობის მახასიათებელი სიდიდე. ასეთ მახასიათებლად პირველ რიგში მივიღოთ შემთხვევითი სიდიდის გადახრის მათემატიკური ლოდინი. მაგრამ, როგორც ვნახეთ, გადახრის მათემატიკური ლოდინი ნულის ტოლია. ცხადია, ასეთი მახასიათებელი სიდიდე ბევრს ვერაფერს მოგვცემს. ამგვარად, შემთხვევითი სიდიდის გადახრის მაგივრად უნდა განვიხილოთ ამ გადახრის აბსოლუტური მნიშვნელობა, მაშინ მისი მათემატიკური ლოდინი უკვე ყოველთვის აღარ იქნება ნულის ტოლი.

შევადგინოთ გადახრის აბსოლუტური მნიშვნელობის განაწილების კანონი. მას ასეთი სახე ექნება:

$$|X-E(X)|\begin{cases}|x_i-E(X)| \\ p_i\end{cases} \quad i=1,\ 2,\ldots,\ n.$$

მაშინ

$$E\{|X-E(X)|\}=\sum_{i=1}^{n}|x_i-E(X)|\,p_i.$$

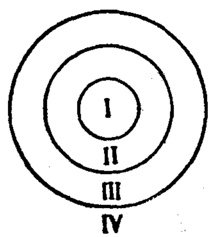
ამ უკანასკნელ გამოსახულებას, ე. ი. გადახრის აბსოლუტური მნიშვნელობის მათემატიკურ ლოდინს, საშუალო აბსოლუტური გადახრა ან, უბრალოდ, საშუალო გადახრა ეწოდება და M_x-ით აღინიშნება. ამგვარად,

$$M_x=E(|X-E(X)|). \tag{3.2}$$

ზემოთ მოცემული ცხრილების მიხედვით X და Y შემთხვევით სიდიდეთა საშუალო გადახრები იქნება:

$$M_x=0{,}5; \quad M_y=50.$$

ვთქვათ, ორმა მსროლელმა უნდა ისროლოს 12-ე ნახაზზე გამოსახულ მიზანში. I არეში მოხვედრისათვის მსროლელს 4 ქულა ეწერება, II არეში მოხვედრისათვის—3 ქულა, III არეში მოხვედრისათვის — 2 ქულა, ხოლო IV არეში მოხვედრისათვის, ე. ი. აცდენისათვის—1 ქულა (გასროლის უფლება ერთ ქულას მაინც აძლევს). ვთქვათ, ერთი მსროლელი 100 გასროლისას საშუალოდ 20-ს ახვედრებს პირველ არეში, მეორეში — 40-ს, მესამეში — 30-ს და მხოლოდ 10-ს აცდენს, მეორე მსროლელი კი ყოველი 100 გასროლისას პირველ არეში საშუალოდ 10-ს ახვედრებს, მეორეში — 40-ს, მესამეში — 40-ს და 10-ს კი აცდენს; მაშინ პირველი და მეორე მსროლელისათვის განაწილების კანონი სათანადოდ იქნება:



ნახ. 12

$$X\begin{cases}1, & 2, & 3, & 4, \\ 0{,}1, & 0{,}3, & 0{,}4, & 0{,}2;\end{cases}$$

$$Y\begin{cases}1, & 2, & 3, & 4, \\ 0{,}1, & 0{,}4, & 0{,}4, & 0{,}1.\end{cases}$$

პირველი მსროლელისათვის საშუალო მნიშვნელობაა $E(X)=2{,}7$, ხოლო მეორისათვის—$E(Y)=2{,}5$.

როგორც ვხედავთ, მათი მათემატიკური ლოდინები ერთმანეთისაგან მცირედ განსხვავდება. გამოვიანგარიშოთ მათი საშუალო გადახრანი:

$$M_x = |1-2{,}7|\,0{,}2 + |2-2{,}7|\,0{,}3 + |3-2{,}7|\,0{,}4 + |4-2{,}7|\,0{,}1 = 0{,}8.$$

$$M_y = |1-2{,}5|\,0{,}1 + |2-2{,}5|\,0{,}4 + |3-2{,}5|\,0{,}4 + |4-2{,}5|\,0{,}1 = 0{,}7.$$

როგორც ჩანს, იმის მიუხედავად, რომ პირველი მსროლელისათვის მათემატიკური ლოდინი ცოტათი მაინც მეტია, ვიდრე მეორე მსროლელისათვის, მეორე მსროლელის მიერ მოცემული მოხვედრანი უფრო ნაკლებად არის გაფანტული საშუალო მნიშვნელობის გარშემო.

საშუალო გადახრის საშუალებით შემთხვევითი სიდიდის დახასიათება პრაქტიკულად ძნელია, ამიტომ გადახრის შესაფასებლად სხვა საზომი შემოაქვთ.

რადგანაც გადახრა $X-E(X)$ შემთხვევითი სიდიდეა, მისი კვადრატი $[X-E(X)]^2$-იც შემთხვევითი სიდიდე იქნება, რომლის განაწილების კანონია:

$$[X-E(X)]^2 \begin{cases} [x_i-E(X)]^2. \\ p_i \end{cases}$$

ხოლო მათემატიკური ლოდინის განმარტების თანახმად,

$$E\,[X-E(X)]^2 = \sum_{i=1}^{n} [x_i-E(X)]^2\, p_i.$$

გადახრის კვადრატის საშუალო მნიშვნელობას შემთხვევითი სიდიდის დისპერსია ეწოდება. აღვნიშნოთ იგი $D(X)$-ით. ამგვარად,

$$D(X) = E\,[X-E(X)]^2. \tag{3.3}$$

თუ X უწყვეტი ტიპის შემთხვევითი სიდიდეა, მაშინ დისპერსიისათვის გვექნება:

$$D(X) = \int_{-\infty}^{\infty} [X-E(X)]^2 f(x)\,dx. \tag{3.4}$$

§ 4. დისპერსიის თვისებები

1. მუდმივის დისპერსია ნულის ტოლია.

ეს აშკარად ჩანს დისპერსიის განმარტებიდან; მუდმივიდან გადახრა ნულის ტოლია, მისი კვადრატიც და საშუალო მნიშვნელობაც ნული იქნება.

2. მუდმივი თანამამრავლი დისპერსიის ნიშნის გარეთ გამოდის კვადრატში ახარისხებული.

ვთქვათ, k მუდმივი რიცხვია და X—შემთხვევითი სიდიდე, მაშინ, განმარტების თანახმად,

$$D(kX)=E\,[kX-E(kX)]^2=E[kX-kE(X)]^2=$$
$$=k^2E\,[X-E(X)]^2=k^2D(X).$$

3. თუ შემთხვევითი სიდიდეები ურთიერთდამოუკიდებელია, მაშინ მათი ჯამის დისპერსია შესაკრებ შემთხვევით სიდიდეთა დისპერსიების ჯამის ტოლია.

ეს თვისება დავამტკიცოთ ორი შესაკრების შემთხვევაში; შესაკრებთა ნებისმიერი რიცხვისათვის მისი განზოგადება არავითარ სიძნელეს არ წარმოადგენს.

ვთქვათ, X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, რომელთაც ცალკე მათემატიკური ლოდინი აქვთ. მაშინ დისპერსიის განმარტების თანახმად,

$$D(X+Y)=E\,[X+Y-E(X+Y)]^2.$$

თუ მარჯვენა მხარეზე მდგომ გამოსახულებაში ჯამის საშუალო მნიშვნელობის თვისებას გამოვიყენებთ და კვადრატულ ფრჩხილებში მოთავსებულ გამოსახულებას კვადრატში ავახარისხებთ, გვექნება:

$$D(X+Y)=E\{[X-E(X)]^2+2[X-E(X)]\,[Y-E(Y)]+$$
$$+[Y-E(Y)]^2\}.$$

უკანასკნელ გამოსახულებაში ისევ ჯამის მათემატიკური ლოდინისა და დამოუკიდებელი ხდომილობათათვის ნამრავლის მათემატიკური ლოდინის თვისებას გამოვიყენებთ და დავწერთ:

$$D(X+Y)=E[X-E(X)]^2+2E[X-E(X)]\,E[Y-E(Y)]+$$
$$+E[Y-E(Y)]^2.$$

დისპერსიის განმარტების თანახმად, გვაქვს

$$E[X-E(X)]^2=D(X),$$
$$E[Y-E(Y)]^2=D(Y),$$

ხოლო, როგორც ვნახეთ, გადახრის საშუალო მნიშვნელობა ყოველთვის ნულის ტოლია. ამიტომ გვექნება:

$$D(X+Y)=D(X)+D(Y).$$

მრავალი შესაკრების შემთხვევაში განზოგადებას მკითხველს ვანდობთ.

4. ცხადია, აგრეთვე, შემდეგი თვისების სამართლიანობა: თუ X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ

$$D(X-Y)=D(X)+D(Y).$$

გადახრის ზომის დასახასიათებლად, რა თქმა უნდა, დისპერსია გამოდგება, რომელიც შემთხვევითი სიდიდის რიცხვით მნიშვნელობათა მათემატიკური ლოდინის გარშემო გაფანტულობაში გვარკვევს, მაგრამ პრაქტიკულად გადახრის ზომის უშუალო საზომად იყენებენ არა დისპერსიას, არამედ კვადრატულ ფესვს დისპერსიიდან, რომელსაც საშუალო კ ვ ა დ რ ა ტ უ ლ გ ა დ ა ხ რ ა ს უწოდებენ.

X შემთხვევითი სიდიდის საშუალო კვადრატული გადახრა σ_x-ით აღვნიშნოთ, ე. ი.

$$\sigma_x = \sqrt{D(X)} = \sqrt{\sum_{i=1}^{n} [x_i - E(X)]^2 p_i}. \tag{4.1}$$

σ_x მახასიათებლის გამოყენება საგრძნობლად ამარტივებს გამოთვლებს და ამიტომ ფართოდაა გავრცელებული პრაქტიკული ამოცანების გადაწყვეტისას.

X შ ე მ თ ხ ვ ე ვ ი თ ი ს ი დ ი დ ი ს ნ ო რ მ ი რ ე ბ უ ლ ი გ ა დ ა ხ რ ა ე წ ო დ ე ბ ა

$$\frac{X - E(X)}{\sqrt{D(X)}}$$

გ ა მ ო ს ა ხ უ ლ ე ბ ა ს.

ნ ო რ მ ი რ ე ბ უ ლ ი გ ა დ ა ხ რ ი ს დ ი ს პ ე რ ს ი ა ყ ო ვ ე ლ თ ვ ი ს ე რ თ ი ს ტ ო ლ ი ა. მართლაც,

$$D\left(\frac{X - E(X)}{\sqrt{D(X)}}\right) = \frac{1}{D(X)} D[X - E(X)] =$$

$$= \frac{1}{D(X)} \{D(X) - D[E(X)]\} = \frac{1}{D(X)} \cdot D(X) = 1.$$

ცხადია, აგრეთვე, რომ ნორმირებული გადახრის მათემატიკური ლოდინი ნულის ტოლია.

გამოვთვალოთ ბერნულის სქემის შემთხვევაში ჩვენთვის სასურველი ხდომილობის მოხდენათა m რიცხვის დისპერსია.

განვიხილოთ n შემთხვევითი სიდიდეები:

$$Z_1, Z_2, \ldots, Z_n.$$

ვიგულისხმოთ, რომ ყოველი მათგანი ემორჩილება განაწილების შემდეგ კანონს:

$$Z_i \begin{cases} 1, & 0, \\ p, & q. \end{cases}$$

როგორც უკვე ვნახეთ, ეს განაწილების კანონი ბერნულის სქემას ეთანხმება.

ჩვენთვის სასურველი ხდომილობის მოხდენათა რიცხვია

$$m=\sum_{i=1}^{n} Z_i.$$

როგორც გამოვთვალეთ,

$$E(m)=np. \tag{4,2}$$

m მეტად საჭირო შემთხვევითი სიდიდეა, განსაკუთრებით გამოყენებით დარგებში. ამიტომ საჭიროა მისი ძირითადი მახასიათებლის ცოდნა.

ვიგულისხმოთ, რომ Z_i სიდიდეები წყვილ-წყვილად დამოუკიდებელია. თუ გამოვიყენებთ ასეთი შემთხვევითი სიდიდეებისათვის ჯამის დისპერსიის თვისებას, მივიღებთ:

$$D(m)=\sum_{i=1}^{n} D(Z_i).$$

რადგანაც ყოველ მათგანს ერთი და იგივე დისპერსია აქვს, ამიტომ

$$D(m)=nD(Z_i).$$

როგორც ვნახეთ, ყოველი Z_i-ის მათემატიკური ლოდინი p-ს ტოლია. თუ შევადგენთ განაწილების კანონს $(Z_i-p)^2$-თვის, გვექნება:

$$(Z_i-p)^2 \begin{cases} (1-p)^2, & p^2 \\ p, & 1-p, \end{cases}$$

საიდანაც

$$D(Z_i)=E(Z_i-p)^2=(1-p^2)\,p+p^2(1-p)=p-p^3+$$
$$+p^2-p^3=p\,(1-p)=pq,$$

ხოლო

$$D(m)=n\cdot D(Z_i)=npq. \tag{4.3}$$

ამგვარად, ჩვენთვის სასურველი ხდომილობის მოხდენათა რიცხვის დისპერსია npq-ს ტოლია.

იმ შემთხვევაში, როდესაც $n\to\infty$, ე. ი. როცა ცდათა რიცხვი დიდდება, n-ის ზრდასთან ერთად m-ის მათემატიკური ლოდინიც იზრდება და ამავე რიგით იზრდება m-ის დისპერსიაც.

სწორედ ამ პრინციპზეა დამყარებული ჩ ე ბ ი შ ე ვ ი ს თეორემის დამტკიცება.

$\frac{m}{n}$ ფარდობითი სიხშირის დისპერსია იქნება:

$$D\left(\frac{m}{n}\right)=\frac{1}{n^2}D(m)=\frac{pq}{n} \quad . \tag{4.4}$$

როგორც ვხედავთ, ჩვენთვის სასურველი ხდომილობის მოსვლათა რიცხვის დისპერსია ცდათა რიცხვის ზრდასთან ერთად იზრდება და, პირიქით — ჩვენთვის სასურველი ხდომილობის ფარდობითი სიხშირის დისპერსია ცდათა რიცხვის ზრდასთან ერთად კლებულობს. რაც უფრო მეტია ცდათა რიცხვი, მით უფრო დიდი იქნება ხდომილობის მოხდენის რიცხვის შემთხვევითი გაფანტულობა თავისი საშუალო np მნიშვნელობის გარშემო. პირიქით, რაც უფრო მეტია ცდათა რიცხვი, მით უფრო ნაკლები იქნება ხდომილობის ფარდობითი სიხშირის შემთხვევითი გაფანტულობა თავისი საშუალო p მნიშვნელობის გარშემო.

ფარდობითი სიხშირის მათემატიკური ლოდინი p-ს ტოლია, რადგანაც

$$E\left(\frac{m}{n}\right)=\frac{1}{n}E(m)=\frac{1}{n}np=p. \tag{4.5}$$

რაც უფრო დიდია ის ინტერვალი, რომელშიც თავის რიცხვით მნიშვნელობებს იღებს შემთხვევითი სიდიდე, ე. ი. რაც უფრო გაფანტულია შემთხვევითი სიდიდის რიცხვითი მნიშვნელობანი, მით უფრო დიდი იქნება დისპერსია. ამგვარად, დისპერსია გაფანტულობის საზომია. დისპერსია გვიჩვენებს შემთხვევითი სიდიდის რიცხვით მნიშვნელობათა განაწილებას მათემატიკური ლოდინის გარშემო.

§ 5. ალბათური გადახრა

ვთქვათ, X არის მანძილი ყუმბარის გასროლის ადგილიდან დაცემის წერტილამდე. ცხადია, X შემთხვევით სიდიდეს წარმოადგენს. ვიგულისხმოთ, რომ დამიზნება აბსოლუტურად ზუსტია; მაშინ საშუალო მნიშვნელობა ტოლი იქნება მანძილისა ყუმბარის გასროლის ადგილიდან მიზნამდე.

თუ მიზნიდან რომელიმე α-ს ტოლ მონაკვეთს გადავზომავთ ორივე მხრიდან და თვით α მცირე სიდიდეა, მაშინ ყუმბარათა მცირე ნაწილი მოხვდება $(-\alpha, \alpha)$ შუალედში, ე. ი. ნაკლებად მოსალოდნელია, რომ $|X-E(X)|$ გადახრა α-ზე მცირე იქნება; უფრო მოსალოდნელია, რომ აღნიშნული გადახრა მეტი იქნება α მცირე რიცხვზე. თუ α-ს, რომელიც ნებისმიერად იყო შერჩეული, თანდათან გავადიდებთ, მაშინ $(-\alpha, \alpha)$ შუალედის შიგნით ყუმბარების თანდათან უფრო მეტი რაო-

დენობა მოხვდება, ხოლო, თუ α-ს საკმაოდ დიდს შევარჩევთ, მაშინ თითქმის ყველა ყუმბარა $(-\alpha, \alpha)$ შუალედის შიგნით მოხვდება.

მაშასადამე, α-ს თანდათანობით გადიდებისას $|X-E(X)|<\alpha$ უტოლობის ალბათობა თანდათან იზრდება, ხოლო $|X-E(X)|>\alpha$ უტოლობის ალბათობა თანდათან მცირდება; α-ს გადიდების დროს შეგვხვდება α-ს ისეთი α_0 მნიშვნელობა, რომ ორივე ზემოთ დაწერილი უტოლობა ერთნაირად მოსალოდნელი იქნება და, თუ მხედველობაში მივიღებთ, რომ $|X-E(X)|=\alpha_0$ ტოლობის ალბათობა უმნიშვნელოდ მცირეა, შეგვიძლია დავწეროთ:

$$P(|X-E(X)|>\alpha_0)=P(|X-E(X)<\alpha_0)=\frac{1}{2}. \tag{5.1}$$

იმ α_0 რიცხვს, რომლისთვის $X-E(X)$ გადახრის აბსოლუტური მნიშვნელობა ერთნაირად შესაძლებელია მასზე მეტიც იყოს და ნაკლებიც, X შემთხვევითი სიდიდის ალბათური გადახრა ეწოდება.

α_0-ის სიდიდე დამოკიდებულია სასროლი იარაღის ხარისხზე და მიზნის გარშემო ყუმბარის მოხვედრათა გაფანტულობის შესწავლაში გვეხმარება. თუ α_0 საკმაოდ მცირეა, ეს იმას ნიშნავს, რომ ყველა გასროლილი ყუმბარის ნახევარი მაინც ახლოს ხვდება მიზანთან; ხოლო თუ α_0 საკმაოდ დიდია, ეს იმას ნიშნავს, რომ ყველა გასროლილი ყუმბარის ნახევარზე მეტი მიზნიდან საკმაოდ შორს ხვდება. ეს კი იმას გვიჩვენებს, რომ ყუმბარათა მოხვედრები ცენტრის გარშემო საკმაოდ დიდად არის გაფანტული.

გამოთვლებისათვის ალბათური გადახრის გამოყენება რთულია და ამიტომ ყოველთვის ცდილობენ ის საშუალო კვადრატული გადახრის საშუალებით გამოხატონ, რაც საგრძნობლად ამარტივებს გამოთვლებს. ალბათური გადახრით განსაკუთრებით საარტილერიო საქმეში სარგებლობენ. მართლაც, იქ მისი გამოყენება შედარებით ადვილია და სხვა მახასიათებელზე უფრო ეფექტურია შემთხვევითი სიდიდის დასახასიათებლად.

§ 6. თეორემები საშუალო კვადრატულ გადახრაზე

საშუალო კვადრატულ გადახრას შემთხვევითი სიდიდის სხვა მახასიათებლებთან (საშუალო გადახრა, ალბათური გადახრა და სხვ.) შედარებით განსაკუთრებული თვისებები აქვს, რითაც იგი მეტად საჭირო მახასიათებლად გვევლინება.

თეორემა I. ურთიერთდამოუკიდებელ შემთხვევით სიდიდეთა ჯამის საშუალო კვადრატული გადახრა უდ-

რის კვადრატულ ფესვს შესაკრებ შემთხვევით სიდიდეთა საშუალო კვადრატული გადახრების კვადრატების ჯამიდან.

დამტკიცება: ვთქვათ, $X_1, X_2, \ldots, X_n$ ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, რომელთა საშუალო კვადრატული გადახრებია შესაბამისად:

$$\sigma_1, \ \sigma_2, \ldots, \ \sigma_n.$$

აღვნიშნოთ

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i,$$

ხოლო X-ის საშუალო კვადრატული გადახრა აღვნიშნოთ σ-თი.

განმარტების თანახმად,

$$\sigma = \sqrt{E[X - E(X)]^2}.$$

როგორც ვნახეთ, დამოუკიდებელ შემთხვევით სიდიდეთა ჯამის დისპერსია შესაკრებ შემთხვევით სიდიდეთა დისპერსიების ჯამის ტოლია

$$D(X) = \sum_{i=1}^{n} D(X_i).$$

აქედან

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

და

$$\sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2}. \tag{6.1}$$

ამით თეორემა დამტკიცებულია.

თეორემა II. n ურთიერთდამოუკიდებელ ერთნაირად განაწილებულ შემთხვევით სიდიდეთა საშუალო არითმეტიკულის საშუალო კვადრატული გადახრა $\sqrt{n}$-ჯერ ნაკლებია, ვიდრე თითოეული ცალკე შემადგენელი სიდიდის საშუალო კვადრატული გადახრა.

დამტკიცება: n გაზომვის შედეგი იყოს

$$X_1, \ X_2, \ldots, \ X_n.$$

ეს მნიშვნელობანი, პირობის თანახმად, ურთიერთდამოუკიდებელია. მათი საშუალო არითმეტიკული აღვნიშნოთ $\overline{X}$-ით:

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}. \qquad (6.2)$$

მაშინ

$$E(\overline{X})=\frac{1}{n}[E(X_1)+E(X_2)+\cdots+E(X_n)].$$

რადგანაც, პირობის თანახმად, $X_1, X_2,\ldots, X_n$ შემთხვევითი სიდიდეები ერთნაირადაა განაწილებული, ამიტომ მათ ერთი და იგივე მათემატიკური ლოდინი ექნება.

ამგვარად,

$$E(\overline{X})=\frac{1}{n}nE(X_i)=E(X_i),$$

ე. ი. საშუალო არითმეტიკულის მათემატიკური ლოდინი იგივე ყოფილა, რაც რომელიმე ცალკე შესაკრების მათემატიკური ლოდინი.

I თეორემის თანახმად, $X_1+X_2+\ldots+X_n$ ჯამის საშუალო კვადრატული გადახრა იქნება:

$$\sigma=\sqrt{\sum_{i=1}^{n}\sigma_i^2}=\sqrt{n\sigma_i^2}=\sigma_i\sqrt{n}, \qquad (6.3)$$

სადაც σ_i ცალკეული გაზომვის შედეგის საშუალო კვადრატული გადახრაა.

$\overline{X}$ საშუალო არითმეტიკულის საშუალო კვადრატული $\sigma_{\overline{x}}$ გადახრა იქნება

$$\sigma_{\overline{x}}=\frac{\sigma}{n}=\frac{\sigma_i}{\sqrt{n}},$$

საიდანაც

$$\sigma_{\overline{x}}\sqrt{n}=\sigma_i.$$

თეორემა დამტკიცებულია.

§ 7. ვარიაციის კოეფიციენტი

როგორც ვიცით, საშუალო კვადრატული გადახრა შემთხვევითი სიდიდის გაფანტულობის საზომს წარმოადგენს, მაგრამ ის მთლიანად ყოველთვის როდი გვაძლევს ნათელ სურათს ამ გაფანტულობის შესახებ. მაგალითად, თუ გვაქვს ორი შემთხვევითი სიდიდე, რომელთაც ერთი და იგივე საშუალო კვადრატული გადახრა აქვთ, მაგრამ მათი საშუალო მნიშვნელობანი საგრძნობლად განსხვავებულია, მაშინ საშუალო კვადრატული გადახრის მოცემით ამ შემთხვევით სიდიდეთა გაფან-

ტულობას მთლიანად ვერ დავახასიათებთ. ამიტომ საჭირო ხდება ე. წ. ვარიაციის კოეფიციენტის შემოყვანა, რომელიც წარმოადგენს საშუალო კვადრატული გადახრის საშუალო მნიშვნელობასთან შეფარდებას და აღინიშნება V-თი:

$$V = \frac{\sigma}{\overline{X}}. \tag{7.1}$$

მაგალითად, ვთქვათ, X შემთხვევითი სიდიდისათვის $\sigma_x = 1{,}6$, ხოლო $\overline{X} = 4{,}8$; Y შემთხვევითი სიდიდისათვის კი $\sigma_y = 1{,}6$, ხოლო $\overline{Y} = 128$, მაშინ გვექნება

$$V_x = \frac{1{,}6}{4{,}8} = 0{,}333,$$

$$V_y = \frac{1{,}6}{128} = 0{,}012.$$

ან, თუ მიღებულ მნიშვნელობებს გამოვხატავთ პროცენტებით, დავწერთ:

$$V_x = 33{,}3\%,$$

$$V_y = 1{,}2\%.$$

ვარიაციის კოეფიციენტი გვარკვევს შემთხვევით სიდიდეთა ფარდობით გაფანტულობაში და მისი დახმარებით შემთხვევით სიდიდეთა გაფანტულობის შედარება შესაძლებელია.

§ 8. მომენტები

ვთქვათ, მოცემულია X შემთხვევითი სიდიდე, რომელიც რიცხვით მნიშვნელობათა სასრულ რაოდენობას იღებს და მისი განაწილების კანონია:

$$X \left\{ \begin{array}{l} x_i \\ p_i \end{array} \right., \quad i = 1, 2, \ldots, n.$$

შევადგინოთ ამ შემთხვევითი სიდიდის სხვადასხვა ხარისხის მათემატიკური ლოდინი, ე. ი. შევადგინოთ შემდეგი სახის გამოსახულებანი:

$$\sum_{i=1}^{n} x_i p_i; \quad \sum_{i=1}^{n} x_i^2 p_i; \quad \sum_{i=1}^{n} x_i^3 p_i; \ldots; \quad \sum_{i=1}^{n} x_i^k p_i,$$

ან საზოგადოდ შევადგინოთ შემდეგი გამოსახულება:

$$\nu_k = \sum_{i=1}^{n} x_i^k p_i. \tag{8.}$$

ამ უკანასკნელ გამოსახულებას X შემთხვევითი სიდიდის k-ური რიგის მომენტი ეწოდება (k საზოგადოდ შეიძლება წილადიც იყოს).

გამოსახულებას

$$\mu_k = E[X - E(X)]^k \tag{8.2}$$

X შემთხვევითი სიდიდის ცენტრალური მომენტი ეწოდება. კერძო შემთხვევაში, როდესაც $k=1$, მივიღებთ:

$$\mu_1 = 0,$$

ხოლო, როდესაც $k=2$-ს,

$$\mu_2 = E[X - E(X)]^2 = D(X).$$

მაღალი რიგის μ_k ცენტრალური მომენტები უმნიშვნელოვანეს როლს ასრულებს თეორიულ ფიზიკაში.

ν_k მომენტებს საწყისი მომენტები ეწოდება, განსხვავებით ცენტრალური მომენტებისაგან. (8.1)-ის ძალით დავწერთ:

$$\nu_k = E(X^k).$$

დავამყაროთ კავშირი საწყის და ცენტრალურ მომენტებს შორის:

$$\nu_1 = E(X),$$

ე. ი. პირველი საწყისი მომენტი მათემატიკური ლოდინია. საზოგადოდ

$$\mu_n = E\,[X - E(X)]^n = E \sum_{k=0}^{n} C_n^k\,[-E(X)]^{n-k}\, X^k =$$

$$= \sum_{k=0}^{n} C_n^k\,[-E(X)]^{n-k}\, E(X^k) = \sum_{k=0}^{n} C_n^k\,[-E(X)]^{n-k}\, \nu_k.$$

მაგრამ, რადგან $\nu_1 = E(X)$, ამიტომ

$$\mu_n = \sum_{k=0}^{n} C_n^k\,(-\nu_1)^{n-k}\nu_k = \sum_{k=0}^{n} (-1)^{n-k}\, C_n^k\, \nu_1^{n-k}\, \nu_k =$$

$$= \sum_{k=0}^{n-2} (-1)^{n-k}\, C_n^k\, \nu_k\, \nu_1^{n-k} + \nu_n - n\nu_1\, \nu_{n-1}. \tag{8.3}$$

აქედან

$$\begin{aligned} \mu_0 &= 1, \\ \mu_1 &= 0, \\ \mu_2 &= \nu_2 - \nu_1^2, \\ \mu_3 &= \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3, \\ \mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4. \end{aligned} \tag{8.4}$$

ეს ~~პირველი~~ მომენტები მნიშვნელოვან როლს ასრულებს სტატის-ტიკაში.

$$m_k = E\,[X-a]^k \tag{8.5}$$

გამოსახულებას, სადაც a მათემატიკური ლოდინია, k-ური რიგის აბსოლუტური მომენტი ეწოდება.

$E\,|X|^k$ გამოსახულება კი იქნება k-ური რიგია საწყისი აბსოლუ-ტური მომენტი.

§ 9. საშუალო არითმეტიკული

ვთქვათ, დაკვირვების შედეგად მიღებული გვაქვს შემდეგი სიდი-დეები:

$$x_1,\ x_2, \ldots,\ x_n.$$

მაშინ მათი საშუალო არითმეტიკული იქნება:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \tag{9.1}$$

ან მოკლედ:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

გავიხსენოთ საშუალო არითმეტიკულის ზოგიერთი თვისება:

1. ჯამის საშუალო არითმეტიკული უდრის შესაკრებთა საშუალო არითმეტიკულების ჯამს, ხოლო სხვაობის საშუალო არითმეტიკული უდრის საკლებისა და მაკლების საშუალო არითმეტიკულების სხვაობას.

დამტკიცება. ვთქვათ, x და y შემთხვევითი სიდიდეებია, ხოლო z—მათი ჯამი

$$z = x + y,$$

მაშინ, განმარტების თანახმად, დავწერთ:

$$z = \frac{1}{n} \sum_{i=1}^{n} z_i,$$

მაგრამ

$$z_i = x_i + y_i,$$

ამიტომ

$$\overline{z} = \frac{1}{n} \sum_{i=1}^{n} (x_i + y_i),$$

ან

$$\bar{z}=\frac{1}{n}\sum_{i=1}^{n} x_i+\frac{1}{n}\sum_{i=1}^{n} y_i=\bar{x}+\bar{y}.$$

ამგვარად, მივიღეთ:

$$\overline{x+y}=\bar{x}+\bar{y}. \tag{9.2}$$

2. თუ გვაქვს x შემთხვევითი სიდიდე და k მუდმივი რიცხვი, მაშინ მათი ჯამის საშუალო არითმეტიკული უდრის x-ის საშუალო არითმეტიკულისა და k მუდმივი რიცხვის ჯამს.

რადგანაც მუდმივის საშუალო არითმეტიკული იგივე მუდმივია, ამიტომ (9.2) ფორმულის გამოყენებით დავწერთ:

$$\overline{x+k}=\bar{x}+\bar{k}=\bar{x}+k.$$

3. მუდმივი თანამამრავლი გამოდის საშუალო არითმეტიკულის ნიშნის გარეთ.

ვთქვათ, x შემთხვევითი სიდიდეა და k — მუდმივი რიცხვი, მათი ნამრავლი აღვნიშნოთ y-ით:

$$y=kx.$$

ცხადია, რომ,

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} kx_i=k\,\frac{1}{n}\sum_{i=1}^{n} x_i=k\bar{x}.$$

ამგვარად, მივიღეთ:

$$\overline{kx}=k\bar{x}.$$

4. საშუალო არითმეტიკულიდან ცალკეულ მნიშვნელობათა გადახრების ჯამი ნულის ტოლია.

დამტკიცება. ვთქვათ, x შემთხვევით სიდიდეს შეუძლია მიიღოს შემდეგი რიცხვითი მნიშვნელობანი

$$x_1,\ x_2, \ldots,\ x_n,$$

რომელთაც ვარიანტებს უწოდებენ; ცალკეული მნიშვნელობიდან, ანუ ვარიანტიდან გადახრა იქნება

$$x_i-\bar{x}.$$

ვიანგარიშოთ შემდეგი ჯამი:

$$\sum_{i=1}^{n}(x_i-\bar{x}).$$

გვაქვს:

$$\sum_{i=1}^{n}(x_i-\bar{x})=\sum_{i=1}^{n}x_i-n\bar{x}=n\cdot\frac{1}{n}\sum_{i=1}^{n}x_i-n\bar{x}=n\bar{x}-n\bar{x}=0,$$

ე. ი.

$$\sum_{i=1}^{n}(x_i-\bar{x})=0.$$

5. საშუალო არითმეტიკულიდან გადახრათა კვადრატების ჯამი სხვა ნებისმიერი რიცხვიდან გადახრათა კვადრატების ჯამზე ნაკლებია.

დამტკიცება. ვთქვათ, x შემთხვევითი სიდიდეა და a რაიმე რიცხვი, მაშინ x შემთხვევითი სიდიდის მიერ მიღებულ რიცხვით მნიშვნელობათა a მუდმივისაგან გადახრათა კვადრატების ჯამი შეიძლება წარმოვიდგინოთ შემდეგნაირად:

$$\sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}[(x_i-\bar{x})-(a-\bar{x})]^2=$$

$$=\sum_{i=1}^{n}\left\{(x_i-\bar{x})^2-2(a-\bar{x})(x_i-\bar{x})+(a-\bar{x})^2\right\}=$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2-2(a-\bar{x})\sum_{i=1}^{n}(x_i-\bar{x})+n(a-\bar{x})^2.$$

მეოთხე თვისების თანახმად დავწერთ:

$$\sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(a-\bar{x})^2.$$

რადგანაც ამ უკანასკნელი ტოლობის მარჯვენა მხარის მეორე შესაკრები დადებითი სიდიდეა, დავასკვნით, რომ

$$\sum_{i=1}^{n}(x_i-\bar{x})^2\leqslant\sum_{i=1}^{n}(x_i-a)^2.$$

§ 10. აწონილი საშუალო

ვთქვათ, x შემთხვევითი სიდიდე ღებულობს შემდეგ მნიშვნელობებს, ე. წ. ვარიანტებს:

$$x_1,\ x_2,\ldots,\ x_n.$$

ვიგულისხმოთ, რომ x შემთხვევითი სიდიდე x_1 მნიშვნელობას ღებულობს n_1-ჯერ, x_2-ს—n_2-ჯერ და ა. შ. x_n-ს—n_n-ჯერ, მაშინ x შემთხვევითი სიდიდის საშუალო არითმეტიკული იქნება:

$$\bar{x}=\frac{n_1x_1+n_2x_2+\cdots+n_nx_n}{n}, \tag{10.1}$$

$$(n_1+n_2+\cdots+n_n=n).$$

მიღებული საშუალო მნიშვნელობა მოკლედ დაიწერება შემდეგნაირად:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}n_ix_i.$$

ასეთ საშუალოს ეწოდება აწონილი საშუალო, ხოლო $n_1, n_2, \cdots n_n$ სიდიდეებს—$x_1, x_2, \ldots, x_n$ სიდიდეთა წონები.

მაგალითი. ვთქვათ, გვაქვს ელექტრონათურათა 4 ჯგუფი. თითოეულ ჯგუფში შესაბამისად არის 5, 10, 15, 20 ნათურა; მათი მუშაობის ხანგრძლიობის დრო პირველი ჯგუფისათვის იყოს 1200 საათი, მეორე ჯგუფისათვის — 1300 საათი, მესამისათვის — 1400 საათი და მეოთხე ჯგუფისათვის — 1500 საათი. მონახეთ აღებულ ნათურათა მუშაობის ხანგრძლიობის დრო.

ა მ ო ხ ს ნ ა. რიცხვები: 5, 10, 15, 20 წარმოადგენს 1200, 1300, 1400, 1500 რიცხვების სათანადო წონებს. ამ მაგალითის ამოსახსნელად გამოვიყენებთ (10.1) ფორმულას და დავწერთ

$$\bar{x}=\frac{5\cdot1200+10\cdot1300+15\cdot1400+20\cdot1500}{50}=1400 \text{ საათი.}$$

მაშასადამე, მოცემულ ნათურათა მუშაობის ხანგრძლიობის საშუალო დრო ყოფილა 1400 საათი.

§ 11. საშუალო ჰარმონიული

ვთქვათ, x შემთხვევითი სიდიდე ღებულობს

$$x_1, x_2, \cdots, x_n$$

რიცხვით მნიშვნელობებს. მათი შებრუნებული სიდიდეები იქნება:

$$\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n},$$

ხოლო ამ შებრუნებულ სიდიდეთა საშუალო არითმეტიკული იქნება:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}.$$

ამ უკანასკნელი გამოსახულების შებრუნებულ სიდიდეს **ეწოდება განხილულ სიდიდეთა საშუალო ჰარმონიული,** ე. ი. თუ საშუალო ჰარმონიულს აღვნიშნავთ $\bar{x}_h$-ით, დავწერთ:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}, \tag{11.1}$$

მაგალითი. მონახეთ 1, 2, 3, 4, 5 რიცხვთა საშუალო ჰარმონიული.

ამოხსნა. საშუალო ჰარმონიულის განმარტების თანახმად დავწერთ:

$$\bar{x}_h = \frac{5}{\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = \frac{300}{137} \simeq 2{,}18.$$

§ 12. საშუალო გეომეტრიული

ვთქვათ, x **შემთხვევითი სიდიდე** ღებულობს რიცხვით მნიშვნელობებს:

$$x_1, x_2, \ldots, x_n.$$

მათი საშუალო **გეომეტრიული** აღვნიშნოთ $\bar{x}_g$-ით, მაშინ ის იქნება n-ური **რიგის** ფესვი ყველა აღებულ რიცხვთა ნამრავლიდან, ე. ი.

$$\bar{x}_g = \sqrt[n]{x_1 \, x_2 \cdots x_n}.$$

მაგალითი. იპოვეთ წინა მაგალითში მოცემული რიცხვების საშუალო გეომეტრიული.

ამოხსნა:

$$\bar{x}_g = \sqrt[5]{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = \sqrt[5]{120} \simeq 2{,}6052.$$

§ 13. გენერალური საშუალო

გენერალური საშუალოს შინაარსი რომ გავარკვიოთ, საჭიროა ვიცოდეთ, თუ რა არის გენერალური ერთობლიობა, ან, როგორც მას უწოდებენ, სტატისტიკური ერთობლიობა.

გენერალური ერთობლიობა ეწოდება ცალკეულ ელემენტთა ისეთ ზოგად ერთობლიობას, რომელიც მთლიანადაა გაერთიანებული რომელიმე ნიშნის მიხედვით, ხოლო სხვა ნიშნების მიხედვით დაყოფილია ცალკეულ ჯგუფებად.

გენერალური ერთობლიობის ცალკეულ ელემენტებს მისი წევრები ეწოდება. იმ ნიშნებს, რომლებითაც გენერალური ერთობლიობა იყოფა ჯგუფებად, ერთობლიობის არგუმენტებს უწოდებენ. გენერალური ერთობლიობის წევრთა რიცხვს მისი მოცულობა ეწოდება. ცალკეულ ჯგუფებში წევრთა რიცხვს არგუმენტში სათანადო ვარიანტთა სიხშირე ეწოდება.

მაგალითი. ვთქვათ, გარკვეულ დროში რომელიმე ქარხანაში მუშათა რაოდენობაა 12000.

აქ გვაქვს გენერალური ერთობლიობა 12000 კაცისაგან შემდგარი, რომლებიც ერთი ნიშნის (ისინი მუშებია) მიხედვით არიან გაერთიანებული. სხვა რომელიმე ნიშნის მიხედვით, მაგალითად, შრომის კატეგორიის მიხედვით ისინი შეიძლება დავყოთ რამდენიმე ჯგუფად: მელითონეებად, მეჩუქურთმეებად, მტვირთავებად და ა. შ. აქ ნიშანი — შრომითი კატეგორია წარმოადგენს გენერალური ერთობლიობის არგუმენტს; ცალკეული მუშები გენერალური ერთობლიობის წევრებია. მუშათა რიცხვი 12000 ერთობლიობის მოცულობას წარმოადგენს. ცალკეული ჯგუფები: მელითონეები, მეჩუქურთმეები, მტვირთავები და ა. შ. წარმოადგენენ „შრომის კატეგორიის“ არგუმენტის ვარიანტებს.

განხილული გენერალური ერთობლიობა სხვა რომელიმე ნიშნის მიხედვით, მაგალითად, მუშათა ასაკის მიხედვით, შესაძლებელია რამდენიმე ჯგუფად დაიყოს. ამ შემთხვევაში ასაკი იქნება გენერალური ერთობლიობის არგუმენტი.

განვიხილოთ კიდევ ასეთი მაგალითი. დავუშვათ, რომ ცნობილია 10000 ჰექტარზე სიმინდთა ერთობლიობა, რომელთა მოსავალი განაწილებულია შემდეგნაირად (ცხრილი 3):

ცხრილი 3

მოსავალი ერთ ჰექტარზე ცენტნერობით	14	14,5	15	15,4	16
ჰექტართა რაოდენობა	900	1300	2100	2500	3200

აქ გენერალური ერთობლიობა (10000 ჰექტარზე სიმინდის მოსავალი) მოსავლიანობის ნიშნის მიხედვით დაყოფილია ჯგუფებად.

მაშასადამე, გენერალური ერთობლიობის არგუმენტს მოსავლიანობა წარმოადგენს. არგუმენტის ვარიანტები კი არის რიცხვები: 14; 14,5; 15; 15,4; 16. ცალკეული ჰექტარი ერთობლიობის წევრია. რიცხვი

10000 ერთობლიობის მოცულობაა. ვარიანტთა სიხშირეები ჯგუფებში სათანადოდ არის: 900, 1300, 2100, 2500, 3200.

გენერალური საშუალო ეწოდება გენერალური ერთობლიობის არგუმენტების საშუალო არითმეტიკულს, რომელიც აწონილია არგუმენტის ვარიანტების სიხშირის მიხედვით.

ამ განმარტებიდან გამომდინარეობს, რომ გენერალური საშუალო აწონილი საშუალოა, ის გამოითვლება (10.1) ფორმულით. ვარიანტთა წონებს მათი სიხშირეები გამოხატავს.

ვთქვათ N არის გენერალური ერთობლიობის ელემენტთა რაოდენობა, რომლის არგუმენტია x.

ვთქვათ, x არგუმენტს აქვს k ვარიანტი:

$$x_1, x_2 \cdots, x_k.$$

ეს იმას ნიშნავს, რომ გენერალური ერთობლიობა ვარიანტთა რიცხვის მიხედვით k ჯგუფად იყოფა. $x_1, x_2, \ldots, x_k$ ვარიანტების ჯგუფთა სიხშირეები სათანადოდ აღვნიშნოთ $n_1, n_2, \ldots, n_k$ სიდიდეებით.

ცხადია,

$$N = n_1 + n_2 + \cdots + n_k.$$

აქ $n_1, n_2, \ldots, n_k$ არის სათანადოდ $x_1, x_2, \ldots, x_k$ ვარიანტთა წონები და ამიტომ გენერალური საშუალო (10.1) ფორმულის მიხედვით იქნება:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{N} = \frac{1}{N} \sum_{i=1}^{k} n_i x_i. \qquad (13.1)$$

ხოლო

$$\frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{N} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \cdots + \frac{x_k}{N} x_k. \qquad (13.2)$$

(13.2) გამოსახულებაში შემავალ

$$\frac{n_1}{N} = p_1, \quad \frac{n_2}{N} = p_2, \ldots, \quad \frac{n_k}{N} = p_k \qquad (13.3)$$

სიდიდეებს, სათანადოდ, ვ ა რ ი ა ნ ტ თ ა ხ ვ ე დ რ ე ბ ს უწოდებენ.

ცხადია, ყველა ხვედრის ჯამი ერთის ტოლია; მართლაც

$$p_1 + p_2 + \cdots p_k =$$

$$= \frac{n_1}{N} + \frac{n_2}{N} + \cdots + \frac{n_k}{N} = \frac{n_1 + n_2 + \cdots + n_k}{N} = \frac{N}{N} = 1.$$

(13.3) ფორმულის საფუძველზე (13.1) ფორმულას წარმოვადგენთ შემდეგნაირად:

$$\bar{x} = p_1 x_1 + p_2 x_2 + \cdots + p_k x_k = \sum_{i=1}^{k} p_i x_i. \qquad (13.4)$$

ამგვარად, მივიღეთ, რომ გენერალური საშუალო ვარიანტების მათ ხვედრებზე ნამრავლთა ჯამის ტოლია.

მაგალითი. წინა მაგალითის მონაცემების მიხედვით განვსაზღვროთ მოსავლიანობის გენერალური საშუალო.

ამოხსნა: 14; 14,5; 15; 15,4; 16 ვარიანტების სიხშირეები იქნება სათანადოდ 900, 1300, 2100, 2500, 3200. ახლა 1 ჰექტარიდან მოსავლიანობის გენერალური საშუალოს გამოვთვლით (13.1) ფორმულით:

$$\bar{x} = \frac{900 \cdot 14 + 1300 \cdot 14,5 + 2100 \cdot 15 + 2500 \cdot 15,4 + 3200 \cdot 16}{10000} =$$

$$= 15,265.$$

ამგვარად, განხილულ შემთხვევაში ჰექტარიდან საშუალო მოსავალი უდრის 15,265 ცენტნერს.

§ 14. შერჩევითი საშუალო

გენერალური ერთობლიობის შესწავლის მიზნით, მისგან შემთხვევითი შერჩევის საფუძველზე გამოყოფენ გარკვეულ ნაწილს, რომელსაც შერჩევითი ერთობლიობა ეწოდება, ხოლო შერჩევითი ერთობლიობის საშუალოს — შერჩევითი საშუალო.

ვთქვათ, გენერალური ერთობლიობის მოცულობა არის N. ვიგულისხმოთ, რომ ჩვენთვის უცნობია ერთობლიობის როგორც x არგუმენტის ვარიანტები, ისე ამ ვარიანტების ხვედრი. ეს უცნობი ვარიანტები და მათი ხვედრები ან, როგორც სტატისტიკის ენით იტყვიან, ერთობლიობის უცნობი განაწილება, განისაზღვრება შემთხვევითი შერჩევის დახმარებით.

ვთქვათ, გენერალური ერთობლიობიდან, რომლის მოცულობაა N, შემთხვევით შერჩევით საფუძველზე გამოვყავით n წევრი. n ერთეული დავყოთ r ჯგუფად, სადაც r არის x არგუმენტის ვარიანტთა რიცხვი. x არგუმენტის ვარიანტები აღვნიშნოთ შემდეგნაირად:

$$x_1, x_2, \ldots, x_r,$$

ხოლო მათი სათანადო სიხშირეები იყოს:

$$n_1, n_2, \ldots, n_r,$$

სადაც

$$n_1+n_2+\dots+n_r=n.$$

ეს იმას ნიშნავს, რომ x_1 ვარიანტიან ჯგუფში იქნება n_1 ელემენტი, x_2 ვარიანტიან ჯგუფში—n_2 ელემენტი და ა. შ.

თუ განვიხილავთ ფარდობებს:

$$\frac{n_1}{n}=v_1,\quad \frac{n_2}{n}=v_2,\dots,\quad \frac{n_r}{n}=v_r \tag{14.1}$$

ისინი იქნებიან

$$x_1,\ x_2,\dots,\ x_r$$

ვარიანტთა ფარდობითი სიხშირეები.

ცხადია, რომ ყველა ფარდობითი სიხშირის ჯამი ერთის ტოლი იქნება, ე. ი.

$$v_1+v_2+\dots+v_r=1.$$

შერჩევითი ერთობლიობის x არგუმენტის საშუალო არითმეტიკული $\ddot{x}$, რომელიც აწონილია არგუმენტის ვარიანტის სიხშირის მიხედვით, (14.1) ფორმულის თანახმად გამოისახება შემდეგნაირად:

$$\ddot{x}=\frac{n_1x_1+n_2x_2+\dots+n_rx_r}{n}, \tag{14.2}$$

საიდანაც

$$\ddot{x}=\frac{n_1}{n}x_1+\frac{n_2}{n}x_2+\dots+\frac{n_r}{n}x_r.$$

თუ მხედველობაში მივიღებთ (14.1) ფორმულას, დავწერთ:

$$\ddot{x}=v_1x_1+v_2x_2+\dots+v_rx_r=\sum_{i=1}^{r}v_ix_i. \tag{14.3}$$

საშუალო $\ddot{x}$-ს, რომელიც გამოსახულია (14.2) და (14.3) ფორმულებით, *შერჩევითი საშუალო* ეწოდება.

(14.3) ფორმულიდან ჩანს, რომ შერჩევითი საშუალო არგუმენტის თითოეული ვარიანტის სათანადო ფარდობით სიხშირეზე ნამრავლთა ჯამის ტოლია.

შემთხვევითი სიდიდის საშუალო მნიშვნელობას უწოდებენ აგრეთვე *სტოქასტურ, ანუ ალბათურ საშუალოს*.

როგორც ვიცით, თუ x შემთხვევითი სიდიდე ღებულობს რიცხვით მნიშვნელობებს

$$x_1,\ x_2,\dots,\ x_n,$$

სათანადო ალბათობებით

$$p_1,\ p_2,\dots,\ p_n,$$

მაშინ საშუალო მნიშვნელობა, ანუ მათემატიკური ლოდინი

$$E(X)=x_1 p_1+x_2 p_2+\ldots+x_n p_n.$$

აქ საშუალო მნიშვნელობა აწონილია ვარიანტთა ალბათობებით და არა ფარდობითი სიხშირეებით.

როგორც ვხედავთ, მათემატიკური ლოდინი და გენერალური საშუალო ერთმანეთს ემთხვევა: გენერალური საშუალოს შემთხვევაში p_i ხვედრი რიცხობრივად მათემატიკური ლოდინის შემთხვევაში p_i ალბათობის ტოლია.

ამგვარად, გენერალური საშუალო და მათემატიკური ლოდინი ერთმანეთს ემთხვევა, მაგრამ უნდა აღინიშნოს, რომ მათემატიკური ლოდინი გვექნება იმ შემთხვევაში, როდესაც საქმე გვაქვს შემთხვევით შერჩევის პროცესებთან.

თეორემა. შ ე რ ჩ ე ვ ი ს დ რ ო ს მ ი ღ ე ბ უ ლ შ ე მ თ ხ ვ ე ვ ი თ ს ი დ ი დ ე თ ა ს ა შ უ ა ლ ო ა რ ი თ მ ე ტ ი კ უ ლ ი ა რ ი ს შ ე რ ჩ ე ვ ი თ ი ს ა შ უ ა ლ ო.

მართლაც, როდესაც გენერალური ერთობლიობიდან შემთხვევით ვარჩევთ n ელემენტს, ვღებულობთ n შემთხვევით სიდიდეს. ვთქვათ, ისინია:

$$x^{(1)},\ x^{(2)},\ldots,\ x^{(n)}.$$

აქ თითოეული შემთხვევითი სიდიდე წარმოადგენს ერთ-ერთ შემდეგ ვარიანტთაგანს:

$$x_1,\ x_2,\ldots,\ x_r.$$

ახლა ცხადია, რომ შერჩევით მიღებულ შემთხვევით სიდიდეთა საშუალო არითმეტიკული შერჩევით საშუალოს გამოსახავს, ე. ი.

$$\frac{x^{(1)}\cdot 1+x^{(2)}\cdot 1+\ldots+x^{(n)}\cdot 1}{n}=\frac{1}{n}x^{(1)}+\frac{1}{n}x^{(2)}+\ldots+\frac{1}{n}x^{(n)}=$$

$$=\frac{n_1}{n}x_1+\frac{n_2}{n}x_2+\cdots+\frac{n_r}{n}x_r=v_1x_1+v_2x_2+\cdots+v_rx_r=\ddot{x}.$$

მაგალითი. ვთქვათ, ზემოგანხილული სიმინდის მოსავლიანობის მაგალითში 10000 ჰექტარიდან შევარჩიეთ 1000 ჰექტარი, რომელიც არგუმენტის ვარიანტების მიხედვით გვაძლევს თითოეულ ჰექტარზე მოსავლიანობის შემდეგ განაწილებას:

$n_1=150$	ჰექტარს,	მოსავლით	თითოეულ	ჰექტარზე	14	ცენტნერი
$n_2=180$	"	"	"	"	14,5	"
$n_3=200$	"	"	"	"	15	"
$n_4=220$	"	"	"	"	15,4	"
$n_5=250$	"	"	"	"	16	"

მონახეთ მოსავლიანობის შერჩევითი საშუალო.

ა მ ო ხ ს ნ ა: შერჩევითი საშუალოს გამოსათვლელად გვქონდა (14.2) და (14.3) ფორმულა. ვისარგებლოთ პირველი მათგანით:

$$\ddot{x}=\frac{150 \cdot 14+180 \cdot 14,5+200 \cdot 15+220 \cdot 15,4+250 \cdot 16}{1000}=15,078.$$

ამგვარად, შერჩევითი საშუალოს დახმარებით ვპოულობთ, რომ ჰექტარზე მოსავალი 15,078 ცენტნერს უდრის. ნამდვილი საშუალო მოსავალი ჰექტარიდან, რომელიც ზემოთ გენერალური საშუალოს დახმარებით იყო მოძებნილი, არის 15,265 ცენტნერი.

მაშასადამე, როგორც ვხედავთ, ჰექტარზე მოსავალი, რომელიც მივიღეთ შერჩევითი საშუალოს გამოთვლით, გარკვეული მიახლოებით შეგვიძლია გენერალურ საშუალოდ მივიღოთ.

განხილული მაგალითიდან ჩანს, თუ რა დიდი მნიშვნელობა აქვს შერჩევით საშუალოს, რადგანაც ის მიახლოებით გენერალურ საშუალოს გამოსახავს. ასეთ შემთხვევაში საჭიროა განისაზღვროს შერჩევითი საშუალოს გენერალურ საშუალოსთან მიახლოების ხარისხი ან სიხშირის ხვედრთან მიახლოების სიდიდე. ამ საკითხს წყვეტს დიდ რიცხვთა კანონის შესახებ ჩებიშევის თეორემა, რომელსაც განვიხილავთ შემდგომ.

თეორემა. შ ე მ თ ხ ვ ე ვ ი თ ს ი დ ი დ ე თ ა მ ა თ ე მ ა ტ ი კ უ რ ი ლ ო დ ი ნ ე ბ ი ს ს ა შ უ ა ლ ო ა რ ი თ მ ე ტ ი კ უ ლ ი მ ა თ ი გ ე ნ ე რ ა ლ უ რ ი ს ა შ უ ა ლ ო ს ტ ო ლ ი ა.

დავუშვათ, რომ გენერალური ერთობლიობის იმ არგუმენტიდან, რომლის ვარიანტებია $x_1, x_2, \ldots, x_k$, შერჩევის წარმოებისას (შერჩევას ვაწარმოებთ „ბურთთა დაბრუნების" სქემის მიხედვით) n ელემენტი მოხვდა. ცხადია, რომ შერჩევაში ყოველი ელემენტის მოხვედრის ალბათობა იქნება ერთი და იგივე. ვთქვათ, შერჩევის შედეგად მივიღეთ შემდეგი შემთხვევითი სიდიდეები:

$$x^{(1)},\ x^{(2)}, \ldots,\ x^{(n)}.$$

რამდენადაც ყოველ შერჩეულ ელემენტს ვარიანტად შეიძლება ჰქონდეს x_1, x_2 და ა. შ. x_k, ამიტომ მათემატიკური ლოდინი თითოეული შემთხვევითი სიდიდისა: $x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(n)}$ იქნება $x_1, x_2, \ldots, x_k$ სიდიდეთა სათანადო ალბათობებზე ნამრავლთა ჯამი:

$$E(X)=\sum_{i=1}^{k} x_i p_i.$$

თუ ყველა შემთხვევით სიდიდეს ერთი და იგივე $E(X)$ მათემატიკური ლოდინი აქვს, მაშინ მათი მათემატიკურ ლოდინთა საშუალო არითმე-

ტიკული იქნება იგივე $E(X)$. მაგრამ, როგორც უკვე ვნახეთ, მათემატიკური ლოდინი რიცხობრივად გენერალურ საშუალოს უდრის. მაშასადამე, განხილულ შემთხვევით სიდიდეთა მათემატიკური ლოდინების საშუალო არითმეტიკული გენერალური საშუალოს ტოლი ყოფილა.

§ 15. გენერალური დისპერსია

როგორც ვნახეთ, გენერალური საშუალო მიახლოებით შერჩევითი საშუალოთი გამოისახება და ამიტომ საჭიროა მათი ურთიერთმიახლოების ხარისხის დადგენა.

ვთქვათ, გვაქვს რომელიმე გენერალური ერთობლიობა, რომლის გენერალური საშუალო არის $\overline{x}$; აღებული ერთობლიობის მოცულობა იყოს N, ხოლო x არგუმენტის ვარიანტები იყოს

$$x_1, x_2, \ldots, x_k,$$

მათი სიხშირეები კი—

$$n_1, n_2, \ldots n_k.$$

ცხადია, რომ საძიებელი მიახლოება მჭიდრო კავშირშია ცალკეულ ვარიანტსა და გენერალურ საშუალოს სხვაობასთან, ან, როგორც იტყვიან, საძიებელი მიახლოება დამოკიდებულია არგუმენტის ვარიანტთა თავისი $\overline{x}$ საშუალო მნიშვნელობის გარშემო გაფანტვის ხარისხზე. საძიებელი ხარისხის გასაზომად, ცხადია, საჭიროა განვიხილოთ სხვაობა

$$x_i - \overline{x},$$

რომელიც წარმოადგენს არგუმენტის ვარიანტთა გენერალურ საშუალოდან გადახრას. ერთი შეხედვით არგუმენტის ვარიანტთა გენერალურ საშუალოდან გადახრათა შესაფასებლად საჭიროა ავიღოთ შემდეგი ჯამი

$$\sum_{i=1}^{k} n_i (x_i - \overline{x}).$$

მაგრამ სინამდვილეში ეს ჯამი არაფერს იძლევა, რადგან ის ნულის ტოლია.

მართლაც, გენერალური საშუალოს განმარტების თანახმად,

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i},$$

საიდანაც

$$\sum_{i=1}^{k} n_i x_i = \overline{x} \sum_{i=1}^{k} n_i = \sum_{i=1}^{k} n_i \overline{x}.$$

მიღებული ტოლობის მარჯვენა მხარეს თუ მარცხნივ გადმოვიტანთ, მივიღებთ:

$$\sum_{i=1}^{k} n_i (x_i - \overline{x}) = 0. \tag{15.1}$$

ცხადია, გადახრის განსასაზღვრავად (15.1) ჯამის ნაცვლად უნდა განვიხილოთ ისეთი ჯამი, რომელიც თავისუფალი იქნება ნიშნების გავლენისაგან. ასეთი იქნებოდა (15.1) ფორმულაში შემავალ შესაკრებთა აბსოლუტური მნიშვნელობანი; მაგრამ, გარკვეული მოსაზრებების გამო, უფრო მიზანშეწონილად ცნეს განიხილონ გადახრათა კვადრატები, ე. ი. იღებენ შემდეგ გამოსახულებას:

$$(x_i - \overline{x})^2.$$

ამის საშუალებით ვარიანტთა გადახრებს ახასიათებენ შემდეგნაირად: განიხილავენ ე. წ. ვარიანტთა გადახრების საშუალო კვადრატს, რომლის მისაღებად ამ ვარიანტთა გადახრების კვადრატთა ჯამს ყოფენ მათ რიცხვზე.

ამგვარად, თუ გადახრის კვადრატის საშუალოს σ_x^2-ით აღვნიშნავთ, მივიღებთ:

$$\sigma_x^2 = \frac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + \ldots + n_k (x_k - \overline{x})^2}{N} =$$

$$= (x_1 - \overline{x})^2 \frac{n_1}{N} + (x_2 - \overline{x})^2 \frac{n_2}{N} + \ldots + (x_k - \overline{x})^2 \frac{n_k}{N} =$$

$$= \sum_{i=1}^{k} (x_i - \overline{x})^2 \frac{n_i}{N}.$$

როგორც წინა შემთხვევაში, თუ აღვნიშნავთ

$$\frac{n_i}{N} = p_i,$$

მაშინ σ_x^2-თვის დავწერთ

$$\sigma_x^2 = \sum_{i=1}^{k} (x_i - \overline{x})^2 p_i. \tag{15.2}$$

σ_x^2-ს ეწოდება x არგუმენტის გ ე ნ ე რ ა ლ უ რ ი დ ი ს პ ე რ ს ი ა.

გამოთვლათა გამარტივების მიზნით (15.2) ტოლობა წარმოვიდგინოთ შემდეგნაირად:

$$\sigma_x^2=\sum_{i=1}^{k}(x_i-\bar{x})^2 p_i=\sum_{i=1}^{k}[(x_i-a)-(\bar{x}-a)]^2 p_i=$$

$$=\sum_{i=1}^{k}[(x_i-a)^2-2(\bar{x}-a)(x_i-a)+(\bar{x}-a)^2]\,p_i=$$

$$=\sum_{i=1}^{k}(x_i-a)^2 p_i-2(\bar{x}-a)\sum_{i=1}^{k}(x_i-a)\,p_i+(\bar{x}-a)^2\sum_{i=1}^{k}p_i,$$

სადაც a ნებისმიერი რიცხვია.

მხედველობაში მივიღოთ, რომ

$$\sum_{i=1}^{k}p_i=1$$

და

$$\sum_{i=1}^{k}(x_i-a)\,p_i=\sum_{i=1}^{k}x_i\,p_i-\sum_{i=1}^{k}a\,p_i=\bar{x}-a\sum_{i=1}^{k}p_i=\bar{x}-a.$$

σ_x^2-თვის მივიღებთ:

$$\sigma_x^2=\sum_{i=1}^{k}(x_i-a)^2 p_i-2(\bar{x}-a)(\bar{x}-a)+(\bar{x}-a)^2=$$

$$=\sum_{i=1}^{k}(x_i-a)^2\,p_i-(\bar{x}-a)^2.$$

ამგვარად გვექნება

$$\sigma_x^2=\sum_{i=1}^{k}(x_i-a)^2\,p_i-(\bar{x}-a)^2. \tag{15.3}$$

აქ a ნებისმიერად იყო აღებული, საზოგადოდ კი, გამოთვლების გამარტივების მიზნით, a ისეთნაირად უნდა იქნეს შერჩეული, რომ x_i-a სხვაობა რაც შეიძლება მინიმალური იყოს.

მაგალითი. განსაზღვრეთ მოსავლიანობის გენერალური დისპერსია ზემოთ განხილული მაგალითის მონაცემებით.

ა მ ო ხ ს ნ ა: აქ გვაქვს 5 ვარიანტი.

$$x_1=14;\ x_2=14{,}5;\ x_3=15,\ x_4=15{,}4;\ x_5=16.$$

ამ ვარიანტთა ხვედრები სათანადოდ იქნება:

$$p_1=0{,}09;\ p_2=0{,}13;\ p_3=0{,}21;\ p_4=0{,}25;\ p_5=0{,}32.$$

წინა მაგალითის ამოხსნისას ვნახეთ, რომ გენერალური საშუალოა $\overline{x}=15{,}265$ ცენტნერი. დისპერსიის საპოვნელად მივმართოთ ფორმულას:

$$\sigma_x^2=(14-15{,}26)^2\cdot 0{,}09+(14{,}5-15{,}26)^2\cdot 0{,}13+$$
$$+(15-15{,}26)^2\cdot 0{,}21+(15{,}4-15{,}26)^2\cdot 0{,}25+$$
$$+(16-15{,}26)^2\cdot 0{,}32=0{,}14+0{,}07+0{,}01+0{,}01+$$
$$+0{,}18=0{,}41.$$

როგორც ვხედავთ, σ_x^2-ის გამოთვლისას კვადრატში უნდა ავიყვანოთ ორნიშნა და სამნიშნა რიცხვები. ამის უფრო მარტივად გამოთვლა შეიძლებოდა (15.3) ფორმულის გამოყენებით. ამისათვის საჭიროა a-ს სათანადო შერჩევა. ვთქვათ, a შევარჩიეთ 15-ის ტოლი, მაშინ σ_x^2-თვის დავწერთ

$$\sigma_x^2=(14-15)^2\cdot 0{,}09+(14{,}5-15)^2 . 0{,}13+(15-15)^2\cdot 0{,}21+$$
$$+(16-15)^2\cdot 0{,}32-(15{,}26-15)^2=0{,}09+0{,}03+$$
$$+0{,}04+0{,}32-0{,}07=0{,}41.$$

ამგვარად, მოსავლიანობის დისპერსია ტოლია 0,41-ის, საშუალო კვადრატული გადახრა σ_x კი იქნება:

$$\sigma_x=\sqrt{0{,}41}\approx 0{,}64.$$

§ 16. შერჩევითი დისპერსია

პრაქტიკაში ხშირია შემთხვევა, როდესაც უცნობია გენერალური საშუალო, მაშინ დისპერსიის მოსანახავად მიმართავენ შერჩევით დისპერსიას და ვარიანტთა სიხშირეებს. შერჩევითი ერთობლიობის დისპერსიას შ ე რ ჩ ე ვ ი თ ი დ ი ს პ ე რ ს ი ა ეწოდება; თუ მას აღვნიშნავთ σ_1^2-ით, ის გამოითვლება შემდეგი ფორმულით:

$$\sigma_1^2=\sum_{i=1}^{k}(x_i-\ddot{x})^2\,v_i. \tag{16.1}$$

(15.3) ფორმულის ანალოგიურად გვექნება

$$\sigma_1^2=\sum_{i=1}^{k}(x_i-a)^2\,v_i-(\ddot{x}-a)^2, \tag{16.2}$$

სადაც $\ddot{x}$ არის შერჩევითი საშუალო, ხოლო v_i — ვარიანტთა სიხშირეები.

მაგალითი. წინა მაგალითის მონაცემებზე დაყრდნობით განსაზღვრეთ მოსავლიანობის შერჩევითი დისპერსია.

ა მ ო ხ ს ნ ა: აქ გვაქვს იგივე ვარიანტები:

$$x_1=14;\ x_2=14{,}5;\ x_3=15;\ x_4=15{,}4;\ x_5=16.$$

ამ ვარიანტების შესაბამისი სიხშირეები იქნება:

$$v_1=0{,}15;\ v_2=0{,}18;\ v_3=0{,}20;\ v_4=0{,}22;\ v_5=0{,}25.$$

როგორც წინათ გამოვთვალეთ, შერჩევითი საშუალოსათვის გვაქვს $\ddot{x}=15{,}08$. შერჩევითი დისპერსიის გამოსათვლელად ვისარგებლოთ (16.1) ფორმულით, a მუდმივი აქაც 15-ის ტოლი იყოს.

მაშინ σ_1^2-თვის მივიღებთ:

$$\sigma_1^2=(14-15)^2\cdot 0{,}15+(14{,}5-15)^2\cdot 0{,}18+(15-15)^2\cdot 0{,}2+$$
$$+(15{,}4-15)^2\cdot 0{,}22+(16-15)^2\cdot 0{,}25-(15{,}08-15)^2=$$
$$=0{,}15+0{,}04+0{,}04+0{,}25-0{,}01=0{,}47.$$

ამგვარად, შერჩევითი დისპერსიაა $\sigma_1^2=0{,}47$, ხოლო გენერალური დისპერსია 0,41-ის ტოლია. როგორც ვხედავთ, შერჩევითი დისპერსია საკმაოდ ახლოს არის გენერალურ დისპერსიასთან.

§ 17. შერჩევითი საშუალოს დისპერსია და მისი კავშირი გენერალურ დისპერსიასთან

შერჩევითი საშუალოს დისპერსია აღვნიშნოთ $\sigma^2_{\ddot{x}}$-ით. შემთხვევითი სიდიდის დისპერსიის განმარტების თანახმად დავწერთ:

$$\sigma^2_{\ddot{x}}=E(\ddot{x}-\bar{x})^2. \tag{17.1}$$

აქ $\ddot{x}$ არის შერჩევითი საშუალო, რომელიც წარმოადგენს შერჩევაში მოხვედრილ ურთიერთდამოუკიდებელ შემთხვევით სიდიდეთა საშუალო არითმეტიკულს.

თუ შერჩევაში მოხვედრილი შემთხვევითი სიდიდეებია:

$$x^{(1)},\ x^{(2)},\dots,\ x^{(n)},$$

მაშინ

$$\ddot{x}=\frac{1}{n}\left[x^{(1)}+x^{(2)}+\cdots+x^{(n)}\right]. \tag{17.2}$$

$\bar{x}$ იქნება იგივე შემთხვევით სიდიდეთა მათემატიკური ლოდინების საშუალო არითმეტიკული

$$\bar{x}=\frac{1}{n}\left\{E[x^{(1)}]+E[x^{(2)}]+\cdots+E[x^{(n)}]\right\}. \tag{17.3}$$

როგორც უკვე იყო ნათქვამი, თითოეული $x^{(1)}, x^{(2)},\ldots, x^{(n)}$ სიდიდისათვის შესაძლებელია იგივე ვარიანტები და ალბათობები გვქონდეს, რაც x-თვის. სწორედ ამის საფუძველზე ყველა მათი მათემატიკური ლოდინი ერთმანეთის ტოლი იქნება, ე. ი.

$$E[x^{(1)}]=E[x^{(2)}]=\ldots=E[x^{(n)}]=E(x)=\bar{x}.$$

მაშინ (17.3) ფორმულა მიიღებს ასეთ სახეს:

$$\bar{x}=\frac{1}{n}\left\{E[x^{(1)}]+E[x^{(2)}]+\cdots+E[x^{(n)}]\right\}=$$

$$=\frac{1}{n}\overbrace{[\bar{x}+\bar{x}+\cdots+\bar{x}]}^{n\text{-ჯერ}}. \tag{17.4}$$

ზემოთ მოყვანილი მოსაზრებით, რომლითაც დავასკვენით მათემატიკურ ლოდინთა ტოლობა, აგრეთვე დავასკვნით დისპერსიათა ტოლობასაც

$$\sigma^2_{x^{(1)}}=\sigma^2_{x^{(2)}}=\ldots=\sigma^2_{x^{(n)}}=\sigma^2_x.$$

ახლა გამოვთვალოთ $\sigma^2_{\ddot{x}}$.

თუ მხედველობაში მივიღებთ (17.2), (17.3) და (17.4) ტოლობებს, (17.1) ფორმულიდან გვექნება

$$\sigma^2_{\ddot{x}}=E(\ddot{x}-\bar{x})^2=E\left\{\frac{1}{n}\left[x^{(1)}+x^{(2)}+\cdots+x^{(n)}-\right.\right.$$

$$\left.\left.-\frac{1}{n}(\bar{x}+\bar{x}+\cdots+\bar{x}\right)\right\}^2.$$

თუ უკანასკნელი გამოსახულების მარჯვენა მხრიდან მათემატიკური ლოდინის ნიშნის გარეთ გამოვიტანთ $\frac{1}{n^2}$, მივიღებთ:

$$\sigma_{\bar{x}}^2=\frac{1}{n^2}E\{[x^{(1)}-\bar{x}]+[x^{(2)}-\bar{x}]+\ldots+[x^{(n)}-\bar{x}]\}^2=$$

$$=\frac{1}{n^2}E\{[x^{(1)}-\bar{x}]^2+[x^{(2)}-\bar{x}]^2+\ldots+[x^{(n)}-\bar{x}]^2+$$

$$+2[x^{(1)}-\bar{x}][x^{(2)}-\bar{x}]+\ldots+2[x^{(n-1)}-\bar{x}[[x^{(n)}-\bar{x}]\}=$$

$$=\frac{1}{n^2}\{E[x^{(1)}-\bar{x}]^2+E[x^{(2)}-\bar{x}]^2+\ldots+E[x^{(n)}-\bar{x}]^2\}+$$

$$+\frac{1}{n^2}2E\{[x^{(1)}-\bar{x}][x^{(2)}-\bar{x}]\}+\ldots+\frac{1}{n^2}2E\{[x^{(n-1)}-\bar{x}][x^{(n)}-\bar{x}]\}.$$

რადგანაც $x^{(1)}, x^{(2)},\ldots, x^{(n)}$ ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, ამიტომ მათთვის ნამრავლის მათემატიკური ლოდინის თვისების გამოყენება შესაძლებელია. ამ თვისების გამოყენების შემდეგ ვნახავთ, რომ უკანასკნელი გამოსახულების მარჯვენა მხარის თანამამრავლთა მათემატიკური ლოდინები ნული იქნება და დაგვრჩება

$$\sigma_{\bar{x}}^2=\frac{1}{n^2}\{E[x^{(1)}-\bar{x}]^2+E[x^{(2)}-\bar{x}]^2+\ldots+E[x^{(n)}-\bar{x}]^2.$$

თუ მხედველობაში მივიღებთ, რომ

$$E[x^{(1)}-\bar{x}]^2=\sigma_{x(1)}^2;\ E[x^{(2)}-\bar{x}]^2=\sigma_{x(2)}^2;\ldots;\ E[x^{(n)}-\bar{x}]^2=\sigma_{x(n)}^2$$

და აგრეთვე ყველა ეს დისპერსია ერთმანეთის ტოლია, გვექნება:

$$\sigma_{\bar{x}}^2=\frac{\sigma_{x(1)}^2+\sigma_{x(2)}^2+\ldots+\sigma_{x(n)}^2=}{n^2}=\frac{n\sigma_x^2}{n^2}=\frac{\sigma_x^2}{n}.$$

ამგვარად, მივიღეთ, რომ

$$\sigma_{\bar{x}}^2=\frac{\sigma_x^2}{n}. \tag{17.5}$$

მაშასადამე, გენერალური დისპერსია უდრის შერჩევითი საშუალოს დისპერსიას გამრავლებულს შერჩევითი ერთობლიობის მოცულობაზე.

(17.5) ფორმულიდან დავწერთ:

$$\sigma_{\bar{x}}=\frac{\sigma_x}{\sqrt{n}}, \tag{17.6}$$

ე. ი. შერჩევითი საშუალოს საშუალო კვადრატული გადახრა $\sqrt{n}$-ჯერ ნაკლებია σ_x საშუალო კვადრატულ გადახრაზე.

§ 18. მედიანა და მოდა

ვთქვათ, შემთხვევით სიდიდეს შეუძლია მიიღოს შემდეგი $2n+1$ რიცხვითი მნიშვნელობანი:

$$x_1, x_2, \ldots, x_n, x_{n+1}, x_{n+2}, \ldots, x_{2n+1}.$$

ვიგულისხმოთ, რომ ზემოთ დაწერილი მიმდევრობა ზრდადია და ელემენტები ერთმანეთისაგან განსხვავებულია.

ცხადია, x_{n+1} აღებული მიმდევრობის შუა ადგილზე იქნება. მიმდევრობის შუა ადგილზე მყოფ წევრს მ ე დ ი ა ნ ა ს უწოდებენ.

განვიხილოთ ისეთი მიმდევრობა, რომლის ელემენტთა რიცხვი კენტია; თუ მიმდევრობის ელემენტთა რიცხვი კენტი არ არის, ვთქვათ, მიმდევრობაში $2n$ ელემენტია

$$x_1, x_2, \ldots x_n, x_{n+1}, \ldots, x_{2n}.$$

მაშინ x_n-ის წინ და x_{n+1}-ის შემდეგ ელემენტთა ტოლი რაოდენობა იქნება. ასეთ შემთხვევაში მედიანად შეიძლება მიღებულ იქნეს ყოველი რიცხვი, რომელიც x_n და x_{n+1}-ს შორის იქნება მოთავსებული; უფრო ხშირად, მედიანად მათ საშუალოს ღებულობენ. ამგვარად, ამ უკანასკნელ შემთხვევაში მედიანა იქნება

$$\frac{1}{2}(x_n+x_{n+1}).$$

თუ მედიანას M_e-ით აღვნიშნავთ, მაშინ, როდესაც მიმდევრობის ელემენტთა რიცხვი კენტია

$$M_e=x_{n+1}, \tag{18.1}$$

ხოლო, როცა ელემენტთა რიცხვი ლუწია

$$M_e=\frac{1}{2}(x_n+x_{n+1}). \tag{18.2}$$

შემთხვევითი სიდიდის მახასიათებელთა ჯგუფს ეკუთვნის აგრეთვე მ ო დ ა, რომელიც აღინიშნება M_0-ით. მოდას აზრი მდგომარეობს შემდეგში: ვთქვათ, გვაქვს x შემთხვევითი სიდიდე, რომელიც სხვადასხვა რიცხვით მნიშვნელობებს ღებულობს. ცდათა სერიების განმეორებისას შესაძლებელია რომელიმე ცალკეული მნიშვნელობა ხშირად განმეორდეს, მაშინ სწორედ მოდა იქნება შემთხვევითი სიდიდის ის რიცხვითი მნიშვნელობა, რომელიც ყველაზე ხშირად გვხვდება.

თუ შემთხვევითი სიდიდის რიცხვით მნიშვნელობათა განაწილების მრუდს ავაგებთ, მაშინ ასეთი მრუდის წვეროს შესაბამისი აბსცისა იქნება მოდა.

ან, მოდა არის შემთხვევითი სიდიდის მიერ მიღებული ის რიცხვითი მნიშვნელობა, რომელსაც ყველაზე მეტი განაწილების სიმკვრივე შეესაბამება.

§ 19. დიაპაზონი

ვთქვათ, x შემთხვევით სიდიდეზე დაკვირვების შედეგად მივიღეთ მნიშვნელობანი:

$$x_1,\ x_2, \ldots,\ x_n. \tag{19.1}$$

ზოგადობის შეუზღუდველად ვიგულისხმოთ, რომ ისინი დალაგებულია ზრდადი მიმდევრობით, ე. ი.

$$x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n.$$

ასე რომ, (19.1) მიმდევრობა წარმოადგენს ე. წ. ვარიაციულ მწკრივს. (19.1) მიმდევრობის უდიდესი ელემენტია x_n, ხოლო უმცირესი კი x_1. მიმდევრობის უდიდეს და უმცირეს მნიშვნელობათა შორის სხვაობას ამ მიმდევრობის დიაპაზონი ეწოდება და აღინიშნება R-ით:

$$R = x_n - x_1.$$

ხშირად R დიაპაზონის მოცემით შემთხვევით სიდიდეზე გარკვეული წარმოდგენის შექმნა შესაძლებელია. დიაპაზონი შედის შემთხვევით სიდიდეთა მახასიათებლების რიცხვში.

§ 20. ორი შემთხვევითი სიდიდის სისტემის რიცხვითი მახასიათებლები

(X, Y) სისტემის (k, s) რიგის საწყისი მომენტი ეწოდება $X^k Y^s$ სიდიდის მათემატიკურ ლოდინს

$$\alpha_{k,s} = E(X^k Y^s).$$

(X, Y) სისტემის k, s რიგის ცენტრალური მომენტი ეწოდება გადახრათა შესაბამისად k-ური და s ხარისხების ნამრავლთა მათემატიკურ ლოდინს.

$$\mu_{k,s} = E\{[X - E(X)]^k\,[Y - E(Y)]^s\}.$$

თუ გამოვიყენებთ მათემატიკური ლოდინის განმარტებას, ორი დისკრეტული ტიპის შემთხვევით სიდიდეთა სისტემის მომენტების გამოსათვლელად მივიღებთ შემდეგ ფორმულას:

$$\alpha_{k,s} = \sum_i \sum_j x_i^k\, y_j^s\, p_{ij}$$

$$\mu_{k,s} = \sum_i \sum_j [x_i - E(X)]^k\,[y_j - E(Y)]^s\, p_{ij}$$

სადაც

$$p_{ij}=P\{X=x_i;\ Y=y_j\}.$$

არის ალბათობა იმისა, რომ (X, Y) სისტემა მიიღებს (x_i, y_j) მნიშვნელობას, ხოლო აჯამვა ხდება X და Y შემთხვევით სიდიდეთა ყველა შესაძლო მნიშვნელობებისათვის.

უწყვეტი ტიპის შემთხვევით სიდიდეთათვის გვექნება:

$$\alpha_{k,s}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s f(x, y)\, dxdy,$$

$$\mu_{k,s}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x-E(X)]^k\, [y-E(Y)]^s f(x, y)\, dxdy,$$

სადაც $f(x, y)$ არის სისტემის განაწილების სიმკვრივე.

როცა $k=1$ და $s=0$, მაშინ მივიღებთ:

$$E(X)=\alpha_{,0}=E(X'Y^0)=E_X$$

თუკი $k=0$ და $s=1$, გვექნება

$$E(Y)=\alpha_{0,1}=E(X^0Y')=E_Y.$$

$E(X)$ და $E(Y)$ მათემატიკურ ლოდინთა ერთობლიობა წარმოადგენს სისტემის მდებარეობის მახასიათებელს. გეომეტრიულად სიბრტყეზე ესენი არიან წერტილის საშუალო კოორდინატები, რომლის გარშემო ხდება (X, Y) წერტილების მნიშვნელობათა გაბნევა.

როცა $k=2$ და $s=0$, მაშინ მივიღებთ

$$\mu_{2,0}=E\{[X-E_X]^2\,[Y-E_Y]^0\}=D(X)=D_X,$$

ხოლო, როცა $k=0$ და $s=2$, გვექნება

$$\mu_{0,2}=E\{(X-E_X)^0\,(Y-E_Y)^2\}=D(Y)=D_Y.$$

D_X და D_Y ახასიათებენ შემთხვევითი წერტილის გაფანტულობას $0X$ და $0Y$ ღერძის მიმართულებით.

სისტემის მახასიათებელთა შორის განსაკუთრებით მნიშვნელოვანია მეორე შერეული ცენტრალური მომენტი:

$$K_{x,y}=\mu_{11}=E\{(X-E_X)\,(Y-E_Y)\},$$

რომელიც წარმოადგენს გადახრათა ნამრავლის მათემატიკურ ლოდინს.

$K_{x,y}$-ს ეწოდება X, Y შემთხვევით სიდიდეთა კორელაციური მომენტი, ანუ სხვანაირად, კავშირის მომენტი.

დისკრეტული ტიპის შემთხვევითი სიდიდეებისათვის კორელაციური მომენტი შემდეგი ფორმულით გამოისახება:

$$K_{x,y}=\sum_i \sum_j (x_i-E_X)(y_i-E_Y)\, p_{ij},$$

ხოლო უწყვეტი ტიპის შემთხვევითი სიდიდეებისათვის კი გვექნება:

$$K_{X,Y}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E_X)(y-E_Y)\, f(x, y)\, dxdy.$$

კორელაციური მომენტი წარმოადგენს ისეთ მახასიათებელს, რომელიც გვიჩვენებს როგორც X და Y სიდიდეთა გაფანტულობას, ისე მათ შორის კავშირსაც.

ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებისათვის კორელაციური მომენტი ნულის ტოლი იქნება.

თუ ორი შემთხვევითი სიდიდის კორელაციური მომენტი ნულისაგან განსხვავებულია, ეს მიგვითითებს მათ შორის დამოკიდებულების არსებობაზე.

კორელაციური მომენტის განმარტებიდან ჩანს, რომ ის არა მარტო შემთხვევით სიდიდეთა შორის დამოკიდებულებას ახასიათებს, არამედ აგრეთვე გვაჩვენებს მათ გაფანტულობასაც. მართლაც, თუ ერთ-ერთი შემთხვევითი სიდიდე X ან Y მცირედ განსხვავდება თავისი მათემატიკური ლოდინისაგან, მაშინ როგორი მჭიდრო კავშირითაც არ უნდა იყვნენ ისინი ერთმანეთთან დაკავშირებულნი, კორელაციური მომენტი პატარა იქნება. ამიტომ შემთხვევით სიდიდეთა შორის კავშირის დასახასიათებლად, კორელაციური მომენტის ნაცვლად, შემოჰყავთ უგანზომილებო მახასიათებელი, კორელაციის კოეფიციენტი, რომელსაც შემდეგ პარაგრაფში განვიხილავთ.

§ 21. რეგრესიისა და კორელაციის კოეფიციენტები

რეგრესიისა და კორელაციის კოეფიციენტები შემთხვევით სიდიდეთა შორის დამოკიდებულების რიცხობრივად გამოსახვის საშუალებას იძლევიან.

ვთქვათ, მოცემულია ორი შემთხვევითი სიდიდე X და Y. მივიღოთ, რომ X-ის რეგრესიის კოეფიციენტი Y-ის მიმართ იქნება

$$\rho(X, Y)=\frac{E(XY)-E(X)E(Y)}{D(Y)}.$$

ანალოგიურად, Y-ის X-ის მიმართ რეგრესიის კოეფიციენტი იქნება:

$$\rho(Y, X)=\frac{E(XY)-E(X)E(Y)}{D(X)}$$

რეგრესიის კოეფიციენტთა საშუალო გეომეტრიულს ეწოდება კორელაციის კოეფიციენტი:

$$R(X, Y)=\sqrt{\rho(X, Y)\rho(Y, X)}=\frac{E(XY)-E(X)E(Y)}{\sqrt{D(X)D(Y)}}.$$

კორელაციის კოეფიციენტს აქვს შემდეგი თვისებები:

1. თუ X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ $R(X, Y)=0$.

მართლაც, ამ შემთხვევაში $E(XY)=E(X)E(Y)$ და $R(X, Y)$-ის გამოსახულება ნულად იქცევა.

2. შემთხვევითი სიდიდის წრფივად გარდაქმნის შედეგად კორელაციის კოეფიციენტი არ იცვლება.

მართლაც, ვთქვათ

$$X=aZ+b,$$

მაშინ, გვექნება

$$E(X)=aE(Z)+b,$$

$$E(XY)=E(aZY+bY)=aE(ZY)+bE(Y),$$

$$D(X)=a^2D(Z).$$

მიღებული მნიშვნელობები შევიტანოთ $R(X, Y)$-ის გამოსახულებაში, გვექნება

$$R(aZ+b; Y)=\frac{aE(ZY)+bE(Y)-[aE(Z)+b]E(Y)}{a\sqrt{DZDY}}=$$

$$=\frac{aE(ZY)-aE(Z)E(Y)}{a\sqrt{D(Z)D(Y)}}=\frac{E(ZY)-E(Z)E(Y)}{\sqrt{D(Z)D(Y)}}.$$

3. თუ X და Y შემთხვევით სიდიდეთა შორის დამოკიდებულება წრფივია, მაშინ კორელაციის კოეფიციენტი ერთის ტოლია.

მივიღოთ, რომ

$$Y=aX+b,$$

მაშინ

$$E(Y)=aE(X)+b$$

$$D(Y)=a^2D(X)$$

$$E(XY)=E(aX^2+bX)=aE(X^2)+bE(X),$$

ხოლო $\rho(X, Y)$ და $\rho(Y, X)$-თვის გვექნება:

$$\rho(X, Y)=\frac{aE(X^2)+bE(X)-a[E(X)]^2+bE(X)}{D(Y)}=$$

$$=a\frac{E(X^2)-[E(X)]^2}{D(Y)}=a\frac{D(X)}{D(Y)}$$

თუ $D(Y)$-ის მნიშვნელობას შევიტანთ უკანასკნელ გამოსახულებაში, მივიღებთ

$$\rho(X, Y)=\frac{aD(X)}{a^2D(X)}=\frac{1}{a}.$$

$\rho(Y, X)$-თვის დავწერთ

$$\rho(Y, X)=\frac{aE(X^2)+bE(X)-a[E(X)]^2-bE(X)}{D(X)}=$$

$$=a\,\frac{E(X^2)-[E(X)]^2}{D(X)}=a.$$

ამგვარად:

$$R(X, Y)=\sqrt{\rho(X, Y)\,\rho(Y, X)}=\sqrt{\frac{1}{a}\cdot a}=1.$$

ს ა ვ ა რ ჯ ი შ ო

1. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის მათემატიკური ლოდინი, თუ მისი განაწილების კანონია.

$$X\begin{cases}2, & -3, & 7, & 8\\ 0{,}1; & 0{,}3; & 0{,}2; & 0{,}4\end{cases}$$

პას. $E(X)=3{,}9$.

2. მონახეთ $Z=2X+5Y$ შემთხვევითი სიდიდის მათემატიკური ლოდინი, თუკი $E(X)=1$ და $E(Y)=-1$.

პას. $E(Z)=-3$.

3. დისკრეტული ტიპის X შემთხვევითი სიდიდე ღებულობს სამ შესაძლო მნიშვნელობას $x_1=1$-ს ალბათობით $p_1=0{,}3$. $x_2=5$-ს ალბათობით $p_2=0{,}4$ და x_3-ს ალბათობით p_3. იპოვეთ x_3 და p_3, თუ $E(X)=5$.

პას. $p_3=0{,}3$; $x_3=9$.

4. X შემთხვევითი სიდიდე ღებულობს რიცხვით მნიშვნელობებს: $x_1=1$, $x_2=2$ და $x_3=-1$. აგრეთვე ცნობილია, რომ $E(X)=7$ და $E(X^2)=2{,}5$. იპოვეთ x_1, x_2 და x_3-ის შესაბამისი ალბათობები p_1, p_2 და p_3.

პას. $p_1=0{,}1$; $p_2=0{,}5$; $p_3=0{,}4$.

5. X შემთხვევითი სიდიდე ღებულობს რიცხვით მნიშვნელობებს: $x_1=1$, $x_2=0$, $x_3=4$. ცნობილია, რომ $E(X)=1{,}4$ და $E(X^2)=5$. იპოვეთ შესაბამისი ალბათობები p_1, p_2 და p_3.

პას. $p_1=0{,}2$; $p_2=0{,}5$; $p_3=0{,}3$.

6. ნაწარმი მოწმდება ვარგისიანობის თვალსაზრისით. ნაწარმი რომ სტანდარტულია, ამის ალბათობა არის 0,9. თითოეული პარტია 5 ნაწარმს შეიცავს. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის მათემატიკური ლოდინი (X არის პარტიათა რაოდენობა, რომელშიაც ზუსტად 4 ნაწარმია სტანდარტული), თუ მოწმდება 50 პარტია.

პას. $E(X)=50\ C_5^4\ 0{,}9^4\cdot 0{,}1\approx 16.$

7. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის მათემატიკური ლოდინი, დისპერსია და საშუალო კვადრატული გადახრა, რომლის განაწილების კანონია:

$$X\left\{\begin{array}{llll} -5, & 2, & 3, & 4 \\ 0{,}4; & 0{,}3; & 0{,}1; & 0{,}2. \end{array}\right.$$

პას. $E(X)=-0{,}3;\ D(X)=15.21;\ \sigma_X\approx 3{,}9.$

8. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის დისპერსია და საშუალო კვადრატული გადახრა, თუ მისი განაწილების კანონია:

$$X\left\{\begin{array}{llll} 131, & 140; & 160, & 180, \\ 0{,}05 & 0{,}1; & 0{,}25; & 0{,}6. \end{array}\right.$$

პას. $D(X)\approx 248{,}35;\ \sigma_X\approx 15{,}77.$

9. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის დისპერსია (X არის A ხდომილობის ორი ურთიერთდამოუკიდებელი ცდის დროს მოხდენათა რიცხვი), თუ A-ს მოხდენის ალბათობა ორივე ცდისას ერთნაირია და $E(X)=1{,}2.$

პას. $D(X)=0{,}48.$

10. დისკრეტული ტიპის X შემთხვევით სიდიდეს შეუძლია მიიღოს მხოლოდ ორი მნიშვნელობა x_1 და x_2 ($x_2>x_1$). ალბათობა იმისა, რომ X მიიღებს მნიშვნელობას x_1, არის 0,6. მონახეთ X-ის განაწილების კანონი, თუ $E(X)=1{,}4;\ D(X)=0{,}24.$

პას. $X\left\{\begin{array}{ll} 1, & 2, \\ 0{,}6; & 0{,}4. \end{array}\right.$

11. დისკრეტული ტიპის X შემთხვევითი სიდიდე ღებულობს მხოლოდ ორ მნიშვნელობას: x_1 და x_2 ($x_1<x_2$). მონახეთ X-ის განაწილების კანონი, თუ $E(X)=2{,}6$ და $\sigma_X=0{,}8.$

პას. $X\left\{\begin{array}{ll} 1, & 3, \\ 0{,}2; & 0{,}8. \end{array}\right.$

12. დისკრეტული ტიპის X შემთხვევითი სიდიდე ღებულობს მხოლოდ 3 მნიშვნელობას: $x_1=1,\ x_2$ და x_3 ($x_1<x_2<x_3$). ალბათობა იმისა, რომ

X მიიღებს მნიშვნელობებს x_1 და x_2-ს, სათანადოდ ტოლია 0,3 და 0,2-ის, მონახეთ X-ის განაწილების კანონი, თუკი $E(X)=2{,}2$ და $D(X)=0{,}76$.

პას. $X \begin{cases} 1, & 2, & 3, \\ 0{,}3; & 0{,}2; & 0{,}5. \end{cases}$

13. მონახეთ პუასონის კანონით განაწილებული X შემთხვევითი სიდიდის მათემატიკური ლოდინი.

პას. $E(X)=\lambda, \quad (\lambda=np)$.

14. იპოვეთ პუასონის კანონით განაწილებული X შემთხვევითი სიდიდის დისპერსია.

პას. $D(X)=\lambda \quad (\lambda=np)$.

15. დისკრეტული ტიპის X შემთხვევითი სიდიდის განაწილების კანონია:

$$X \begin{cases} 1, & 2, \\ 0{,}3; & 0{,}7. \end{cases}$$

იპოვეთ პირველი, მეორე, მესამე და მეოთხე რიგის საწყისი მომენტები.

პას. $\nu_1=1{,}7;\ \nu_2=3{,}1;\ \nu_3=5{,}9;\ \nu_4=11{,}5$.

16. დისკრეტული ტიპის X შემთხვევითი სიდიდის განაწილების კანონია:

$$X \begin{cases} 1, & 2, & 4, \\ 0{,}1; & 0{,}3; & 0{,}6. \end{cases}$$

იპოვეთ პირველი ოთხი რიგის ცენტრალური მომენტი.

პას. $\mu_1=0;\ \mu_2=1{,}29;\ \mu_3=-0{,}888,\ \mu_4=2{,}7777$.

17. დისკრეტული ტიპის X შემთხვევითი სიდიდის განაწილების კანონია:

$$X \begin{cases} 3, & 5, \\ 0{,}2; & 0{,}8. \end{cases}$$

იპოვეთ პირველი ოთხი რიგის ცენტრალური მომენტები.

პას. $\mu_1=0;\ \mu_2=0{,}64;\ \mu_3=-0{,}12;\ \mu_4=1{,}33$.

18. უწყვეტი ტიპის X შემთხვევითი სიდიდის განაწილების სიმკვრივეა:

$$f(x)=a(1+x^2)^{-\frac{n+1}{2}},$$

სადაც $n>1$ მთელი დადებითი რიცხვია. იპოვეთ X-ის მათემატიკური ლოდინი და დისპერსია.

პას. $E(X)=0;\ D(X)=\dfrac{1}{n-2} \quad (n>2)$.

19. იპოვეთ ნორმალურად განაწილებული შემთხვევითი სიდიდის მოდა და მედიანა.

პას. მოდაც და მედიანაც ერთი და იგივეა და ემთხვევა ნორმალურად განაწილებულ შემთხვევითი სიდიდის მათემატიკურ ლოდინს. ე. ი.

$$M_0 = M_e = E(X) = a.$$

20. ξ შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x) = \begin{cases} a\cos x, & \text{როცა} \quad |x| \leqslant \dfrac{\pi}{2}, \\ 0, & \text{როცა} \quad |x| > \dfrac{\pi}{2}. \end{cases}$$

იპოვეთ ξ შემთხვევითი სიდიდის მათემატიკური ლოდინი და დისპერსია.

პას. $E(\xi) = 0$, $D(\xi) = \dfrac{\pi^2 - 8}{4}$.

21. მონახეთ $Z = 4X - 5Y$ შემთხვევითი სიდიდის მათემატიკური ლოდინი, თუ ვიცით, რომ $E(X) = 1$ და $E(Y) = -1$.

პას. $E(Z) = -1$.

22. მონახეთ $Z = 5X - 2Y + 1$ შემთხვევითი სიდიდის დისპერსია, თუ $D(X) = D(Y) = 2$.

პას. $D(Z) = 58$.

23. მონახეთ $Z = 3(X - 2) + Y$ შემთხვევითი სიდიდის დისპერსია, თუ $D(X) = 1$ და $D(Y) = 5$.

პას. $D(Z) = 14$.

24. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის დისპერსია და საშუალო კვადრატული გადახრა, თუ ვიცით მისი განაწილების კანონი:

$$X \begin{cases} -5, & 2, & 3, & 4 \\ 0{,}4; & 0{,}3; & 0{,}1; & 0{,}2. \end{cases}$$

პას. $D(X) = 15{,}21$; $\sigma_X = 3{,}9$.

25. მონახეთ დისკრეტული ტიპის X შემთხვევითი სიდიდის დისპერსია, თუ ის წარმოადგენს რაიმე A ხდომილობის 5 დამოუკიდებელი ცდის დროს მოხდენათა რიცხვს და ცალკეული ცდის დროს $P(A) = 0{,}2$.

პას. $p = 0{,}8$.

26. დაამტკიცეთ, რომ მეორე რიგის ცენტრალური მომენტი ყოველთვის ნაკლებია სხვა ჩვეულებრივ მეორე რიგის მომენტზე.

ე. ი. აჩვენეთ, რომ

$$E[X-E(X)]^2<E(X-a)^2,$$

სადაც $a \neq E(X)$.

27. აჩვენეთ, რომ მესამე რიგის ცენტრალური მომენტი μ_3 დაკავშირებულია საწყის მომენტებთან შემდეგი თანაფარდობით:

$$\mu_3=\nu_3-3\nu_1\nu_2+2\nu_1^3.$$

28. აჩვენეთ, რომ მეოთხე რიგის ცენტრალური მომენტი μ_4 დაკავშირებულია საწყის მომენტებთან შემდეგი თანაფარდობით:

$$\mu_4=\nu_4-4\nu_3\nu_1+6\nu_1^2\nu_2-3\nu_1^4.$$

29. თუ X და Y ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, დაამტკიცეთ, რომ:

$$D(XY)=D(X)D(Y)+[E(Y)]^2D(X)+[E(X)]^2D(Y).$$

VI თავი

ნორმალური განაწილების კანონი

§ 1. ნორმალური განაწილების კანონი და მისი თვისებები

განაწილების კანონთა ისეთ ტიპს, რომელსაც ემორჩილება შემთხვევით სიდიდეთა საგრძნობი უმრავლესობა, წარმოადგენს ე. წ. ნორმალური განაწილების კანონი.

თუ გვაქვს შემთხვევით სიდიდეთა ისეთი მიმდევრობა, რომელთაგან თითოეული წევრი ცალ-ცალკე ყველას ჯამთან შედარებით უმნიშვნელოა, ვიტყვით, რომ აღებული მიმდევრობა ნორმალურ კანონს ემორჩილება.

მაგალითად, შემთხვევითი ცდომილებანი ნორმალური განაწილების კანონს ემორჩილება. თუ გვინდა გასაზომი სიდიდის ჭეშმარიტი მნიშვნელობის დადგენა, ამისათვის მას მრავალჯერ ვზომავთ. თუ მხედველობაში არ მივიღებთ სისტემატურ ცდომილებებს, ცალკეული გაზომვის შემთხვევაში ცდომილება მცირე სიდიდეს წარმოადგენს. თუ ყველა გაზომვის შედეგად მიღებულ ცდომილებებს განვიხილავთ, მაშინ ცალკეული გაზომვისას მიღებული ცდომილება, ყველა ცდომილებათა ჯამთან შედარებით, უმნიშვნელო სიდიდე იქნება და შეგვიძლია დავასკვ-

ნათ, რომ შემთხვევითი ცდომილებანი ნორმალური განაწილების კანონს ემორჩილებიან.

ვიტყვით, რომ შემთხვევითი სიდიდე ნორმალური განაწილების კანონს ემორჩილება, თუკი მისი განაწილების ფუნქციას აქვს ასეთი სახე:

$$F(x)=k\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\,dt, \tag{1.1}$$

სადაც k, σ და a მუდმივი სიდიდეებია, რომელთაგან k და σ დადებითია.

ნორმალური განაწილების სიმკვრივეს ექნება შემდეგი სახე:

$$f(x)=ke^{-\frac{(x-a)^2}{2\sigma^2}}. \tag{1.2}$$

იმ შემთხვევაში, როდესაც $a=0$ და $\sigma=1$, გვექნება ძირითადი ნორმალური განაწილების კანონი.

როგორც ვიცით, თუ $f(x)$ განაწილების სიმკვრივეა,

$$\int_{-\infty}^{\infty} f(x)dx=1;$$

ნორმალური კანონის შემთხვევაში გვექნება:

$$k\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx=1.$$

შემოვიღოთ აღნიშვნა:

$$\frac{x-a}{\sigma}=z;$$

მაშინ

$$dx=\sigma dz$$

და გვექნება

$$k\sigma\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz=1.$$

როგორც ცნობილია,

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz=\sqrt{2\pi},$$

ამიტომ

$$k=\frac{1}{\sigma\sqrt{2\pi}}.$$

თუ k მუდმივის მიღებულ მნიშვნელობას ნორმალური განაწილების სიმკვრივის გამოსახულებაში ჩავსვამთ, გვექნება:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}. \tag{1.3}$$

ძირითადი ნორმალური განაწილების სიმკვრივე იქნება:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \tag{1.4}$$

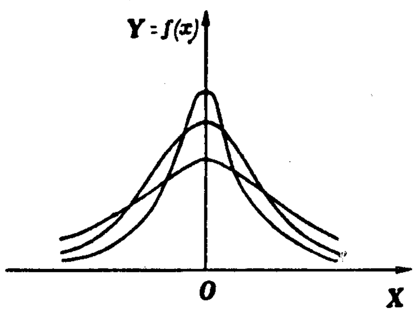
განაწილების სიმკვრივე $x=a$ წერტილზე აღწევს მაქსიმუმს, რომელიც $\frac{1}{\sigma\sqrt{2\pi}}$ სიდიდის ტოლია.

განაწილების სიმკვრივის გამომსახველ მრუდს, ცხადია, ექნება ორი გადაღუნვის წერტილი $x=a\pm\sigma$ წერტილებზე.

როდესაც $x=\pm\infty$, მაშინ განაწილების სიმკვრივის მრუდისათვის აბსცისათა ღერძი ასიმპტოტს წარმოადგენს.

რადგანაც შემთხვევითი სიდიდე, რომელიც ნორმალური განაწილების კანონს ემორჩილება, რიცხვით მნიშვნელობათა უამრავ რაოდენობას ღებულობს, ამიტომ უკეთესი და უფრო მოსახერხებელიც არის ნორმალური განაწილების კანონის გრაფიკულად გამოსახვა.

ვიგულისხმოთ, რომ $a=0$; მაშინ σ-ს სხვადასხვა მნიშვნელობისათვის განაწილების სიმკვრივის მრუდს ექნება შემდეგი სახე (ნახ. 13).



ნახ. 13.

მე-13 ნახაზზე ნაჩვენებ განაწილების სიმკვრივის მრუდთაგან ყველაზე დიდი მაქსიმუმის მქონე მრუდს ყველაზე მცირე σ შეესაბამება. პირიქით, ყველაზე დაბალი მაქსიმუმის მქონე მრუდის შესაბამისი σ ყველაზე დიდი იქნება.

საზოგადოდ, თუ a მუდმივი დადებით მნიშვნელობებს ღებულობს, განაწილების სიმკვრივის მრუდს 13-ე ნახაზზე ნაჩვენები ფორმა ექნება, მხოლოდ გადაწეული იქნება მარჯვნივ, თუ $a<0$, მაშინ მრუდი მარცხნივ იქნება გადაწეული.

იმ შემთხვევაში, როდესაც $a=0$, რაც უფრო მცირეა σ-ს მნიშვნელობა, მით უფრო დიდი იქნება $(-\alpha, \alpha)$ შუალედში შემთხვევითი სიდიდის მიერ მიღებული შესაძლო რიცხვითი მნიშვნელობის მოხვედრის ალბათობა.

ამგვარად, σ წარმოადგენს სიდიდეს, რომელიც გვიჩვენებს შემთხვევითი სიდიდის მიერ მიღებული შესაძლო რიცხვითი მნიშვნელობების გაფანტულობას.

როგორც ვნახეთ, σ-ს სხვადასხვა მნიშვნელობისათვის სხვადასხვა მაქსიმუმის მქონე მრუდებს ვიღებთ, რომელთაც საერთო თვისებები აქვს. მაგალითად, ყოველ ასეთ მრუდთაგანს მხოლოდ ერთი მაქსიმუმის წერტილი ექნება, თითოეული მრუდი სიმეტრიულად არის განლაგებული მაქსიმუმის შესაბამისი ორდინატების მიმართ, თითოეული მათგანი მაქსიმუმის წერტილიდან თანაბრად ეცემა და აქვს მარჯვნივ და მარცხნივ გადაღუნვის წერტილები.

როგორც ვნახეთ, განაწილების სიმკვრივის მრუდის ქვემოთ მოთავსებული ფართობი ალბათობას გამოსახავს. ეს ფართობი არის ალბათობა იმისა, რომ შემთხვევითი სიდიდე რომელიმე თავის შესაძლო რიცხვით მნიშვნელობას მიიღებს. ეს კი აუცილებელი ხდომილობის ალბათობაა, რაც ერთის ტოლია. ამგვარად, რა სახეც არ უნდა ჰქონდეს განაწილების სიმკვრივის მრუდს, მის ქვემოთ მოთავსებული ფართობი ყოველთვის ერთის ტოლი იქნება.

სხვადასხვა განაწილების სიმკვრივის მრუდებს შორის განსხვავება იმაში მდგომარეობს, რომ ისინი სხვადასხვანაირადაა განაწილებული შუალედების მიმართ. ნორმალური განაწილების კანონებისათვის არსებითი მნიშვნელობა აქვს იმას, თუ ფართობის რა ნაწილია მოთავსებული შემთხვევითი სიდიდის უალბათესი მნიშვნელობის მახლობლად და რა ნაწილია ამ უალბათესი მნიშვნელობიდან შედარებით შორს. ისეთი განაწილების სიმკვრივის მრუდების შემთხვევაში, რომელთაც საშუალო კვადრატული გადახრა მცირე აქვთ, შემთხვევითი სიდიდის შესაძლო რიცხვითი მნიშვნელობანი ძირითადად თავმოყრილი იქნება უალბათესი მნიშვნელობის გარშემო, ხოლო შედარებით უმნიშვნელო ნაწილი უალბათესი მნიშვნელობიდან შორს მოთავსდება.

ნორმალური განაწილების შემთხვევაში მისი სიმეტრიულობის გამო უალბათესი მნიშვნელობა ყოველთვის ემთხვევა შემთხვევითი სიდიდის საშუალო მნიშვნელობას.

ახლა გავარკვიოთ ნორმალური განაწილების სიმკვრივის გამოსახულებაში შემავალი a და σ^2 მუდმივების ალბათური ბუნება.

პირველ რიგში ვიპოვოთ ნორმალური კანონით განაწილებული შემთხვევითი სიდიდის მათემატიკური ლოდინი.

როგორც ვიცით, უწყვეტი ტიპის შემთხვევითი სიდიდეებისათვის

$$E(x)=\int_{-\infty}^{\infty} xf(x)\,dx.$$

ამ გამოსახულებაში შევიტანოთ $f(x)$ ნორმალური განაწილების სიმკვრივის მნიშვნელობა, გვექნება:

$$E(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}}\,dx.$$

შემოვიღოთ აღნიშვნა:

$$\frac{x-a}{\sigma}=z.$$

მაშინ გვექნება:

$$E(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z+a)\,e^{-\frac{z^2}{2}}\,dz=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z\,e^{-\frac{z^2}{2}}\,dz+$$

$$+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz.$$

როგორც წინათ გვქონდა

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz=\sqrt{2\pi},$$

ხოლო

$$\int_{-\infty}^{\infty} z\,e^{-\frac{z^2}{2}}\,dz=-e^{-\frac{z^2}{2}}\Big|_{-\infty}^{\infty}=0.$$

მაშასადამე, მივიღებთ:

$$E(x)=a. \tag{1.5}$$

ამგვარად, ნორმალური განაწილების სიმკვრივის გამოსახულებაში შემავალი a მუდმივი, ნორმალური კანონით განაწილებული შემთხვევითი სიდიდის მათემატიკური ლოდინი ყოფილა.

ახლა ვიპოვოთ ნორმალური კანონით განაწილებული შემთხვევითი სიდიდის დისპერსია.

როგორც ვიცით, უწყვეტი ტიპის შემთხვევითი სიდიდეებისათვის

$$D(x)=\int_{-\infty}^{\infty}[x-E(x)]^2 f(x)\,dx.$$

ჩვენს შემთხვევაში $E(x)=a$. შევიტანოთ $f(x)$-ის მნიშვნელობა $D(x)$-ის გამოსახულებაში, გვექნება:

$$D(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

აქაც აღვნიშნოთ:

$$\frac{x-a}{\sigma}=z.$$

მაშინ

$$D(x)=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2 e^{-\frac{z^2}{2}}dz=\frac{\sigma^2}{\sqrt{2\pi}}\left(-z e^{-\frac{z^2}{2}}\Big|_{-\infty}^{\infty}+\right.$$

$$\left.+\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz\right)=\frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi}=\sigma^2,$$

ე. ი.

$$D(x)=\sigma^2. \tag{1.6}$$

ამგვარად, ნორმალური კანონით განაწილებული შემთხვევითი სიდიდის დისპერსია ყოფილა σ^2, ხოლო σ კი იქნება ნორმალურად განაწილებული შემთხვევითი სიდიდის საშუალო კვადრატული გადახრა.

ვთქვათ, x შემთხვევითი სიდიდე ისეთი ნორმალური კანონითაა განაწილებული, რომლის მათემატიკური ლოდინია a და საშუალო კვადრატული გადახრა კი σ.

ვიპოვოთ ალბათობა იმისა, რომ $x-a$ გადახრის აბსოლუტური მნიშვნელობა k რიცხვს არ აღემატება

$$P(|x-a|<k)=?$$

თუ X შემთხვევითი სიდიდე განაწილებულია ნორმალური კანონით, მაშინ ალბათობა იმისა, რომ ის მოხვდება ძალიან მცირე $(x,\ x+dx)$ შუალედში, გამოისახება შემდეგნაირად:

$$P\{x<X<x+dx\}=f(x)\,dx=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

საზოგადოდ, ალბათობა იმისა, რომ X მოხვდება $(x_1,\ x_2)$ შუალედში, იქნება:

$$P(x_1<X<x_2)=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

ვთქვათ,

$$\frac{x-a}{\sigma}=t, \quad \text{მაშინ} \quad dx=\sigma dt;$$

$$t_1=\frac{x_1-a}{\sigma} \quad \text{და} \quad t_2=\frac{x_2-a}{\sigma}, \tag{1.7}$$

ხოლო

$$P(x_1<X<x_2)=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{t^2}{2}}dt=\Phi(t_2)-\Phi(t_1).$$

თუ x_1 და x_2-ს შევცვლით (1.7) თანაფარდობიდან t_1 და t_2-ით, დავწერთ:

$$P(x_1<X<x_2)=P(\sigma t_1+a<X<\sigma t_2+a)=$$
$$=P(\sigma t_1<X-a<\sigma t_2)=\Phi(t_2)-\Phi(t_1).$$

ამგვარად მივიღეთ, რომ

$$P(\sigma t_1<X-a<\sigma t_2)=\Phi(t_2)-\Phi(t_1).$$

თუკი

$$-t_1=t_2=t,$$

მაშინ გვექნება:

$$P(|X-a|<t)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx=2\Phi(t).$$

II ცხრილის დახმარებით ვიპოვით, რომ

როცა $t=1$, მაშინ $2\Phi(t)=0,683$,
როცა $t=2$, მაშინ $2\Phi(t)=0,954$,
როცა $t=3$, მაშინ $2\Phi(t)=0,997$.

ხდომილობა, რომლის ალბათობა აღემატება 0,997, ითვლება პრაქტიკულად აუცილებელ ხდომილობად. ამიტომ, თუ შემთხვევითი სიდიდე ნორმალური კანონითაა განაწილებული, მაშინ მისი მათემატიკური ლოდინისაგან გადახრა რომ არ აღემატება გასამკეცებულ საშუალო კვადრატულ გადახრას, პრაქტიკულად აუცილებლად ითვლება.

ნორმალური განაწილების a და σ^2 პარამეტრების მოცემა მის სახეს სავსებით განსაზღვრავს.

§ 2. ნორმალური განაწილების მრუდის აგება ემპირიული მონაცემებით

მაგალითისათვის ავიღოთ მთლიანი რკინიგზის ქსელის დატვირთვის ყოველდღიური მონაცემები [1].

[1] მაგალითი აღებულია წიგნიდან: Б. С. Ястремский, Математическая статистика, Госстатиздат, 1956.

არ არის სავალდებული, რომ დღეთა რაოდენობა ძალიან დიდი იყოს. უნდა ვეცადოთ, ისე შევარჩიოთ დღეთა რაოდენობა, რომ მათში არ მოხვდეს ისეთი დღეები, რომლებსაც სეზონიდან სეზონზე გადავყავართ და ამით მკვეთრად იცვლება დატვირთვათა სიდიდე მთელი რკინიგზის ქსელზე. დიდ სადგურებზე დატვირთვა შეიძლება საგრძნობლად იცვლებოდეს, ხოლო პატარა სადგურებზე კი—უმნიშვნელოდ. ალბათობის განაწილების ხასიათი სხვადასხვა სადგურზე შესაძლებელია ერთმანეთისაგან სავსებით განსხვავებული იყოს. მაგრამ საერთო ჯამი ცალკეულ სადგურთა დატვირთვების ჯამისაგან შედგება, რომელთა რიცხვი საკმაოდ დიდია და რომელთა დატვირთვა ერთმანეთისაგან დამოუკიდებელია. ამიტომ აქ თითქმის მკაცრად სრულდება ნორმალობისათვის საჭირო პირობები.

მე-4 ცხრილში წარმოდგენილია დატვირთვა 80 დღის განმავლობაში 1939 წლის ივნის-აგვისტოში. ცხრილში ნაჩვე ებია დატვირთვის ნორმირებულ მნიშვნელობათა გადაანგარიშება. გადაანგარიშების დროს გამოყენებულია ორი ემპირიული მახასიათებელი: საშუალო და ვირთვა ($\bar{x}=98,3$) და საშუალო კვადრატული გადახრა ($\sigma_x=2,78$). მათი გამოთვლა ადრე შესწავლილი გვაქვს. აქაც ვნახოთ, თუ როგორ მიიღება ისინი.

ცხრილი 4

დატვირთვა (გეგმის პროცენტებით)	დღეთა რიცხვი s_i	ზედა საზღვრის ნორმირებული მნიშვნელობა t	$\Phi(t)$	p_i	$\bar{s}_i$	x_i	$\varphi(t)$	y_i
91— 93	2	—1,91	0,028	0,028	2	—2,27	0,030	1,3
93— 95	7	—1,19	0,117	0,089	7	—1,55	0,120	6,91
95— 97	17	—0,47	0,319	0,202	16	—0,83	0,283	16,30
97— 99	22	0,25	0,599	0,280	22	—0,11	0,396	22,81
99—101	19	0,97	0,834	0,235	19	0,61	0,331	19,07
101—103	10	1,69	0,954	0,120	10	1,33	0,165	9,50
103—105	2	2,41	0,992	0,038	3	2,05	0,049	2,82
105	1		1,000	0,008	1	2,77	0,009	0,52
Σ	80			1,000	80			

$\bar{x}$-ის გამოსათვლელად საჭიროა ავიღოთ დატვირთვათა შესაბამის ინტერვალებს შუა მნიშვნელობანი. ისინი გადავამრავლოთ შესაბამისად s_i დღეთა რიცხვზე, მიღებული შედეგები შევკრიბოთ და გავყოთ დღეთა საერთო რაოდენობაზე—80-ზე. მართლაც,

$$\bar{x}=\frac{1}{80}(2.92+7.94+17.96+22.98+19.100+10.102+$$
$$+2.104+1.106)=98,3.$$

σ_x-ის მისაღებად, საჭიროა ვიპოვოთ $\overline{x^2}$, მას გამოვაკლოთ $\bar{x}^2$ და მიღებული გამოსახულებიდან ამოვიღოთ კვადრატული ფესვი:

$$x=\frac{1}{80}(2\cdot 92^2+7\cdot 94^2+17\cdot 96^2+22\cdot 98^2+$$

$$+19\cdot 100^2+10\cdot 102^2+2\cdot 104^2+1\cdot 106^2)=9670,6,$$

ხოლო

$$\bar{x}^2=9662,89,$$

მაშინ

$$\sigma_x=\sqrt{9670,60-9662,89}=\sqrt{7,71}=2,78.$$

t ნორმირებული გადახრების საანგარიშებლად ინტერვალების ბოლო მნიშვნელობებს უნდა გამოაკლდეს $\bar{x}$ და მიღებული შედეგები გაიყოს σ_x-ზე:

$$t=\frac{X-98,3}{2,78},$$

სადაც X ინტერვალთა ბოლოების მნიშვნელობანია. მაგალითად, პირველ სტრიქონში გვექნება:

$$\frac{93-98,3}{2,78}=-1,91$$

და ა. შ. შეივსება t ს მნიშვნელობანი.

$\Phi(t)$-ს მნიშვნელობებს ვიპოვით I ცხრილის დახმარებით. მაგალითად, როცა $t=0,25$, მაშინ ცხრილით ვიპოვით: $\Phi(t)=0,599$, ხოლო როცა $t=-1,91$, მაშინ $\Phi(-1,91)=1-\Phi(1,91)$, ხოლო ცხრილის დახმარებით კი მივიღებთ $\Phi(1,91)=0,972$, მაშასადამე,

$$\Phi(-1,91)=1-0,972=0,028$$

და ა. შ.

თუ ყველა $\Phi(t)$-ს მის წინა მნიშვნელობას გამოვაკლებთ, მივიღებთ p_i ალბათობებს იმისას, რომ დატვირთვა მოთავსებულია გარკვეულ ჯგუფში

$$p_i=\Phi(t_{i+1})-\Phi(t_i),\quad (i=0,\ 1,\ 2,\ldots,\ 8),$$

$$p_0=\Phi(t_1)-\Phi(t_0)=0,028-0=0,028,$$

$$p_1=\Phi(t_2)-\Phi(t_1)=0,117-0,028=0,089$$

და ა. შ.

იმისათვის, რომ ამათგან მივიღოთ რიგი „თეორიული“ სიხშირეებისა $\bar{s}_i$, რომელთაც ვადარებთ s_i-ებს, საჭიროა p_i მნიშვნელობანი გადავამრავლოთ 80-ზე. მაგალითად, $0,028\cdot 80=2$; $0,089\cdot 80=7$ და ა. შ.

როგორც ვხედავთ, s_i და $\bar{s}_i$ სიდიდეთა შორის თითქმის ყველგან

თანმთხვევას აქვს ადგილი. განსხვავება გვხვდება მხოლოდ ერთი ერთეულით, უკანასკნელის წინა და ზემოდან მესამე სტრიქონში.

იმისათვის, რომ ავაგოთ სათანადო გრაფიკი, საჭიროა ვიცოდეთ განაწილების სიმკვრივე $\varphi(x)$ თითოეული ინტერვალის შუა წერტილისათვის.

x არგუმენტები იანგარიშება შემდეგნაირად: ინტერვალის შუა მნიშვნელობებს აკლდება $\overline{x}$ და იყოფა σ_x-ზე. მაგალითად, პირველი სტრიქონისათვის

$$x_1=\frac{92-98{,}3}{2{,}78}=-2{,}27,$$

ხოლო მისი შესაბამისი განაწილების სიმკვრივე იქნება:

$$\varphi(-2{,}27)=\varphi(2{,}27)=0{,}030.$$

ამ უკანასკნელს ვიპოვით I ცხრილის მიხედვით.

ანალოგიურად,

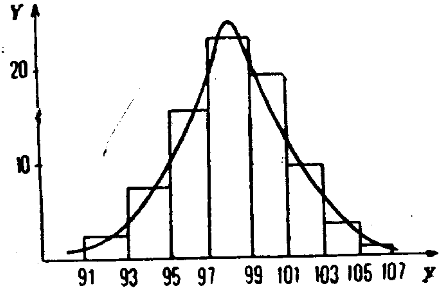
$$x_2=\frac{94-98{,}3}{2{,}78}=-1{,}55;$$

ხოლო $\varphi(-1{,}55)=\varphi(1{,}55)=0{,}120$ და ა. შ.

ყოველი ინტერვალის სიგრძეა ორი ერთეული ($h=2$). თითოეულ ინტერვალში არის $2:2{,}78=0{,}72$ ნორმირებული ერთეული. სწორედ ამ სიდიდეზე უნდა გამრავლდეს $\varphi(x)$, რათა მივიღოთ ფარდობითი სიხშირის „თეორიული" სიმკვრივე. სიხშირის „თეორიული" სიმკვრივის მისაღებად კი ეს უკანასკნელი უნდა გამრავლდეს 80-ზე. ამგვარად, ნორმალური განაწილების მრუდის ორდინატი ინტერვალებს შუა მნიშვნელობათათვის, რომლებსაც ვადარებთ s_i ფარდობით სიხშირეებს, გამოითვლება შემდეგი ფორმულით:

$$y_i=\varphi(x)\cdot\frac{h}{\sigma}\sum s_i=\varphi(x)\,.\,0{,}72\cdot 80=57{,}6\,\varphi(x).$$

უშუალო აგება მოგვცემს შემდეგ სურათს (ნახ. 14):



ნახ. 14.

როცა $x = -2{,}27$, მაშინ $\varphi(x) = 0{,}03$, ხოლო

$$y = 57{,}6 \cdot 0{,}03 = 1{,}73.$$

ასევე, როცა $x = -1{,}55$, მაშინ $\varphi(x) = 0{,}12$, ხოლო

$$y = 57{,}6 \cdot 0{,}12 = 6{,}91$$

და ა. შ.

მე-14 ნახაზზე ერთდროულად მოცემულია როგორც „თეორიული", ისე კიბისებრი მრუდები, რომლებიც მოცემულ სიხშირეებს შეესაბამება.

მოცემულ სიხშირეთა შესაბამის კიბისებრ მრუდს ჰისტოგრამას უწოდებენ. დისკრეტული განაწილების შემთხვევაში ემპირიული სიხშირეები კიბეების ნაცვლად წერტილებს გვაძლევენ, რომელთაც სწორის მონაკვეთებით აერთებენ. ამ შემთხვევაში მიღებულ ტეხილს **სიხშირეთა პოლიგონს** უწოდებენ.

ზემოთ მოყვანილი მაგალითი იმითაა შესანიშნავი, რომ იდეალურად ეთანადება თეორიული განაწილება ემპირიულს, რომელიც ნორმალობის პირობებს აკმაყოფილებს.

შემთხვევითი სიდიდეები, რომლებიც ნორმალურ კანონს ემორჩილება ან ახლოსაა ნორმალურ კანონთან, საკმაოდ ბევრი გვხვდება ბუნებაში. ისინი განსაკუთრებით გვხვდებიან ბიოლოგიურ მოვლენებში, წარმოებათა ცდომილებების შესწავლისას, გაზომვათა ცდომილებებში და ა. შ. როგორც ვნახეთ, თუ ვიცით, რომ განაწილება ნორმალურ კანონს ემორჩილება, მაშინ მისი აგება შეიძლება მხოლოდ ორი პარამეტრის მოცემით. იგი გამოიყენება რიგი მნიშვნელოვანი პრაქტიკული ამოცანების გადაწყვეტისას.

მაგალითად, თუ ვიცით, რომ ადამიანების ფეხსაცმელთა ზომები განაწილებულია ნორმალური კანონით, მაშინ საშუალო და საშუალო კვადრატული ცდომილების მოცემით შესაძლებელია ავაგოთ ყველა განაწილება, ე. ი. შესაძლებელია გამოვთვალოთ, თუ რამდენია ისეთი ადამიანი, რომელსაც ამა თუ იმ ზომის ფეხსაცმელი სჭირდება.

§ 3. ასიმეტრია და ექსცესი

შემთხვევითი სიდიდის განაწილების პირველი საწყისი მომენტია მისი მათემატიკური ლოდინი

$$\nu_1 = E(X),$$

ხოლო მეორე ცენტრალური მომენტი—მისი დისპერსიაა—

$$\mu_2 = D(X).$$

ასევე შეიძლება მესამე და მეოთხე რიგის მომენტთათვისაც კონკრეტული ალბათური შინაარსის მიცემა.

სტატისტიკურ განაწილებათა დახასიათების მიზნით განვიხილოთ ნორმალური განაწილების ცენტრალური მომენტები.

როგორც ვიცით, ნორმალური განაწილების მრუდის განტოლებაა:

$$y=\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{3.1}$$

სადაც a არის X შემთხვევითი სიდიდის მათემატიკური ლოდინი, ხოლო σ^2—მისი დისპერსია.

შემთხვევითი სიდიდის ცენტრალური მომენტი, როდესაც ის დისკრეტულ მნიშვნელობებს ღებულობს, განმარტებული გვქონდა შემდეგნაირად:

$$\mu_k=E[X-E(X)]^k,$$

იმ შემთხვევაში, როდესაც შემთხვევითი სიდიდე ღებულობს უწყვეტ მნიშვნელობებს, ცენტრალური მომენტი განისაზღვრება შემდეგნაირად:

$$\mu_k=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)^k e^{-\frac{(x-a)^2}{2\sigma^2}}dx. \tag{3.2}$$

შემოვიღოთ აღნიშვნა:

$$\frac{x-a}{\sigma}=t. \tag{3.3}$$

t-ს ნორმირებული შემთხვევითი სიდიდე ეწოდება, თუ X შემთხვევითი სიდიდის მათემატიკური ლოდინი არის a, ხოლო დისპერსია—σ^2. როგორც არ უნდა იყოს განაწილებული X შემთხვევითი სიდიდე, ყოველთვის გვექნება:

$$E(t)=0,\ \sigma_t^2=E(t^2)=1.$$

(3.3) ფორმულის გათვალისწინებით (3.2)-დან მივიღებთ.

$$\mu_k=\frac{\sigma^k}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^k e^{-\frac{t^2}{2}}dt. \tag{3.4}$$

თუ ამ უკანასკნელი გამოსახულების მარჯვენა მხარეში შევასრულებთ ნაწილობით ინტეგრებას, მივიღებთ:

$$\mu_k=\frac{(k-1)}{\sqrt{2\pi}}\sigma^k\int_{-\infty}^{\infty}t^{k-2}e^{-\frac{t^2}{2}}dt. \tag{3.5}$$

(3.5) ტოლობა დავწეროთ შემდეგნაირად:

$$\mu_k=(k-1)\,\sigma^2\,\frac{\sigma^{k-2}}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}t^{k-2}e^{-\frac{t^2}{2}}\,dt. \tag{3.6}$$

მაგრამ, (3.4) ფორმულის თანახმად,

$$\frac{\sigma^{k-2}}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}t^{k-2}e^{-\frac{t^2}{2}}\,dt=\mu_{k-2}.$$

ამგვარად, (3.6) ტოლობიდან მივიღებთ შემდეგ რეკურენტულ ფორმულას:

$$\mu_k=(k-1)\,\sigma^2\,\mu_{k-2}. \tag{3.7}$$

ცხადია, რომ როცა $k=1$, მაშინ (3.7) ფორმულიდან ვღებულობთ:

$$\mu_1=0;$$

როცა $k=1$-ს, (3.4) ფორმულიდან ისევ ვღებულობთ:

$$\mu_1=0;$$

როცა $k=3$-ს, მაშინ (6) ფორმულის ძალით ვღებულობთ:

$$\mu_3=2\sigma^2\,\mu_1=0.$$

ასევე, როცა $k=5$, მაშინ

$$\mu_5=4\sigma^2\,\mu_3=0$$

და ა. შ. მივიღებთ, რომ ყველა კენტი რიგის ცენტრალური მომენტი ნორმალური განაწილების შემთხვევაში ნულის ტოლია:

$$\mu_{2r+1}=0, \tag{3.8}$$

სადაც r ღებულობს ნებისმიერ მთელ მნიშვნელობას.

ან აგრეთვე, თუ (3.7) ფორმულაში k-ს შევცვლით $k-2$-ით, გვექნება:

$$\mu_{k-2}=(k-3)\,\sigma^2\,\mu_{k-4}.$$

თუ ამას (3.7) ფორმულაში შევიტანთ, მივიღებთ:

$$\mu_k=(k-1)\,(k-3)\,\sigma^4\,\mu_{k-4}. \tag{3.9}$$

თუ (3.7) ფორმულაში k-ს შევცვლით $k-4$-ით, გვექნება:

$$\mu_{k-4}=(k-5)\,\sigma^2\,\mu_{k-6},$$

თუ მიღებულს შევიტანთ (3,9) ფორმულაში, დავწერთ:

$$\mu_k=(k-1)\,(k-3)\,(k-5)\,\sigma^6\,\mu_{k-6}.$$

და თუ ამ პროცესს გავაგრძელებთ იმ შემთხვევაში, როდესაც k კენტია, მივიღებთ:

$$\mu_k = (k-1)(k-3)\dots 2\sigma^{k-1}\mu_1.$$

როგორც ვნახეთ, $\mu_1=0$ და ამიტომ k-ს კენტი მნიშვნელობისათვის ყოველთვის ყველა კენტი რიგის ცენტრალური მომენტი ნორმალური განაწილების შემთხვევაში ნულის ტოლია.

ვთქვათ, k ლუწია

$$k=2r;$$

მაშინ ანალოგიურად მივიღებთ:

$$\mu_{2r}=(2r-1)(2r-3)\dots 3\cdot 1\,\sigma^{2r}\mu_0.$$

(3.4) ტოლობის ძალით დავწერთ:

$$\mu_0=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t_2}{2}}\,dt=1,$$

რადგანაც, როგორც ცნობილია,

$$\int_{-\infty}^{\infty} e^{-\frac{t_2}{2}}\,dt=\sqrt{2\pi}.$$

ამგვარად, ლუწი რიგის ცენტრალური მომენტებისათვის ნორმალური განაწილების შემთხვევაში გვექნება:

$$\mu_{2r}=(2r-1)(2r-3)\dots 3\cdot 1\cdot\sigma^{2r}.$$

კერძოდ,

$$\mu_2=D(x)=\sigma^2,$$

თუ ამას შევიტანთ ზემოთ მოყვანილ μ_{2r} გამოსახულებაში, მივიღებთ:

$$\mu_{2r}=1\cdot 3\dots(2r-3)(2r-1)\,\mu_2^r. \qquad (3.10)$$

აქვე შეიძლება შევნიშნოთ, რომ არა მარტო ნორმალური განაწილების შემთხვევაში, არამედ ყოველგვარი განაწილებისას, რომელიც მათემატიკური ლოდინის მიმართ სიმეტრიულია, ყველა კენტი რიგის ცენტრალური მომენტი ნულის ტოლი იქნება.

უკანასკნელი მოსაზრების გამო, თუ მაგალითად, რომელიმე განაწილებისათვის მესამე რიგის ცენტრალური მომენტი განსხვავებულია ნულისაგან, ეს იმის მაჩვენებელი იქნება, რომ მოცემული განაწილება ასიმეტრიულია.

რომ მივიღოთ ზომის ერთეულისაგან თავისუფალი ასიმეტრიულობის საზომი, უკეთესია ავიღოთ ნორმირებული შემთხვევითი სიდიდე

$$t=\frac{x-a}{\sigma}.$$

წინასწარ გამოვთვალოთ t^3-ის მათემატიკური ლოდინი. გვაქვს:

$$E(t^3)=E\left[\left(\frac{x-a}{\sigma}\right)^3\right]=\frac{E\,(x-a)^3}{\sigma^3},$$

მაგრამ, ცენტრალური მომენტის განმარტების თანახმად,

$$\mu_3=E(x-a)^3$$

და, მაშასადამე, დავწერთ:

$$E(t^3)=\frac{\mu_3}{\sigma^3}. \qquad (3.11)$$

მიღებულ სიდიდეს ეწოდება x შემთხვევითი სიდიდის განაწილების ასიმეტრია და აღინიშნება S_k-თი.

ამგვარად, ასიმეტრიისათვის მივიღეთ

$$S_k=\frac{\mu_3}{\sigma^3}.$$

თუ n საკმაოდ დიდია და შერჩევას ვღებულობთ ნორმალური ერთობლიობიდან, მაშინ ასიმეტრიის საშუალო კვადრატული გადახრა σ_{S_k} მიახლოებით გამოისახება შემდეგნაირად:

$$\sigma_{S_k}=\sqrt{\frac{6}{n}}.$$

განაწილების მრუდების თეორიაში ხშირად სარგებლობენ β_1 სიდიდით, რომელიც S_k ასიმეტრიის კვადრატის ტოლია:

$$\beta_1=S_k^2=\frac{\mu_3^2}{\sigma^6}=\frac{\mu_3^2}{\mu_2^3},$$

საიდანაც

$$S_k=\sqrt{\beta_1}.$$

კვადრატული ფესვი აიღება μ_3 ნიშნის მიხედვით. შემთხვევითი სიდიდის განაწილების ასიმეტრია დადებითი სიდიდე იქნება, თუ განაწილების მოდა მათემატიკური ლოდინის მარცხნივ მდებარეობს, ხოლო წინააღმდეგ შემთხვევაში ასიმეტრია უარყოფითი იქნება. უარყოფით ასიმეტრიას მარჯვენა ასიმეტრიას უწოდებენ, დადებითს კი — მარცხენა ასიმეტრიას.

ახლა განვიხილოთ ნორმალური განაწილების მეოთხე რიგის ცენტრალური მომენტი. არანორმირებული შემთხვევითი სიდიდისათვის, (3.10) ფორმულის თანახმად, მეოთხე რიგის ცენტრალური მომენტი იქნება:

$$\mu_4 = 3\mu_2^2. \tag{3.12}$$

თუ შემთხვევითი სიდიდე ნორმირებულია, მაგალითად,

$$t = \frac{x-a}{\sigma}\text{-თვის},$$

ნახ. 15.

მეოთხე რიგის ცენტრალური მომენტი იქნება:

$$E\left(\frac{x-a}{\sigma}\right)^4 = \frac{E\,(x-a)^4}{\sigma^4} = \frac{\mu_4}{(\sigma^2)^2} = \frac{\mu_4}{\mu_2^2}.$$

ეს კი (3.12) ტოლობის თანახმად, უდრის სამს, ე. ი. ნორმირებული შემთხვევითი სიდიდისათვის გვექნება:

$$\frac{\mu_4}{\mu_2^2} = 3.$$

ამ ტოლობის მარცხენა მხარეს აღნიშნავენ β_2-ით.

როგორც ვხედავთ, β_2 განყენებული რიცხვია, რომლის კონკრეტულ შინაარსს ვნახავთ შემდგომ.

განვიხილოთ ორი სიმეტრიული განაწილება, რომელთაგან ერთი ნორმალურია. ვიგულისხმოთ, რომ მათ აქვთ ერთნაირი მეოთხე რიგის ცენტრალური μ_4 მომენტი, ხოლო მეორე რიგის მომენტები განსხვავებული აქვთ. ვთქვათ, ნორმალური განაწილებისათვის მეორე რიგის მომენტი არის μ_2, ხოლო მეორე განაწილებისათვის—μ_2'.

ვთქვათ,

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \beta_2' = \frac{\mu_4}{\mu_2'^2}.$$

დავუშვათ, რომ

$$\mu_2 > \mu_2'. \tag{3.13}$$

მაშინ გვექნება

$$\beta_2' = \frac{\mu_4}{\mu_2'^2} > \frac{\mu_4}{\mu_2^2} = 3,$$

რადგანაც ნორმალური განაწილებისათვის $\beta_2 = 3$.

(3.13) უტოლობა გვიჩვენებს, რომ ნორმალურ განაწილებას აქვს მეტი დისპერსია, ვიდრე მეორე არანორმალურ განაწილებას; ეს იმას ნიშნავს, რომ მეორე განაწილებაში, ნორმალურ განაწილებასთან შედარებით, მათემატიკური ლოდინიდან გადახრა უნდა შევხვდეს ხშირად, ხოლო დიდი გადახრები — იშვიათად; მაშასადამე, მეორე განაწილებას ნორმალურთან შედარებით ექნება უფრო მაღალი წვერო. ასეთ შემთხვევაში ამბობენ, რომ მეორე განაწილება მაღალწვეროიანი ან მახვილწვეროიანია.

სიდიდეს

$$E = \beta_2 - 3$$

განაწილების ექსცესი ეწოდება.

ცხადია, რომ მეორე განაწილების შემთხვევაში ექსცესი დადებითი იქნება.

თუკი

$$\mu_2' > \mu_2,$$

მაშინ

$$E < 0.$$

ამ შემთხვევაში მეორე განაწილების მრუდს უფრო დაბალი და ბრტყელი წვერო ექნება, ვიდრე ნორმალური განაწილების მრუდს. თუ მიღებული მრუდი იქნება დაბალი ან ბრტყელწვეროიანი, მაშინ მეორე მრუდის ექსცესი უარყოფითი იქნება.

უნდა შევნიშნოთ, რომ მასშტაბის ცვლილებით ყოველთვის შეგვიძლია მივაღწიოთ იმას, რომ ორ განაწილებას ერთი და იმავე მეოთხე რიგის მომენტი ჰქონდეს. ამგვარად, ზემოთ ჩატარებული მსჯელობა ზოგადად შეიძლება ჩაითვალოს და მივდივართ იმ დასკვნამდე, რომ ექსცესი

$$E = \beta_2 - 3,$$

სადაც

$$\beta_2 = \frac{\mu_4}{\mu_2^2},$$

წარმოადგენს განხილული განაწილების წვეროს, ნორმალური განაწილების წვეროსთან შედარებით, განსაკუთრებული აგებულების მახასიათებელს; სახელდობრ, როცა $E > 0$, გვაქვს მახვილი მაღალი წვერო, როცა $E < 0$ — დაბალი, ბრტყელი წვერო, ხოლო როცა $E = 0$, გვაქვს ნორმალური განაწილება.

n-ის დიდი მნიშვნელობისათვის E ექსცესის საშუალო კვადრატული გადახრა მიახლოებით გამოისახება შემდეგნაირად:

$$\sigma_E = \sqrt{\frac{24}{n}}.$$

იგულისხმება, რომ განხილული განაწილება ნორმალურიდან n მოცულობის შერჩევას წარმოადგენს.

ამგვარად, მესამე და მეოთხე რიგის მომენტები განხილული განაწილების ასიმეტრიულობის და ექსცესის დასახასიათებლად გამოდგება.

სავარჯიშო

1. X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის მათემატიკური ლოდინია 30 და დისპერსია 10. იპოვეთ ალბათობა იმისა, რომ X მოთავსებული იქნება (10, 50) შუალედში.

მ ი თ ი თ ე ბ ა. ისარგებლეთ ფორმულით:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right),$$

სადაც

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

პას. $p(10 < x < 50) = 0{,}9544$.

2. ნორმალურად განაწილებული X შემთხვევითი სიდიდის მათემატიკური ლოდინია 10, ხოლო დისპერსია 2. იპოვეთ ალბათობა იმისა, რომ X მიიღებს მნიშვნელობას (12; 14) შუალედში.

პას. $P(12 < x < 14) = 0{,}1359$.

3. X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის საშუალო კვადრატული გადახრაა 5 მმ. იპოვეთ იმ ინტერვალის სიგრძე, რომლის შიგნით X ხვდება 0,9973-ის ტოლი ალბათობით.

პას. $6\sigma = 30$ მმ.

4. X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის მათემატიკური ლოდინია 10. ალბათობა იმისა, რომ X მოხვდება (10; 20) ინტერვალში, არის 0,3. რას უდრის ალბათობა იმისა, რომ X მოხვდება (0; 10) ინტერვალში?

პას. $P(0 < x < 10) = P(10 < x < 20) = 0{,}3$.

5. X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის მათემატიკური ლოდინია 50. ალბათობა იმისა, რომ X მოხვდება (20; 40) შუალედში, არის 0,4. რას უდრის ალბათობა იმისა, რომ X მოხვდება (60; 80) შუალედში?

პას. $P(20<x<40)=P(60<x<80)=0{,}4$.

6. X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის მათემატიკური ლოდინია 10 და საშუალო კვადრატული გადახრა კი 5. მონახეთ ის ინტერვალი, რომელშიაც X მოხვდება 0,9973-ის ტოლი ალბათობით.

პას. $(a-3\sigma;\ a+3\sigma)=(-5,\ 25)$.

VII თავი

დიდ რიცხვთა კანონი

ვიტყვით, რომ $X_1, X_2, \ldots, X_n$ შემთხვევითი სიდიდეები ემორჩილება დიდ რიცხვთა კანონს, თუ მათი საშუალო არითმეტიკული, როცა მათი რიცხვი საკმაოდ დიდია, ახლოს არის მათ მათემატიკურ ლოდინთა საშუალო არითმეტიკულთან.

ეს იმას ნიშნავს, რომ, თუ შემთხვევით სიდიდეთა რაოდენობა საკმაოდ დიდია, მაშინ მათი საშუალო არითმეტიკული, რომელიც შემთხვევით სიდიდეს წარმოადგენს, თითქმის მუდმივი ხდება.

დიდ რიცხვთა კანონის შესახებ სხვადასხვა დებულებათა დასადგენად საჭიროა წინასწარ განვიხილოთ ჩებიშევის უტოლობა, რომელსაც ეყრდნობა ყველა მათი დასაბუთება.

§ 1. ჩებიშევის უტოლობა

ვთქვათ, X შემთხვევითი სიდიდეა, რომლის კვადრატის მათემატიკური ლოდინი არსებობს, და a რაიმე დადებითი მუდმივი რიცხვია, ამ შემთხვევაში ადგილი ექნება ჩებიშევის შემდეგ უტოლობას:

$$P(X>a)\leqslant\frac{E(X^2)}{a^2}. \tag{1.1}$$

მართლაც, ვთქვათ, შემთხვევითი სიდიდის განაწილების კანონია:

$$X\begin{cases} x_1, & x_2, \ldots, & x_n \\ p_1, & p_2, \ldots, & p_n. \end{cases}$$

მაშინ X^2-ის განაწილების კანონი იქნება:

$$X^2 \begin{cases} x_1^2, & x_2^2, \dots, & x_n^2 \\ p_1, & p_2, \dots, & p_n, \end{cases}$$

საიდანაც, მათემატიკური ლოდინის განმარტების თანახმად, დავწერთ:

$$E(X^2) = \sum_{i=1}^{n} x_i^2 p_i. \tag{1.2}$$

თუ ჯამში შევინარჩუნებთ მხოლოდ ისეთ x_i სიდიდეებს, რომლებიც აღემატება a-ს, მაშინ, ცხადია, ადგილი ექნება შემდეგ უტოლობას:

$$E(X^2) \geqslant \sum_{x_i > a} x_i^2 p_i, \tag{1.3}$$

მაგრამ რადგანაც $a > 0$, ამიტომ, თუ $x_i > a$, მაშინ $x_i^2 > a^2$ და, თუ უკანასკნელ უტოლობაში x_i^2-ს შევცვლით a^2-ით, ამით (1.3) უტოლობა უფრო გაძლიერდება, ე. ი.

$$E(X^2) \geqslant \sum_{x_i > a} a^2 p_i,$$

ხოლო, რადგანაც აჯამვა a-ზე დამოუკიდებელია, დავწერთ

$$E(X^2) \geqslant a^2 \sum_{x_i > a} p_i. \tag{1.4}$$

ჯამის აქსიომის თანახმად

$$\sum_{x_i > a} p_i = P(X > a).$$

თუ ამას გავითვალისწინებთ, (1.4) უტოლობა მიიღებს შემდეგ სახეს

$$E(X^2) \geqslant a^2 P(X > a),$$

საიდანაც

$$P(X > a) \leqslant \frac{E(X^2)}{a^2}. \tag{1.5}$$

ეს კი დასამტკიცებელი ჩებიშევის უტოლობაა.

ახლა განვიხილოთ $X - E(X)$ გადახრა, რომელიც აგრეთვე შემთხვევითი სიდიდეა, და შევაფასოთ შემდეგი ალბათობა:

$$P\{|X - E(X)| > \varepsilon\}.$$

მიღებული ჩებიშევის უტოლობის თანახმად დავწერთ:

$$P\{|X - E(X)| > \varepsilon\} \leqslant \frac{E[X - E(X)]^2}{\varepsilon^2} = \frac{D(X)}{\varepsilon^2}.$$

მაგრამ, რადგანაც

$$P\{|X-E(X)|<\varepsilon\}=1-P\{|X-E(X)|>\varepsilon\},$$

თუ მიღებულ (1.5) უტოლობას გამოვიყენებთ, დავწერთ:

$$P\{|X-E(X)|<\varepsilon\}\geqslant 1-\frac{D(X)}{\varepsilon^2}. \tag{1.6}$$

ცხადია, აქ იგულისხმება, რომ X შემთხვევით სიდიდეს აქვს სასრული დისპერსია.

(1.6) უტოლობაც ჩებიშევის უტოლობის ერთ-ერთ სახეს წარმოადგენს.

მიღებულ უტოლობათა დახმარებით დავამტკიცოთ დიდ რიცხვთა კანონის შესახებ ზოგიერთი დებულება.

§ 2. ბერნულის თეორემა

ვთქვათ, ვატარებთ n-ჯერ ურთიერთდამოუკიდებელ ცდებს, ჩვენთვის სასურველი ხდომილობის მოხდენათა რიცხვი აღვნიშნოთ m-ით, ხოლო ცალკეული ცდის დროს ჩვენთვის სასურველი ხდომილობის ალბათობა იყოს p.

როგორც ვნახეთ,

$$E(m)=np,\quad E\left(\frac{m}{n}\right)=p,$$
$$D(m)=npq,\quad D\left(\frac{m}{n}\right)=\frac{pq}{n}. \tag{2.1}$$

ბერნულის თეორემა მდგომარეობს შემდეგში:

როგორიც არ უნდა იყოს $\varepsilon>0$ მცირე რიცხვი,

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n}-p\right|<\varepsilon\right\}=1. \tag{2.2}$$

დამტკიცება: თუ გამოვიყენებთ ჩებიშევის უტოლობის (1.6) სახეს, $\frac{m}{n}$ შემთხვევითი სიდიდისათვის დავწერთ:

$$P\left\{\left|\frac{m}{n}-E\left(\frac{m}{n}\right)\right|<\varepsilon\right\}=P\left\{\left|\frac{m}{n}-p\right|<\varepsilon\right\}\geqslant 1-$$

$$-\frac{D\left(\frac{m}{n}\right)}{\varepsilon^2}=1-\frac{pq}{n\varepsilon^2}.$$

ახლა, თუ გადავალთ ზღვარზე, როდესაც $n \to \infty$, გვექნება:

$$\lim_{n \to \infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \geqslant 1.$$

მაგრამ, რადგან ალბათობა არ შეიძლება აღემატებოდეს 1-ს, ამიტომ დავწერთ:

$$\lim_{n \to \infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

ამით დამტკიცებულია ბერნულის თეორემა, რომელიც გვეუბნება, რომ ცდათა რიცხვის გადიდებით ფარდობითი სიხშირე უახლოვდება ცალკეული ცდის დროს ხდომილობის ალბათობას.

§ 3. ჩებიშევის თეორემა

დიდ რიცხვთა კანონის ზოგადი დებულება ჩებიშევს ეკუთვნის და მისი დამტკიცება ჩებიშევის უტოლობას ეყრდნობა.

ჩებიშევის თეორემა მდგომარეობს შემდეგში:

ვთქვათ,

$$z_1, z_2, \dots, z_n \tag{3.1}$$

წარმოადგენს წყვილ-წყვილად დამოუკიდებელ შემთხვევით სიდიდეთა მიმდევრობას, რომელთაც შემოსაზღვრული დისპერსიები აქვთ. შევადგინოთ (3.1) მიმდევრობის საშუალო არითმეტიკული

$$\frac{1}{n} \sum_{i=1}^{n} z_i, \tag{3.2}$$

რომელიც თვითონ შემთხვევით სიდიდეს წარმოადგენს. დიდ რიცხვთა კანონის მიზანია სწორედ ამ საშუალო არითმეტიკულის ხასიათის შესწავლა.

შევადგინოთ აგრეთვე (3.1) მიმდევრობის მათემატიკურ ლოდინთა საშუალო არითმეტიკული

$$\frac{1}{n} \sum_{i=1}^{n} E(z_i). \tag{3.3}$$

იგი მუდმივ სიდიდეს წარმოადგენს.

დიდ რიცხვთა კანონის ჩებიშევის თეორემაში მტკიცდება, რომ n-ის საკმაოდ დიდი მნიშვნელობი-

სათვის $\frac{1}{n}\sum_{i=3}^{n} z_i$ თითქმის აუცილებლად რაგინდ ახლოს

არის $\frac{1}{n}\sum_{i=1}^{n} E(z_i)$-თან.

რა თქმა უნდა, ეს მტკიცება შემთხვევით სიდიდეთა მცირე რაოდენობისათვის არ იქნება სამართლიანი.

ამგვარად, ჩებიშევის თეორემა გვეუბნება, რომ, როცა $n \to \infty$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| > \varepsilon\right\} \to 0. \tag{3.4}$$

ე. ი. ალბათობა იმისა, რომ (3.1) მიმდევრობის შემთხვევით სიდიდეთა საშუალო არითმეტიკულსა და იმავე სიდიდეების მათემატიკურ ლოდინთა საშუალო არითმეტიკულს შორის სხვაობის აბსოლუტური მნიშვნელობა მეტი იქნება რაგინდ მცირე დადებით რიცხვზე, რაგინდ ახლო იქნება ნულთან, როცა n საკმაოდ დიდია.

დამტკიცება. (3.4) გამოსახულება გადავწეროთ შემდეგნაირად:

$$P\left\{\left|\sum_{i=1}^{n}[z_i - E(z_i)]\right| > n\varepsilon\right\} \underset{n\to\infty}{\to} 0.$$

თუ მარცხენა მხარისათვის ჩებიშევის უტოლობას გამოვიყენებთ, დავწერთ:

$$P\left\{\left|\sum_{i=1}^{n}[z_i - E(z_i)]\right| > n\varepsilon\right\} \leqslant \frac{E\left\{\sum_{i=1}^{n}[z_i - E(z_i)]\right\}^2}{n^2\varepsilon^2}.$$

ამ გამოსახულების მარჯვენა მხარის მრიცხველში მრავალწევრის კვადრატის საშუალო მნიშვნელობა გვაქვს და მას გადმოვწერთ შემდეგნაირად:

$$E\left\{\sum_{i=1}^{n}[z_i - E(z_i)]^2 + \sum_{i\neq j=1}^{n}[z_i - E(z_i)]\,[z_j - E(z_j)]\right\}.$$

ეს კი, რადგანაც შემთხვევითი სიდიდეები ურთიერთდამოუკიდებელია, წარმოვადგინოთ ასე:

$$\sum_{i=1}^{n} E\,[z_i - E(z_i)]^2 + \sum_{i\neq j=1}^{n} E[z_i - E(z_i)]\,E[z_j - E(z_j)].$$

რადგანაც გადახრის საშუალო მნიშვნელობა ყოველთვის ნულია, ამიტომ გვექნება

$$P\left\{\left|\sum_{i=1}^{n}[z_i-E(z_i)]>n\varepsilon\right.\right\}\leqslant\frac{\sum\limits_{i=1}^{n}E[z_i-E(z_i)]^2}{n^2\varepsilon^2}=\frac{\sum\limits_{i=1}^{n}D(z_i)}{n^2\varepsilon^2}.$$

თუ აქ შემთხვევით სიდიდეთა დისპერსიების შემოსაზღვრულობის პირობას გამოვიყენებთ, რომელიც ნიშნავს შემდეგს: არსებობს ისეთი მუდმივი C, რომ ყველა $D(z_i)<C$, მაშინ მივიღებთ:

$$P\left\{\left|\sum_{i=1}^{n}[z_i-E(z_i)]\right|>n\varepsilon\right\}\leqslant\frac{nC}{n^2\varepsilon^2}=\frac{C}{n\varepsilon^2}.$$

ცხადია, ε მცირე დადებითი სიდიდეა, მაგრამ ფიქსირებულია, ამიტომ, როდესაც $n\to\infty$, მაშინ უკანასკნელი გამოსახულების მარჯვენა მხარე ნულისაკენ მიისწრაფვის.

ამგვარად,

$$\lim_{n\to\infty}P\left\{\left|\sum_{i=1}^{n}[z_i-E(z_i)]\right|>n\varepsilon\right\}\leqslant 0.$$

რადგანაც ალბათობა არაუარყოფითი სიდიდეა, ამიტომ გვექნება:

$$\lim_{n\to\infty}P\left\{\left|\sum_{i=1}^{n}[z_i-E(z_i)]\right|>n\varepsilon\right\}=0$$

ან, რაც იგივეა,

$$\lim_{n\to\infty}P\left\{\left|\sum_{i=1}^{n}[z_i-E(z_i)]\right|<n\varepsilon\right\}=1.$$

ამით დიდ რიცხვთა კანონის შესახებ ჩებიშევის თეორემა დამტკიცებულია.

§ 4. მარკოვის თეორემა

ვთქვათ, $x_1, x_2, \ldots, x_n$ ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეებია, რომელთაც აქვთ სასრული მათემატიკური ლოდინი და დისპერსია. გარდა ამისა, დავუშვათ, რომ

$$\frac{1}{n^2}\sum_{i=1}^{n}D(x_i)\underset{n\to\infty}{\to}\infty. \tag{4.1}$$

მაშინ, როგორიც არ უნდა იყოს მუდმივი $\varepsilon>0$, ადგილი ექნება ტოლობას:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-\frac{1}{n}\sum_{i=1}^{n} E(x_i)\right|<\varepsilon\right\}=1. \tag{4.2}$$

დ ა მ ტ კ ი ც ე ბ ა. თუ აქ გამოვიყენებთ ჩებიშევის უტოლობის (1.6) ფორმას, დავწერთ:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-\frac{1}{n}\sum_{i=1}^{n} E(x_i)\right|<\varepsilon\right\}\geqslant 1-\frac{D\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\}}{\varepsilon^2}$$

მაგრამ, რადგანაც შემთხვევითი სიდიდეები ურთიერთდამოუკიდებელია, ამიტომ მივიღებთ:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-\frac{1}{n}\sum_{i=1}^{n} E(x_i)\right|<\varepsilon\right\}\geqslant 1-\frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} D(x_i).$$

პირობის თანახმად

$$\frac{1}{n^2}\sum_{i=1}^{n} D(x_i)\underset{n\to\infty}{\to} 0,$$

ამიტომ დავწერთ:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-\frac{1}{n}\sum_{i=1}^{n} E(x_i)\right|<\varepsilon\right\}\geqslant 1,$$

მაგრამ, რადგანაც ალბათობა არ შეიძლება ერთზე მეტი იყოს, საბოლოოდ მივიღებთ:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-\frac{1}{n}\sum_{i=1}^{n} E(x_i)\right|<\varepsilon\right\}=1, \tag{4.3}$$

რითაც მარკოვის თეორემა დამტკიცებულია.

§ 5. ჩებიშევის თეორემის ლიაპუნოვისეული ფორმულირება

ლაპლასის თეორემის შემთხვევაში გვქონდა:

$$P\left\{-t\sqrt{\frac{pq}{n}}<\frac{m}{n}-p<t\sqrt{\frac{pq}{n}}\right\}=2\cdot\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx. \tag{5.1}$$

ეს გადავწეროთ შემდეგნაირად:

$$P\left(\left|\frac{m}{n}-p\right|<t\sqrt{\frac{pq}{n}}\right)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx=2\Phi(t) \qquad (5.2)$$

ახლა t ისეთნაირად შევარჩიოთ, რომ $\Phi(t)$ იყოს დაახლოებით ერთის ტოლი, და n იმდენად დიდი ავიღოთ, რომ ადგილი ჰქონდეს უტოლობას:

$$t\sqrt{\frac{pq}{n}}<\varepsilon,$$

სადაც ε წინასწარ აღებული რაგინდ მცირე დადებითი რიცხვია. მიღებული შეზღუდვების საფუძველზე დავწერთ:

$$P\left(\left|\frac{m}{n}-p\right|\leqslant\varepsilon\right)\simeq 1.$$

ეს კი ბერნულის თეორემის შინაარსს გამოხატავს. ამგვარად, ბერნულის თეორემა ლაპლასის თეორემის კერძო შემთხვევა ყოფილა.

ჩებიშევის თეორემაში ნაჩვენებია, რომ საძიებელი ალბათობა გარკვეულ სიდიდეზე მეტია, მაგრამ აქ ყველაფერი როდია ნათელი. საჭიროა ვიცოდეთ, თუ რას უდრის საძიებელი ალბათობა. ეს ამოცანა ლიაპუნოვმა ამოხსნა. მან გვიჩვენა, რომ ჩებიშევის უტოლობაში მიღებული უტოლობის ალბათობა რიცხობრივად ლაპლასის ინტეგრალით გამოისახება, ე. ი. დაამტკიცა, რომ

$$P\left\{\left|\tilde{x}-\bar{x}\right|\leqslant t\frac{\sigma_x}{\sqrt{n}}\right\}=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx, \qquad (5.3)$$

თუ n უსაზღვროდ იზრდება. აქ $\tilde{x}$ შერჩევითი ერთობლიობის, ხოლო $\bar{x}$ გენერალური ერთობლიობის საშუალოა.

უკანასკნელი თანაფარდობა წარმოადგენს სწორედ ჩებიშევის თეორემის ლიაპუნოვისეულ ფორმულირებას.

თვითონ ლიაპუნოვის კლასიკური ზღვარითი თეორემის შინაარსი კი მდგომარეობს შემდეგში:

ვთქვათ, მოცემულია ურთიერთდამოუკიდებელ შემთხვევით სიდიდეთა მიმდევრობა:

$$X_1,\ X_2, \ldots,\ X_n.$$

n-ის საკმაოდ დიდი მნიშვნელობისათვის აღებული მიმდევრობის ჯამის განაწილების კანონი რაგინდ ახლოს იქნება ნორმალური განაწილების კანონთან, თუკი ცალკეული შესაკრები მიმდევრობის ყველა წევრის ჯამთან შედარებით უმნიშვნელოა.

ლიაპუნოვის თეორემას რომ ჰქონდეს ადგილი, ამისათვის საკმარისია $X_1, X_2, \ldots, X_n$ სიდიდეთა გადახრა მათივე მათემატიკური ლოდინებიდან—$a_1, a_2, \ldots, a_n$ ზემოდან შემოსაზღვრული იყოს; ე. ი.

$$Max\,|X_i - a_i| < M,$$

ხოლო ყველას დისპერსია კი ქვემოდან იყოს შემოსაზღვრული; ე. ი. უნდა არსებობდეს ისეთი დადებითი C რიცხვი, რომ მიმდევრობის ყველა შემთხვევითი სიდიდისათვის გვქონდეს

$$E(X_i - a_i)^2 > C.$$

ს ა ვ ა რ ჯ ი შ ო

1. ისარგებლეთ ჩებიშევის უტოლობით და შეაფასეთ ალბათობა იმისა, რომ $|X - E(X)| < 0{,}1$, თუკი $D(X) = 0{,}001$.

პას. $p \geqslant 0{,}9.$

2. მოცემულია $P(|X - E(X)| < \varepsilon) \geqslant 0{,}9;$

$$D(X) = 0{,}004$$

მონახეთ ε.

პას. $\varepsilon = 0{,}2.$

3. ისარგებლეთ ჩებიშევის უტოლობით და შეაფასეთ ალბათობა იმისა, რომ X შემთხვევითი სიდიდე თავის მათემატიკური ლოდინისაგან გადაიხრება არა ნაკლებ გასამკეცებულ საშუალო კვადრატული გადახრით.

პას. $P(|X - E(X)| < 3\sigma) \geqslant 1 - \dfrac{\sigma^2}{9\sigma^2} = \dfrac{8}{9}.$

4. შეაფასეთ ალბათობა იმისა, რომ X შემთხვევითი სიდიდე თავისი მათემატიკური ლოდინისაგან გადაიხრება არა ნაკლებ გაორკეცებული საშუალო კვადრატული გადახრით.

პას. $P(|X - E(X)| \geqslant 2\sigma) \leqslant \dfrac{\sigma^2}{4\sigma^2} = \dfrac{1}{4}.$

5. მოწყობილობას აქვს 10 ერთმანეთისაგან დამოუკიდებელი ელემენტი, თითოეულის გაფუჭების ალბათობა რაიმე T დროის განმავლობაში არის 0,05. შეაფასეთ ალბათობა იმისა, რომ T დროში გაფუჭებათა რიცხვისა და მისი საშუალო მნიშვნელობის სხვაობის აბსოლუტური მნიშვნელობა ნაკლები იქნება 2-ზე ან არანაკლები იქნება 2-ზე.

პას. $P(|X - E(X)| < 2) \geqslant 1 - \dfrac{0{,}475}{4} = 0{,}12$

და $\qquad P(|X - E(X)| \geqslant 2) \leqslant 1 - 0{,}12 = 0{,}88.$

6. ცალკეული ცდის დროს A ხდომილობის ალბათობაა 0,5. შეაფა-

სეთ ალბათობა იმისა, რომ A-ს მოხდენათა რიცხვი X მოთავსებულ იქნება (40; 60) შუალედში, თუ ცდათა რიცხვია 100.

პას. $P(|X-E(X)|<\varepsilon)\geqslant 0{,}75$

აქ $$E(X)=50;\ D(X)=25,\ \varepsilon=10.$$

7. დისკრეტული ტიპის X შემთხვევითი სიდიდის განაწილების კანონია.

$$X\begin{cases}0{,}3; & 0{,}6\\ 0{,}2; & 0{,}8\end{cases}$$

შეაფასეთ ალბათობა იმისა, რომ $|X-E(X)|<0{,}2$.

პას. $P\{|X-E(X)|<0{,}2\}\geqslant 0{,}64$.

8. $X_1, X_2, \ldots, X_n, \ldots$ შემთხვევით სიდიდეთა მიმდევრობის განაწილების კანონია.

$$X_n\begin{cases}-n\alpha, & 0, & n\alpha\\ \dfrac{1}{2n^2}; & 1-\dfrac{1}{n^2}; & \dfrac{1}{2n^2}\end{cases}.$$

გამოიყენება თუ არა ამ მიმდევრობისათვის ჩებიშევის თეორემა?

პას. გამოიყენება.

9. $X_1, X_2, \ldots, X_n, \ldots$ შემთხვევით სიდიდეთა მიმდევრობის განაწილების კანონია:

$$X_n\begin{cases}-n\alpha, & 0, & n\alpha\\ \dfrac{1}{2^n}, & 1-\dfrac{1}{2^{n-1}}, & \dfrac{1}{2^n}\end{cases}.$$

ამ მიმდევრობისათვის გამოიყენება თუ არა ჩებიშევის თეორემა?

პას. გამოიყენება.

10. $X_1, X_2, \ldots X_n, \ldots$ შემთხვევით სიდიდეთა მიმდევრობის განაწილების კანონია:

$$X_n\begin{cases}-\sqrt{3}, & 0, & \sqrt{3}.\\ \dfrac{1}{3}; & \dfrac{1}{3}; & \dfrac{1}{3}.\end{cases}$$

გამოიყენება თუ არა ამ მიმდევრობისათვის ჩებიშევის თეორემა?

პას. გამოიყენება, რადგანაც $E(X_n)=0$ და $D(X_n)=2$.

11. შემთხვევით სიდიდეთა $X_1, X_2, \ldots X_n, \ldots$ მიმდევრობას აქვ ერთი და იგივე განაწილების ფუნქცია

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\operatorname{arc\,tg}\frac{x}{2}.$$

გამოიყენება თუ არა ასეთი მიმდევრობისათვის ჩებიშევის თეორემა?

პას. არა, რადგანაც მათემატიკური ლოდინი არ არსებობ

$$[E(X_n)=\infty].$$

ნაწილი მეორე

მათემატიკური სტატისტიკა

თ ა ვ ი I

უმცირეს კვადრატთა ხერხი

§ 1. უმცირეს კვადრატთა ხერხის არსი

ვთქვათ, რომელიმე y სიდიდის გაზომვის შედეგად მივიღეთ მნიშვნელობანი:

$$y_1, \ y_2, \ \ldots, \ y_n,$$

რომლებიც სხვა რომელიმე x სიდიდის შემდეგ

$$x_1, \ x_2, \ \ldots, \ x_n$$

მნიშვნელობებს შეესაბამებიან.

ამოცანა იმაში მდგომარეობს, რომ ვიპოვოთ $y=f(x)$ ისეთი ფუნქცია, რომელიც $x=x_1$; $x=x_2$; $\ldots$; $x=x_n$ მნიშვნელობათათვის შესაბამისად ღებულობს ემპირიულ $y_1, y_2, \ldots, y_n$ მნიშვნელობებისაგან რაც შეიძლება მცირედ განსხვავებულ მნიშვნელობებს.

გეომეტრიულად ეს იმას ნიშნავს, რომ გავატაროთ ისეთი $y=f(x)$ ბრტყელი წირი, რომელიც რაც შეიძლება ახლოს გაივლის

$$(x_1, \ y_1), \ (x_2, \ y_2), \ \ldots, \ (x_n, \ y_n)$$

წერტილებთან.

რა თქმა უნდა, რადგანაც ყოველი გაზომვა გარკვეულ ცდომილებას შეიცავს, ამიტომ ვერ მოვითხოვთ, რომ $y=f(x)$ მრუდი ყველა ზემოხსენებულ წერტილზე გადიოდეს.

$y=f(x)$ ფუნქციის სახის განსაზღვრა ჩვენზეა დამოკიდებული, თუკი რაიმე მოსაზრებით მისი სახე წინასწარ არ არის დადგენილი.

$f(x)$ ფუნქციას საერთოდ აძლევენ რაც შეიძლება მარტივ სახეს, მას შეიძლება მრავალწევრის ან წილად წრფივი ფუნქციის ან სხვა რომელიმე მარტივი ფუნქციის სახე ჰქონდეს.

უმცირეს კვადრატთა ხერხის არსი იმაში მდგომარეობს, რომ $y=f(x)$ ფუნქციის სათანადო შერჩევით, ჯამი

$$S=\sum_{i=1}^{n}[y_i-f(x_i)]^2$$

უნდა იყოს რაც შეიძლება მცირე, ე. ი. გამოთვლილ $f(x_1)$, $f(x_2)\ldots$, $f(x_n)$ სიდიდეებიდან ემპირიულ $y_1, y_2, \ldots, y_n$ მონაცემების გადახრათა კვადრატების ჯამი უნდა გავხადოთ მინიმალური.

ფაქტიურად $y_i=f(x_i)$ არის i-ურ წერტილის შესაბამის $f(x)$ ფუნქციის ორდინატსა და (x_i, y_i) წერტილის ორდინატებს შორის სხვაობა.

$y=f(x)$ ფუნქციის პარამეტრების განსაზღვრა, თუ თვით ფუნქციის სახე ცნობილია, არავითარ სიძნელეს არ წარმოადგენს.

ვთქვათ, $f(x)$ წრფივად არის სამ პარამეტრზე დამოკიდებული:

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+a_3\varphi_3(x). \tag{1.1}$$

იგულისხმება, რომ $\varphi_1(x)$, $\varphi_2(x)$ და $\varphi_3(x)$ ფუნქცია ცნობილია. უნდა ვიპოვოთ უმცირესი მნიშვნელობა ჯამისა:

$$S=\sum_{i=1}^{n}[y_i-a_1\varphi_1(x_i)-a_2\varphi_2(x_i)-a_3\varphi_3(x_i)]^2. \tag{1.2}$$

ამისათვის S სიდიდის კერძო წარმოებულები პარამეტრების მიმართ გავუტოლოთ ნულს, გვექნება:

$$\frac{\partial S}{\partial a_1}=-2\sum_{i=1}^{n}[y_i-a_1\varphi_1(x_i)-a_2\varphi_2(x_i)-a_3\varphi_3(x_i)]\varphi_1(x_i)=0,$$

$$\frac{\partial S}{\partial a_2}=-2\sum_{i=1}^{n}[y_i-a_1\varphi_1(x_i)-a_2\varphi_2(x_i)-a_3\varphi_3(x_i)]\varphi_2(x_i)=0,$$

$$\frac{\partial S}{\partial a_3}=-2\sum_{i=1}^{n}[y_i-a_1\varphi_1(x_i)-a_2\varphi_2(x_i)-a_3\varphi_3(x_i)]\varphi_3(x_i)=0.$$

მივიღეთ წრფივ განტოლებათა სისტემა, რომლის ამოხსნა a_1, a_2 და a_3 პარამეტრის მიმართ თავისუფლად შეიძლება. ამ სისტემის ამოხსნის შედეგად მიღებული a_1, a_2, a_3 პარამეტრების მნიშვნელობებს თუ შევიტანთ (1.1) ტოლობაში, ე. ი. $f(x)$-ის გამოსახულებაში

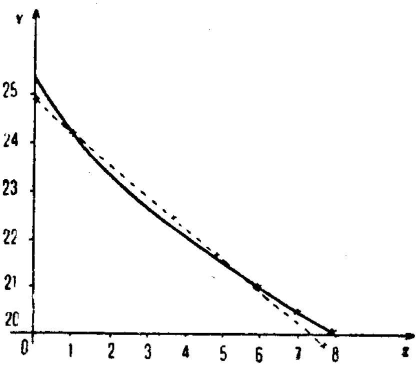
და გამოვთვლით (1.2) ჯამს, ვნახავთ, რომ S ჯამს მართლაც უმცირესი მნიშვნელობა ექნება.

მაგალითი. ვთქვათ, x-ის მოცემულ მნიშვნელობათა შესაბამისი y-ის მნიშვნელობანი წარმოდგენილია შემდეგ ცხრილში:

ცხრილი 5

x	0	1	2	3	4	5	6	7	8
y	25,0	24,2	23,5	22,9	22,4	21,4	20,8	20,3	20,0

x და y შორის დამოკიდებულება გრაფიკულად შემდეგნაირად გამოისახება:



ნახ. 16.

ამ გრაფიკის მიხედვით ვიტყვით, რომ მიღებული მრუდი შეგვიძლია მესამე რიგის პარაბოლით შევცვალოთ:

$$y=f(x)=a+bx+cx^2+dx^3$$

ანდა, უფრო უხეშად, მრუდი შეგვიძლია შევცვალოთ

$$y=f(x)=a+bx \tag{1.3}$$

წრფით·

გამოთვლების გამარტივების მიზნით განვიხილოთ ეს შემთხვევა. პირველ რიგში უნდა ვიანგარიშოთ (1.3) ტოლობაში შემავალი a და b პარამეტრი· ამისათვის გამოვიყენოთ უმცირეს კვადრატთა ხერხი.

განვიხილოთ შემდეგი ჯამი:

$$S=\sum_{x=0}^{8}(y_x-a-bx)^2. \tag{1.4}$$

რომ ვიპოვოთ S-ის მინიმალური მნიშვნელობა, ამისათვის საჭიროა მისი კერძო წარმოებულები a და b-ს მიმართ ნულს გავუტოლოთ, მაშინ გვექნება:

$$\frac{\partial S}{\partial a}=-2\Sigma(y_x-a-bx)=0,$$

$$\frac{\partial S}{\partial b}=-2\Sigma(y_x-a-bx)x=0.$$

ანდა, რაც იგივეა:

$$9a+b\Sigma x=\Sigma y_x, \tag{1.5}$$

$$a\Sigma x+b\Sigma x^2=\Sigma xy_x.$$

გვაქვს ორუცნობიან განტოლებათა სისტემა; მისი ამოხსნის შედეგად მიღებული a და b-ს მნიშვნელობანი ჩასმული (1.4) ტოლობაში S-ს მიანიჭებს მინიმალურ მნიშვნელობას.

a და b-ს გამოსათვლელად შევადგინოთ შემდეგი ცხრილი:

ცხრილი 6

x	y_x	x^2	xy_x
0	25,0	0	0,0
1	24,2	1	24,2
2	23,5	4	47,0
3	22,9	9	68,7
4	22,2	16	88,8
5	21,4	25	107,0
6	20,8	36	124,8
7	20,3	49	142,1
8	20,0	64	160,0
36	200,3	204	762,6

გამოთვლილი ჯამები შევიტანოთ (1.5) ტოლობაში, მივიღებთ:

$$9a+36b=200,3,$$

$$36a+204b=762,6,$$

საიდანაც

$$a=24,82; \quad b=-0,64.$$

ახლა a და b-ს მნიშვნელობანი შევიტანოთ (1.3) ტოლობაში, გვექნება

$$f(x)=24,82-0,64x. \tag{1.6}$$

თუ ამ უკანასკნელში x-ს მივცემთ სათანადო მნიშვნელობებს, მივიღებთ y-ის შესაბამის მნიშვნელობებს. რომლებსაც ემპირიულად მიღებულ y_x მნიშვნელობებისაგან განსხვავებულად Y_x-ით აღვნიშნავთ. თუ ემპირიულად მიღებულ y_x-ების და (1.6) ტოლობით გამოთვლილ Y_x-ების სხვაობათა კვადრატების ჯამს ვიანგარიშებთ, ის იქნება სწორედ S-ის მნიშვნელობა. S-ის გამოსაანგარიშებლად შევადგინოთ შემდეგი ცხრილი:

ცხრილი 7

x	y_x	$-0{,}64x$	Y_x	$y_x - Y_x$	$(y_x - Y_x)^2$
0	25,00	0	24,82	0,18	0,0324
1	24,20	—0,64	24,18	0,02	0,0004
2	23,50	—1,28	23,54	0,04	0,0016
3	22,90	—1,92	22,90	0	0
4	22,20	—2,56	22,26	0,06	0,0036
5	21,40	—3,20	21,62	0,22	0,0484
6	20,80	—3,84	20,98	0,18	0,0324
7	20,30	—4,48	20,34	0,04	0,0016
8	20,00	—5,12	19,70	0,30	0,0900
					0,2104

ამგვარად, მივიღეთ, რომ $S=0{,}2104$.

ზემოთ მოცემულ ნახაზზე

$$y=24{,}82-0{,}64\,x$$

წრფე პუნქტირითაა ნაჩვენები.

უფრო ზუსტი ფორმულით რომ გვესარგებლა, მაშინ პარამეტრთა რიცხვი გაიზრდებოდა, გამოთვლა უფრო გაგვირთულდებოდა, მაგრამ საბოლოოდ მიღებული მრუდი უფრო მიუახლოვდებოდა ემპირიულ მრუდს.

როგორც ვხედავთ, უმცირეს კვადრატთა ხერხით სარგებლობისას გამოთვლები მით უფრო რთულდება, რაც უფრო მეტია როგორც დაკვირვებათა რიცხვი, ისე აღებულ განტოლებაში შემავალი პარამეტრების რაოდენობა.

§ 2. საშუალო არითმეტიკულის როლი

მაგალითად, ვთქვათ რაიმე სიდიდის გაზომვის შედეგად მივიღეთ შემდეგი მნიშვნელობები მმ-ით:

125, 136, 129, 132, 128.

მოითხოვება ისეთი ზომის მონახვა, რომელიც რაც შეიძლება ახლოს იქნება გაზომვის შედეგად მიღებულ ხუთივე სიდიდესთან, ან სხვანაირად რომ ვთქვათ, აღებული ზომა რაც შეიძლება მცირედ უნდა იყოს გადახრილი მათგან.

როგორც ვნახეთ, უმცირეს კვადრატთა ხერხის შემთხვევაში სიდიდეთა „სიახლოვე" და „დაშორება" ერთმანეთისაგან მათი სხვაობის კვადრატით გაიზომება. მაგრამ, რადგანაც მხედველობაში გვაქვს ისეთი რიცხვის მონახვა, რომელიც რაც შეიძლება ახლოს იქნება რიცხვთა მწკრივთან, ეს მოთხოვნა იმას ნიშნავს, რომ საძებნ რიცხვსა და მწკრივის ყველა წევრს შორის სხვაობის კვადრატების ჯამი იყოს რაც შეიძლება მცირე.

ამგვარად, ამოცანა ჩამოყალიბდება შემდეგნაირად: მოვნახოთ ისეთი რიცხვი, რომ აღებულ რიცხვებთან მისი სხვაობათა კვადრატების ჯამი იყოს მინიმალური.

დავუშვათ, რომ ასეთი რიცხვია a, მაშინ ამ რიცხვიდან გადახრათა კვადრატების ჯამი იქნება:

$$\varphi(a)=(a-125)^2+(a-136)^2+(a-129)^2+(a-132)^2+(a-128)^2.$$

გავადიფერენციალოთ მიღებული ჯამი a-თი და მიღებული გამოსახულება გავუტოლოთ ნულს, გვექნება:

$$2(a-125)+2(a-136)+2(a-129)+2(a-132)+2(a-128)=0$$

ან

$$a+a+a+a+a=125+136+129+132+128,$$

საიდანაც

$$a=\frac{125+136+129+132+128}{5}=130.$$

ამგვარად, მივიღეთ, რომ a-ს საძებნი მნიშვნელობა ყოფილა გაზომვის შედეგად მიღებულ მნიშვნელობათა საშუალო არითმეტიკული.

იმავე შედეგის ზოგადად ჩვენება შეიძლება. ვთქვათ, მოცემული გვაქვს $y_1, y_2, \dots, y_n$ გაზომვის შედეგად მიღებული მნიშვნელობანი. a რიცხვიდან მათი გადახრათა კვადრატების ჯამი იქნება:

$$s=\sum^{n}(a-y_i)^2.$$

ავიღოთ s-ის წარმოებული a-თი, მივიღებთ:

$$\frac{\partial s}{\partial a}=2\sum_{i-1}^{n}(a-y_i).$$

მიღებული კერძო წარმოებული გავუტოლოთ ნულს, გვექნება:

$$\sum_{i=1}^{n} (a-y_i)=0,$$

ან

$$\sum_{i=1}^{n} a=\sum_{i=1}^{n} y_i,$$

რაც იგივეა,

$$na=\sum_{i=1}^{n} y_i.$$

აქედან

$$a=\frac{1}{n}\sum_{i=1}^{n} y_i.$$

ამგვარად, აქაც მივიღეთ, რომ საძებნი რიცხვი ყოფილა გაზომვის შედეგად მიღებულ მნიშვნელობათა საშუალო არითმეტიკული.

§ 3. უმცირეს კვადრატთა ხერხის გამოყენების ერთი კონკრეტული მაგალითი

განვიხილოთ შემდეგი მაგალითი. მნიშვნელოვან ინტერესს წარმოადგენს მოსახლეობის მოთხოვნის დიფერენცირება ასაკისა და სქესთა ჯგუფების მიხედვით. თუ ვიცით მოთხოვნათა შემადგენლობის განსხვავება ასაკისა და სქესის მიხედვით, შესაძლებელია წინასწარ განვსჭვრიტოთ მოთხოვნის ცვლილება მოსახლეობის შემადგენლობის მოსალოდნელი ცვლილების საფუძველზე და ა. შ. ამ შემთხვევაში პირდაპირი დაკვირვების ჩატარება შეუძლებელია, რადგანაც არ შეიძლება მოვითხოვოთ, რომ ოჯახში ყოველი პროდუქტის მიმართ ფიქსირებული იქნეს მოთხოვნა ოჯახის ყოველ წევრზე ცალ-ცალკე. მიუხედავად ამისა, უმცირეს კვადრატთა ხერხის გამოყენებით შესაძლებელია დასმული ამოცანის გადაწყვეტა.

მაგალითისათვის განვიხილოთ რძე, რომლის მოთხოვნა ასაკისდა მიხედვით საგრძნობლად იცვლება. თუ შეუძლებელია უშუალოდ დადგინდეს ოჯახში რამდენ რძეს ხარჯავენ ბავშვები და რამდენს—მოზრდილები, მაშინ ამის დადგენა საჭირო იქნება სხვა გზით, კერძოდ, ოჯახზე მთლიანი მონაცემების მიხედვით.

თუ ამოვწერთ იმ რძის რაოდენობას თითოეულ ოჯახზე, რომელსაც ისინი ხარჯავენ, მაშინ უბრალო დაკვირვებითაც შევატყობთ, რომ იმ ოჯახში უფრო მეტი რძე იხარჯება, სადაც უფრო მეტი ბავშვია. რა თქმა უნდა, ცალ-ცალკე უნდა განვიხილოთ მოსახლეობათა დაჯგუფება პროფესიის მიხედვით. წინააღმდეგ შემთხვევაში ზოგიერთი სხვა ფაქტორის მოქმედება საგრძნობი იქნება.

y-ით აღვნიშნოთ ოჯახში ერთ სულზე რძის მოთხოვნის რაოდენობა. შეიძლება ვიფიქროთ, რომ მოზრდილთათვის მოთხოვნის დიფერენცირებულ სიდიდეს მივიღებთ, თუ y-ის საშუალო სიდიდეს ვიანგარიშებთ ისეთი ოჯახებისათვის, სადაც ბავშვები არ არიან, მაგრამ ეს არ იქნება მართებული. მოზრდილთათვის რძის მოთხოვნის რაოდენობა უნდა გავიგოთ საშუალოდ იმ ოჯახების ჩათვლითაც, რომლებშიც ბავშვებიცაა. ასეთ ოჯახებში მოთხოვნა უფრო მეტია, რადგანაც, თუ ოჯახში რძეს ბავშვებისათვის ყიდულობენ, მაშინ იგულისხმება, რომ ის შედის აგრეთვე მოზრდილთა საკვებშიც.

ჩვენთვის ცნობილი y სიდიდე არის რძის რაოდენობა ლიტრობით გაყოფილი ოჯახის წევრების რიცხვზე, ე. ი. ოჯახში სულზე მოთხოვნის საშუალო. ადვილად დაითვლება ოჯახების მიხედვით მოზრდილთა და მოზარდთა რაოდენობა. მოზრდილთა რაოდენობა აღვნიშნოთ v-თი, ხოლო მოზარდებისა u-თი, მოზარდთა საშუალო მოთხოვნა იყოს y_u, ხოლო მოზრდილთა კი y_v. მაშინ სულზე მოთხოვნა საშუალოდ იქნება:

$$y=\frac{y_v v+y_u u}{u+v}.$$

ამ გამოსახულების მარჯვენა მხარის მრიცხველს დავუმატოთ და გამოვაკლოთ $y_v u$ სიდიდე, გვექნება:

$$y=\frac{y_v v+y_v u+y_u u-y_v u}{u+v}.$$

ან, რაც იგივეა

$$y=y_v+(y_u-y_v)\frac{u}{u+v}.$$

შემოვიღოთ აღნიშვნა

$$\frac{u}{u+v}=x.$$

მაშინ დავწერთ

$$y=y_v+(y_u-y_v)x, \tag{3.1}$$

სადაც x არის მოზარდთა რაოდენობის საერთო რაოდენობასთან ფარდობა.

როგორც ვხედავთ, მოცემულ y_u და y_v-თვის y სიდიდე წარმოადგენს წრფივ ფუნქციას ბავშვთა ფარდობითი x რიცხვის მიმართ.

$y_u - y_v$ სხვაობა წარმოადგენს ამ წრფივი დამოკიდებულების საკუთხო კოეფიციენტს.

იმ შემთხვევაში, როდესაც დიფერენცირებული საშუალოები y_u და y_v ყველა ოჯახისათვის ერთი და იგივეა (სინამდვილეში, ცხადია, ისინი იცვლება იმასთან დამოკიდებულებით, თუ რა ხნის არიან თვითონ ბავშვები და ა. შ.), მაშინ ამ წრფივი ფუნქციის გამომსახველი (3.1) განტოლება ძალაში იქნებოდა ყველასათვის.

მაშასადამე, სხვა თანაბარ პირობებში, თუ ოჯახებს ერთმანეთს შევადარებთ და თვალყურს ვადევნებთ — რა სწრაფად იზრდება საერთო y მაჩვენებელი (იმისდა მიხედვით, თუ როგორ იზრდება ბავშვთა ფარდობითი x რიცხვი), გავიგებთ $y_u - y_v$ სხვაობას. აქედან კი შეგვიძლია მივიღოთ მოთხოვნის დიფერენცირებული ზომები.

წარმოდგენა რომ მივიღოთ y-ის ცვლილებაზე x-ის ცვლილებასთან დაკავშირებით, სხვა თანაბარ პირობებში არ შეიძლება ცალკეული ოჯახების შედარება. ამისათვის საჭიროა შევქმნათ ოჯახთა ისეთი ჯგუფები, რომელთაც ბავშვთა ფარდობითი x რიცხვი სხვადასხვა აქვთ და ამ ჯგუფების მიხედვით გამოვთვალოთ საშუალო y სიდიდეები.

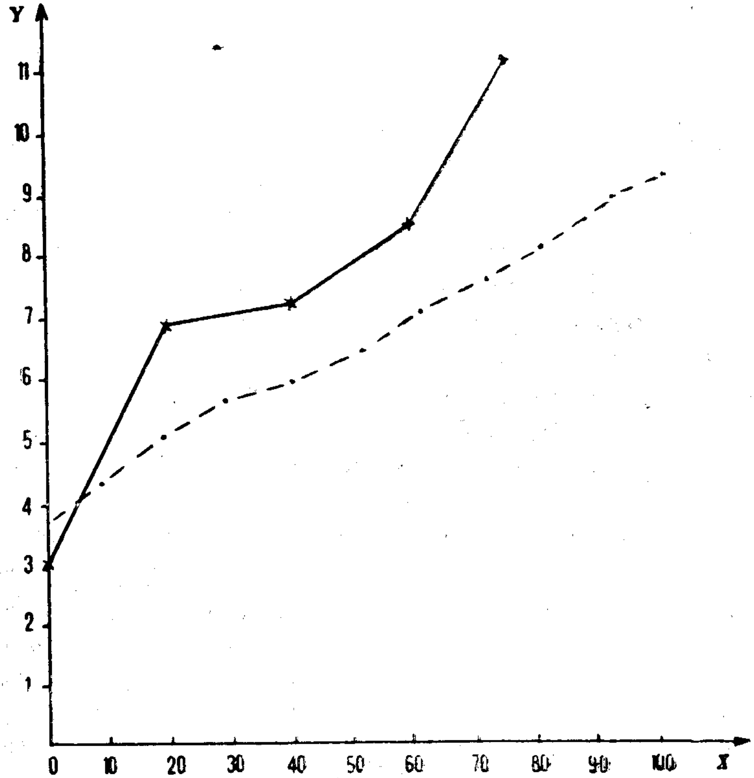
დავუშვათ, რომ მოვახდინეთ ოჯახთა ასეთნაირი დაჯგუფება: პირველ ჯგუფში მოთავსებულია უბავშვო ოჯახები, მეორე ჯგუფში — ისეთი ოჯახები, სადაც ბავშვები არიან, მაგრამ მათი რაოდენობა 23%-ს არ აღემატება (1 ბავშვი მოზრდილების 5 ან უფრო მეტი რაოდენობიდან, 2 ბავშვი 9 მოზრდილიდან და ა. შ.), მესამე ჯგუფში მოთავსებულია ისეთი ოჯახები, სადაც ბავშვთა რაოდენობა მთელი რაოდენობის 23-დან 46% მდე შეადგენს და ა. შ.

შედეგები ნაჩვენებია მე-8 ცხრილში.

ცხრილი 8

ჯგუფები ბავშვთა პროცენტების მიხედვით	ბავშვთა რაოდენობა (პროცენტობით) x	რძის რაოდენობა (ლიტრობით) y
0	0	3,0
0—23	19	6,8
23—46	40	7,1
46—69	60	8,8
62—92	74	11,2

როგორც მე-8 ცხრილიდან ჩანს, საეჭვოა, რომ x და y შორის დამოკიდებულება წრფივი იყოს.



ნახ. 17.

თუ დავაკვირდებით მე-17 ნახაზზე მოცემულ გრაფიკს, ცხადია, რომ წრფივი დამოკიდებულებიდან გადახრა არ გამოხატავს რძის მოთხოვნის კანონზომიერ ცვლილებას ბავშვთა რაოდენობასთან დაკავშირებით. მაგალითად, მესამე ჯგუფში $y=7,1$ დიდად არ განსხვავდება მეორე ჯგუფის $y=6,8$-საგან. რამდენადაც ჯგუფები შედგება კონკრეტული ოჯახებისაგან, ამიტომ შესაძლებელი იყო, რომ მესამე ჯგუფში მოხვედრილიყო ისეთი ოჯახები, რომლებსაც, რაიმე მიზეზის გამო, ჩვენთვის საინტერესო საკითხთან არა აქვთ საერთო; მათთვის რძის მოხმარების რაოდენობა შეიძლება მცირე იყოს და ეს უფრო მეტ გავლენას ახდენს, ვიდრე ბავშვთა რაოდენობა მეორე ჯგუფის ოჯახებში. შესაძლებელია მესამე ჯგუფში შემთხვევით ისეთი

ოჯახები მოხვდნენ, რომლებიც ერთ ქუჩაზე ცხოვრობენ, სადაც რძის მაღაზია დროებით შეკეთების პროცესშია ამ დროს და რძე არ იყიდება. რომ მივიღოთ რძის რაოდენობის ჯგუფიდან ჯგუფში კანონზომიერი ზრდა, საჭიროა ასეთი შემთხვევითი ფაქტორები უკუვაგდოთ და მაშინ მივიღებთ იმ წრფივ განტოლებას, რომელზედაც უკვე გვქონდა ლაპარაკი. ამისათვის ვეძიოთ ისეთი წრფივი ფუნქცია, რომელიც, რაც შეიძლება კარგად ასახავს რეალურ მონაცემებს. რამდენადაც ამის მოსანახავად უმცირეს კვადრატთა ხერხის გამოყენება გვინდა, ამიტომ მოვითხოვთ, რომ ამით გამოთვლილი სულზე მოთხოვნილებათა მნიშვნელობის თითოეული ჯგუფის ფაქტიური მნიშვნელობისაგან გადახრათა კვადრატების ჯამი იყოს მინიმალური.

დავუშვათ, რომ ამ წრფივი ფუნქციის განტოლებაა

$$\overline{y}_x = a_0 + a_1 x.$$

მაშინ ის ჯამი, რომლის მინიმუმს ვეძებთ, იქნება:

$$s = \Sigma(a_0 + a_1 x - y)^2,$$

სადაც y არის ჯგუფების მიხედვით ფაქტიური მნიშვნელობა, ხოლო აჯამვა ხდება ყველა ჯგუფის მიხედვით; x არის ბავშვთა პროცენტული შემადგენლობა ჯგუფში.

განხილული მაგალითისათვის s ჯამი იქნება:

$$s = (a_0 - 3{,}0)^2 + (a_0 + 19\, a_1 - 6{,}8)^2 + (a_0 + 40 a_1 + 7{,}1)^2 +$$
$$+ (a_0 + 60 a_1 - 8{,}8)^2 + (a_0 + 74 a_1 - 11{,}2)^2.$$

s-ის მნიშვნელობა დამოკიდებულია იმაზე, თუ როგორ შევარჩევთ a_0 და a_1-ს. როგორც უკვე ვნახეთ, იმისათვის, რომ მათ s-ს მიანიჭონ მინიმუმი, საჭიროა ავიღოთ კერძო წარმოებულები მათ მიმართ და ისინი გავუტოლოთ ნულს, ამოვხსნათ მიღებული სისტემა და ამონახსნები ჩავსვათ s-ის გამოსახულებაში. მიღებული მნიშვნელობა იქნება მინიმალური s-თვის.

მაშასადამე,

$$2\Sigma(a_0 + a_1 x - y) = 0,$$
$$2\Sigma(a_0 + a_1 x - y)x = 0,$$

ან

$$\Sigma(a_0 + a_1 x - y) = 0,$$
$$\Sigma(a_0 + a_1 x - y)x = 0.$$

უკანასკნელი გამოსახულებები გადავწეროთ შემდეგნაირად:

$$\Sigma a_0+\Sigma a_1x=\Sigma y,$$
$$\Sigma a_0x+\Sigma a_1x^2=\Sigma xy,$$

ან

$$na_0+a_1\Sigma x=\Sigma y,$$
$$a_0\Sigma x+a_1\Sigma x^2=\Sigma xy,$$

სადაც n შესაკრებთა რიცხვია და ჩვენი შემთხვევისათვის უდრის 5-ს.

უკანასკნელ განტოლებებს ეწოდება უმცირეს კვადრატთა ხერხის ნორმალური განტოლებანი.

ამ ნორმალურ განტოლებებში შემავალი ჯამების მნიშვნელობანი ჩვენი მაგალითისათვის მოცემულია მე-9 ცხრილში.

ცხრილი 9

x	y	x^2	xy
0	3,0	0	0
19	6,8	361	129,2
40	7,1	1600	284,0
60	8,8	3600	528,0
74	11,2	5476	828,8
$\Sigma x=193$	$\Sigma y=36,9$	$\Sigma x^2=11037$	$\Sigma xy=1770,0$

თუ მიღებულ მნიშვნელობებს ჩავსვამთ ნორმალურ განტოლებაში, გვექნება:

$$\left.\begin{aligned}5\,a_0+193a_1&=36,9.\\ 193a_0+11037\ a_1&=1770,0.\end{aligned}\right\}$$

ამ სისტემის ამოხსნა გვაძლევს

$$a_0=3,66;\ a_1=0,0964,$$

ამგვარად, წრფივი ფუნქციისათვის მივიღეთ გამოსახულება:

$$\overline{y}_x=3,66+0,0964\ x.$$

კერძოდ, თუ $x=0$, ე. ი ბავშვთა რაოდენობა ნულია, მაშინ ვღებულობთ მაჩვენებელს მოზრდილთათვის, რომელიც იქნება 3,66. თუ $x=100$, ე. ი. მოზრდილები არ არიან, მაშინ ვღებულობთ მაჩვენებელს ბავშვებისათვის, რომელიც ტოლია 13,3-ის.

უმცირეს კვადრატთა ხერხით წრფივი ფუნქციის მონახვას ხშირად გაწრფივებას უწოდებენ.

ფუნქციის განტოლებას, რომელიც მიიღება გაწრფივების შედეგად, უწოდებენ y-ის და x-ის მიმართ კორელაციური დამოკიდებულების განტოლებას

$$y_x = 3{,}66 + 0{,}0564x$$

წრფივ ფუნქციას შეესაბამება მე-17 ნახაზზე პუნქტირით გავლებული წრფე.

სავარჯიშო

1. ვთქვათ X და Y შემთხვევით სიდიდეთა შორის დამოკიდებულება მოცემულია ცხრილით

X	0,	2	4,	6,	8,	10,	12,	14,	16
Y	9,0;	8,9;	8,4;	8,3;	7,8;	7,5;	7,1;	6,9;	6,8

X და Y შორის დამოკიდებულება გამოსახეთ გრაფიკულად. მიღებული გრაფიკის მიხედვით დაადგინეთ $y=\varphi(x)$ მრუდის სახე. უმცირეს კვადრატთა ხერხის გამოყენებით დაადგინეთ ამ მრუდის განტოლებაში შემავალი პარამეტრების მნიშვნელობანი. შეადგინეთ S ჯამი და იპოვეთ მისი მინიმალური მნიშვნელობა.

თავი II

შერჩევითი მეთოდი

§ 1. შესავალი

მოცემული ცდების ან დაკვირვებების მიხედვით მასობრივ მოვლენებზე და პროცესებზე მეცნიერულად დასაბუთებული დასკვნების მიღების მეთოდების გამონახვა არის მათემატიკური სტატისტიკის ერთ-ერთი ძირითადი ამოცანა. მიღებული დასკვნები ეხებიან არა ცალკეული ცდის ან დაკვირვების შედეგს, არამედ წარმოადგენენ მოცემული პროცესის ალბათური მახასიათებლების შესახებ გარკვეულ მტკიცებას, სახელდობრ, ალბათობებზე, განაწილების კანონებზე, მათემატიკურ ლოდინებზე და სხვათა შესახებ გარკვეული დასკვნის გაკეთებას.

შერჩევითი მეთოდი ერთ-ერთი მნიშვნელოვანი მეთოდია მათემა-

ტიკურ სტატისტიკაში, რომლის საფუძველსაც დიდ რიცხვთა კანონი წარმოადგენს.

სრული წარმოდგენა რომ შევიქმნათ გენერალურ ერთობლიობაზე, ამისათვის საჭიროა მისი ყოველი ელემენტის სათითაოდ შემოწმება. იმ შემთხვევაში, როდესაც გენერალური ერთობლიობის მოცულობა საკმაოდ დიდია, პრაქტიკულად ძნელია მთლიანი შემოწმების ჩატარება. ასეთ შემთხვევაში აწარმოებენ გენერალური ერთობლიობის გარკვეული ნაწილის ე. წ. შერჩევითი ერთობლიობის შემოწმებას, რის შედეგად გარკვეულ წარმოდგენას იქმნიან გენერალურ ერთობლიობაზე.

ხშირად, მაშინაც კი შეუძლებელია გენერალური ერთობლიობის მთლიანი შემოწმება, როდესაც მისი მოცულობა არც ისე დიდია. მაგალითად, თუ რომელიმე ელემენტის შემოწმების შემდეგ ის მწყობრიდან გამოდის, მაშინ ცხადია, მთლიან შემოწმებას აზრი არა აქვს. ასეთ შემთხვევაში შერჩევითი ერთობლიობის შემოწმებას მიმართავენ და ამით წარმოდგენას იქმნიან გენერალურ ერთობლიობაზე.

შერჩევითი მეთოდის არსი სწორედ იმაში მდგომარეობს, რომ გენერალური ერთობლიობიდან სავსებით შემთხვევით არჩევენ შერჩევით ერთობლიობას, რომლის ყოველ ელემენტს ცალ ცალკე ამოწმებენ, რის შედეგად ანგარიშობენ შერჩევითი ერთობლიობის მახასიათებლებს და ამით გარკვეულ წარმოდგენას იქმნიან თვითონ გენერალურ ერთობლიობაზე.

შერჩევითი ერთობლიობის გამოყოფა შესაძლებელია სხვადასხვანაირად მოხდეს· თუ შერჩევას ისე ვაწარმოებთ, რომ მასში ერთხელ მოხვედრილი ელემენტი განმეორებითაც შეიძლება მოხვდეს, მაშინ მას განმეორებითი შერჩევა ეწოდება, ხოლო თუ ერთხელ შერჩევაში მოხვედრილი ელემენტის ხელმეორედ მასში მოხვედრა გამორიცხულია, მაშინ განუმეორებელი შერჩევა გვექნება. მიღებულ შერჩევას ეწოდება რეპრეზენტატული (წარმომადგენლური), თუ ის საკმარისად კარგად წარმოადგენს გენერალური ერთობლიობის პროპორციას. შერჩევის რეპრეზენტატულობა ხშირად შერჩევის შემთხვევითობით მიიღწევა.

განმეორებითი შერჩევის დროს შესაძლებელია გენერალური ერთობლიობის რომელიმე ელემენტი რამდენჯერმე იქნეს შემოწმებული, ხოლო განუმეორებელი შერჩევის დროს ასეთი შემთხვევა გამორიცხულია. ამიტომ განუმეორებელი შერჩევის დროს მიღებული შედეგები უფრო ზუსტი და საიმედო იქნება, ვიდრე ის შედეგები, რომლებიც მიიღება განმეორებითი შერჩევის შემთხვევაში.

შერჩევითი მეთოდის გამოყენებისას შერჩევითი ერთობლიობის მახასიათებლები გარკვეულ წარმოდგენას გვიქმნის გენერალური ერთობლიობის მახასიათებლებზე. დიდ რიცხვთა კანონის ძალით, რომ

ლიც შერჩევითი მეთოდის საფუძველს წარმოადგენს, რაც უფრო დიდია შერჩევითი ერთობლიობის მოცულობა, მით უფრო ზუსტად ასახავს მისი მახასიათლებლები გენერალური ერთობლიობის მახასიათებლებს.

თუ გენერალური ერთობლიობა ელემენტთა საკმაოდ დიდ რაოდენობას შეიცავს, მაშინ ამ ერთობლიობიდან ყოველთვის შეიძლება გამოიყოს ელემენტთა გარკვეული ჯგუფი, რომელთაც იგივე მახასიათებლები აქვს, რაც მთლიანად გენერალურ ერთობლიობას.

ვთქვათ, უნდა შევამოწმოთ ფეხსაცმლის ფაბრიკაში დამზადებული ფეხსაცმელი. ფეხსაცმელი შეიძლება იყოს როგორც ვარგისი, ისე უვარგისი. შესამოწმებელია 1000 წყვილი ფეხსაცმელი.

ვთქვათ, 1000-ვე წყვილი შევამოწმეთ მთლიანად და გაირკვა, რომ 950 წყვილი ვარგისია, 50 კი—უვარგისი. ახლა ვთქვათ, 1000 წყვილიდან შემთხვევით ვარჩევთ 100 წყვილს და მათ სათითაოდ ვამოწმებთ. შემოწმების შედეგად, ვთქვათ ვნახეთ, რომ 95 წყვილი ფეხსაცმელი ვარგისია და 5 წყვილი კი უვარგისი. ამ შერჩევითი ერთობლიობის შემოწმების შედეგად შეგვიძლია დავასკვნათ, რომ, რადგანაც შერჩევით ერთობლიობაში, რომლის მოცულობა უდრის 100-ს, 5 წუნდებულია. მთლიან გენერალურ ერთობლიობაში, რომლის მოცულობა ათჯერ აღემატება შერჩევითი ერთობლიობის მოცულობას — 100-ს, იქნება 10.5=50 წუნდებული. სინამდვილეში მთლიანმა შემოწმებამ აჩვენა, რომ გენერალური ერთობლიობა 50 წუნდებულს შეიცავს. როგორც ვხედავთ, შერჩევითი ერთობლიობის ისეთნაირი შერჩევა შეიძლება, რომ მისი შესწავლით საკმაოდ ზუსტი წარმოდგენა შევიქმნათ გენერალურ ერთობლიობაზე. ამოცანა სწორედ ის არის, თუ როგორ მივაღწიოთ ისეთ შერჩევას, რომელიც რაც შეიძლება ზუსტად ასახავს გენერალურ ერთობლიობას.

შერჩევით მეთოდს აქვს ორი ძირითადი პრინციპი: ა) გენერალური ერთობლიობის ელემენტების შერჩევა დამოუკიდებელია შესასწავლი ობიექტის მნიშვნელობისაგან და ბ) გენერალური ერთობლიობის ელემენტებს შერჩევის ერთობლიობაში მოხვედრის ერთი და იგივე ალბათობები აქვს.

პირველი პრინციპი ნიშნავს ელემენტის შერჩევას გენერალური ერთობლიობიდან ისე, რომ მას არავითარი კავშირი არა ჰქონდეს ობიექტის შესასწავლ თვისებასთან. ამის განხორციელება მაშინ შეიძლება, თუ მაგალითად, გენერალური ერთობლიობის ყოველ ელემენტს თავისი ნომერი აქვს და ეს ნომრები ცალ-ცალკე ერთი და იგივე ზომის მუყაოს ქაღალდზეა დაწერილი, რის შემდეგ მათ მოვათავსებთ რაიმე ყუთში, გულდასმით ავურევთ და სრულიად შემთხვევით ამოვიღებთ რომელიმე მათგანს.

მეორე პრინციპს ადგილი აქვს როგორც განმეორებით, ისე განუმეორებელი შერჩევის შემთხვევაში.

მართლაც, ვთქვათ გვაქვს განმეორებითი შერჩევა. გენერალური ერთობლიობიდან ყოველი ცდის დროს ვიღებთ მხოლოდ ერთ ელემენტს, ვამოწმებთ მას და უკან ვაბრუნებთ. სავსებით ერთსა და იმავე პირობებში ვატარებთ ასეთ ცდას n-ჯერ. გენერალური ერთობლიობის მოცულობა იყოს N. ყოველი ცდის დროს შესაძლებელია N სხვადასხვა შემთხვევა გვქონდეს, ხოლო ყოველ ელემენტს n-ჯერ შეუძლია მოხვდეს შერჩევაში. მაშასადამე, ყოველი ელემენტის შერჩევაში მოხვედრის ალბათობა ერთი და იგივე იქნება, სახელდობრ, თუ ამ ალბათობას p-თი აღვნიშნავთ, გვექნება:

$$p=\frac{n}{N}.$$

ვთქვათ, გვაქვს განუმეორებელი შერჩევა. დავუშვათ, გენერალური ერთობლიობა N ელემენტს შეიცავს, ხოლო შერჩევითი ერთობლიობა კი — n ელემენტს. ასეთი შერჩევა N-ელემენტიანი გენერალური ერთობლიობიდან შეიძლება ავიღოთ იმდენი, რამდენიც იქნება ჯუფთება N ელემენტიდან n-ად, ე. ი.

$$C_N^n=\frac{N!}{n!(N-n)!}.$$

გენერალური ერთობლიობის რომელიმე ელემენტი დავაფიქსიროთ. ცხადია, ის შეიძლება მოხვდეს რომელიმე ერთ-ერთ შერჩევაში. ისეთ შერჩევას, რომელშიც ეს დაფიქსირებული ელემენტია, მივიღებთ შემდეგნაირად: გენერალური ერთობლიობიდან გამოვყოფთ ამ ფიქსირებულ ელემენტს და დარჩენილი $N-1$ ელემენტისაგან შევადგენთ ყველა შესაძლო შერჩევას $n-1$ ელემენტისაგან. ასეთ შერჩევათა რიცხვი იქნება:

$$C_{N-1}^{n-1}=\frac{(N-1)!}{(n-1)!(N-n)!},$$

თითოეულ მათგანს დავუმატებთ ფიქსირებულ ელემენტს. ამგვარად, იმ შერჩევათა რიცხვი, რომლებიც შეიცავს დაფიქსირებულ ელემენტს, იქნება:

$$\frac{(N-1)!}{(n-1)!\ (N-n)!}.$$

მაშასადამე, განუმეორებელი შემთხვევითი შერჩევის დროს, როცა გენერალური ერთობლიობის მოცულობა არის N, ხოლო შერჩევითი

ერთობლიობის კი n, არსებობს შერჩევის $\frac{N!}{n!\,(N-n)!}$ ერთნაირად შესაძლებელი შემთხვევა და გენერალური ერთობლიობის ყოველ ელემენტს აქვს შესაძლებლობა $\frac{(N-1)!}{(n-1)!\,(N-n)!}$ შემთხვევაში მოხვდეს შერჩევაში. ე. ი. ეს უკანასკნელი რიცხვი წარმოადგენს შერჩევაში ნებისმიერი ელემენტის მოხვედრის ხელშემწყობ შემთხვევათა რიცხვს. თუ ალბათობის კლასიკურ განსაზღვრას გამოვიყენებთ, მივიღებთ, რომ გენერალური ერთობლიობის ნებისმიერი ერთ-ერთი ელემენტის შერჩევაში მოხვედრის ალბათობა იქნება:

$$\frac{(N-1)}{(n-1)!\,(N-n)!} : \frac{N!}{n!\,(N-n)!} = \frac{n}{N}.$$

როგორც ვხედავთ, გვექნება განმეორებითი თუ განუმეორებელი შერჩევა, გენერალური ერთობლიობის რომელიმე ელემენტის მოხვედრის ალბათობა შერჩევაში ერთი და იგივეა და ის $\frac{n}{N}$ სიდიდის ტოლია.

შერჩევითი ერთობლიობის პრინციპები პრაქტიკაში ელემენტების შერჩევის ობიექტურობის გარანტიას იძლევა და ყოველ შემთხვევაში გამორიცხულია შერჩევის ტენდენციურად, წინასწარ განზრახვით ჩატარება.

შერჩევითი და გენერალური ერთობლიობის ელემენტებზე დაკვირვების შედეგებს შორის თანადობას რ ე პ რ ე ზ ე ნ ტ ა ტ ი ვ ო ბ ა ეწოდება.

ამბობენ, შერჩევითი ერთობლიობა რეპრეზენტივობს გენერალურ ერთობლიობას, ე. ი. გენერალური ერთობლიობიდან შერჩეული ნაწილი მას წარმოგვიდგენს შემცირებული მასშტაბით.

შერჩევის სხვადასხვა სახე არსებობს. თუ გენერალური ერთობლიობიდან პირდაპირ აწარმოებენ შერჩევას, შეიძლება გვქონდეს უბრალო შემთხვევითი განუმეორებელი ან განმეორებითი შერჩევა.

უბრალო შემთხვევითი შერჩევა გვექნება მაშინ, როდესაც გენერალური ერთობლიობიდან ირჩევენ თითო-თითო ელემენტს. თუ შემოწმების შემდეგ ამოღებულ ელემენტს უკანვე აბრუნებენ, გვექნება განმეორებითი შერჩევა. წინააღმდეგ შემთხვევაში განუმეორებელ შერჩევასთან გვექნება საქმე.

თუ გენერალური ერთობლიობის დანაწილება შესაძლებელია, მაშინ გვექნება ტიპიური, მექანიკური ან სერიული შერჩევა.

ტიპიური შერჩევისას ელემენტების შერჩევა ხდება არა მთლიანად

გენერალური ერთობლიობიდან, არამედ მისი თითოეული ნაწილებიდან. მაგალითად, თუ რაიმე დეტალი მზადდება რამდენიმე დაზგაზე, მაშინ შერჩევა ხდება არა მთლიანი გენერალური ერთობლიობიდან, არამედ ცალკ-ცალკე ყველა დაზგიდან.

მექანიკური შერჩევისას გენერალურ ერთობლიობას მექანიკურად ყოფენ იმდენ ნაწილად, რამდენი ელემენტიცაა საჭირო შემოწმებისათვის და თითოეული ნაწილიდან იღებენ თითო ელემენტს.

სერიული ეწოდება ისეთ შერჩევას, რომლის დროსაც გენერალური ერთობლიობიდან ელემენტებს არჩევენ არა თითო-თითოდ, არამედ „სერიებით", რომელთა მთლიანი შემოწმება ხდება. თუ დეტალი მზადდება რამდენიმე დაზგა-ავტომატზე, რამდენიმე მათგანის დამზადებულ დეტალებს მთლიანად ამოწმებენ.

ხშირად მიმართავენ ე. წ. კომბინირებულ მეთოდს, რომელშიც მონაწილეობს შერჩევის ზემოთ ჩამოთვლილი ყველა სახე, სახელდობრ გენერალურ ერთობლიობას ყოფენ ერთი და იგივე მოცულობის ტოლ სერიებად, შემდეგ უბრალო შერჩევით იღებენ რამდენიმე სერიას და ბოლოს ყველა სერიიდან უბრალო შემთხვევითი შერჩევის გზით იღებენ ცალკე ელემენტებს.

§ 2. დაკვირვების ცდომილებანი

აბსოლუტურად ზუსტად არავითარი გაზომვის წარმოება არ შეიძლება. ასევე აბსოლუტურად ზუსტ შედეგს არავითარი დაკვირვება არ იძლევა.

გაზომვის ან დაკვირვების შედეგი ყოველთვის რაღაც ცდომილებას შეიცავს, რომელიც განმეორებითი გაზომვის ან დაკვირვების დროს ვლინდება. იმისდა მიხედვით, თუ როგორი განსხვავება იქნება ცალკეულ გაზომვათა შედეგებს შორის, შეიძლება ცდომილების სიდიდეზე ვილაპარაკოთ.

ცდომილებანი დამოკიდებულია სხვადასხვა ფაქტორზე. ცდომილება შეიძლება გამოწვეული იყოს ხელსაწყოს, მზომელის ან სხვა გარეშე მიზეზით.

საზოგადოდ, ამა თუ იმ სიდიდის გაზომვისას ან რაიმე დაკვირვებისას მიღებული ცდომილებანი შესაძლებელია დავანაწილოთ სამ ტიპად: სისტემატურ, უხეშ და შემთხვევით ცდომილებებად.

ს ი ს ტ ე მ ა ტ უ რ ი ცდომილება ყოველი ცალკეული გაზომვისას მეორდება. მისი მიზეზი და ხასიათი ცნობილია. სისტემატური ცდომილებათა უკუგდება შესაძლებელია თავიდანვე ან საბოლოო შედეგის დაზუსტებისას.

ცხადია, რომ სისტემატური ცდომილების მათემატიკური ლოდინი ნულისაგან განსხვავებული სიდიდე იქნება.

უ ხ ე შ ცდომილებათა გამოცნობა ადვილია, რადგანაც ისინი აშკარად ყალბ შედეგებს გამოსახავენ.

გაზომვის შედეგი უხეშ ცდომილებად ჩაითვლება, თუ მისი სიდიდე დანარჩენ გაზომვათა ან დაკვირვების შედეგად მიღებულ მნიშვნელობათა მათემატიკური ლოდინისაგან გასამკეცებულ საშუალო კვადრატულ გადახრაზე მეტი სიდიდით განსხვავდება.

ამ უკანასკნელ დებულებას ის აზრი აქვს, რომ საზოგადოდ, ტექნიკაში დაშვების საზღვრებად გასამკეცებულ საშუალო კვადრატულ გადახრას ღებულობენ, რაც ნიშნავს შემდეგს: თუ გაზომვის ან დაკვირვების შედეგად მიღებული მნიშვნელობანი ნორმალური განაწილების კანონს ემორჩილება, მაშინ ალბათობა იმისა, რომ გადახრის აბსოლუტური მნიშვნელობა ნაკლები იქნება გასამკეცებულ საშუალო კვადრატულ გადახრაზე, 0,997-ს უდრის.

ეს ჩაიწერება შემდეგნაირად:

$$P\{\,|\,x-E(x)\,|\,<3\sigma_x\}=\Phi(3)=0{,}997, \qquad (2.1)$$

სადაც $\Phi(a)$ ნორმალური ინტეგრალია, რომლის მნიშვნელობა მოინახება II ცხრილის საშუალებით.

უკანასკნელი ტოლობა გადავწეროთ შემდეგნაირად:

$$P\,|\,x-E(x)\,|\,\geqslant 3\sigma_x\,\}=1-\Phi(3)=0{,}003. \qquad (2.2)$$

ეს იმის მაჩვენებელია, რომ ისეთი სიდიდეები, რომელთა გადახრის აბსოლუტური მნიშვნელობა გასამკეცებულ საშუალო კვადრატულ გადახრას აღემატება, ყოველ ათასში საშუალოდ არა უმეტეს სამჯერ შეგვხვდება.

ისმება კითხვა: საკმარისია თუ არა ასეთი სიზუსტე პრაქტიკული ამოცანების გადაწყვეტისას? ცხადია, ეს თვითონ დასმული ამოცანის ხასიათზე იქნება დამოკიდებული. უფრო ხშირად, პრაქტიკული ამოცანების გადაწყვეტისას ეს სიზუსტე სავსებით საკმარისია.

ამგვარად, თუ რომელიმე ცალკეული გაზომვის ან დაკვირვების შედეგად მიღებული ცდომილება $3\sigma_x$-ს აღემატება, მაშინ ასეთი გაზომვისას ან დაკვირვებისას მიღებული მნიშვნელობა მხედველობაში არ მიიღება და გასაზომი სიდიდის ან დაკვირვებათა შედეგების რეალური მნიშვნელობის დადგენა გაზომვის ან დაკვირვების სხვა შედეგის საშუალებით ხდება.

თუ გაზომვების ან დაკვირვებების ჩატარების შემდეგ როგორც სისტემატურ, ისე უხეშ ცდომილებებს გამოვრიცხავთ, ისეთი მნიშვნე-

ლობანი დაგვრჩება, რომლებიც მხოლოდ შემთხვევით ცდომილებებს შეიცავს.

შ ე მ თ ხ ვ ე ვ ი თ ცდომილებას სხვადასხვა უცნობი მიზეზი იწვევს, რომელთა წინასწარ მხედველობაში მიღება შეუძლებელია.

გაზომვის ან დაკვირვების შედეგის გადახრა მათემატიკური ლოდინიდან შემთხვევით ცდომილებას წარმოადგენს, ხოლო, როგორც ცნობილია, გადახრის საშუალო მნიშვნელობა ყოველთვის ნულის ტოლია, ამიტომ შემთხვევითი ცდომილების მათემატიკური ლოდინი ნულის ტოლი იქნება.

გასაზომი ან დასაკვირვებელი სიდიდის ნამდვილი მნიშვნელობა აღვნიშნოთ a-თი, ხოლო გაზომვის ან დაკვირვების შედეგი კი — x-ით, მაშინ

$$\Delta x = x - a \tag{2.3}$$

სხვაობას ა ბ ს ო ლ უ ტ უ რ ი ც დ ო მ ი ლ ე ბ ა ეწოდება. ამგვარად, გაზომვის ან დაკვირვების დროს მიღებული ცდომილება აბსოლუტური ცდომილება იქნება.

გაზომვის ან დაკვირვების შედეგის შესაფასებლად მარტო აბსოლუტური ცდომილების ცოდნა არ იქნება საკმარისი· მართლაც, თუ ვიტყვით, რომ რაიმე სიდიდის სიგრძის გაზომვისას ცდომილება 5 მმ-ია, ამით ვერ დავასკვნით, რამდენად კარგად არის გაზომვა ჩატარებული; მაგალითად, თუ გასაზომი სიდიდის სიგრძე 5 კმ-ია, მაშინ მიღებული შედეგი ძალიან კარგი იქნება, მაგრამ თუ გასაზომი სიდიდის სიგრძე 15 მმ-ია, მაშინ გაზომვის შედეგი უვარგისი იქნება. ამგვარად, აბსოლუტური ცდომილების ცოდნასთან ერთად თვით გასაზომი სიდიდის საორიენტაციო მნიშვნელობის ცოდნაა საჭირო.

გაზომვის ან დაკვირვების შედეგის ხარისხის შესაფასებლად იყენებენ ფარდობით ცდომილებას, რომელიც აბსოლუტური ცდომილების გასაზომი სიდიდის ჭეშმარიტ მნიშვნელობასთან ფარდობას წარმოადგენს.

ცხადია, რომ ფარდობითი ცდომილება ყოველთვის განყენებული რიცხვით გამოისახება.

რაც მცირე იქნება ფარდობითი ცდომილება, მით უფრო ზუსტი იქნება გაზომვის ან დაკვირვების შედეგი და ამიტომ ფარდობითი ცდომილება შედეგის სიზუსტის ზომად არის მიღებული.

ვთქვათ, რომელიმე a სიდიდის გაზომვის შედეგად მივიღეთ შემდეგი მნიშვნელობანი:

$$x_1,\ x_2\ \cdots,\ x_n.$$

ვიგულისხმოთ, რომ ეს მნიშვნელობანი ერთმანეთისაგან წყვილ-წყვილად დამოუკიდებელი შემთხვევითი სიდიდეებია და ისინი მხოლოდ შემთხვევით ცდომილებებს შეიცავენ.

ამოცანა იმაში მდგომარეობს, რომ მათემატიკური სტატისტიკის მეთოდების მომარჯვებით და გაზომვის შედეგად მიღებული n შესაძლო მნიშვნელობის მიხედვით დავადგინოთ გასაზომი სიდიდის მნიშვნელობა.

როგორც უკვე ვნახეთ, რადგანაც გაზომვის შედეგები მხოლოდ შემთხვევით ცდომილებებს შეიცავენ, ცალკეულ გაზომვათა შედეგების მათემატიკური ლოდინები ერთმანეთის ტოლი იქნება და გასაზომი სიდიდის ჭეშმარიტ მნიშვნელობას ემთხვევა, ე. ი.

$$E(x_1)=E(x_2)=\ldots=E(x_n)=a.$$

ამ უკანასკნელი გარემოების გამოყენებით დავასკვნით, რომ გაზომვის შედეგად მიღებულ მნიშვნელობათა საშუალო არითმეტიკულის მათემატიკური ლოდინი a-ს ტოლი იქნება, ე. ი. შეგვიძლია დავწეროთ:

$$E(\overline{x})=a. \tag{2.4}$$

მართლაც, როგორც ვიცით, $\overline{x}$ საშუალო არითმეტიკული გამოისახება შემდეგნაირად:

$$\overline{x}=\frac{x_1+x_2+\cdots+x_n}{n},$$

მაშინ

$$E(\overline{x})=\frac{1}{n}[E(x_1)+E(x_2)+\cdots+E(x_n)]=\frac{1}{n}\cdot n\cdot a=a.$$

ამის საფუძველზე შეგვიძლია შემდეგი მიახლოებითი ტოლობის დაწერა:

$$a\simeq\overline{x}. \tag{2.5}$$

ამგვარად, გასაზომი სიდიდის ნამდვილ მნიშვნელობად შეგვიძლია მიახლოებით მივიღოთ გაზომვის შედეგად მიღებულ მნიშვნელობათა საშუალო არითმეტიკული; ეს მიახლოება მით უფრო ზუსტი იქნება, რაც მეტი იქნება როგორც ცალკეულ გაზომვათა სიზუსტე, ისე გაზომვათა რაოდენობა.

§ 3. რეპრეზენტატივობის ცდომილება

შერჩევით მეთოდს თან ახლავს ამ მეთოდისათვის დამახასიათებელი ცდომილება, რომელსაც **რეპრეზენტატივობის ცდომილება** ეწოდება.

ვთქვათ, გენერალური ერთობლიობის საშუალო არის a, ხოლო დისპერსია კი—σ^2. a და σ^2 სიდიდეები განხილული გენერალური ერთობლიობისათვის შესაძლებელია ჩვენთვის უცნობი იყოს, მაგრამ, რა თქმა უნდა, ისინი სავსებით განსაზღვრული სიდიდეებია.

გენერალური ერთობლიობიდან ავიღოთ ერთი და იგივე მოცულობის შერჩევითი ერთობლიობა რამდენჯერმე ერთი და იმავე ხერხით. ასეთ დროს სხვადასხვა შერჩევით ერთობლიობებს შეიძლება ჰქონდეს განსხვავებული ელემენტები და ამიტომ შესაძლებელია მათ განსხვავებული $\overline{x}$ საშუალო არითმეტიკული და s^2 ემპირიული დისპერსია ექნეს.

$\overline{x}$ საშუალოთა და s^2 ემპირიულ დისპერსიათა შორის სხვაობა სწორედ იმას გვიჩვენებს, რომ შერჩევითი ერთობლიობის $\overline{x}$ და s^2 მახასიათებლებსა და გენერალური ერთობლიობის a და σ^2 მახასიათებლებს შორის განსხვავება არსებობს.

შერჩევითი და გენერალური ერთობლიობის მახასიათებელთა მნიშვნელობებს შორის სხვაობას ეწოდება რ ე პ რ ე ზ ე ნ ტ ა ტ ი ვ ო ბ ი ს ც დ ო მ ი ლ ე ბ ა.

რეპრეზენტატივობის ცდომილების სიდიდე იმაზეა დამოკიდებული, თუ რამდენად კარგად წარმოადგენს შერჩევითი ერთობლიობა გენერალურს. ცხადია, რომ რეპრეზენტატივობის ცდომილებას ადგილი არ ექნება გენერალური ერთობლიობის მთლიანად შემოწმების დროს.

რეპრეზენტატივობის ცდომილება, ცხადია, შერჩევითი მეთოდის მათემატიკური თეორიის ძირითად ობიექტს წარმოადგენს. სწორედ მის სიდიდეზეა დამოკიდებული შერჩევითი ერთობლიობის შესწავლით მიღებულ შეფასებათა როგორც სიზუსტე, ისე საიმედოობა.

§ 4. შერჩევის სიზუსტესა და მოცულობას შორის დამოკიდებულება

ვთქვათ, მოცემულია გენერალური ერთობლიობა, რომლის მოცულობაა N. ვიგულისხმოთ, რომ მისი ელემენტები ერთმანეთისაგან განსხვავებულია და გადანომრილია 1-დან N-მდე. შერჩევითი ერთობლიობის მოცულობა იყოს n. დავუშვათ, რომ შერჩევაში მოხვედრილი ელემენტებია: $x_1, x_2, \ldots x_n$, მაშინ გენერალური ერთობლიობის საშუალო იქნება:

$$a = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad (4.1)$$

ხოლო შერჩევითი საშუალო არითმეტიკული იქნება:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{4.2}$$

$\overline{x}$ არის შემთხვევითი სიდიდე, რომელიც შერჩევიდან შერჩევამდე იცვლება და, როგორც უკვე ვნახეთ, მისი მათემატიკური ლოდინი ემთხვევა გენერალურ საშუალოს

$$E(\overline{x}) = a.$$

ეს იმას ნიშნავს, რომ

$$E(\overline{x} - a) = 0.$$

ე. ი. გადახრის მათემატიკური ლოდინი ნულია· თვით გადახრა შემთხვევით ცდომილებას წარმოადგენს.

თუ

$$E(\overline{x}) \neq a,$$

მაშინ $\overline{x} - a$ გადახრის მათემატიკური ლოდინი ნულისაგან იქნება განსხვავებული და გვიჩვენებს სისტემატური ცდომილების სიდიდეს.

$\overline{x}$ და a-ს შორის სიახლოვის შეფასება შეიძლება ჩებიშევის უტოლობის დახმარებით:

$$P\{\ |\overline{x} - a| \leqslant \varepsilon\} > 1 - \frac{\sigma^2_{\overline{x}}}{\varepsilon^2},$$

სადაც $\sigma_{\overline{x}}$ არის შერჩევითი ერთობლიობის საშუალო კვადრატული გადახრა.

როგორც ვიცით, განმეორებით შერჩევის შემთხვევაში

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n},$$

ხოლო განუმეორებელი შერჩევის დროს კი —

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right).$$

შერჩევითი საშუალო არითმეტიკულის საშუალო კვადრატულ გადახრას a-ს $\overline{x}$-ით განსაზღვრის საშუალო კვადრატული ცდომილება ეწოდება.

განმეორებითი შერჩევის დროს

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}},$$

ხოლო განუმეორებელი შერჩევის დროს კი —

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}.$$

როგორც ვხედავთ, შერჩევის სიზუსტე დამოკიდებულია გენერალური ერთობლიობის საშუალო კვადრატულ გადახრაზე, შერჩევის n მოცულობასა და $\sqrt{1-\frac{n}{N}}$ სიდიდეზე, რომელიც განმეორებითი შერჩევისას ერთის ტოლია.

პრაქტიკულად შერჩევითი ერთობლიობის მოცულობა გენერალური ერთობლიობის მოცულობის მეათედს არ აღემატება, იმ შემთხვევაში, როდესაც

$$\frac{n}{N}=0{,}1, \text{ მაშინ } 1-\frac{n}{N}=0{,}9 \text{ და } \sqrt{1-\frac{n}{N}}=0{,}95, \text{ თუ}$$

$$\frac{n}{N}<0{,}1, \text{ მაშინ } \sqrt{1-\frac{n}{N}}>0{,}95.$$

საშუალო კვადრატული ცდომილების სიდიდე განმეორებითი და განუმეორებელი შერჩევის შემთხვევაში თითქმის ერთი და იგივეა, და, მაშასადამე, შერჩევითი ერთობლიობის საშუალოს სიზუსტე შერჩევის მოცულობიდან კვადრატული ფესვის პირდაპირპროპორციულია.

როგორც ვხედავთ, სიზუსტის გასადიდებლად გაცილებით უფრო მეტად უნდა გაიზარდოს შერჩევის მოცულობა.

შერჩევის სიზუსტე გენერალური ერთობლიობის საშუალო კვადრატული გადახრის უკუპროპორციულია, ე. ი., რაც უფრო მეტია გაფანტულობა ცენტრის გარშემო გენერალურ ერთობლიობაში, მით უფრო ნაკლებია სიზუსტე და მით უფრო მეტია $a=\bar{x}$ ტოლობის საშუალო კვადრატული ცდომილება.

§ 5. საშუალო კვადრატული ცდომილება

$a=\bar{x}$ ტოლობის საშუალო კვადრატულ ცდომილებას და გენერალური ერთობლიობის საშუალო კვადრატულ გადახრას შორის, როგორც ვნახეთ, საჟანადო კავშირი არსებობს. მაგრამ ხშირ შემთხვევაში გენერალური ერთობლიობის საშუალო კვადრატული გადახრა უცნობია და ამიტომ მისი შეცვლა გვიხდება შერჩევითი ერთობლიობის საშუალო კვადრატული გადახრით, რომელიც s-ით აღინიშნება.

როგორც ვიცით, შერჩევითი დისპერსია გამოისახება შემდეგნაირად:

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

ეს უკანასკნელი წარმოვადგინოთ ასე:

$$s^2=\frac{1}{n}\sum_{i=1}^{n}[(x_i-a)-(\overline{x}-a)]^2=\frac{1}{n}\sum_{i=1}^{n}[(r_i-a)^2-$$

$$-2(\overline{x}-a)(x_i-a)+(\overline{x}-a)^2]=\frac{1}{n}\sum_{i=1}^{u}(x_i-a)^2-$$

$$-2\frac{1}{n}(\overline{x}-a)\sum_{i=1}^{n}(x_i-a)+\frac{1}{n}\cdot n(\overline{x}-a)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2-$$

$$-2\cdot\frac{1}{n}(\overline{x}-a)\cdot n(\overline{x}-a)+(\overline{x}-a)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2-(\overline{x}-a)^2.$$

შერჩევითი ერთობლიობის დისპერსია შემთხვევით სიდიდეს წარმოადგენს. ის შერჩევიდან შერჩევამდე იცვლება. ვიპოვოთ მისი მათემატიკური ლოდინი:

$$Es^2=\frac{1}{n}E\sum_{i=1}^{n}(x_i-a)^2-E(\overline{x}-a)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}E(x_i-a)^2-E(\overline{x}-a)^2,$$

მაგრამ, რადგანაც

$$E(\overline{x}-a)^2=E\left(\frac{x_1-a}{n}+\frac{x_2-a}{n}+\dots+\frac{x_n-a}{n}\right)^2=$$

$$=\frac{1}{n^2}\,]E(x_1-a)^2+E(x_2-a)^2+\dots+E(x_n-a)^2\,]=$$

$$=\frac{1}{n^2}\cdot n\sigma_x^2=\frac{\sigma_x^2}{n},$$

ხოლო

$$\sum_{i=1}^{n}E(x_i-a)^2=n\sigma_x^2,$$

ამიტომ დავწერთ:

$$Es^2=\frac{1}{n}n\sigma_x^2-\frac{\sigma_x^2}{n}=\sigma_x^2-\frac{\sigma_x^2}{n}=\sigma_x^2\left(1-\frac{1}{n}\right)=\sigma_x^2\left(\frac{n-1}{n}\right).$$

ამგვარად მივიღეთ, რომ შერჩევითი ერთობლიობის დისპერსიის მათემატიკური ლოდინი ყოფილა:

$$E\,s^2=\frac{n-1}{n}\sigma_x^2. \tag{5.1}$$

როგორც ვხედავთ, შერჩევითი ერთობლიობის დისპერსიის მათემატიკური ლოდინი არ ემთხვევა გენერალური ერთობლიობის დისპერსიას:

$$E\,s^2=\sigma_x^2-\frac{\sigma_x^2}{n}.$$

განსხვავებაა $\frac{\sigma_x^2}{n}$ სიდიდე. ამიტომ შერჩევითი დისპერსია წარმოადგენს გენერალური ერთობლიობის დისპერსიის მიახლოებით მნიშვნელობას, რომლის მუდმივი ცდომილება $\frac{1}{n}$ σ_x^2-ის ტოლია.

§ 6. შერჩევის ცდომილებისა და მოცულობის დადგენა

საჭიროა შერჩევითი ერთობლიობის მოცულობის ისეთნაირად არჩევა, რომ მან უზრუნველყოს a პარამეტრის $\overline{x}$-ის საშუალებით და σ-ს s-ის საშუალებით რაც შეიძლება ზუსტი და საიმედო შეფასება.

ლიაპუნოვის თეორემის თანახმად დავწერთ:

$$P\left\{-z_a\leqslant\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}<z_a\right\}\simeq 2\Phi(z_a)=\alpha.$$

ამ ფორმულის გამოყენების დროს აუცილებელია გენერალური დისპერსიით სარგებლობა, მაგრამ ეს დისპერსია ყოველთვის არ იქნება ჩვენთვის ცნობილი. იმ შემთხვევაში, როდესაც გენერალური დისპერსია უცნობია, მას ვცვლით s^2 ემპირიული დისპერსიით.

თუ ვისარგებლებთ ბერნულის თეორემით

$$P\left\{\left|\frac{m}{n}-P\right|<z\sqrt{\frac{pq}{n}}\right\}=2\Phi(z)=\frac{2}{\sqrt{2\pi}}\int_0^z e^{-\frac{t^2}{2}}dt, \tag{6.1}$$

ამის დაგვჭირდება $z\sqrt{\frac{pq}{n}}$ გამოსახულების დახმარებით ფარდობითი სიხშირის ცალკეული ხდომილობის p ალბათობისაგან გადახრის სიდიდის განსაზღვრა.

როდესაც p უცნობია, მის მაგივრად $\frac{m}{n}$ ფარდობით სიხშირეს ღებულობენ, ხოლო $q=1-\frac{m}{n}$ და ფარდობითი სიხშირის p ალბათობისაგან გადახრის სიდიდის განსასაზღვრავად ღებულობენ შემდეგ გამოსახულებას:

$$z\sqrt{\frac{m}{n^2}\left(1-\frac{m}{n}\right)}=\frac{z}{n}\sqrt{\frac{m(n-m)}{n}}. \tag{6.2}$$

უკანასკნელი გამოსახულება შერჩევის ცდომილების ზღვარს წარმოადგენს. ის აღვნიშნოთ Δ-თი, მაშინ გვექნება:

$$\Delta=\frac{z}{n}\sqrt{\frac{m(n-m)}{n}}=z\sqrt{\frac{pq}{n}}. \tag{6.3}$$

ბურთების არდაბრუნების სქემის შემთხვევაში ბერნულის თეორემა ჩაიწერება შემდეგნაირად:

$$P\left\{-z\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}<\frac{m}{n}-p<z\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}\right\}=$$

$$=\frac{2}{\sqrt{2\pi}}\int_0^z e^{-\frac{t^2}{2}}dt. \tag{6.4}$$

აქ შერჩევის ცდომილების ზღვარი იქნება:

$$\Delta=z\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}, \tag{6.5}$$

სადაც N გენერალური ერთობლიობის, ხოლო n შერჩევითი ერთობლიობის მოცულობანია.

იმავე ბურთების არდაბრუნების სქემის დროს ჩებიშევის თეორემა ჩაიწერება შემდეგნაირად:

$$P\left\{|\overline{x}-a|<z\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}\right\}=\frac{2}{\sqrt{2\pi}}\int_0^z e^{-\frac{t^2}{2}}dt. \tag{6.6}$$

აქ შერჩევის ცდომილების ზღვარი იქნება:

$$\Delta = z\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}. \qquad (6.7)$$

ბურთთა დაბრუნების სქემის შემთხვევაში ჩებიშევის თეორემის გამოყენებისას შერჩევის ცდომილების ზღვარი იქნება:

$$\Delta = z\frac{\sigma}{\sqrt{n}}. \qquad (6.8)$$

შევნიშნოთ, რომ (6.4) და (6.6) ფორმულებით სარგებლობენ ბურთთა არდაბრუნების სქემის შემთხვევაში, თუ $\frac{n}{N}$ ფარდობა არ არის მცირე სიდიდე.

თუ ეს ფარდობა მცირეა, ვთქვათ, ის არის 0,1, მაშინ (6.4) და (6.6) ფორმულებით მოცემული სიზუსტე მცირედ იქნება განსხვავებული (6.1) და (6.2) ფორმულებით მოცემული სიზუსტეებისაგან, ამიტომ, თუ $\frac{n}{N}$ ფარდობა მცირეა, მაშინ უმჯობესია სიმარტივის გამო (6.1) და (6.2) ფორმულებით სარგებლობა.

ახლა დავსვათ ასეთი საკითხი: რას უდრის ალბათობა, რომელიც შერჩევის მოცემულ ცდომილებას არ აღემატება. ამის გასარკვევად განვიხილოთ შემდეგი მაგალითი:

ვთქვათ, გვაქვს 3000 ერთეულისაგან შემდგარი რაიმე გენერალური ერთობლიობა, საიდანაც 1000 ერთეულის შემოწმებით გამოირკვა, რომ 25 წუნდებულია. რას უდრის ალბათობა იმისა, რომ მთელი გენერალური ერთობლიობა 4%-ზე მეტ წუნდებულს არ შეიცავს?

აქ გვაქვს ბურთების არდაბრუნების სქემა. ვისარგებლოთ (6.4) ფორმულით. მხედველობაში მივიღოთ ის გარემოება, რომ $\frac{n}{N}=\frac{1}{3}$ ფარდობა არც ისე მცირე სიდიდეა.

წუნდებულთა რაოდენობა $m=25$. მაშასადამე, წუნდებულის ფარდობითი სიხშირე შერჩევით ერთობლიობაში იქნება:

$$\frac{m}{n}=\frac{25}{1000}=0,025,$$

გენერალური ერთობლიობისათვის წუნდებულთა ხვედრი ჩვენთვის უცნობია, ამიტომ შეგვიძლია დავუშვათ, რომ ის დაახლოებით შერჩევით ერთობლიობაში წუნდებულის ფარდობითი სიხშირის ტოლია:

$$p=\frac{m}{n}=0,025.$$

მაშინ

$$q=1-p=0{,}975.$$

(6.5) ფორმულის დახმარებით გამოვთვალოთ ცდომილების ზღვარი; რადგანაც, პირობის თანახმად, წუნდებულის რაოდენობა გენერალური ერთობლიობის 0,04-ს არ უნდა აღემატებოდეს, ამიტომ ცდომილების ზღვარი ტოლი იქნება შემდეგი სხვაობისა:

$$\frac{4}{100}-\frac{1}{40}=0{,}015.$$

მეორე მხრივ, (6.5) ფორმულიდან ჩანს, რომ ცდომილების ზღვარისათვის გვექნება:

$$\Delta=z\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}.$$

მაშასადამე, შეგვიძლია შემდეგი ტოლობის დაწერა:

$$z\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}=0{,}015.$$

ამ უკანასკნელ ტოლობაში შევიტანოთ p, q, n და N-ის მნიშვნელობანი, გვექნება:

$$z\sqrt{\frac{1\,.\,39\,.\,2}{40\,.\,40\,.\,1000\,.\,3}}=0{,}015,$$

საიდანაც

$$z=3{,}72.$$

ახლა $\Phi(z)$ ფუნქციათა ცხრილიდან ვიპოვით, რომ საძიებელი ალბათობა არის

$$p=\Phi(3{,}72)=0{,}9998.$$

როგორც ვხედავთ, მიღებული ალბათობა ძალიან ახლოს არის ერთთან. ამიტომ საკმაოდ დიდი რწმენით შეგვიძლია ვამტკიცოთ, რომ გენერალური ერთობლიობა 4%-ზე მეტ წუნდებულს არ შეიცავს.

ახლა მოცემული n-თვის განვსაზღვროთ შერჩევის სიზუსტე.

აქაც განვიხილოთ შემდეგი მაგალითი. 500 000 ჰექტარიანი ფართობის ერთ ჰექტარზე პურის საშუალო მოსავლის განსასაზღვრავად 5000 ჰექტარზე ჩატარებულია შერჩევითი გამოკვლევა; ვთქვათ, მოსავალი განაწილებულია შემდეგნაირად (ცხრილი 10).

ცხრილი 10

მოსავალი ჰექტარზე ცენტნერობით	11,5	11,8	12	12,4	12,8	13,2
ჰექტართა რიცხვი	500	600	900	800	1000	1200

ვიპოვოთ შერჩევითი გამოკვლევის ცდომილების ზღვარი. (6.7) ფორმულის თანახმად ცდომილების ზღვარი იქნება:

$$\Delta = z\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}.$$

განხილული მაგალითის შემთხვევაში

$$\frac{n}{N}=0{,}01,$$

მაშასადამე,

$$\sqrt{1-\frac{n}{N}}=\sqrt{0{,}99},$$

ეს კი მცირედ განსხვავდება ერთისაგან, ამიტომ ამ განსხვავებას შეიძლება ყურადღება არ მივაქციოთ და ცდომილების ზღვარის განსასაზღვრავად გამოვიყენოთ (6.8) ფორმულა.

ვიანგარიშოთ შერჩევითი საშუალო $\overline{x}$:

$$\overline{x}=\frac{1}{5000}(11{,}5\cdot 500+11{,}8\cdot 600+12\cdot 900+12{,}4\cdot 800+$$

$$+12{,}8\cdot 1000+13{,}2\cdot 1200)=12{,}44.$$

ამგვარად, მიახლოებით შეიძლება მივიღოთ, რომ პურის საშუალო მოსავალი ჰექტარზე 12,44 ცენტნერია.

ახლა განვსაზღვროთ ცდომილება (6.8) ფორმულით. რადგანაც ჩვენთვის უცნობია σ^2 გენერალური დისპერსია, ამიტომ Δ ცდომილების ზღვარის საანგარიშებლად ის უნდა შევცვალოთ s^2 შერჩევითი დისპერსიით.

პირველად ვიპოვოთ ფარდობითი სიხშირეები. აღვნიშნოთ ისინი სათანადოდ v_1, v_2, v_3, v_4, v_5 და v_6-ით.

$$v_1=0{,}1;\ v_2=0{,}12;\ v_3=0{,}18;\ v_4=0{,}16;\ v_5=0{,}20;\ v_6=0{,}24.$$

ახლა ვიანგარიშოთ s^2 შერჩევითი დისპერსია.

„ყალბ ნულად“ მივიღოთ $c=12{,}4$. s^2-თვის დავწეროთ:

$$s^2=(11,5-12,4)^2\cdot 0,1+(11,8-12,4)^2\cdot 0,12+(12-12,4)^2\cdot 0,19+$$
$$+0+(12,8-12,4)^2\cdot 0,20+(13,2-12,4)^2\cdot 0,24-$$
$$-(12,44-12,4)^2=0,337,$$

საიდანაც

$$s=\sqrt{0,337}\,.$$

(6.8) ფორმულაში σ-ს მაგივრად შევიტანოთ მისი მიახლოებითი მნიშვნელობა s. გვექნება:

$$\Delta=z\,\frac{\sqrt{0,337}}{\sqrt{5000}}=0,0082\,z;$$

თუ $z=1$, მაშინ Δ ცდომილების ზღვარი იქნება

$$\Delta=0,0082.$$

ამ ცდომილების ალბათობას ვიპოვით $\Phi(z)$ ფუნქციის ცხრილის დახმარებით, გვექნება:

$$\Phi(1)=0,68269.$$

როგორც ვხედავთ, ცდომილების ალბათობა საგრძნობლად განსხვავდება ერთისაგან, ამიტომ არ შეიძლება ვამტკიცოთ, რომ შერჩევით მიღებული მოსავალი ნამდვილად, ჰექტარიდან 0,0082 ცენტნერის სიზუსტით, ჰექტარზე საშუალო მოსავალს გვაძლევს.

როცა $z=3$, მაშინ ცდომილება იქნება:

$$\Delta=0,0246.$$

ამ ცდომილების ალბათობაა

$$p=\Phi(3)=0,9973.$$

ეს ალბათობა ახლოს არის ერთთან და ამიტომ შეიძლება ვამტკიცოთ, რომ შერჩევით მიღებული მოსავალი სინამდვილეში, 0,0246 ცენტნერის სიზუსტით, ჰექტარზე საშუალო მოსავალს წარმოადგენს, ე. ი. ნამდვილი საშუალო მოსავალი ჰექტარზე მოთავსდება 12,415 და 12,465 ცენტნერებს შორის.

ახლა მოცემული სიზუსტის მიხედვით განვსაზღვროთ შერჩევითი ერთობლიობის მოცულობა.

ამ სახის ამოცანების ამოსახსნელად ნაჩვენები უნდა იყოს ის სიზუსტე, რომელიც სასურველია, და აგრეთვე მოცემული უნდა იყოს ის ალბათობა, რომლითაც შეიძლება გარანტირებული ვიყოთ.

ცდომილების ზღვარი Δ-თი გვაქვს აღნიშნული; მივცეთ z-ს სხვადასხვა მნიშვნელობა და ვუჩვენოთ სათანადო ალბათობები. ასე მაგა-

ლითად, თუ გვინდა, რომ მოცემული სიზუსტის ალბათობა 0,9995-ის ტოლი იყოს, მაშინ უნდა ავიღოთ $z=3{,}5$.

დასმული ამოცანის ამოსახსნელად ვისარგებლოთ (6.1) ფორმულით, მაშინ Δ ცდომილების ზღვარი იანგარიშება (6.8) ფორმულით

$$\Delta=z\frac{\sigma}{\sqrt{n}},$$

საიდანაც ვპოულობთ

$$n=\frac{z^2\sigma^2}{\Delta^2}, \tag{6.9}$$

თუ ვისარგებლებთ (6.2) ფორმულით, მაშინ ცდომილების ზღვარი (6.3) ფორმულის თანახმად იქნება:

$$\Delta=z\sqrt{\frac{pq}{n}},$$

საიდანაც n-თვის მივიღებთ:

$$n=\frac{z^2pq}{\Delta^2}. \tag{6.10}$$

მაგალითისათვის, ვთქვათ, მოსავლის დისპერსია $\sigma^2=0{,}337$ ცენტნერს შეადგენს. რას უდრის შერჩევის მოცულობა, რომ 0,9973-ის ტოლი ალბათობით შეგვეძლოს მტკიცება იმისა, რომ მოსავლის განსაზღვრის ცდომილება 0,00821 ცენტნერს არ აღემატება.

ვთქვათ, გვაქვს ბურთების დაბრუნების სქემა, მაშინ შეგვიძლია ვისარგებლოთ (6.9) ფორმულით. რადგანაც მოცემული ალბათობა უდრის 0,9973-ს, ხოლო $\Phi(3)=0{,}9973$, ამიტომ $z=3$. პირობის თანახმად $\Delta=0{,}00821$, მაშინ $\Delta^2=0{,}0000674$. თუ მიღებულ მნიშვნელობებს ჩავსვამთ n-ის გამოსახულებაში, გვექნება:

$$n=\frac{3^2\cdot 0{,}337}{0{,}0000674}=45000.$$

§ 7. სტატისტიკური განაწილება და ემპირიული განაწილების ფუნქცია

ვთქვათ გენერალური ერთობლიობიდან ვიღებთ n მოცულობის შერჩევას. დავუშვათ, რომ შერჩევის ელემენტებია $x_1, x_2, \ldots x_n$, რომლებიც ზოგადობის შეუზღუდველად ვიგულისხმოთ, რომ დალაგებულია ზრდადი მიმდევრობით, ე. ი. მიღებული მიმდევრობა წარმოადგენს ვარიაციულ მწკრივს. ვიგულისხმოთ, რომ x_1 შერჩევაში მოხვდა n_1-ჯერ, x_2—n_2-ჯერ და ა. შ. x_k—n_k-ჯერ. ცხადია,

$$\sum_{i=1}^{k} n_i = n.$$

დაკვირვების შედეგად მიღებულ სიდიდეებს ვ ა რ ი ა ნ ტ ე ბ ს უწოდებენ, ვარიანტთა რიცხვს—ს ი ხ შ ი რ ე ე ბ ს, ხოლო მათ შეფარდებას შერჩევის მოცულობასთან ფარდობითი სიხშირე ეწოდება.

შერჩევის სტატისტიკური განაწილება ეწოდება ისეთ ცხრილს, რომლის პირველ სტრიქონში მოცემულია ვარიანტთა მნიშვნელობანი, ხოლო მეორე სტრიქონში კი—შესაბამისი სიხშირეები ან ფარდობითი სიხშირეები.

სტატისტიკური განაწილების წარმოდგენა შეიძლება აგრეთვე ინტერვალების მიმდევრობისა და მათი შესაბამისი სიხშირეთა სახით. აქ სიხშირედ მიიღება ინტერვალში მოხვედრილ სიხშირეთა ჯამი.

სტატისტიკურ განაწილებაში ესმით შემთხვევითი სიდიდის შესაძლო მნიშვნელობათა და შესაბამის ალბათობებს შორის თანადობა, ხოლო მათემატიკურ სტატისტიკაში ასეთი თანადობა მყარდება დაკვირვების შედეგად მიღებულ ვარიანტებსა და მათ სიხშირეებს შორის.

მაგალითად: დავუშვათ, რომ შერჩევის მოცულობა არის 50, მიღებულ ვარიანტთა მნიშვნელობანია 1, —1,2 და 5, ხოლო მათი სიხშირეები—შესაბამისად 5, 10, 20, 15. შევადგინოთ სტატისტიკური განაწილების ცხრილი. მას ექნება შემდეგი სახე:

$$\begin{matrix} x_i \\ n_i \end{matrix} \left\{ \begin{matrix} 1, & -1, & 2, & 5. \\ 5, & 10, & 20, & 15. \end{matrix} \right.$$

ანდა, თუ განაწილებას ჩავწერთ ფარდობითი სიხშირეებით, გვექნება:

$$\begin{matrix} x_i \\ \dfrac{n_i}{n} \end{matrix} \left\{ \begin{matrix} -1, & +1, & 2, & 5 \\ 0{,}2; & 0{,}1; & 0{,}4; & 0{,}3; \end{matrix} \right.$$

მიღებული განაწილების შესამოწმებლად მეორე სტრიქონის ფარდობითი სიხშირეები უნდა შევკრიბოთ, რაც ეტოლება ერთს, მართლაც,

$$0{,}1 + 0{,}2 + 0{,}4 + 0{,}3 = 1.$$

რადგანაც შერჩევითი ერთობლიობის შესწავლით ხდება გენერალურ

ერთობლიობაზე გარკვეული წარმოდგენის შექმნა, ამიტომ პირველ რიგში უნდა შევძლოს შერჩევითი ერთობლიობის განაწილების, მისი მახასიათებლების და ამ მახასიათებელთა განაწილების გამოთვლა.

ვთქვათ, გამოსარკვევია რომელიმე დაზგის და მისი მოწყობილობის სიზუსტე. ვიგულისხმოთ, რომ დაზგაზე რომელიმე დეტალის დამზადების პროცესი ძირითადად უცვლელია. დეტალის თავისი რომელიმე მახასიათებლიდან გადახრა იქნება შემთხვევითი სიდიდე. აღვნიშნოთ იგი X-ით. ვთქვათ, n ჩატარებული დაკვირვების შედეგია $x_1, x_2, \ldots, x_n$. მიღებული მნიშვნელობანი შეგვიძლია მივიღოთ შერჩევითი ერთობლიობის ელემენტებად, რომელთა შესწავლით წარმოდგენა უნდა შევიქმნათ დაზგაზე დამზადებული ყველა დეტალის ერთობლიობაზე.

დავუშვათ, გენერალური ერთობლიობა განაწილებულია $F(x)$ განაწილების კანონის მიხედვით, რომელსაც თეორიულ განაწილებას უწოდებენ. განაწილების ფუნქციის განმარტების თანახმად:

$$F(x)=P\{X<x\}.$$

დავუშვათ, x რიცხვთა ღერძის რომელიმე წერტილია. n_x-ით აღვნიშნოთ ამავე ღერძზე x-ის მარცხნივ მოხვედრილ შერჩევითი ერთობლიობის ელემენტთა რაოდენობა; ცხადია, $\frac{n_x}{n}$ იქნება X შემთხვევით სიდიდეზე დაკვირვების შედეგად მიღებული ფარდობითი სიხშირე, რომელიც არ აღემატება x-ს. ცხადია, ეს ფარდობითი სიხშირე იქნება x-ის ფუნქცია და ის აღვნიშნოთ $S_n(x)$-ით.

$S_n(x)$-ს შერჩევითი განაწილების ფუნქციას ან ემპირიული განაწილების ფუნქციას უწოდებენ. ამგვარად დავწერთ

$$S_n(x)=\frac{n_x}{n}.$$

$S_n(x)$ წარმოადგენს თეორიული განაწილების $F(x)$-ის შეფასებას, რომელიც ეთანადება გენერალური ერთობლიობის განაწილების ფუნქციას.

თუ შემთხვევითი ხდომილობის ალბათობის შეფასება ხდება ფარდობითი სიხშირის საშუალებით, ასევე $F(x)$-ის შეფასება შეიძლება $S_n(x)$-ის დახმარებით-

ფორმალურად $S_n(x)$-ს განაწილების ფუნქციის ყველა თვისება აქვს; მისი მნიშვნელობა ალბათობებით კი არ გამოისახება, არამედ აღებული შერჩევისათვის $X<x$ უტოლობის ფარდობითი სიხშირეებით. სახელდობრ:

$$S_n(x) = \begin{cases} 0, & \text{როცა} \quad x < x_1 \\ \frac{k}{n}, & \text{როცა} \quad x_k \leqslant x < x_{k+1} \\ 1, & \text{როცა} \quad x > x_n, \end{cases}$$

სადაც დაკვირვების შედეგად მიღებული მნიშვნელობანი დალაგებულია ზრდადი მიმდევრობით.

$$x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n,$$

ასეთ მიმდევრობას, როგორც ვიცით, ვარიაციული მწკრივი ეწოდება.

განაწილების ფუნქციის მნიშვნელობა დამოკიდებულია შერჩევით მნიშვნელობათა მდებარეობაზე X ღერძზე და არა მათი შერჩევის რიგზე. ასე რომ, ორ შერჩევას, რომელთაც აქვთ ელემენტთა ერთნაირი შემადგენლობა და სხვადასხვა რიგი, ექნებათ ერთი და იგივე $S_n(x)$ განაწილების ფუნქცია.

რადგანაც $S_n(x)$ წარმოადგენს $F(x)$-ის შეფასებას, ბუნებრივია $S_n(x)$-ის მახასიათებლები მივიღოთ $F(x)$-ის მახასიათებლების შეფასებად.

შერჩევითი მახასიათებლების გამოთვლა არავითარ სიძნელეს არ წარმოადგენს.

თუ ვიანგარიშებთ დაკვირვების შედეგად მიღებულ სიდიდეთა საშუალო არითმეტიკულს

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

ეს იქნება X შემთხვევითი სიდიდის $E(X)$ მათემატიკური ლოდინის შეფასება.

ასე ანალოგიურად შერჩევითი ერთობლიობის დისპერსია, მედიანა, მოდა და სხვ. იქნებიან გენერალური ერთობლიობის დისპერსიის, მედიანის, მოდის და სხვათა შეფასებანი. ცხადია, ემპირიული მახასიათებლები განსხვავებული იქნებიან თეორიული მახასიათებლებისაგან, მაგრამ მათი განსხვავების სიდიდის შეფასება ყოველთვის შესაძლებელია.

$S_n(x)$ ემპირიული განაწილების ფუნქციას აქვს ყველა ის თვისება, რაც ახასიათებს $F(x)$ თეორიული განაწილების ფუნქციას. მართლაც, თვით ემპირიული განაწილების ფუნქციის განმარტებიდან გამომდინარეობს, რომ

1) ემპირიული განაწილების ფუნქციის მნიშვნელობა მოთავსებულია [0,1] სეგმენტში;

2) $S_n(x)$ არის არაკლებადი ფუნქცია;

3) თუ x_1 უმცირესი ვარიანტია, მაშინ $S_n(x)=0$, როცა $x \leqslant x_1$; თუ x_k უდიდესია, მაშინ $S_n(x)=1$, როცა $x > x_k$.

ამგვარად, $S_n(x)$ ემპირიული განაწილების ფუნქცია ემსახურება თეორიული განაწილების ფუნქციის შეფასების საქმეს.

ზემოთ განხილული მაგალითის შემთხვევაში ვიპოვოთ ემპირიული განაწილების სახე.

შერჩევის მოცულობაა 50. უმცირესი ვარიანტია -1, მაშასადამე,

$$S_n(x)=0, \quad \text{როცა } x \leqslant -1.$$

მნიშვნელობა $x<1$, სახელდობრ $x_1=-1$ შემთხვევას ადგილი ჰქონდა 10-ჯერ, მაშასადამე,

$$S_n(x)=\frac{10}{50}=0{,}2; \text{ როცა } -1<x\leqslant 1.$$

მნიშვნელობანი $x<2$, სახელდობრ $x_1=-1$ და $x_2=1$ გვქონდა $5+10=15$ შემთხვევაში. მაშასადამე,

$$S_n(x)=\frac{15}{50}=0{,}3; \text{ როცა } 1<x\leqslant 2.$$

მნიშვნელობანი $x<5$, სახელდობრ $x_1=-1$, $x_2=1$ და $x_3=2$ გვქონდა $5+10+20=35$-ჯერ, მაშასადამე

$$S_n(x)=\frac{35}{50}=0{,}7; \text{ როცა } 2<x\leqslant 5.$$

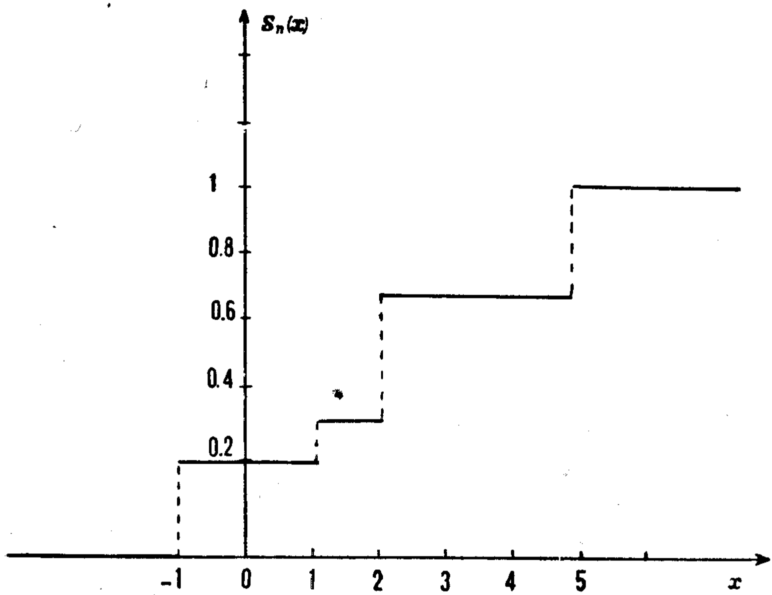
რადგანაც $x=25$ არის უდიდესი ვარიანტის მნიშვნელობა, მაშინ

$$S_n(x)=1, \quad \text{როცა } x>5.$$

ამგვარად საძებნ ემპირიულ ფუნქციას ექნება სახე:

$$S_n(x)=\begin{cases} 0, & \text{როცა } x\leqslant -1, \\ 0{,}2, & \text{როცა } -1<x\leqslant 1, \\ 0{,}3, & \text{როცა } 1<x\leqslant 2, \\ 0{,}7, & \text{როცა } 2<x\leqslant 5, \\ 1, & \text{როცა } x>5. \end{cases}$$

თუ ავაგებთ $S_n(x)$ ფუნქციის გრაფიკს, მას ექნება შემდეგი სახე:



ნახ. 18.

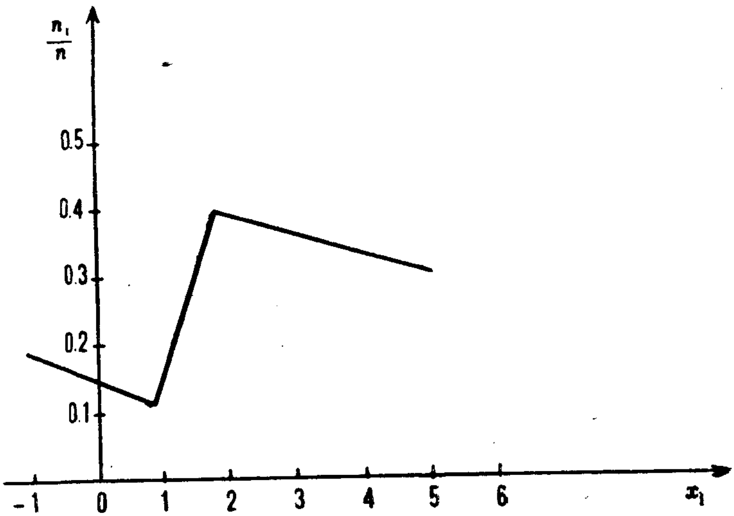
§ 8. პოლიგონი და ჰისტოგრამა

თვალსაჩინოებისათვის ხშირად აგებენ სტატისტიკური განაწილების სხვადასხვა გრაფიკს. ჩვენ განვიხილავთ ორ მათგანს:

1. პოლიგონი. სიხშირეთა პოლიგონს უწოდებენ ტეხილს, რომლის მონაკვეთები აერთებენ (x_i, n_i) წერტილებს $(i=1, 2,\dots k)$. პოლიგონის ასაგებად აბსცისათა ღერძზე გადაზომავენ x_i ვარიანტებს, ხოლო ორდინატთა ღერძზე კი მათ შესაბამის სიხშირეებს. (x_i, n_i) წერტილებს აერთებენ წრფის მონაკვეთით და ღებულობენ სიხშირეთა პოლიგონს.

ანალოგიურად, ფარდობითი სიხშირის პოლიგონი ეწოდება ტეხილს, რომლის მონაკვეთები აერთებენ $\left(x_i; \frac{n_i}{n}\right)$ წერტილებს. ფარდობითი სიხშირის პოლიგონის ასაგებად აბსცისათა ღერძზე გადაზომავენ ვარიანტებს, ხოლო ორდინატთა ღერძზე კი შესაბამის $\frac{n_i}{n}$ ფარდობით სიხშირეებს. $\left(x_i; \frac{n_i}{n}\right)$ წერტილებს აერთებენ წრფის მონაკვეთებით და ღებულობენ ფარდობითი სიხშირის პოლიგონს.

წინა პარაგრაფში განხილული მაგალითისათვის ავაგოთ ფარდობითი სიხშირის პოლიგონი. მას ექნება შემდეგი სახე:



ნახ. 19.

2. ჰ ი ს ტ ო გ რ ა მ ა. უწყვეტი ვარიანტის შემთხვევაში მიზანშეწონილია ჰისტოგრამის აგება. ამ მიზნით იმ ინტერვალს, რომელშიც მოთავსებულია ყველა ვარიანტის მნიშვნელობა, ჰყოფენ h სიგრძის მქონე რამდენიმე ინტერვალად და ანგარიშობენ ყველა კერძო ინტერვალისათვის მასში მოხვედრილ სიხშირეთა ჯამს.

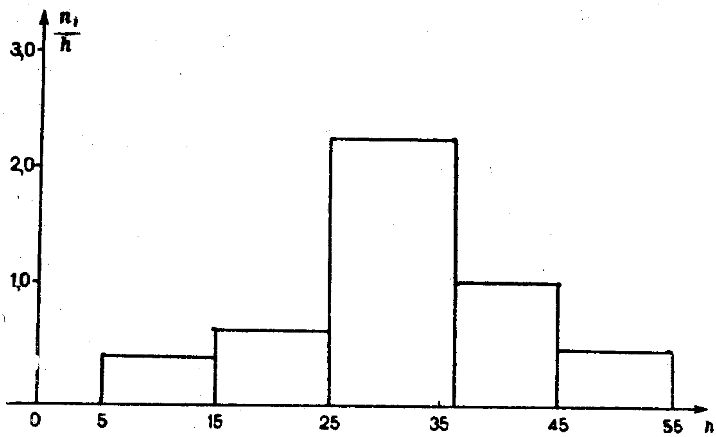
სიხშირის ჰისტოგრამას უწოდებენ კიბისებურ ფიგურას, რომელიც ისეთი მართკუთხედებისაგან შედგება, რომელთა ფუძეს წარმოადგენს h სიგრძის კერძო ინტერვალები; ხოლო სიმაღლე კი $\frac{n_i}{h}$ სიხშირის სიმკვრივის ტოლია. სიხშირის ჰისტოგრამის ასაგებად აბსცისათა ღერძზე გადაზომავენ კერძო ინტერვალებს და მათ ზემოთ $\frac{n_i}{h}$ მანძილის დაშორებით ავლებენ მონაკვეთებს, რომლებიც აბსცისათა ღერძის პარალელურია.

i-ური მართკუთხედის კერძო ფართობი იქნება $h \cdot \frac{n_i}{h} = n_i$. ეს არის ვარიანტის სიხშირეთა ჯამი i-ურ ინტერვალში. მაშასადამე, სიხშირის ჰისტოგრამის ფართობი ყველა სიხშირეთა ჯამის, ე. ი. შერჩევის მოცულობის ტოლია.

ნახ. 20-ზე აგებულია სიხშირის განაწილების ჰისტოგრამა, რომელიც შეესაბამება ქვემოთ მოყვანილი ცხრილის მონაცემებს. აღებულია $n=50$ და $h=10$.

ცხრილი 11

კერძო ინტერვ.	კერძო ინტერვალში ვარიანტის სიხშირეთა ჯამი	სიხშირის სიმკვრივე
5—15	2	0,4
15—25	6	0,6
25—35	24	2,4
35—45	12	1,2
45—55	6	0,6



ნახ. 20.

ფარდობითი სიხშირის ჰისტოგრამა ეწოდება კიბისებურ ფიგურას, რომელიც ისეთი მართკუთხედებისაგან შედგება, რომელთა ფუძეა h სიგრძის კერძო ინტერვალები, ხოლო სიმაღლეა $\frac{n_i}{nh}$ ფარდობითი სიხშირის სიმკვრივეები.

ფარდობითი სიხშირის ჰისტოგრამის ასაგებად აბსცისათა ღერძზე გადაზომავენ კერძო ინტერვალებს, ხოლო მათ ზევით გაჰყავთ $\frac{n_i}{nh}$ მანძილით დაშორებული პარალელური მონაკვეთები.

i-ური კერძო მართკუთხედის ფართობი უდრის $h \cdot \frac{n_i}{n \cdot h} = \frac{n_i}{n}$.

ეს არის იმ ვარიანტის ფარდობითი სიხშირე, რომელიც ხდება i-ურ ინტერვალში.

მაშასადამე, ფარდობითი სიხშირის ჰისტოგრამის ფართობი ყველა ფარდობითი სიხშირეთა ჯამის, ე. ი. ერთის ტოლია.

შერჩევით ერთობლიობის მახასიათებელთა განაწილების კანონების შესწავლა მათემატიკური სტატისტიკის ერთ-ერთ ძირითად ამოცანას წარმოადგენს იმდენად, რამდენადაც მხოლოდ მათი შესწავლით შეიძლება გენერალური ერთობლიობის შეფასების სიზუსტის და საიმედოობის დახასიათება.

მიღებული შეფასების სიზუსტე და საიმედოობა ხასიათდება ალბათობით იმისა, რომ შესაფასებელი სიდიდისათვის მიღებულ მნიშვნელობასა და ჭეშმარიტ მნიშვნელობას შორის სხვაობა არ აღემატება მოცემულ საზღვარს. მაგალითად, თუ საიმედოობის დონედ მივიღებთ 0,997 ალბათობას და ვიგულისხმებთ, რომ შეფასების ჭეშმარიტი მნიშვნელობისაგან გადახრა ისეთ ნორმალური განაწილების კანონს ემორჩილება, რომლის საშუალო მნიშვნელობა ნულია და დისპერსია 1, მაშინ შეფასების სიზუსტე დახასიათდება გადახრის ზედა საზღვრით, რომელიც 3σ-ს ტოლია.

მართლაც, როგორც ცნობილია

$$P\{|x-\overline{x}|<k\}=\Phi\left(\frac{k}{\sigma}\right).$$

თუ დავუშვებთ, რომ $k=3\sigma$ ს, მივიღებთ

$$P\{|x-\overline{x}|<3\sigma\}=\Phi(3).$$

თუ $\Phi(3)$-ის მნიშვნელობას ვიპოვით ნორმალური განაწილების ცხრილიდან, გვექნება

$$P\{|x-\overline{x}|<3\sigma\}=0{,}997.$$

პრაქტიკულად ეს იმას ნიშნავს, რომ გადახრა საშუალო მნიშვნელობიდან, რომელიც 3σ-ს აღემატება, ყოველი 1000 შემთხვევიდან გვხვდება არა უმეტეს 3-ჯერ. მისაღებია თუ არა ასეთი სიზუსტე? ცხადია, ეს ამოცანის შინაარსზეა დამოკიდებული. ისე კი, ხშირ შემთხვევაში, ასეთი უზრუნველყოფა სავსებით საკმარისია.

შეფასების სიზუსტე და საიმედოობა ერთმანეთთან განუყრელად დაკავშირებულია და ორი სხვადასხვა შეფასების სიზუსტის შედარება შეგვიძლია მხოლოდ მაშინ, როცა მათ ერთი და იგივე საიმედოობა აქვთ.

თუ მოცემული გვაქვს დაკვირვების შედეგები, რომლებიც ქმნიან შერჩევით ერთობლიობას და მოგვეთხოვება მათი მეშვეობით შევაფასოთ გენერალური ერთობლიობის პარამეტრები და დავადგინოთ შეფასების სიზუსტე, მაშინ ცხადია, არსებობს დასმული ამოცანის ამოხსნათა უსასრულო სიმრავლე, რადგანაც შესაძლებელია შერჩევით მნიშვნელობათაგან უამრავი ფუნქციის აღება, რომელთა გამოყენება შესაფასებლად შეიძლება.

ამიტომ არსებითია არა მარტო შეფასების პოვნა, არამედ საჭიროა მათი თვისებათა შედარება, რათა მივიღოთ საიმედოობის ერთნაირი დონისათვის რაც შეიძლება ზუსტი შეფასება.

§ 9. შერჩევითი მახასიათებლების პოვნა

ვთქვათ შერჩევითი ერთობლიობის ელემენტებია

$$X_1, \ X_2, \ldots, \ X_n,$$

მაშინ შერჩევითი საშუალო იქნება

$$\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}.$$

თუ შერჩევას ერთი და იგივე გენერალური ერთობლიობიდან ვაწარმოებთ, ცხადია შერჩევაში მოხვედრილი ელემენტები ურთიერთდამოუკიდებელნი იქნებიან და ემორჩილებიან ერთსა და იმავე განაწილების კანონს.

გენერალური ერთობლიობის მათემატიკური ლოდინი იყოს a. ცხადია, რომ

$$E(X_1)=E(X_2)=\cdots=E(X_n)=a,$$

მაშინ

$$E(\overline{X})=\frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n}=a.$$

ამის საფუძველზე ვწერთ:

$$\overline{X}\approx a.$$

გამოვთვალოთ $\overline{X}$-ის დისპერსია, გვექნება:

$$D(\overline{X})=\frac{1}{n^2}\Big[\ D(X_1)+D(X_2)+\cdots+D(X_n)\Big],$$

მაგრამ, რადგანაც

$$D(X_1)=D(X_2)=\cdots=D(X_n)=\sigma_x^2,$$

ამიტომ

$$D(\overline{X})=\frac{n\sigma_x^2}{n^2}=\frac{\sigma_x^2}{n},$$

საიდანაც

$$\sigma_{\overline{x}}=\sqrt{D(\overline{X})}=\frac{\sigma_x}{\sqrt{n}}.$$

დიდ რიცხვთა კანონის თანახმად $\overline{X}$ ახლოს არის a-თან, როცა n დიდია და ბუნებრივია ის გამოდგება a-ს შესაფასებლად.

შერჩევითი პარამეტრები, რომლებიც ახლოს არიან განაწილების კანონის რომელიმე პარამეტრთან, იქნებიან მათი შეფასებები.

შერჩევითი დისპერსიის გამოთვლა ხდება შემდეგი ფორმულით:

$$S^2=m_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

თუ $\overline{X}$-ის მათემატიკური ლოდინი ემთხვევა გენერალური ერთობლიობის მათემატიკურ ლოდინს—a-ს, შერჩევითი დისპერსიის S^2-ის მათემატიკური ლოდინი არ ემთხვევა გენერალური ერთობლიობის დისპერსიას σ_x^2-ს. თუ n საკმაოდ დიდია, მათ შორის განსხვავება არ იქნება საგრძნობი, ხოლო მცირე n-ის შემთხვევაში S^2-ის შესწორება იქნება საჭირო. შესწორება ხდება მისი $\frac{n}{n-1}$ მამრავლზე გამრავლებით. ამგვარად, შესწორებული დისპერსია იქნება

$$\overline{S}^2=\frac{n}{n-1}S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

საიდანაც

$$\overline{S}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

სავარჯიშო

1. ააგეთ ემპირიული განაწილების ფუნქციის გრაფიკი, თუ:

ა) $\left.\begin{matrix} x_i \\ n_i \end{matrix}\right\}\begin{matrix} 4, & 7, & 9, & 12 \\ 5, & 6, & 11, & 18. \end{matrix}$

ბ) $\left\{\begin{matrix} x_i \\ n_i \end{matrix}\right.\begin{matrix} 2, & 10, & 16, & 23 \\ 11, & 15, & 14, & 10. \end{matrix}$

2. ააგეთ სიხშირის და ფარდობითი სიხშირის პოლიგონები, თუ:

ა) $\begin{array}{l} x_i \\ n_i \end{array} \left\} \begin{array}{lllll} 1, & 4, & 7, & 10, & 15, \\ 8, & 7, & 15, & 18, & 12. \end{array} \right.$

ბ) $\begin{array}{l} x_i \\ n_i \end{array} \left\} \begin{array}{llll} 2, & 3, & 13, & 19, \\ 7, & 17, & 16, & 30. \end{array} \right.$

3. ააგეთ განაწილების სიხშირის და ფარდობითი სიხშირის ჰისტოგრამის გრაფიკი შემდეგი მონაცემების მიხედვით, სადაც პირველ სვეტში კერძო ინტერვალებია მოცემული, ხოლო მეორეში—კერძო ინტერვალში ვარიანტის სიხშირეთა ჯამი.

1.

3—6	7
6—9	11
9—12	20
12—15	12

2.

5—11	1
11—17	5
17—23	19
23—29	11
29—33	4

4. ვთქვათ, შერჩევის ელემენტებია: 1, 3, 5, 7, რომელთა შესაბამისი სიხშირეებია: 2, 5, 6, 7. იპოვეთ ფარდობითი სიხშირის განაწილება.

პას. $\dfrac{m}{n} \left\{ \begin{array}{llll} 1, & 3, & 5, & 7 \\ 0{,}1; & 0{,}25; & 0{,}3; & 0{,}35 \end{array} \right.$

5. შერჩევა მოცემულია სიხშირეთა განაწილების სახით:

$$m \left\{ \begin{array}{lllll} 2, & 3, & 6, & 9, & 11 \\ 10, & 8, & 12, & 11, & 9. \end{array} \right.$$

იპოვეთ ფარდობითი სიხშირის განაწილება.

პას. $\dfrac{m}{n} \left\{ \begin{array}{lllll} 2, & 3, & 6, & 9, & 11 \\ 0{,}2 & 0{,}16, & 0{,}24, & 0{,}22, & 0{,}18. \end{array} \right.$

6. შერჩევა მოცემულია სიხშირეთა განაწილების სახით:

$$m \left\{ \begin{array}{lll} 1, & 3, & 7, \\ 17, & 14, & 19. \end{array} \right.$$

მონახეთ ემპირიული განაწილების ფუნქცია.

პას.

$$S_n(x) = \begin{cases} 0, & \text{როცა } x \leqslant 1, \\ 0{,}34, & \text{როცა } 1 < x \leqslant 3 \\ 0{,}62, & \text{როცა } 3 < x \leqslant 7 \\ 1, & \text{როცა } x > 7. \end{cases}$$

7. შერჩევის განაწილებაა:

$$X \begin{cases} 2, & 4, & 5, \\ 0,2; & 2,5, & 9,3. \end{cases}$$

იპოვეთ ემპირიული განაწილების ფუნქცია.

პას.

$$S_n(x) = \begin{cases} 0, & \text{როცა} \quad x \leqslant 2 \\ 0,2, & \text{როცა} \quad 2 < x \leqslant 4 \\ 0,7, & \text{როცა} \quad 4 < x < 5 \\ 1, & \text{როცა} \quad x < 5. \end{cases}$$

8. შერჩევა მოცემულია სიხშირეთა შემდეგი განაწილებით:

$$m \begin{cases} 2,5 & 7,8, \\ 1,3, & 2,4. \end{cases}$$

იპოვეთ ემპირიული განაწილების ფუნქცია.

პას.

$$S_n(x) = \begin{cases} 0, & \text{როცა} \quad x \leqslant 2 \\ 0,1, & \text{როცა} \quad 2 < x \leqslant 5 \\ 0,4, & \text{როცა} \quad 5 < x \leqslant 7 \\ 0,6, & \text{როცა} \quad 7 < x \leqslant 8 \\ 1, & \text{როცა} \quad x > 8. \end{cases}$$

9. შერჩევა მოცემულია სიხშირეთა შემდეგი განაწილებით:

$$m \begin{cases} 4, & 7, & 8 \\ 5, & 2, & 3. \end{cases}$$

იპოვეთ ემპირიული განაწილების ფუნქცია.

პას.

$$S_n(x) = \begin{cases} 0, & \text{როცა} \quad x \leqslant 4 \\ 0,4, & \text{როცა} \quad 4 < x \leqslant 7 \\ 0,7, & \text{როცა} \quad 7 < x \leqslant 8 \\ 1, & \text{როცა} \quad x > 8. \end{cases}$$

10. შერჩევა მოცემულია სიხშირეთა შემდეგი განაწილებით:

$$m \begin{cases} 1, & 6, & 11 \\ 2, & 5, & 3. \end{cases}$$

იპოვეთ ემპირიული განაწილების ფუნქცია და ააგეთ მისი გრაფიკი.

პას.

$$S_n(x) = \begin{cases} 0, & \text{როცა} \quad x \leqslant 1 \\ 0,2, & \text{როცა} \quad 1 < x \leqslant 6 \\ 0,7, & \text{როცა} \quad 6 < x \leqslant 11 \\ 1, & \text{როცა} \quad x > 11 \end{cases}$$

11. შერჩევა მოცემულია სიხშირეთა შემდეგი განაწილებით:

$$m\begin{cases}1, & 5, & 7, & 10, & 12\\ 20, & 25, & 15, & 35, & 5.\end{cases}$$

ააგეთ სიხშირეთა პოლიგონი და იპოვეთ ფარდობითი სიხშირის განაწილება.

პას. $$\frac{m}{n}\begin{cases}1, & 5, & 7, & 10, & 12\\ 0{,}2; & 0{,}25; & 0{,}15; & 0{,}35; & 0{,}05.\end{cases}$$

12. მოცემულია ცხრილი:

2—5	9
5—8	10
8—11	25
11—14	6,

სადაც პირველ სვეტში კერძო ინტერვალებია მოცემული, ხოლო მეორეში—ვარიანტთა სიხშირის ჯამებია კერძო ინტერვალებში. ააგეთ სიხშირეთა ჰისტოგრამა და იპოვეთ ფარდობითი სიხშირის განაწილება:

13. ააგეთ სიხშირეთა პოლიგონი, თუ სიხშირეთა განაწილებაა

$$m\begin{cases}2, & 3, & 5, & 6\\ 10, & 15, & 5, & 20.\end{cases}$$

14. ააგეთ სიხშირეთა პოლიგონი, თუ სიხშირეთა განაწილებაა

$$m\begin{cases}1, & 5, & 9, & 13\\ 25, & 37, & 16, & 22.\end{cases}$$

15. ააგეთ ფარდობით სიხშირეთა პოლიგონი, თუ ის მოცემულია განაწილებით:

$$\frac{m}{n}\begin{cases}5, & 7, & 11, & 16\\ 0{,}1; & 0{,}3; & 0{,}4; & 0{,}2.\end{cases}$$

მითითება: აბსცისათა ღერძზე გადაზომეთ ვარიანტთა მნიშვნელობანი, ორდინატთა ღერძზე შესაბამისი ფარდობითი სიხშირეები და შეაერთეთ მიღებული წერტილები.

16. ააგეთ სიხშირეთა ჰისტოგრამა, თუ მოცემულია $n=100$ მოცულობის შერჩევის შემდეგი განაწილება:

ინტერვალის ნომერი	კერძო ინტერვალი	ინტერვალის სიხშირეთა ჯამი	სიხშირეთა სიმკვრივე
I	1—5	10	2,5
2	5—9	20	5
3	9—13	50	12,5
4	13—17	12	3
5	17—21	8	2

17. ააგეთ ფარდობით სიხშირეთა ჰისტოგრამა, თუ შერჩევის განაწილებაა:

ინტერვალის ნომერი	კერძო ინტერვ.	სიხშირეთა ჯამი
1	10—15	2
2	15—20	4
3	20—25	8
4	25—30	4
5	30—35	2

მ ი თ ი თ ე ბ ა: პირველ რიგში მონახეთ ფარდობითი სიხშირეები და მათი შესაბამისი სიმკვრივეები ცალკეული ინტერვალებისათვის.

18. მონახეთ შერჩეული ერთობლიობის ისეთი მინიმალური მოცულობა, რომლისათვისაც 0,975-ის ტოლი საიმედოობით გენერალური ერთობლიობის a საშუალოს შეფასების სიზუსტე შერჩევითი საშუალოთი იქნება $\Delta=0,3$, თუ ნორმალურად განაწილებული გენერალური ერთობლიობის საშუალო კვადრატული გადახრა $\sigma=1,2$.

პას. $\Delta=t\cdot\dfrac{\sigma}{\sqrt{n}}$, სადაც $n=\dfrac{t^2\sigma^2}{\Delta^2}=81$.

19. მონახეთ შერჩევითი ერთობლიობის ისეთი მინიმალური მოცულობა, რომლისთვისაც 0,925 ტოლი საიმედოობით გენერალური ერთობლიობის a საშუალოს შეფასების სიზუსტე შერჩევითი საშუალოთი იქნება $\Delta=0,2$, თუ ნორმალურად განაწილებული გენერალური ერთობლიობის საშუალო კვადრატული გადახრა $\sigma=1,5$.

პას. $n=179$.

თავი III

განაწილების პარამეტრების სტატისტიკური შეფასება

§. 1 საკითხის დასმა

ვთქვათ გვაქვს რაიმე გენერალური ერთობლიობა, რომლის მოცულობა შესაძლებელია უსასრულოც იყოს. აქედან სავსებით შემთხვევით ვიღებთ შერჩევით ერთობლიობას. მას ვსწავლობთ სტატისტიკური

მეთოდებით და ამით წარმოდგენას ვიქმნით გენერალურ ერთობლიობასა და მის მახასიათებლებზე.

მოცემული შერჩევითი განაწილების კანონის შეფასება, რომელსაც პარამეტრიზაცია ეწოდება, მათემატიკური სტატისტიკის არსებით პრობლემას წარმოადგენს. როცა გავიგებთ შესასწავლი სიდიდის განაწილების კანონს, მაშინვე შეგვიძლია ანალიზის მომარჯვებით პრაქტიკულად დასმული ამოცანების ამოხსნა; სახელდობრ, შესაძლებელია მასობრივი პროცესების შედეგთა შედარება და წინასწარმეტყველება.

პრაქტიკული ამოცანების დროს თეორიული განაწილების ფორმა შეიძლება ცნობილი იყოს. მაგალითად, თუ განვიხილავთ გაზომვათა შედეგად მიღებულ ცდომილებებს, მაშინ ისინი, ლიაპუნოვის თეორემის თანახმად, ნორმალური კანონით იქნებიან განაწილებულნი. თუ განვიხილავთ რაიმე ქვემეხიდან რომელიმე მიზანში სროლის შედეგად მოხვედრათა განაწილებას, ისინი ემორჩილებიან ორგანზომილებიან ნორმალური განაწილების კანონს. უამრავი ფიზიკური და ბიოლოგიური სქემები ემორჩილებიან პუასონის ან მაჩვენებლიანი განაწილების კანონს და ა. შ. ასეთ შემთხვევებში შეგვიძლია ვიგულისხმოთ, რომ თეორიული განაწილების კანონი ეკუთვნის რომელიმე ოჯახს, რომელიც ერთ ან რამდენიმე პარამეტრზეა დამოკიდებული. სახელდობრ, ნორმალური კანონისათვის ზუსტად რომ იყოს ცნობილი a და σ პარამეტრების მნიშვნელობა, ხოლო პუასონის კანონისათვის a პარამეტრის მნიშვნელობა, მაშინ განაწილების კანონი ასეთ შემთხვევაში ზუსტად იქნებოდა განსაზღვრული. როგორც ვხედავთ, შესასწავლი სიდიდის განაწილების კანონის დადგენა დაიყვანება განაწილების უცნობი პარამეტრების მნიშვნელობათა დადგენაზე, ე. ი. პარამეტრიზაციამდე. a მათემატიკური ლოდინის და σ საშუალო კვადრატული გადახრის მნიშვნელობათა ცოდნა, ზემოთ ჩამოყალიბებული ამოცანების შემთხვევაში, რომლებიც ნორმალურ კანონთან არიან დაკავშირებული, შესაძლებლობას იძლევა ამოვხსნათ მრავალი საინჟინრო და ეკონომიური ამოცანები ზუსტი დამუშავებისა და ზუსტი გაზომვების სფეროში.

შერჩევითი ერთობლიობის საშუალო არითმეტიკული დიდ რიცხვთა კანონის ძალით წარმოადგენს ნორმალური განაწილების კანონის a მათემატიკური ლოდინის შეფასებას. ასევე შერჩევითი ერთობლიობის s^2 დისპერსია არის გენერალური ერთობლიობის σ^2 დისპერსიის შეფასება. სწორედ ამ შეფასებათა საშუალებით, სათანადო მიახლოებით, შესაძლებელია მოცემული ამოცანისათვის ალბათობათა თეორიული განაწილების ფუნქციის მონახვა.

ვთქვათ, უცნობი განაწილება k პარამეტრზეა დამოკიდებული, ხოლო შერჩევის მოცულობა n-ის ტოლია. დავუშვათ, შერჩევითი ერთობლიობის ელემენტებია $x_1, x_2, \ldots, x_n$, რომლებიც ურთიერთდამოუკიდებელ შემთხვევით სიდიდეებს წარმოადგენენ, ხოლო უცნობი პარამეტრები აღვნიშნოთ $\theta_1, \theta_2, \ldots, \theta_k$-თი. ამ უცნობი პარამეტრების შეფასება ნიშნავს $x_1, x_2, \ldots, x_n$ შერჩევითი ერთობლიობის ელემენტების ისეთი $\overline{\theta}(x_1, x_2, \ldots, x_n), \ldots, \overline{\theta}_k(x_1, x_2, \ldots, x_n)$ ფუნქციების მონახვას, რომლებიც დაახლოებით საძებნი პარამეტრების მნიშვნელობებს გამოხატავენ. ამასთან დარწმუნებული უნდა ვიყოთ, რომ n-ის საკმაოდ დიდი მნიშვნელობისათვის მაინც $\overline{\theta}_1(x_1, x_2, \ldots, x_n), \ldots, \overline{\theta}_k(x_1, x_2, \ldots, x_n)$ შემთხვევით სიდიდეთა განაწილება მჭიდროდ კონცენტრირებულია $\theta_1, \theta_2, \ldots, \theta_k$ უცნობი პარამეტრების მახლობლობაში და, ამგვარად, რაც შეიძლება მცირე შეცდომის დაშვებით შეგვიძლია შემდეგი მიახლოებითი ტოლობების დაწერა:

$$\begin{aligned}
\overline{\theta}_1(x_1, x_2, \ldots, x_n) &\cong \theta_1, \\
\overline{\theta}_2(x_1, x_2, \ldots, x_n) &\cong \theta_2, \\
\cdot \quad \cdot \quad \cdot \quad &\cdot \quad \cdot \\
\overline{\theta}_k(x_1, x_2, \ldots, x_n) &\cong \theta_k.
\end{aligned}$$

სტატისტიკურ შეფასებათა თავმოყრა შესაფასებელი პარამეტრების ირგვლივ ჩატარებული შეფასების საიმედოობას და სიზუსტეს ახასიათებენ.

ცხადია, რომ ჩვენთვის საინტერესო ოჯახის ყოველი $F_x(x, \theta_1, \theta_2, \ldots, \theta_n)$ განაწილების ფუნქციისათვის ყოველი $\overline{\theta}_i(x_1, x_2, \ldots, x_n)$ შესაძლო სტატისტიკური შეფასების განაწილების კანონი სავსებით განსაზღვრულია. სინამდვილეში, $\overline{\theta}_i(x_1, x_2, \ldots, x_n)$ ფუნქციები ქმნიან შემთხვევით სიდიდეთა სისტემას, რომლებიც უშუალოდ განისაზღვრებიან $x_1, x_2, \ldots x_n$ შერჩევის შესაძლო ანდა შემთხვევითი შედეგებით.

თავის მხრივ, ამ უკანასკნელთა განაწილების კანონი მთლიანად განისაზღვრება X გამოსავალი შემთხვევითი სიდიდის განაწილების კანონით, რომლის განაწილების ფუნქციაა $F_x(x, \theta_1, \theta_2, \ldots, \theta_k)$. ეს იმიტომ, რომ, როგორც ვიცით, $x_1, x_2, \ldots, x_n$ არიან ურთიერთდამოუკიდებელნი და წარმოადგენენ X შემთხვევით სიდიდეზე დაკვირვების შედეგებს, რომლებიც ცალ-ცალკე განაწილებული არიან $F_x(x, \theta_1, \theta_2, \ldots, \theta_k)$ კანონით.

აგრეთვე, ცხადია, რომ თითოეული $\theta_i (i=1, 2, \ldots, k)$ შეფასების განაწილების კანონი დამოკიდებულია $\theta_1, \theta_2, \ldots, \theta_k$ პარამეტრების მნიშვნელობაზე, რომლებზედაც დამოკიდებულია $F_x(x, \theta_1, \theta_2, \ldots, \theta_k)$

ფუნქცია. კარგად შერჩეული შეფასებისათვის ეს დამოკიდებულება ისეთი უნდა იყოს, რომ შეძლებისდაგვარად გააადვილოს თვითონ პარამეტრების მიახლოებით განსაზღვრის ამოცანა.

ამა თუ იმ სტატისტიკური შეფასების შერჩევითი ერთობლიობის განაწილების კანონის ცოდნა შესაძლებლობას გვაძლევს გამოვთვალოთ შესაფასებელი პარამეტრიდან მისი გადახრის ალბათობა, მათემატიკური ლოდინი და დისპერსია. ხოლო ამ უკანასკნელთა დახმარებით შესაძლებელია სტატისტიკური შეფასების მიმართ მოთხოვნილებანი უფრო ზუსტად ჩამოყალიბდეს.

პირველი მოთხოვნილება, რომელსაც უნდა აკმაყოფილებდეს რაციონალურად შერჩეული $\overline{\theta}(x_1, x_2, \ldots, x_n)$ შეფასება კონკრეტული θ პარამეტრის მიმართ, მდგომარეობს იმაში, რომ შერჩევის n მოცულობის უსაზღვროდ ზრდის დროს შეფასება ალბათობით იკრიბებოდეს შესაფასებელი პარამეტრისაკენ.

ალბათობით კრებადობის არსი შემდეგში მდგომარეობს: ვთქვათ, მოცემული გვაქვს შემთხვევით სიდიდეთა

$$z_1, z_2, \ldots, z_n. \tag{1.1}$$

მიმდევრობა. ვიგულისხმოთ, რომ t_n შემთხვევითი სიდიდეები წარმოადგენენ (1.1) მიმდევრობის პირველი n წევრის რაღაც მოცემულ სიმეტრიულ ფუნქციას:

$$t_n = f_n(z_1, z_2, \ldots, z_n).$$

თუ არსებობს $a_1, a_2, \ldots, a_n, \ldots$ მუდმივთა ისეთი მიმდევრობა, რომ ნებისმიერი $\varepsilon > 0$-თვის

$$\lim_{n \to \infty} \mathrm{P}\{\,|\,t_n - a_n\,| < \varepsilon\,\} = 1, \tag{1.2}$$

მაშინ ვიტყვით, რომ (1.1) მიმდევრობა f_n მოცემული ფუნქციით ემორჩილება დიდ რიცხვთა კანონს.

თუ (1.2) თანაფარდობაში ყველა a_n უდრის ერთსა და იმავე a სიდიდეს, მაშინ ამბობენ, რომ z_n შემთხვევითი სიდიდეები ალბათობით იკრიბებიან a-კენ. ამ ტერმინებში (1.2) ნიშნავს, რომ $t_n - a_n$ ალბათობით ნულისკენ იკრიბება.

მაშასადამე, თუ θ შესაფასებელი პარამეტრია და $\overline{\theta}(x_1, x_2, \ldots, x_n)$ მისი მოსალოდნელი შეფასებაა, მაშინ ეს უკანასკნელი იქნება θ-ს ნამდვილი შეფასება, თუკი ალბათობით

$$\overline{\theta}(x_1, x_2, \ldots x_n) \underset{n \to \infty}{\longrightarrow} \theta.$$

ანდა, სხვანაირად რომ ვთქვათ, ნებისმიერი $\varepsilon > 0$-თვის

$$P\{|\overline{\theta}(x_1, x_2, \ldots, x_n)-\theta| < \varepsilon\} \underset{n\to\infty}{\to} 1. \tag{1 3}$$

ე. ი. ალბათობა იმისა, რომ $\overline{\theta}$ სტატისტიკურ შეფასებასა და შესაფასებელ θ პარამეტრს შორის სხვაობის აბსოლუტური მნიშვნელობა ნაკლები იქნება რაგინდ მცირე $\varepsilon>0$ სიდიდეზე, რაგინდ ახლოს არის ერთთან, როდესაც დაკვირვებათა რიცხვი n საკმაოდ დიდია.

თუ საშუალო არითმეტიკულს განვიხილავთ როგორც მათემატიკური ლოდინის სტატისტიკურ შეფასებას, ვნახავთ, რომ საშუალო არითმეტიკული ალბათობით იკრიბება მათემატიკური ლოდინისაკენ. მართლაც, თუ განვიხილავთ ერთნაირად განაწილებულ შემთხვევით სიდიდეებს $X_1, X_2, \ldots, X_n$, რომელთაც ერთი და იგივე მათემატიკური ლოდინი აქვთ, ჩებიშევის თეორემის თანახმად დავწერთ:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i-a\right| < \varepsilon\right\} = 1.$$

აქ, როგორც ვხედავთ, (1.3) პირობა სრულდება და, მაშასადამე, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_1$ არის a მათემატიკური ლოდინის სტატისტიკური შეფასება, რომელიც ალბათობით კრებადია a-საკენ.

ის შეფასებანი, რომლებიც (1.3) პირობას აკმაყოფილებენ, იწოდებიან ძალდებულ შეფასებებად.

სტატისტიკური შეფასების ძალდებულება იმ შემთხვევაში, როდესაც შერჩევის მოცულობა საკმაოდ დიდია, უზრუნველყოფს მის პრაქტიკულ სიახლოვეს შესაფასებელ პარამეტრთან. სამწუხაროდ, n-ის მცირე მნიშვნელობისათვის, იმ შემთხვევაში, როდესაც შეფასება ძალდებულია, არავითარი დასკვნის გაკეთება არ შეიძლება მისი სათანადო პარამეტრის მიახლოებითი მნიშვნელობის განსასაზღვრავად გამოყენების შესახებ.

მეორე მოთხოვნა, რომელიც ედება სტატისტიკურ შეფასებას, არის მისი *გადაუადგილებლობა*, რაც ნიშნავს მასში სისტემატური ცდომილების არარსებობას.

ვიტყვით, რომ $\overline{\theta}(x_1, x_2, \ldots, x_n)$ არის θ-ს გადაუადგილებელი შეფასება, თუკი

$$E[\overline{\theta}(x_1, x_2, \ldots, x_n)] = \theta, \tag{1.4}$$

ე. ი. გადაუადგილებელი შეფასების მათემატიკური ლოდინი შესაფასებელი პარამეტრის ტოლი უნდა იყოს.

შეფასებას ეწოდება დადებითად გადაადგილებული, თუ

$$E[\overline{\theta}(x_1, x_2, \ldots, x_n)] > \theta,$$

ხოლო შეფასებას ეწოდება უარყოფითად გადაადგილებული, თუ

$$E[\overline{\theta}(x_1, x_2, \ldots, x_n)] < \theta.$$

ამგვარად, გადაუადგილებელი შეფასების გაბნევის ცენტრი ემთხვევა შესაფასებელი პარამეტრის მნიშვნელობას. საზოგადოდ, ყოველი ძალდებული შეფასება როდი იქნება გადაუადგილებელი; შესაძლებელია მისი მათემატიკური ლოდინი არ ემთხვეოდეს θ-ს, მხოლოდ n-ის გადიდებით უსაზღვროდ უახლოვდებოდეს მას. n-ის სასრულო მნიშვნელობისათვის ძალდებული შეფასება შესაძლებელია გადაადგილებული იყოს ხან ერთ, ხან მეორე მხარეს.

შეფასების გადაუადგილებლობის მოთხოვნა განსაკუთრებით მნიშვნელოვანია, როცა შერჩევის მოცულობა n არ არის დიდი. ამ შემთხვევაში გვიხდება სხვადასხვა პატარა სერიების დროს მიღებულ დაკვირვებათა შედეგების გასაშუალოება და მეტად მნიშვნელოვანია დარწმუნებული ვიყოთ იმაში, რომ სისტემატურ ცდომილობებს არა აქვს ადგილი.

გაფანტულობის ცენტრისათვის გადაუადგილებელ შეფასებას წარმოადგენს საშუალო არითმეტიკული.

მართლაც, თუ გვაქვს ერთნაირად განაწილებული შემთხვევითი სიდიდეები $X_1, X_2, \ldots, X_n$, რომელთაც ერთი და იგივე a-ს ტოლი მათემატიკური ლოდინი აქვთ, მაშინ

$$E(\overline{X}) = E\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{na}{n} = a.$$

გადაადგილებული შეფასების მაგალითად შეიძლება დავასახელოთ შერჩევითი ერთობლიობის s^2 დისპერსია.

როგორც ვიცით,

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2.$$

ავარჩიოთ „ყალბ ნულად“ $E(X) = a$, ვიგულისხმოთ, რომ შემთხვევითი სიდიდეები ერთნაირად არიან განაწილებული, მაშინ,

$$s^2=\frac{1}{n}\sum_{i=1}^{n}[x_i-a-(\overline{X}-a)]^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2-\frac{1}{n}2(\overline{X}-a)\sum_{i=1}^{n}(x_i-a)+\frac{1}{n}\sum_{i=1}^{n}(\overline{X}-a)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2-\frac{1}{n}2(\overline{X}-a)\sum_{i=1}^{n}(x_i-a)+(\overline{X}-a)^2.$$

ახლა ვიპოვოთ s^2-ის მათემატიკური ლოდინი, გვექნება:

$$Es^2=\frac{1}{n}\sum_{i=1}^{n}E(x_i-a)^2-\frac{1}{n}2(\overline{X}-a)\sum_{i=1}^{n}E(x_i-a)+E(\overline{X}-a)^2.$$

მაგრამ რადგან შემთხვევით სიდიდეებს ერთი და იგივე დისპერსია და მათემატიკური ლოდინი აქვთ, დავწერთ:

$$Es^2=\frac{1}{n}n\sigma^2-2\cdot\frac{1}{n}(\overline{X}-a)nE(\overline{X}-a)+E(\overline{X}-a)^2=$$

$$=\sigma^2-E(\overline{X}-a)^2.$$

აგრეთვე,

$$E(\overline{X}-a)^2=E\left(\frac{x_1-a}{n}+\frac{x_2-a}{n}+\dots+\frac{x_n-a}{n}\right)^2=$$

$$=\frac{1}{n^2}[E(x_1-a)^2+E(x_2-a)^2+\dots+E(x_n-a)^2]=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

ამიტომ Es^2-თვის მივიღებთ:

$$Es^2=\sigma^2-\frac{\sigma^2}{n}=\sigma^2\left(1-\frac{1}{n}\right)=\sigma^2\left(\frac{n-1}{n}\right). \tag{1.5}$$

ამგვარად, როგორც ვხედავთ, ემპირიული დისპერსია ყოფილა გადაადგილებული და გადაადგილების ზომა ყოფილა $\frac{n-1}{n}$.

როცა ვაფასებთ σ^2-ს s^2-ის საშუალებით, რადგანაც უარყოფითი გადაადგილება გვაქვს, ჩვენ ვუშვებთ სისტემატურ ცდომილებას, რომელიც $\frac{\sigma^2}{n}$-ის ტოლია, რაც, რა თქმა უნდა, n-ის დიდი მნიშვნელობისათვის მცირე სიდიდეს წარმოადგენს და მისი უგულებელყოფა შეიძლება.

$s^2 = m_2$ ალბათობით კრებადია σ^2-კენ, ამიტომ, ცხადია, ის იქნება ძალდებული შეფასება.

აგრეთვე s არის σ-ს ძალდებული, მაგრამ—გადაადგილებული შეფასება. n-ის მცირე მნიშვნელობისათვის ეს გადაადგილება იქნება საგრძნობი.

ხანდახან ძალდებული, გადაუადგილებელი შეფასების აგება შეიძლება სხვადასხვანაირად. ასე მაგალითად, ნორმალური კანონის შემთხვევაში გაფანტულობის ცენტრის შესაფასებლად $\overline{X}$ საშუალო არითმეტიკულთან ერთად შეიძლება განვიხილოთ ემპირიული m_e მედიანა. ეს უკანასკნელიც იქნება როგორც ძალდებული, ისე გადაუადგილებელი შეფასება გაფანტულობის ცენტრისათვის.

ვთქვათ, მოცემულია ორი $\overline{\theta}_1$ და $\overline{\theta}_2$ გადაუადგილებელი შეფასება θ პარამეტრისათვის. მათ შორის უპირატესობა უნდა მივანიჭოთ იმას, რომლისთვისაც ალბათობის განაწილების მასა უფრო კონცენტრირებულია θ-ს ახლოს.

გაფანტულობის საზომად ჩვეულებრივად დისპერსია მიიღება

$$D(\overline{\theta}) = \sigma^2 = E(\overline{\theta} - \theta)^2,$$

სადაც, რადგანაც შეფასება გადაუადგილებელია,

$$E(\overline{\theta}) = \theta.$$

პრაქტიკულად θ პარამეტრისათვის ორ $\overline{\theta}_1$ და $\overline{\theta}_2$ შეფასებებს შორის უკეთესია ის, რომლის დისპერსია უფრო მცირეა. მაშინ შეიძლება დაისვას საკითხი ისეთი გადაუადგილებელი ძალდებული შეფასების მონახვის შესახებ, რომელსაც ყველაზე პატარა დისპერსია აქვს. ყველაზე ზოგად შემთხვევაში შეიძლება ვაჩვენოთ, რომ n მოცულობის მქონე შერჩევის საფუძველზე შეფასებული დისპერსია არა ნაკლებია ზოგიერთ ქვედა საზღვარზე. ერთი პარამეტრის შემთხვევაში ნებისმიერი გადაუადგილებელი შეფასების დისპერსიისათვის ადგილი აქვს შემდეგ უტოლობას:

$$\sigma^2_{\overline{\theta}} \geqq \frac{1}{n\,\psi(\theta)}, \tag{1.6}$$

სადაც $\psi(\theta)$ დადებითი სიდიდეა, რომელიც შემდეგნაირად გამოისახება:

$$\psi(\theta) = E\left\{\left[\frac{\partial \ln \varphi(x, \theta)}{\partial \theta}\right]^2\right\}. \tag{1.7}$$

ანდა მისი გამოსახვა ეკვივალენტური სიდიდით ასედაც შეიძლება

$$\psi(\theta) = -E\left[\frac{\partial^2 \ln \varphi(x, \theta)}{\partial \theta^2}\right], \tag{1.8}$$

სადაც n შერჩევის მოცულობაა, $\varphi(x, \theta)$ არის უწყვეტი განაწილების შემთხვევაში შემთხვევითი სიდიდის სიმკვრივე, დისკრეტული განაწილების შემთხვევაში

$$\varphi(x, \theta) = P\{X = x;\ \theta\}.$$

(1.7), (1.8) გამოსახულებებში დგანან იმ სიდიდეთა მათემატიკური ლოდინები, რომლებსაც მივიღებთ, თუ

$$\left[\frac{\partial \ln \varphi(x, \theta)}{\partial\theta}\right]^2 \text{ და } \frac{\partial^2 \ln \varphi(x,\theta)}{\partial\theta^2}$$

ფუნქციებში პირველ არგუმენტს შევცვლით X შემთხვევითი სიდიდით.

თუ არსებობს ისეთი გადაუადგილებელი შეფასება $\overline{\theta}$, რომლისთვისაც $\sigma^2_{\overline{\theta}}$ დისპერსია აღწევს (1.6) უტოლობის მარჯვენა მხარის ტოლ ქვედა საზღვარს, მაშინ მას ეწოდება ე ფ ე ქ ტ უ რ ი შეფასება.

ეფექტური შეფასება ყოველთვის იქნება ძალდებული და ითვლება საუკეთესო შეფასებად.

ეფექტური შეფასების დისპერსია აღვნიშნოთ $\sigma^2_{\theta э}$-თი, სხვა რომელიმე გადაუადგილებელი შეფასების დისპერსია იყოს $\sigma^2_{\overline{\theta}}$, მაშინ ამ უკანასკნელი შეფასების შედარებითი ეფექტიანობა, ბუნებრივია, გაიზომება

$$\psi(\overline{\theta}) = \frac{\sigma^2_{\theta э}}{\sigma^2_{\overline{\theta}}} \tag{1.9}$$

დისპერსიათა ფარდობით.

როგორც ვხედავთ, ნებისმიერი გადაუადგილებელი შეფასების შედარებითი ეფექტიანობა $\psi(\overline{\theta})$ არ აღემატება ერთს.

თუ განვიხილავთ შერჩევის ნორმალური გენერალური ერთობლიობიდან, a მათემატიკური ლოდინის ეფექტური შეფასება იქნება საშუალო არითმეტიკული $\overline{x}$, ხოლო m_e მედიანის შედარებითი ეფექტიანობა საკმაოდ დიდი მოცულობის შერჩევისათვის მიახლოებით შემდეგნაირად გამოისახება

$$\frac{\sigma^2_{\overline{x}}}{\sigma^2_{m_e}} = \frac{2}{\pi} = 0,6366.$$

პრაქტიკულად ეს იმას ნიშნავს, რომ განაწილების a ცენტრი n დაკვირვების დროს m_e მედიანით იმავე სიზუსტით განისაზღვრება, როგორითაც $\frac{2}{\pi}\, n = 0,6366\, n$ დაკვირვებისას $\overline{x}$ საშუალო არითმეტიკულით.

ზოგიერთი შემთხვევისათვის მოვნახოთ სხვადასხვა შეფასება:

1. ვთქვათ, გვაქვს ბურთების დაბრუნების სქემა. შერჩევის მოცულობა იყოს n, ავიღოთ ასეთ შერჩევათა r რაოდენობა. ცალკეული ცდის დროს ჩვენთვის სასურველი A ხდომილობის ალბათობა იყოს p, ხოლო ცდათა სერიის დროს A ხდომილობის სიხშირედ მივიღოთ m, რომელიც შემთხვევით სიდიდეს წარმოადგენს.

ცხადია, რომ m ბინომიალური განაწილების კანონს ემორჩილება. იგულისხმება, რომ შერჩევითი ერთობლიობის მოცულობა n ცნობილია. ვიპოვოთ p პარამეტრის ეფექტური შეფასება ბინომიალური კანონისათვის. ე. ი., თუ ცნობილია n მოცულობის r შერჩევის შედეგი, საჭიროა ვიპოვოთ ცალკეული ცდის დროს A ხდომილობის p ალბათობის ეფექტური შეფასება.

ვთქვათ, თითოეული n მოცულობის r შერჩევიდან i-ური შერჩევის დროს მივიღეთ $\frac{k_1}{n}$ ფარდობითი სიხშირე, რომელიც შეიძლება p-ს მიახლოებით შეფასებად ჩავთვალოთ.

$\frac{k_1}{n}, \frac{k_2}{n}, \ldots, \frac{k_n}{n}$ ფარდობითი სიხშირეებისაგან დამოუკიდებლად r უნდა გამოვიყენოთ p პარამეტრის უფრო ზუსტად შესაფასებლად.

(1.9) ფორმულის დახმარებით დავწეროთ:

$$r\,\psi(\theta)=-rE\left[\frac{\partial^2 \ln \varphi(x,\theta)}{\partial\theta^2}\right]=-rE\left[\frac{\partial^2 \ln \varphi(x,p)}{\partial p^2}\right]=$$

$$=-rE\left[\frac{\partial^2 \ln C_n^m p^m q^{n-m}}{\partial p^2}\right]=-rE\left\{\frac{\partial^2}{\partial p^2}\left[\ln C_n^m+m\ln p+\right.\right.$$

$$\left.\left.+(n-m)\ln(1-p)\right]\right\}=-rE\left[\frac{\partial}{\partial p}\left(\frac{m}{p}-\frac{n-m}{1-p}\right)\right]=$$

$$=-rE\left[-\frac{m}{p^2}-\frac{n-m}{(1-p)^2}\right]=rE\left[\frac{m}{p^2}+\frac{n-m}{q^2}\right]=$$

$$=r\left[\frac{E(m)}{p^2}+\frac{E(n-m)}{q^2}\right].$$

მაგრამ, როგორც ცნობილია, ბინომიალური განაწილების დროს

$$E(m)=np.$$

ამიტომ

$$r\,\psi(\theta)=r\left[\frac{np}{p^2}+\frac{n-np}{q^2}\right]=r\left(\frac{n}{p}+\frac{n}{q}\right)=$$

$$=r\,\frac{n(p+q)}{pq}=\frac{nr}{pq}.$$

ამგვარად, ბინომიალური კანონის დროს (1.7) ის ძალით გვექნება

$$\sigma_{\frac{2}{p}} \geq \frac{pq}{nr}. \tag{1.10}$$

ე. ი. p პარამეტრის შესაძლო სტატისტიკური შეფასებისათვის, როცა n-ის მნიშვნელობა ცნობილია, დისპერსიის ქვედა საზღვარია.

$$\frac{pq}{rn}.$$

შევამოწმოთ, არის თუ არა p-თვის ეფექტური გადაუადგილებელი შეფასება

$$t = \frac{\overline{x}}{n} = \frac{m_1 + m_2 + \cdots + m_r}{rn}.$$

სტატისტიკური შეფასება, სადაც $m_1, m_2 \ldots, m_r$ არიან A ხდომილობის გამოჩენათა რიცხვი r სერიაში შესაბამისად (იგულისხმება, რომ თითოეული შერჩევის მოცულობაა n და საქმე გვაქვს ბურთების დაბრუნების სქემასთან).

ვიპოვოთ t-ს მათემატიკური ლოდინი და დისპერსია:

$$E(t) = E\left(\frac{m_1 + m_2 + \cdots + m_r}{rn}\right) = \frac{1}{rn}[E(m_1) + E(m)_2 + \cdots + E(m_r)].$$

მაგრამ, რადგანაც $m_1, m_2, \ldots, m_r$ ერთნაირად არიან განაწილებული, მათ ყველას ერთი და იგივე მათემატიკური ლოდინი აქვთ და ის np-ს ტოლია, ამიტომ დავწერთ

$$E(t) = \frac{1}{rn} \cdot rnp = p.$$

ეს იმას გვეუბნება, რომ t-ს სტატისტიკური შეფასება არის p-ს გადაუადგილებელი შეფასება.

$$D(t) = \sigma_t^2 = \frac{1}{r^2 n^2} D\left(\sum_{i=1}^{r} m_i\right),$$

მაგრამ, რადგანაც $m_1, m_2, \ldots, m_r$ ურთიერთდამოუკიდებელი არიან და თითოეულის დისპერსია უდრის npq-ს, ამიტომ გვექნება:

$$D(t) = \sigma_t^2 = \frac{1}{r^2 n^2} rnpq = \frac{pq}{rn}.$$

როგორც ვხედავთ, მივიღეთ იგივე გამოსახულება, რომელიც დგას (1.10) უტოლობის მარჯვენა მხარეში. ეს კი იმას ნიშნავს, რომ

$t=\frac{\overline{x}}{n}$ არის p პარამეტრის ეფექტური გადაუადგილებელი შეფასება, რომლისაც შერჩევის n მოცულობა ცნობილია.

2. ვთქვათ, X შემთხვევითი სიდიდე ემორჩილება პუასონის კანონს. შევამოწმოთ—პირველი საწყისი მომენტი $a_1=\overline{x}$ არის თუ არა გადაუადგილებელი ეფექტური შეფასება პუასონის კანონის a პარამეტრისათვის.

(1. 9) ფორმულის ძალით დავწერთ

$$\sigma_{\theta}^2 \geqq \frac{1}{n\,\psi(\theta)}=\frac{1}{-nE\left[\frac{\partial^2 \ln \varphi(x,\ \theta)}{\partial\theta^2}\right]}=-\frac{1}{n\,E\left[\frac{\partial^2}{\partial a^2}\left(\ln\frac{a^m e^{-a}}{m!}\right)\right]}=$$

$$=-\frac{1}{nE\left[\frac{\partial^2}{\partial a^2}(m\ln a-a-\ln m!)\right]}=-\frac{1}{nE\left[-\frac{m}{a^2}\right]}=\frac{a^2}{nE(m)}.$$

მაგრამ, როგორც ვიცით, პუასონის კანონის შემთხვევაში $E(m)=a$, ამიტომ

$$\sigma_{\theta}^2 \geqq \frac{1}{n\psi(\theta)}=\frac{a}{n}. \tag{1.11}$$

ამგვარად $\frac{a}{n}$ წარმოადგენს a პარამეტრის შესაძლო სტატისტიკური შეფასებისათვის დისპერსიის ქვედა საზღვარს.

ახლა ვიპოვოთ $a_1=\overline{x}$-ის დისპერსია:

$$D(a_1)=D(\overline{x})=D\left(\frac{X_1+X_2+\dots+X_n}{n}\right)=$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\,nD(X_i)=\frac{1}{n}\sigma_x^2.$$

მაგრამ, როგორც ცნობილია, პუასონის კანონის შემთხვევაში $\sigma_x^2=a$. ამიტომ დავწერთ: $D(a_1)=\frac{a}{n}$.

მივიღეთ იგივე გამოსახულება, რომელიც დგას (1. 11) მარჯვენა მხარეზე; ეს კი იმას ნიშნავს, რომ $a_1=\overline{x}$ არის პუასონის კანონის შემთხვევაში a პარამეტრის გადაუადგილებელი ეფექტური შეფასება.

§ 2. საშუალოსა და დისპერსიის შეფასება

პირველ რიგში განვიხილოთ ის შემთხვევა, როდესაც გენერალური ერთობლიობის მოცულობა N სასრულია. შერჩევითი ერთობლიობის მოცულობა იყოს n, გენერალური ერთობლიობის ელემენტები აღვნიშნოთ სათანადოდ $x_1, x_2, \ldots, x_N$-ით. ცალკეული ელემენტის შერჩევაში მოხვედრა ერთნაირად მოსალოდნელია. სიმარტივისათვის ვიგულისხმოთ, რომ გენერალური ერთობლიობის ელემენტები ერთმანეთისაგან განსხვავებული არიან.

გენერალური ერთობლიობის საშუალო აღვნიშნოთ a-თი. ასე რომ,

$$a=\frac{1}{N}\sum_{i=1}^{N}x_i,$$

ხოლო გენერალური ერთობლიობის დისპერსია

$$\mu_2=\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-a)^2.$$

შერჩევითი ერთობლიობის საშუალო იყოს $\overline{x}$, დისპერსია — s^2.

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

ვნახოთ, თუ როგორ ზუსტად აფასებს $\overline{x}$ გენერალური ერთობლიობის a საშუალოს. იგულისხმება, რომ ნებისმიერი C_n^N ჯუფთებას, რომლის შედგენა შეიძლება N ელემენტიდან n-n-ად, აქვს შერჩევაში მოხვედრის ერთნაირი შესაძლებლობა. ასევე $x_1, x_2, \ldots, x_N$ ელემენტები ტოლი p_1 ალბათობით შეიძლება მოხვდნენ შერჩევაში. p_1 ალბათობა შემდეგნაირად გამოისახება:

$$p_1=\frac{n}{N}.$$

მართლაც, ისეთი ჯუფთებათა რიცხვი n ელემენტისაგან, რომელიც შეიცავს რომელიმე ფიქსირებულ ელემენტს, იქნება C_{N-1}^{n-1}, რადგანაც, თუ ამ ელემენტს მოვაცილებთ ყველა n-ელემენტიან ჯუფთებას, მივიღებთ რომელიმე $(n-1)$-ელემენტიან ჯუფთებას, რომელიც შედგენილია გენერალური ერთობლიობის დანარჩენი $N-1$ ელემენტებისაგან; ამიტომ

$$p_1 = \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{(N-1)!}{(n-1)!\,(N-n)!} : \frac{N!}{n!\,(N-n)!} =$$

$$= \frac{(N-1)!\,n!(N-n)!}{N!(n-1)!\,(N-n)!} = \frac{n}{N}.$$

ანალოგიურად, ერთდროულად ნებისმიერი ორი ელემენტის შერჩევაში მოხვედრის ალბათობა იქნება.

$$p_2 = \frac{C_{N-2}^{n-2}}{C_N^n} = \frac{n(n-1)}{N(N-1)}.$$

ყოველი ელემენტის x_i მნიშვნელობას შევუსაბამოთ ისეთი α_i შემთხვევითი სიდიდე, რომელიც ღებულობს მხოლოდ ორ რიცხვით მნიშვნელობას 0-ს, ან 1-ს, იმისდა მიხედვით, ხვდება თუ არა x_i ელემენტი შერჩევაში. თუ ის შერჩევაში ხვდება, α_i ღებულობს მნიშვნელობას 1-ს, წინააღმდეგ შემთხვევაში—0-ს.

ცხადია, რომ შერჩევითი საშუალოს წარმოდგენა შემდეგნაირად შეიძლება

$$\overline{x} = \frac{x_1\alpha_1 + x_2\alpha_2 + \cdots + x_N\alpha_N}{n} = \frac{1}{n}\sum_{i=1}^{N} x_i\alpha_i.$$

თუ რომელიმე შერჩევის შედეგად მივიღეთ ელემენტები $x_{i_1}, x_{i_2}, \ldots, \ldots, x_{i_n}$, მაშინ ეს იმას ნიშნავს, რომ $\alpha_{i_1} = \alpha_{i_2} = \ldots = \alpha_{i_n} = 1$ და α-ს დანარჩენი მნიშვნელობანი ტოლია ნულის. ამგვარად, ამ შემთხვევაში

$$x = \frac{x_{i_1} + x_{i_2} + \cdots + x_{i_n}}{n}.$$

ახლა ვიპოვოთ $\overline{x}$-ის მათემატიკური ლოდინი და დისპერსია:

$$E(\overline{x}) = \frac{1}{n}\sum_{i=1}^{N} E(x_i\,\alpha_i).$$

რადგანაც დაკვირვების შედეგად მიღებული x_i მნიშვნელობანი მუდმივ სიდიდეებს წარმოადგენენ, გვექნება

$$E(\overline{x}) = \frac{1}{n}\sum_{i=1}^{N} x_i\,E(\alpha_i).$$

მაგრამ, როგორც α_i შემთხვევითი სიდიდეების განსაზღვრიდან ჩანს, მათი განაწილების კანონი შემდეგია:

$$\alpha_i \quad \begin{cases} 1, & 0 \\ p_1, & 1-p_1. \end{cases}$$

აქედან კი, მათემატიკური ლოდინის განმარტების თანახმად, დავწერთ

$$E(\alpha_i)=p_1.$$

ამგვარად, $E(\overline{x})$-თვის ვღებულობთ

$$E(\overline{x})=\frac{1}{n}\,p_1\sum_{i=1}^{N}x_i=a.$$

როგორც ვხედავთ, $\overline{x}$ საშუალო არითმეტიკული არის a გენერალური საშუალოს გადაუადგილებელი შეფასება.

ახლა გამოვთვალოთ $\overline{x}$-ის დისპერსია, რადგანაც

$$\frac{1}{n}\sum_{i=1}^{N}\alpha_i=1,$$

ამიტომ

$$\overline{x}-a=\frac{1}{n}\sum_{i=1}^{N}x_i\alpha_i-a\cdot\frac{1}{n}\sum_{i=1}^{N}\alpha_i=\frac{1}{n}\sum_{i=1}^{N}(x_i-a)\alpha_i,$$

საიდანაც $D(\overline{x})$ დისპერსიისათვის დავწერთ:

$$D(\overline{x})=E(\overline{x}-a)^2=E\left[\frac{1}{n}\sum_{i=1}^{N}(x_i-a)\alpha_i\right]^2=$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{N}(x_i-a)\alpha_i\right]^2=$$

$$=\frac{1}{n^2}E\sum_{i=1}^{N}(x_i-a)^2\alpha_i^2+2\sum_{i\pm j}(x_i-a)(x_j-a)\,\alpha_i\alpha_j\Bigg]=$$

$$=\frac{1}{n^2}\sum_{i=1}^{N}(x_i-a)^2E\alpha_i^2+\frac{1}{n^2}\,2\sum_{i\pm j}(x_i-a)(x_j-a)E(\alpha_i\alpha_j).$$

ცხადია, α_i^2 და $\alpha_i\alpha_j$-ს განაწილების კანონია:

$$\alpha_i^2 \quad \begin{cases} 1, & 0 \\ p_1, & 1-p_1, \end{cases},$$

$$\alpha_i\,\alpha_j \left\{ \begin{array}{cc} 1, & 0 \\ p_2, & 1-p_2 \end{array} \right. ,$$

ამიტომ

$$E(\alpha_i^2)=p_1=\frac{n}{N}.$$

და

$$E(\alpha_l\alpha_j)=p_2=\frac{n(n-1)}{N(N-1)}.$$

მიღებული მნიშვნელობანი შევიტანოთ $D(\overline{x})$-ის გამოსახულებაში, გვექნება:

$$D(\overline{x})=\frac{1}{n}\sum_{i=1}^{N}\frac{1}{N}(x_i-a)^2+\frac{n-1}{N(N-1)n}\sum_{i\neq j}(x_i-a)(x_j-a).$$

მაგრამ, რადგანაც

$$\sum_{i\pm j}(x_i-a)(x_j-a)=\sum_{i=1}[(x_i-a)\sum_{i\pm j}(x_j-a)]=$$

$$=\sum_{i=1}^{N}\left\{(x_i-a)\left[\sum_{j=1}^{N}(x_j-a)-(x_i-a)\right]\right\}=$$

$$=\sum_{i=1}^{N}\left\{(x_i-a)\left[\sum_{i=1}^{N}x_j-\mathrm{N}a-(x_i-a)\right]\right\}=$$

$$=\sum_{i=1}^{N}\left\{(x_i-a)\left[N\frac{1}{n}\sum_{j=1}^{N}x_j-Na-(x_i-a)\right]\right\}=$$

$$=\sum_{i=1}^{N}\{(x_l-a)[\mathrm{N}a-Na-(x_i-a)]\}=$$

$$=-\sum_{i=1}^{N}(x_l-a)^2=-N\frac{1}{N}\sum_{j=1}^{N}(x_i-a)^2=-N\sigma^2,$$

ამიტომ

$$D(\overline{x})=\frac{\sigma^2}{n}-\frac{n-1}{N-1}\cdot\frac{\sigma^2}{n},$$

საიდანაც

$$\sigma_{\bar{x}}^2 = D(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{1-\frac{n}{N}}{1-\frac{1}{N}} \cdot \frac{\sigma^2}{n}.$$

აქედან კი საშუალო კვადრატული გადახრისათვის დავწერთ:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{1-\frac{n}{N}}{1-\frac{1}{N}}} \, . \tag{2.1}$$

თუკი X შემთხვევითი სიდიდე ისეთი ბუნებისაა, რომ მას შეუძლია მხოლოდ ორი მნიშვნელობის მიღება (1 და 0-ის) და გვაინტერესებს გენერალური ერთობლიობის რა ნაწილს შეადგენს 1-ის ტოლი მნიშვნელობის მქონეთა რაოდენობა, მაშინ ანალოგიურ შეფასებებთან გვექნება საქმე.

ვთქვათ $x_1 = x_2 = \ldots = x_\mu = 1$ და $x_{\mu+1} = \ldots = x_N = 0$, ასე რომ,

$$a = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{\mu}{N} = p$$

იქნება გენერალურ ერთობლიობაში 1-ის ტოლი მნიშვნელობის მქონე ელემენტთა წილი.

აგრეთვე

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \frac{\mu}{N} \right)^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \frac{2}{N} \frac{\mu}{N} \sum_{i=1}^{N} x_i + \frac{1}{N} N \frac{\mu^2}{N^2} =$$

$$= \frac{\mu}{N} - 2\frac{\mu^2}{N^2} + \frac{\mu^2}{N^2} = \frac{\mu}{N} - \frac{\mu^2}{N^2} = \frac{\mu}{N}\left(1 - \frac{\mu}{N}\right) = p(1-p).$$

აქედან კი

$$\sigma = \sqrt{p(1-p)}.$$

σ-ს მიღებული მნიშვნელობა ჩავსვათ $\sigma_{\bar{x}}$-ის (2.1) გამოსახულებაში და მხედველობაში მივიღოთ, რომ ამ შემთხვევაში $\bar{x}$ ემთხვევა შერჩევის $\frac{m}{n}$ ფარდობით სიხშირეს, გვექნება:

$$\sigma_{m/n} = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{1-\frac{n}{N}}{1-\frac{1}{N}}} \, .$$

a-ს ან, რაც იგივეა, p-ს განსაზღვრის სიზუსტე და საიმედოობა მოცემული მოცულობის მქონე შერჩევის შემთხვევაში $\sigma_{\bar{x}}$-ით ან, რაც იგივეა, $\sigma_{m/n}$-ით ხასიათდება.

როგორც ვხედავთ, $\sigma_{\bar{x}}$ და $\sigma_{m/n}$ სიდიდეები გაუმეორებელი შერჩევის დროს განსხვავდებიან განმეორებითი შერჩევისას მათი სათანადო მნიშვნელობებისაგან

$$c=\sqrt{\frac{N-n}{N-1}}=\sqrt{\frac{1-\frac{n}{N}}{1-\frac{1}{N}}}$$

მამრავლით, რომელიც N-ის დიდი მნიშვნელობისათვის 1-თან ახლოს არის.

მაგალითად, თუ შერჩევითი ერთობლიობა გენერალურის 10%-ია, ე. ი. $\frac{n}{N}=0{,}1$, მაშინ $0{,}95=\sqrt{0{,}9}<c<1$ და ამიტომ $\sigma_{\bar{x}}$ შეადგენს 0,95-დან 1-მდე თავის მნიშვნელობას განმეორებითი შერჩევის დროს.

საზოგადოდ, a-ს $\bar{x}$-ით შეფასების სიზუსტე, როგორც ეს (2.1)-დან გამომდინარეობს, ნაკლებადაა დამოკიდებული გენერალური ერთობლიობის N მოცულობაზე და გაცილებით უფრო დამოკიდებულია შერჩევითი ერთობლიობის n მოცულობაზე. n-ის გადიდებით სიზუსტე იზრდება $\sqrt{n}$ რიგის სიდიდით, რადგანაც ამდენივე ფარდობით მცირდება მაშინ საშუალო კვადრატული გადახრა $\sigma_{\bar{x}}$.

$\sigma_{\bar{x}}$-ის n-თან ასეთ დამოკიდებულებას იქამდის მივყავართ, რომ პრაქტიკულად შერჩევითი მოცულობანი არ იქნებიან დიდი, რადგანაც მათი შემდგომი გადიდება დიდად მაინც ვერ ზრდის შეფასების სიზუსტეს, ხოლო შემდგომი უფრო მეტი სიზუსტის მისაღებად საჭიროა შეუდარებლად დიდი მოცულობის შერჩევითი ერთობლიობის აღება.

საშუალო მნიშვნელობის და დისპერსიის შეფასებები მივიღეთ იმ შემთხვევისათვის, როდესაც გენერალური ერთობლიობის მოცულობა სასრული რიცხვი იყო. ახლა განვიხილოთ შემთხვევა, როდესაც შერჩევას ვაწარმოებთ ისეთი გენერალური ერთობლიობიდან, რომლის მოცულობა განუსაზღვრელად დიდია.

ამ შემთხვევაში შერჩევითი მოცულობაც n აგრეთვე ვიგულისხმოთ საკმაოდ დიდად, რათა ზღვრული თეორემების თავისუფლად გამოყენების შესაძლებლობა გვქონდეს.

როგორც ცნობილია, განსახილავი შემთხვევის დროს უცნობი პარამეტრების სტატისტიკურად შეფასებისათვის საკმარისია ამ შეფასებათა ძალდებულობა.

იმ შემთხვევაში, როდესაც დაკვირვებათა რიცხვი საკმაოდ დიდია, თეორიული განაწილების უცნობ პარამეტრთა შესაფასებლად შესაძლებელია უშუალოდ ემპირიული მომენტების გამოყენება, რომელთა გამოთვლა ხდება შერჩევის ერთობლიობის მნიშვნელობათა მიხედვით. შეფასებათა მიღების ასეთ ხერხს **მომენტთა მეთოდი** ეწოდება.

მომენტთა მეთოდის ძალდებულობას ამტკიცებს დიდ რიცხვთა კანონის შესახებ ხინჩინის თეორემა, რომლის არსი შემდეგში მდგომარეობს:

ვთქვათ $z_1, z_2, \ldots z_n$ არიან ერთნაირად განაწილებული ურთიერთდამოუკიდებელი შემთხვევითი სიდიდეები, რომელთა მათემატიკური ლოდინია რაიმე a სასრული სიდიდე, მაშინ რაგინდ მცირე $\varepsilon > 0$ რიცხვისათვის

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} z_i - a\right| < \varepsilon\right\} = 1.$$

მართლაც, r რიგის მომენტი

$$a_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

არის $z_k = X_k^r$ სიდიდეთა საშუალო არითმეტიკული, რომლებიც ერთნაირად არიან განაწილებული და აქვთ ერთი და იგივე $Ez_k = EX_k^r = \nu_r$ მათემატიკური ლოდინი. ამგვარად, თუ არსებობს თეორიული განაწილების ν_r მომენტი, როგორი ტიპის შემთხვევით სიდიდეებთანაც არ უნდა გვქონდეს საქმე, ის წარმოდგენილი იქნება ან ჯამის, ან ინტეგრალის სახით და იქნება აბსოლუტურად კრებადი, მაშინ r რიგის ემპირიული a_r საწყისი მომენტი ალბათობით იკრიბება იმავე რიგის ν_r თეორიული საწყისი მომენტისაკენ. ამის ანალოგიურად r რიგის ემპირიული ცენტრალური მომენტი m_r ალბათობით იკრიბება იმავე რიგის თეორიული ცენტრალური μ_r მომენტისაკენ· ეს კი იმას ნიშნავს, რომ ემპირიული მომენტები წარმოადგენენ სათანადო თეორიული მომენტების ძალდებულ შეფასებას.

რადგანაც $X_k^r (k=1, 2, \ldots, n)$ სახის შემთხვევითი სიდიდეები $\sum_{k=1}^{n} X_k^r$ ჯამთან შედარებით მცირეა, ამიტომ ლიაპუნოვის თეორემის თანახმად, დაკვირვებათა საკმაოდ დიდი რიცხვისათვის ემპირიული მომენტების შერჩევითი განაწილება, რომლებიც წარმოადგენენ X_k^r სახის სიდიდეთა საშუალო არითმეტიკულს, ახლოს იქნება ნორმალურ განა-

წილებასთან, რომლის ცენტრი იქნება $\mu_r = EX'_k$ და დისპერსია —

$$-D\left(\frac{1}{n}\sum_{k=1}^{n} X'_k\right).$$

$\overline{x}$ განვიხილოთ, როგორც განაწილების პირველი რიგის საწყისი მომენტის ν_1-ის შეფასება, რომელიც მათემატიკურ ლოდინს წარმოადგენს.

როგორც ზემოთ ვნახეთ, $E\overline{x}=\nu_1$, ე. ი. საშუალო არითმეტიკული წარმოადგენს განაწილების პირველი რიგის საწყისი მომენტის ν_1-ის გადაუადგილებელ შეფასებას. ამასთან $\overline{x}$ წარმოადგენს აგრეთვე ძალდებულ შეფასებას. ამასთან, $\overline{x}$ წარმოადგენს აგრეთვე ძალდებულ შეფასებასაც ν_1-თვის.

x-ის შერჩევითი განაწილება, რომელიც ნებისმიერი გენერალური ერთობლიობიდან მიიღება, თუ შერჩევის მოცულობა n საკმაოდ დიდია, საკმარისად ახლოს იქნება ისეთ ნორმალურ განაწილების კანონთან, რომლის პარამეტრებია

$$E(\overline{x})=\nu_1=a, \qquad D(\overline{x})=\sigma^2_{\overline{x}}=\frac{\sigma^2}{n}, \quad \sigma_{\overline{x}}=\frac{\sigma}{\sqrt{n}}.$$

ამ შემთხვევაში თვითონ σ^2 გენერალური დისპერსია უცნობია და მისი შეფასება ხდება ემპირიული დისპერსიით, რომელიც წარმოადგენს მეორე რიგის ემპირიულ ცენტრალურ მომენტს, $m_2=s^2$. როგორც ზემოთ ვნახეთ, თუ შევარჩევთ „ყალბ ნულს“ $c=\nu_1$ წერტილში, შემდეგი იგივეობა გვექნება:

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\nu_1)^2-(\overline{X}-\nu_1)^2.$$

ამიტომ

$$Es^2=\frac{1}{n}\sum_{i=1}^{n}E(x_i-\nu_1)^2-E(\overline{X}-\nu_1)^2.$$

მაგრამ $x_1, x_2, \ldots, x_n$ არიან ერთი და იგივე X შემთხვევითი სიდიდის შესაძლო მნიშვნელობანი და ამიტომ

$$E(x_1-\nu_1)^2=\ldots=E(x_n-\nu_1)^2=E(X-\nu_1)^2=D(X)=\sigma^2.$$

აგრეთვე, როგორც ცნობილია

$$E(\overline{X}-\nu_1)^2=D(\overline{X})=\frac{\sigma^2}{n},$$

ამიტომ $E(s^2)$-თვის დავწერთ:

$$E(s^2)=\frac{1}{n}n\sigma^2-\frac{\sigma^2}{n} = \sigma^2\left(1-\frac{1}{n}\right).$$

ამგვარად, აქაც მივიღეთ

$$E(m_2)=E(s^2)=\sigma^2\cdot\frac{n-1}{n}. \tag{2.2}$$

აქედან ჩანს, რომ ემპირიული დისპერსია არ არის თეორიულის გადაუადგილებელი შეფასება.

რომ მივიღოთ თეორიული $D(X)=\mu_2$ დისპერსიის გადაუადგილებელი შეფასება, აუცილებელია ემპირიული s^2 დისპერსიის მაგივრად ავიღოთ შესწორებული დისპერსია $\overline{s^2}$.

$$\overline{s^2}=\frac{n}{n-1}s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

ეს უკანასკნელი იქნება თეორიული μ_2 დისპერსიის გადაუადგილებელი შეფასება. მართლაც, თუ (2.2)-ს მივიღებთ მხედველობაში, გვექნება:

$$E(\overline{s^2})=\frac{n}{n-1}E(s^2)=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2=\mu_2.$$

ამგვარად, საბოლოოდ დავწერთ

$$E(\overline{s^2})=\mu_2=D(x)=\sigma_x^2.$$

ეს კი იმას ნიშნავს, რომ $\overline{s^2}$ არის თეორიული მეორე რიგის ცენტრალური მომენტის ან, რაც იგივეა, თეორიული დისპერსიის, გადაუადგილებელი შეფასება.

ამ პარაგრაფში მიღებულ შეფასებებს წერტილოვან შეფასებებს უწოდებენ, რადგანაც შემფასებელი მახასიათებლები (სტატისტიკები) წერტილებს წარმოადგენენ. შემდეგ პარაგრაფში ამისგან განსხვავებით პარამეტრების შეფასებას მოვახდენთ ნდობის ინტერვალების გამოყენებით.

§ 3. პარამეტრების შეფასება ნდობის ინტერვალებით

შერჩევითი ერთობლიობის შესწავლის შედეგად მიღებული შეფასებანი, ცხადია, მიახლოებითი იქნება. ამიტომ მიღებულ შეფასებას მხოლოდ მაშინ ექნება გარკვეული აზრი, როდესაც გვეცოდინება შეფასების შესაძლო ცდომილებათა საზღვრები ან, სხვანაირად რომ ვთქვათ, გვეცოდინება ის ინტერვალი, რომლისთვისაც გარკვეული

ალბათობით შეგვიძლია ვამტკიცოთ, რომ ის ფარავს შესაფასებელი პარამეტრის მუდმივ მნიშვნელობას.

ვთქვათ, მოცემულია ნორმალური გენერალური ერთობლიობა და გვინდა მისი დაჯგუფების ცენტრის, a საშუალო მნიშვნელობის შეფასება, როდესაც σ საშუალო კვადრატული გადახრა ცნობილია. როგორც ზემოთ ვნახეთ, a-ს მიახლოებით შეფასებად შესაძლებელია მივიღოთ შერჩევითი ერთობლიობის $\bar{x}$ საშუალო არითმეტიკული. აგრეთვე ვნახეთ, რომ $\bar{x}$ შემთხვევითი სიდიდე ისეთ ნორმალურ კანონს ემორჩილება, რომლის მათემატიკური ლოდინია $E_{\bar{x}}=a$ და დისპერსია კი $\sigma^2_{\bar{x}}=\frac{\sigma^2_x}{n}$.

თუ განვიხილავთ $\bar{x}$ ნორმალურად განაწილებული შემთხვევითი სიდიდის a დაჯგუფების ცენტრიდან ნორმირებულ გადახრას

$$u=\frac{\bar{x}-a}{\sigma/\sqrt{n}},$$

ის ისეთი ნორმალური კანონით იქნება განაწილებული, რომლის მნიშვნელობა პარამეტრებზე სრულებით არაა დამოკიდებული. ასეთი განაწილების სიმკვრივე შემდეგნაირად გამოისახება:

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\}.$$

ამიტომ ჩვენ ყოველთვის შეგვიძლია ადვილად განვსაზღვროთ u სიდიდის ნებისმიერ მოცემულ ინტერვალში მოხვედრის ალბათობა. კერძოდ, თუ ავიღებთ (α, β) ინტერვალს, როგორც ცნობილია,

$$P\{\alpha<u<\beta\}=\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta \exp\left\{-\frac{u^2}{2}\right\}du=\Phi(\beta)-\Phi(\alpha)=\gamma.$$

γ ალბათობას „ნდობის ალბათობა“ ეწოდება და, ჩვეულებრივად, მას საჭირო საიმედოობის დონედ წინასწარ ასახელებენ. თუ მხედველობაში მივიღებთ u-ს მნიშვნელობას, უკანასკნელი გამოსახულება შემდეგნაირად შეიძლება გადავწეროთ:

$$P\left\{a<\frac{\bar{x}-a}{\frac{\sigma}{\sqrt{n}}}<\beta\right\}=P\left\{\alpha\frac{\sigma}{\sqrt{n}}<\bar{x}-a<\beta\frac{\sigma}{\sqrt{n}}\right\}=\gamma,$$

ან, რაც იგივეა,

$$P\left\{a+\alpha\frac{\sigma}{\sqrt{n}}<\bar{x}<a+\beta\frac{\sigma}{\sqrt{n}}\right\}=\gamma. \tag{3.1}$$

როგორც აქედან ჩანს, $\left(a+\alpha\frac{\sigma}{\sqrt{n}},\ a+\beta\frac{\sigma}{\sqrt{n}}\right)$ ინტერვალი იმ არეს წარმოადგენს, რომელშიაც $\overline{x}$-ის მოხვედრა გარანტირებულია მოცემული ალბათობის დონით. ამ არეს a-ს მოცემული მნიშვნელობისათვის, მოცემული ალბათობის დონით, ეწოდება $\overline{x}$-ის პრაქტიკულად შესაძლო მნიშვნელობათა არე.

რადგანაც

$$\overline{x}-a<\beta\frac{\sigma}{\sqrt{n}}$$

უტოლობა ტოლძალოვანია

$$\overline{x}-\beta\frac{\sigma}{\sqrt{n}}<a$$

უტოლობის, ხოლო

$$\alpha\frac{\sigma}{\sqrt{n}}<\overline{x}-a$$

ტოლძალოვანია

$$a<\overline{x}-\alpha\frac{\sigma}{\sqrt{n}}$$

უტოლობის, ამიტომ, (3.1) შეგვიძლია შემდეგნაირად გადავწეროთ:

$$P\left\{\overline{x}-\beta\frac{\sigma}{\sqrt{n}}<a<\overline{x}-\alpha\frac{\sigma}{\sqrt{n}}\right\}=\gamma. \tag{3.2}$$

დავუშვათ, რომ $\alpha=-\beta$, მაშინ

$$\left(\overline{x}-\beta\frac{\sigma}{\sqrt{n}},\quad \overline{x}-\alpha\frac{\sigma}{\sqrt{n}}\right).$$

ინტერვალის სიგრძე იქნება

$$\overline{x}-\alpha\frac{\sigma}{\sqrt{n}}-\overline{x}+\beta\frac{\sigma}{\sqrt{n}}=2\beta\frac{\sigma}{\sqrt{n}}.$$

ეს უკანასკნელი, როგორც ვხედავთ, დამოკიდებულია γ „ნდობის ალბათობაზე", გენერალური ერთობლიობის ცნობილ σ საშუალო კვადრატულ გადახრაზე და n შერჩევითი ერთობლიობის მოცულობაზე. ინტერვალის ცენტრის $\overline{x}$-ის მნიშვნელობა და მასთან, თვით ინტერვალის მდებარეობა $\overline{x}$ ღერძზე შემთხვევითია და სხვადასხვა შერჩევის დროს სხვადასხვაა.

(3.2) ტოლობა იმას ნიშნავს, რომ, თუ ჩვენ ავიღებთ სხვადასხვა შერჩევით ერთობლიობას და თითოეულისათვის განვსაზღვრავთ ინტე-

γ – სიმეტრიული ნიშანი

რვალის საზღვრებს სათანადო $\overline{x}$-ის დახმარებით, მაშინ, თუ $\overline{x}$-ზე ინტერვალთა განსასაზღვრავად, რომელთა სიგრძეა $2\beta\frac{\sigma}{\sqrt{n}}$, მრავალჯერ ჩავატარებთ ცდებს, $100\gamma\%$ შემთხვევა იქნება ისეთი, რომლებიც დაფარავენ a პარამეტრის ჩვენთვის უცნობ მნიშვნელობას.

ვთქვათ $\gamma=0,95$, $\sigma=2$ და $n=9$, მაშინ ცხრილი X-ის მიხედვით მოვნახავთ $q_z=1-\gamma=1-0,95=0,05$ მნიშვნელობისათვის ან, რაც იგივეა, $q=5$-თვის, რომ $z_q=\beta=1,96$. ინტერვალისათვის გვექნება საზღვრები: $\overline{x}-1,31$ და $\overline{x}+1,31$, სადაც $1,31\cong 1,96\cdot\frac{2}{\sqrt{9}}$.

0,95-ის ტოლი ნდობის ალბათობით გვექნება:

$$\overline{x}-1,31<a<\overline{x}+1,31.$$

ამასთან, ალბათობას ვაკავშირებთ ინტერვალის საზღვრებთან, რომლებიც შემთხვევითი შერჩევით განისაზღვრებიან და არა a პარამეტრთან, რომელიც, საზოგადოდ, შემთხვევაზე არ არის დამოკიდებული. ამიტომ პრინციპული უხეში შეცდომა იქნებოდა, $\overline{x}$ სიდიდის ნაცვლად ჩაგვესვა შერჩევით მიღებული მონაცემები, მაგალითად, $\overline{x}=4,12$ და გვემტკიცებია, რომ $2,81<a<5,43$ უტოლობის ალბათობა 0,95-ის ტოლია.

სინამდვილეში, თუ a მუდმივი სიდიდეა, მაშინ უკანასკნელი ტოლობა ან აუცილებელია და მას აქვს ალბათობა 1, თუ a სინამდვილეში ნაჩვენებ ინტერვალშია, ანდა შეუძლებელია და აქვს ნულის ტოლი ალბათობა, თუ a ამ ინტერვალის გარეთაა. ამ უკანასკნელ უტოლობასთან ჩვენ ვერავითარ სხვა ალბათობას ვერ დავაკავშირებთ. ამგვარად, კარგად უნდა გავარჩიოთ ერთმანეთისაგან ნდობის ინტერვალები, რომლებსაც ცვალებადი საზღვრები აქვთ, და მათი კონკრეტული წარმომადგენლები, რომლებსაც რომელიმე ცალკეული შერჩევის დროს ვღებულობთ.

a პარამეტრის ყოველ მნიშვნელობას ჩვენ ვუსაბამებთ $\Delta_a=(a+\alpha\sigma/\sqrt{n},\ a+\beta\sigma/\sqrt{n})$ ინტერვალს, რომელშიაც მოთავსებულია მოცემული a-თვის $\overline{x}$-ის პრაქტიკულად შესაძლო მნიშვნელობა.

ვთქვათ, რომელიმე შერჩევისას მივიღეთ $\overline{x}$ მნიშვნელობა, მაშინ a-ს ყველა მნიშვნელობა შეიძლება ორ კატეგორიად დავანაწილოთ: პირველში მოვათავსოთ a-ს მნიშვნელობანი, რომელთათვისაც სათანადო Δ_a ინტერვალები ფარავენ $\overline{x}$ მიღებულ მნიშვნელობას, ასე რომ, ასეთი a-თათვის ადგილი აქვს უტოლობას:

$$a+\alpha\frac{\sigma}{\sqrt{n}}<\bar{x}<a+\beta\frac{\sigma}{\sqrt{n}}. \tag{3.3}$$

მეორე კატეგორიას მივაკუთვნოთ a-ს ის მნიშვნელობანი, რომელთათვის Δ_a ინტერვალი $\bar{x}$-ს არ შეიცავს.

პირველი კატეგორიის a-თა მნიშვნელობა მოცემულ შერჩევას ეთანხმება, ხოლო მეორე კატეგორიის — არ ეთანხმება.

(3.3) უტოლობიდან გამომდინარეობს, რომ პირველი კატეგორიის მნიშვნელობანი a ღერძზე ქმნიან

$$\bar{x}-\beta\frac{\sigma}{\sqrt{n}}<a<\bar{x}-\alpha\frac{\sigma}{\sqrt{n}}$$

ინტერვალს. მაშასადამე, ცდის შედეგად მიღებული ნდობის ინტერვალის საზღვრები შესაძლებელია განხილულ იქნეს როგორც a მნიშვნელობათა საზღვრები, რომლებიც ეთანხმებიან ცდის მონაცემებს, ხოლო a-ს ის მნიშვნელობანი, რომლებიც ამ ინტერვალში არ ხვდებიან, შესაძლებელია განხილულ იქნენ ისეთებად, რომლებიც არ ეთანხმებიან ცდის მონაცემებს.

თუ ცნობილია u სიდიდის განაწილების სიმკვრივე $\varphi(u)$, ჩვენ შეგვიძლია სხვადასხვა ხერხით (α, β) ნდობის ინტერვალის შერჩევა ისე, რომ $P(\alpha<u<\beta)=\gamma$. ამის გამო, თუ გვაქვს

$$\bar{x}-\beta\frac{\sigma}{\sqrt{n}}<a<\bar{x}-\alpha\frac{\sigma}{\sqrt{n}}$$

სახის ნდობის ინტერვალი, a პარამეტრისათვის მიღებული $(\beta-\alpha)\frac{\sigma}{\sqrt{n}}$ სიგრძეს ექნება სხვადასხვა მნიშვნელობა.

ჩვენ მიერ ზემოთ განხილულ მაგალითში გვქონდა $\alpha=-\beta=-1,96$, აგრეთვე $\sigma=2$, $n=9$, მაშინ

$$(\beta-\alpha)\frac{\sigma}{\sqrt{n}}=(1,96+1,96)\,\frac{2}{3}\cong 2,61.$$

ახლა დავუშვათ $\alpha=-2,3$, მაშინ, თუ სხვა პირობები უცვლელი იქნება, (3.2) ტოლობის დახმარებით დავწერთ:

$$P\left\{\bar{x}-\beta\cdot\frac{2}{3}<a<\bar{x}+2,3.\frac{2}{3}\right\}=P\left(-\frac{2}{3}\beta<a-\right.$$

$$\left.-\bar{x}<2,3\cdot\frac{2}{3}\right)=P\left\{-\beta<\frac{a-\bar{x}}{\frac{2}{3}}<2,3\right\}=$$

$$= P\left\{-2{,}3 < \frac{\overline{x}-a}{\frac{2}{3}} < \beta\right\} = \Phi(\beta) - \Phi(-2{,}3) = \Phi(\beta) + \Phi(2{,}3).$$

მაგრამ, რადგანაც განხილულ მაგალითში $\gamma = 0{,}95$, გვექნება:

$$\Phi(\beta) + \Phi(2{,}3) = 0{,}95,$$

საიდანაც

$$\Phi(\beta) = 0{,}95 - \Phi(2{,}3).$$

თუ ვისარგებლებთ დართული ცხრილით (ცხრ. I), ვნახავთ, რომ

$$\Phi(2{,}3) = 0{,}4892.$$

მაშასადამე, დავწერთ:

$$\Phi(\beta) = 0{,}95 - 0{,}4892 = 0{,}4608.$$

იმავე დართული ცხრილით ვიპოვით, რომ

$$\beta = 1{,}76.$$

ამ შემთხვევაში ინტერვალის სიგრძე იქნება:

$$(\beta - \alpha)\frac{\sigma}{\sqrt{n}} = (1{,}76 + 2{,}3)\cdot\frac{2}{3} = 2{,}77,$$

ე. ი. ამ შემთხვევაში სიგრძე ყოფილა ცოტა უფრო მეტი, ვიდრე პირველ შემთხვევაში.

ბუნებრივია, α და β ისეთნაირად უნდა შეირჩეს, რომ $(\beta - \alpha)$ ინტერვალის სიგრძე იყოს მინიმალური. გეომეტრიულად, ცხადია, უფრო ვიწრო ინტერვალის ცენტრი იქნება a, რომელიც წარმოადგენს $(0,1)$ პარამეტრებით ნორმალურად განაწილებული შემთხვევითი სიდიდის განაწილების სიმკვრივის მრუდის წვეროს შესაბამის აბსცისას, ე. ი. უნდა ავიღოთ $\alpha = -\beta$. ამგვარად, შედარებით ვიწრო 95%-იანი ნდობის ინტერვალი ემთხვევა

$$\left(\overline{x} - 1{,}96\frac{\sigma}{\sqrt{n}},\ \overline{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$

ინტერვალს.

თუ ავიღებთ სხვა ნდობის ალბათობას, ვთქვათ $\gamma = 0{,}99$, მაშინ, ისევე, ცხრილი x-ის დახმარებით, მოვნახავთ სხვა ნდობის ინტერვალს. ამ ალბათობის შესაბამისი ნდობის ინტერვალთა შორის ყველაზე პატარა ინტერვალი იქნება:

$$\left(\overline{x} - 2{,}58\frac{\sigma}{\sqrt{n}},\ \overline{x} + 2{,}58\frac{\sigma}{\sqrt{n}}\right).$$

ნდობის ინტერვალთა აგება ზოგად შემთხვევაში შემდეგნაირად ხდება:

ვთქვათ, მოცემულია რომელიმე ფუნქცია u, რომელიც დამოკიდებულია შერჩევის მონაცემებზე და შესაფასებელ θ პარამეტრზე. ვიგულისხმოთ, რომ ფუნქცია θ-ზე დამოკიდებულია უწყვეტად და მონოტონურად. დავუშვათ, რომ $u(x_1, x_2, \ldots, x_n; \theta)$ განაწილება დამოუკიდებელია როგორც შესაფასებელ პარამეტრზე, ისე რომელიმე სხვა უცნობ პარამეტრზე. ჩვენ მიერ განხილულ მაგალითში ასეთი ფუნქცია იყო $\overline{x}-a$, რომლის განაწილება დამოუკიდებელია a პარამეტრზე, თუ მოთხოვნილი პირობები სრულდება. მაშინ γ ნდობის ალბათობის ყოველი დონისათვის ჩვენ შეგვიძლია, თუ ვისარგებლებთ u სიდიდის ცნობილი განაწილებით, მოვნახოთ ისეთი α_γ და β_γ ზღვრები, რომლებიც აღებული შერჩევისაგან დამოუკიდებელნი არიან, და ადგილი ექნება ტოლობას:

$$P(\alpha_\gamma < u < \beta_\gamma) = \gamma.$$

მოცემული γ-ის ამ ზღვართა შერჩევა ცალსახად არ განისაზღვრება, მაგრამ

$$\alpha_\gamma < u(x_1, x_2, \cdots x_n; \theta) < \beta_\gamma$$

უტოლობა, რადგანაც ის θ არგუმენტის მიმართ მონოტონურია და უწყვეტი, შეიძლება ამოიხსნას θ-ს მიმართ და ჩაიწეროს შემდეგი სახით:

$$\theta_1(x_1, x_2, \ldots, x_n) < \theta < \theta_2\ (x_1, x_2, \ldots, x_n),$$

სადაც $\theta_1(x_1, x_2, \ldots, x_n)$ და $\theta_2=(x_1\ x_2, \ldots, x_n)$ ემპირიული ფუნქციებია, რომლებიც აგრეთვე დამოუკიდებელი არიან α_γ და β_γ სიდიდეებზე. სწორედ (θ_1, θ_2) ინტერვალი იქნება მოცემული γ ნდობის— ალბათობით ნდობის ინტერვალი.

ნდობის ინტერვალთა ხერხის განზოგადება შეიძლება მაშინაც, როცა შეფასება რამდენიმე პარამეტრზეა დამოკიდებული. მაგალითად, ნორმალური განაწილების შემთხვევაში შეფასება ხდება ერთდროულად ორი—a და σ^2 პარამეტრის. ამ შემთხვევაში შეფასების პრობლემის ამოხსნა ნიშნავს შერჩევის მონაცემებით ისეთი ნდობის ბრტყელი არის აგებას, რომელიც მოცემული ალბათობით დაფარავს იმ წერტილს, რომელიც ეთანადება გენერალური ერთობლიობის უცნობი პარამეტრების მნიშვნელობებს: $a=a_0$ და $\sigma^2=\sigma_0^2$.

§ 4. ნდობის ინტერვალის დადგენა სტიუდენტის განაწილების დახმარებით

პრაქტიკული მიზნებისათვის სავალდებულო არ არის გასაზომი სიდიდის ჭეშმარიტი მნიშვნელობის პოვნა; საჭიროა იმ საზღვრების ცოდნა, რომელთა შიგნით ეს მნიშვნელობა იქნება მოთავსებული.

ვიგულისხმოთ, რომ გაზომვის შედეგად მიღებული მნიშვნელობები ისეთი ნორმალური განაწილების კანონს ემორჩილებიან, რომლისთვისაც მათემატიკური ლოდინი საძიებელი სიდიდის ჭეშმარიტი მნიშვნელობის ტოლია, ხოლო დისპერსია σ_x^2-ის.

მაშინ, როგორც ცნობილია, გაზომვის შედეგად მიღებულ შესაძლო მნიშვნელობებს ერთი და იგივე მათემატიკური ლოდინი და დისპერსია ექნებათ:

$$E(x_1)=E(x_2)=\ldots=E(x_n)=a$$

$$D(\iota)=E(x_1-a)^2=E(x_2-a)^2=\ldots=E(x_n-a)^2=\sigma_x^2.$$

აქ σ_x საშუალო კვადრატულ გადახრას წარმოადგენს, რომელიც ყოველი ცალკეული გაზომვის შემთხვევაში ერთსა და იმავე მუდმივ მნიშვნელობას ინარჩუნებს. ძალიან ხშირ შემთხვევაში σ_x-ის მნიშვნელობა შესაძლებელია უცნობიც იყოს.

ამ შემთხვევაში, როგორც ვნახეთ

$$E\overline{x}=a$$

ე. ი.

$$a\simeq\overline{x} \tag{4.1}$$

(4.1) მიახლოებითი ტოლობა გასაზომი სიდიდის ჩვენთვის უცნობი მნიშვნელობის დადგენის შესაძლებლობას იძლევა. საჭიროა შევაფასოთ, თუ რამდენად ზუსტია ეს მიახლოებითი ტოლობა.

ასეთი შეფასების აზრი შემდეგში მდგომარეობს:

ავიღოთ ისეთი $\alpha(0<\alpha<1)$, რომ ხდომილობა, რომელსაც ცალკეული ცდის დროს $1-\alpha$-ს ტოლი ან მასზე ნაკლები ალბათობა ექნება, პრაქტიკულად შეუძლებლად, ანდა ისეთ ხდომილობად ჩავთვალოთ, რომელსაც იშვიათად ექნება ადგილი. თუ რა სიზუსტე მოეთხოვება (4. 1) მიახლოებით ტოლობას, იმისდა მიხედვით არჩევენ α-ს მნიშვნელობას. მაგალითად, იღებენ $\alpha=0{,}95$; $\alpha=0{,}99$; $\alpha=0{,}999$. შემდეგ გულისხმობენ, რომ გასაზომი სიდიდის შესაძლო მნიშვნელობანი ერთისა და იმავე განაწილების კანონს ემორჩილებიან.

საძიებელია ისეთი $\varepsilon>0$, რომ

$$\overline{x}-\varepsilon<a<\overline{x}+\varepsilon$$

უტოლობის ალბათობა α-ს ტოლი იყოს, ე. ი. ადგილი უნდა ჰქონდეს შემდეგ ტოლობას:

$$P(\overline{x}-\varepsilon < a < \overline{x}+\varepsilon)=\alpha. \qquad (4.2)$$

თუ შესაძლებელია ისეთი ε-ის მონახვა, რომელიც დამოკიდებული იქნება n-ზე, α-ზე, x_i სიდიდეებზე და (4.2) ტოლობას აკმაყოფილებს, მაშინ ვიტყვით, რომ α არის (4,1) მიახლოებითი ტოლობის საიმედოობა, ხოლო ε—ამავე მიახლოებითი ტოლობის სიზუსტე.

α-ს—ნდობის ალბათობას უწოდებენ, $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$ შუალედს—ნდობის ინტერვალს, ხოლო $\overline{x}-\varepsilon$ და $\overline{x}+\varepsilon$ სიდიდეებს—ნდობის საზღვრებს.

ვთქვათ, ჩვენ მიერ განხილული გაზომვის შედეგად მიღებული მნიშვნელობათა განაწილების კანონი დამოკიდებულია a-ზე და კიდევ სხვა უცნობ პარამეტრებზე, მაშინ ε სიზუსტე მათზე არ იქნება დამოკიდებული, ე. ი. (4.2) ტოლობა ყოველთვის შესრულდება, როგორიც არ უნდა იყოს a და სხვა უცნობი პარამეტრები.

(4.1) მიახლოებითი ტოლობის შეფასება შეიძლება ცდომილებათა თეორიის ერთი ძირითადი დაშვებისა და მათემატიკური სტატისტიკის ერთ-ერთი მეტად მნიშვნელოვანი განაწილების კანონის დახმარებით.

ცდომილებათა თეორიაში ერთ-ერთი ძირითადი დაშვება იმაში მდგომარეობს, რომ დაკვირვების შედეგად მიღებული მნიშვნელობანი $x_1, x_2, \ldots, x_n$, როგორც შემთხვევითი სიდიდეები, უნდა ემორჩილებოდნენ ერთსა და იმავე ნორმალური განაწილების კანონს, რომლის მათემატიკური ლოდინი a-ს ტოლია, ხოლო საშუალო კვადრატული გადახრა σ_x-ის.

ერთ-ერთი მნიშვნელოვანი განაწილების კანონი, რომელსაც ვიყენებთ (4. 1) მიახლოებითი ტოლობის შესაფასებლად, არის სტიუდენტის განაწილების კანონი.

ცდომილებათა თეორიის ძირითადი დაშვების საფუძველზე შესაძლებელია გაზომვის შედეგად მიღებულ მნიშვნელობათათვის ვუჩვენოთ, რომ დაკვირვებათა რიცხვის ნებისმიერი მნიშვნელობისათვის ($n \geqslant 2$)

$$t=\frac{\overline{x}-a}{S_{\overline{x}}}. \qquad (4.3)$$

სიდიდის განაწილების სიმკვრივე იქნება:

$$S(t, k)=C_k\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \qquad (4.4)$$

სადაც C_k დამოკიდებულია მხოლოდ k მუდმივზე, ხოლო $k=n-1$.

(4.3) ტოლობაში შემავალი $S_{\overline{x}}$ შემდეგნაირად განისაზღვრება:

$$S_{\overline{x}}=\frac{S}{\sqrt{n}}, \qquad S=\sqrt{\frac{\Sigma(x_i-\overline{x})^2}{n-1}}. \qquad (4.5)$$

$S(t;\, k)$ არის სტიუდენტის განაწილების სიმკვრივე, ხოლო S—შესწორებული საშუალო კვადრატული გადახრა. (4.4) ტოლობა დაგვეხმარება

$$-t_\alpha < t < t_\alpha$$

უტოლობის ალბათობის ზუსტად განსაზღვრაში, სახელდობრ, თუ დაკვირვებათა რიცხვი $n \geqslant 2$, როგორიც არ უნდა იყოს σ_x, t_α-ს ($t_\alpha > 0$) ნებისმიერი მნიშვნელობისათვის, ადგილი ექნება შემდეგ ტოლობას:

$$P(-t_\alpha<t<t_\alpha)=\int\limits_{t_\alpha}^{-t_\alpha} S(t;\, k)dt. \tag{4.6}$$

თუ ამ ტოლობის მარცხენა მხარეს α-თი აღვნიშნავთ და მხედველობაში მივიღებთ, რომ $S(t;\, k)$ განაწილების სიმკვრივე ლუწი ფუნქციაა, მაშინ დავწერთ:

$$\alpha=P(-t_\alpha<t<t_\alpha)=2\int\limits_{0}^{-t_\alpha} S(t;\, k)dt. \tag{4.7}$$

თუ გავიხსენებთ t-ს მნიშვნელობას, შეგვიძლია დავწეროთ:

$$-t_\alpha S_{\overline{X}} < \overline{x} - \alpha < t_\alpha S_{\overline{X}}.$$

ამგვარად, (4,7) ტოლობის ანალოგიურად გვექნება:

$$P(\overline{x}-t_\alpha S\overline{x}<a<\overline{x}+t_\alpha S_{\overline{X}})=2\int\limits_{0}^{t_\alpha} S(t;\, k)dt. \tag{4.8}$$

როგორც (4.4) ტოლობიდან ჩანს, სტიუდენტის განაწილება მხოლოდ t და k–ზეა დამოკიდებული, ამიტომ, როდესაც α ალბათობაა მოცემული, მაშინ, (4.7) ტოლობით ყოველთვის ვიპოვით t_α-ს ისეთ დადებით მნიშვნელობას, რომელიც დამოკიდებული იქნება მხოლოდ α და n-ზე. მაშასადამე, თუ შემოვიღებთ აღნიშვნას

$$\varepsilon = t_\alpha S_{\overline{X}}, \tag{4.9}$$

მაშინ (4.8) ტოლობის ძალით გვექნება:

$$P(\overline{x}-\varepsilon<a<\overline{x}+\varepsilon)=\alpha, \tag{4.10}$$

სადაც ε დამოკიდებული იქნება n, α და x_i სიდიდეებზე (ეს უკანასკნელები ε-ში შედიან $S_{\overline{X}}$-ის საშუალებით).

მიღებული (4.10) ტოლობა საშუალებას გვაძლევს ზუსტად შევაფასოთ (4.1) მიახლოებითი ტოლობა.

ზემოთ, გასაზომი სიდიდის მიახლოებით მნიშვნელობად გაზომვათა საშუალო არითმეტიკული მივიღეთ იმ დაშვებით, რომ ამ საშუალოს მა-

თემატიკური ლოდინი გასაზომი სიდიდის a ჭეშმარიტი მნიშვნელობის ტოლია. (4.10) ტოლობა შესაძლებლობას გვაძლევს სხვა გზითაც დავადასტუროთ ასეთი მიახლოებითი ტოლობით სარგებლობის მართებულობა.

სახელდობრ, (4.10) ტოლობაზე დაყრდნობით შეიძლება ვუჩვენოთ, რომ თითქმის ერთის ტოლი ალბათობით $\overline{x}$ რაგინდ მცირედ იქნება განსხვავებული a-გან, თუკი x_i სიდიდეების მიმართ ადგილი აქვს ჩვენს ძირითად დაშვებას.

როგორც ზემოთ გვქონდა, ძირითადი დაშვების შემთხვევაში გაზომვის შედეგად მიღებულ სიდიდეებს ერთი და იგივე დისპერსია ექნებათ და მაშინ S^2-ის მათემატიკური ლოდინი σ_x^2-ის ტოლი იქნება:

$$E(S^2)=\sigma_x^2. \tag{4.11}$$

მართლაც,

$$E(S^2)=E\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{n-1}\right]=\frac{1}{n-1}E\left\{\sum_{i=1}^{n}[(x_i-a)-(\overline{x}-a)]^2\right\}=$$

$$=\frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i-a)^2-2\sum_{i=1}^{n}(x_i-a)(\overline{x}-a)+\sum_{i=1}^{n}(\overline{x}-a)^2\right]=$$

$$=\frac{1}{n-1}\left]\sum_{i=1}^{n}E(x_i-a)^2-2\sum_{i=1}^{n}E(x_i-a)(\overline{x}-a)+\sum_{i=1}^{n}E(\overline{x}-a)^2\right]=$$

$$=\frac{1}{n-1}\left[n\sigma_x^2-2(\overline{x}-a)\sum_{i=1}^{n}E(x_i-a)+nE(\overline{x}-a)^2\right].$$

ახლა გამოვთვალოთ აქ შემავალი გამოსახულებები:

$$(\overline{x}-a)\sum_{i=1}^{n}E(x_i-a)=(\overline{x}-a)\,nE(\overline{x}-a)=nE(\overline{x}-a)^2, \tag{4.12}$$

$$E(\overline{x}-a)^2=E\left(\frac{x_1-a}{n}+\frac{x_2-a}{n}+\dots+\frac{x_n-a}{n}\right)^2=$$

$$=\frac{1}{n^2}[E(x_1-a^2]+\dots+E(x_n-a)^2]=\frac{n\sigma_x^2}{n^2}=\frac{\sigma_x^2}{n}. \tag{4.13}$$

(4.13) ტოლობის მნიშვნელობა შევიტანოთ (4.12) ტოლობაში, გვექნება:

$$(\overline{x}-a)\sum_{i=1}^{n}E(x_i-a)=\sigma_x^2. \tag{4.14}$$

(4. 13) და (4.14) ტოლობათა მნიშვნელობანი შევიტანოთ $E(S^2)$-ის გამოსახულებაში, გვექნება:

$$E(S^2)=\frac{1}{n-1}\left(n\sigma_x^2-2n\frac{\sigma_x^2}{n}+n\frac{\sigma_x^2}{n}\right)=\sigma_x^2.$$

ამგვარად, მივიღეთ:

$$E(S^2)=\sigma_x^2, \tag{4.15}$$

საიდანაც შემდეგი მიახლოებითი ტოლობის დაწერა შეგვიძლია:

$$S^2\simeq\sigma_x^2. \tag{4.16}$$

მარტივად შეიძლება იმის ჩვენება, რომ

$$\lim_{n\to\infty} S_{\bar{x}}=\lim_{n\to\infty}\frac{S}{\sqrt{n}}=0, \tag{4.17}$$

ე. ი., როდესაც $n\to\infty$, გვექნება

$$\varepsilon=t_\alpha S_{\bar{x}}\to 0. \tag{4.18}$$

ეს კი იმას ნიშნავს, რომ, რაც მეტი იქნება გაზომვათა რიცხვი, მით უფრო ზუსტი იქნება (4.1) მიახლოებითი ტოლობის სიზუსტე.

მაგალითი. ვთქვათ, რაიმე x სიდიდის გაზომვის შედეგად მივიღეთ შემდეგი მნიშვნელობანი:

20,22; 20,16; 20,39; 20,27; 20,11; 20,18; 20,15; 20,08; 20,33. მაშინ,

$$\bar{x}=20+\frac{1}{9}\sum_{i=1}^{9}(x_i-20)=20+\frac{1,89}{9}=20+0,21=20,21,$$

საიდანაც,

$$\bar{x}-20=0,21.$$

$\bar{x}$-ის გამოთვლის დროს პირობით ნულად 20 არის მიჩნეული. (4.5) ფორმულის ძალით, რადგანაც $n=9$-ს, მივიღებთ,

$$S_{\bar{x}}=\frac{S}{3},$$

ხოლო

$$S_{\bar{x}}^2=\frac{S^2}{9},$$

$$S^2=\frac{\sum_{i=1}^{9}(x_i-20,21)^2}{8}.$$

თუ ამ უკანასკნელით ვისარგებლებთ, გვექნება:

$$S_{\bar{x}}^{2}=\frac{1}{72}\sum_{i=1}^{9}(x_i-20,21)^2.$$

ტოლობის მარჯვენა მხარეში შემავალი ჯამის მნიშვნელობა მოცემულია შემდეგ ცხრილში.

ცხრილი 12

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
20,22	0,01	0,0001
20,16	—0,5	0,0025
20,39	0,18	0,0324
20,27	0,06	0,0036
20,11	—0,10	0,0100
20,18	—0,03	0,0009
20,15	—0,06	0,0036
20,08	—0,13	0,0169
20,33	—0.12	0,0144
	0	0,0844

ამგვარად,

$$S_{\bar{x}}^{2}=\frac{1}{72}\cdot 0.0844\simeq 0.0012,$$

საიდანაც

$$S_{\bar{x}}=\sqrt{0,0012}\simeq 0,0346.$$

ვთქვათ, სიმტკიცე $\alpha=0,99$, მაშინ წიგნის ბოლოს მოყვანილი VIII ცხრილით ვიპოვით, რომ, როცა $k=n-1=8$ და $\alpha=0,99$, $t_\alpha=3,36$. აქედან სიზუსტე

$$\varepsilon=t_\alpha S_{\bar{x}}=0,1163.$$

ამგვარად, $\alpha=0,99$-ის ტოლი ალბათობით ვამტკიცებთ, რომ $20,21-0,116<a<20,21+0,116$,

ე. ი.

$$20,094<a<20,326.$$

α რომ აგვეღო 0,999, მაშინ ცხრილში ვნახავდით, რომ

$$t_\alpha=5,04;$$

ამ შემთხვევაში სიზუსტე

$$\varepsilon=t_\alpha S_{\bar{x}}=0,174$$

და ნდობის ინტერვალი იქნება:

$$20,036<a<20,384.$$

თუ ნდობის ალბათობა $\alpha=0,99$, მაშინ ნდობის ინტერვალის სიგრძეა 0,232, ხოლო, როდესაც სიმტკიცე $\alpha=0,999$, მაშინ ნდობის ინტერვალის სიგრძე უდრის 0,348-ს, ე. ი. როდესაც სიმტკიცეს ვზრდით, ნდობის ინტერვალიც დიდდება.

თუ დაკვირვებათა რიცხვს გავზრდით, მაშინ, ცხადია, ნდობის ინტერვალიც უფრო დაზუსტდება.

§ 5. შერჩევითი დისპერსიის განაწილება და გენერალური დისპერსიის შეფასებისათვის ნდობის არის აგება

დავუშვათ, რომ გენერალური ერთობლიობა ნორმალური განაწილების კანონს ემორჩილება და შევაფასოთ σ^2 პარამეტრი. ამისათვის, როგორც ვნახეთ, საჭიროა ისეთი ფუნქციის მონახვა, რომელიც დამოკიდებული იქნება შერჩევით ერთობლიობაზე, შესაფასებელ პარამეტრზე და, რომლის განაწილება ყველა სხვა უცნობი პარამეტრისაგან დამოუკიდებელია. ცხადია, ამ მიზნით უნდა განვიხილოთ ემპირიული s^2 დისპერსია, რომელიც დამოკიდებულია შერჩევით ერთობლიობაზე და მისი კავშირი σ^2 თეორიულ დისპერსიასთან დამოკიდებულია შესაფასებელი პარამეტრის მნიშვნელობაზე.

როგორც ცნობილია, ნორმალურად განაწილებული ნორმირებული ურთიერთდამოუკიდებელი შემთხვევით სიდიდეთა კვადრატების ჯამი განაწილებულია χ^2 კანონით, რომლის განაწილების სიმკვრივეა:

$$\varphi_{\chi^2}(x)=\begin{cases} 0, & \text{როცა } x\leqq 0. \\ \dfrac{1}{\Gamma\left(\dfrac{k}{2}\right)2^{k/2}}x^{(k/2)-1}e^{-x/2}, & \text{როცა } x>0. \end{cases}$$

დავუშვათ, რომ X შემთხვევითი სიდიდე განაწილებულია ნორმალური კანონით, რომლის განაწილების სიმკვრივეა

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

ვთქვათ, შერჩევითი ერთობლიობის ელემენტებია: $x_1, x_2, \ldots, x_n$, რომელთა საშუალო არითმეტიკულია $\overline{x}$, ხოლო დისპერსია—s^2. როგორც ვიცით, $\overline{x}$ და s^2 ურთიერთდამოუკიდებელნი არიან. s^2 წარმოვიდგინოთ შემდეგნაირად:

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n}\sum_{i=1}^{n}[(x_i-a)-(\overline{x}-a)]^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2-(\overline{x}-a)^2,$$

საიდანაც დავწერთ:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2=s^2+(\overline{x}-a)^2.$$

უკანასკნელი ტოლობის ორივე მხარე გავყოთ $\frac{\sigma^2}{n}$-ზე, გვექნება

$$\sum_{i=1}^{n}\left(\frac{x_i-a}{\sigma}\right)^2=\frac{ns^2}{\sigma^2}+\left(\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}\right)^2. \tag{5.1}$$

შემოვიღოთ აღნიშვნა:

$$\frac{ns^2}{\sigma^2}=T,$$

$$\left(\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}\right)^2=H,$$

$$\sum_{i=1}^{n}\left(\frac{x_i-a}{\sigma}\right)^2=L,$$

ხოლო (5.1) ტოლობის ორივე მხარის მაწარმოებელი ფუნქციები იყოს

$$m_{T+H}(t) \text{ და } m_L(t),$$

ასე რომ,

$$m_{T+H}(t)=m_L(t).$$

როგორც ცნობილია, x შემთხვევითი სიდიდის მაწარმოებელი ფუნქცია ეწოდება e^{xt}-ს მათემატიკურ ლოდინს, ე. ი.

$$m(t)=Ee^{xt}.$$

რადგანაც $\overline{x}$ და s^2 ურთიერთდამოუკიდებელნი არიან, ამიტომ T და H ურთიერთდამოუკიდებელია და დავწერთ:

$$m_T(t)\cdot m_H(t)=m_L(t). \tag{5.2}$$

რადგანაც L არის ურთიერთდამოუკიდებელი ნორმალურად განაწილებულ n ნორმირებულ შემთხვევით სიდიდეთა კვადრატების ჯამი, ამიტომ მისი მაწარმოებელი ფუნქცია იქნება χ^2 სიდიდის მაწარმოებელი ფუნქცია n თავისუფლების ხარისხით:

$$m_{\chi^2}(t)=m_L(t)=(1-2t)^{-n/2}.$$

ზუსტად ასევე, რადგანაც H წარმოადგენს მხოლოდ ერთ ნორმალურად განაწილებული ნორმირებული შემთხვევითი სიდიდის კვადრატს, მისი მაწარმოებელი ფუნქცია იქნება:

$$m_H(t)=(1-2t)^{-1/2}.$$

მაშინ, (5.2) ტოლობის ძალით დავწერთ:

$$m_T(t)=\frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}=(1-2t)^{-(n-1)/2}. \tag{5.3}$$

აქედან იმ დასკვნის გაკეთება შეიძლება, რომ, თუ შერჩევა წარმოებს გენერალური ერთობლიობიდან, რომელიც ისეთი ნორმალური კანონითაა განაწილებული, რომლის პარამეტრებია a და σ^2, მაშინ $T=\frac{ns^2}{\sigma^2}$ სიდიდე იქნება განაწილებული ისეთი χ^2 კანონით, რომლის თავისუფლების ხარისხია $k=n-1$.

ჯამი $n s^2=\sum_{i=1}^{n}(x_i-\overline{x})^2$ განხილულ შემთხვევაში წარმოადგენს n ურთიერთდამოკიდებულ $y_i=x_i-\overline{x}$ შემთხვევით სიდიდეთა კვადრატების ჯამს; ეს დამოკიდებულება წრფივია და შემდეგნაირად გამოიხატება:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}(x_i-\overline{x})=0 .$$

ეს კავშირი განაწილების კანონში თავისუფლების ხარისხს ერთი ერთეულით ამცირებს.

$s^2=\frac{\sigma^2 T}{n}$ სიდიდის განაწილების სიმკვრივის მისაღებად χ^2-ის განაწილების სიმკვრივის გამოსახულებაში x უნდა შევცვალოთ $\frac{n}{\sigma^2}x$-ით, გვექნება:

$$\varphi_s^2(x)=\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}\left(\frac{n}{\sigma^2}x\right)^{(n-1)/2-1}\exp\left\{-\frac{nx}{2\sigma^2}\right\}\frac{n}{\sigma^2}=$$

$$=\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n}{2\sigma^2}\right)^{(n-1)/2}x^{(n-3)/2}\exp\left\{-\frac{nx}{2\sigma^2}\right\},$$

აქედან ადვილად მიიღება s^2-ის სხვადასხვა რიგის მომენტები. როგორც ვიცით, როცა თავისუფლების ხარისხია k, მაშინ

$$E(\chi^2)=k \quad \text{და} \quad \sigma^2_{\chi^2}=D(\chi^2)=2k,$$

და, მაშასადამე, როცა $k=n-1$, მაშინ გვექნება:

$$E(T)=n-1 \quad \text{და} \quad \sigma_T^2=2(n-1).$$

მაგრამ, რადგან $T=\frac{ns^2}{\sigma^2}$, საიდანაც $s^2=\frac{\sigma^2}{n}T$, ამიტომ:

$$E(s^2)=E\left(\frac{\sigma^2 T}{n}\right)=\frac{\sigma^2}{n}E(T)=\frac{n-1}{n}\sigma^2,$$

აგრეთვე

$$D(s^2)=D\left(\frac{\sigma^2 T}{n}\right)=\frac{\sigma^4}{n^2}D(T)=\frac{2(n-1)}{n^2}\sigma^4,$$

საიდანაც

$$\sigma_s^*=\frac{\sigma^2}{n}\sqrt{2(n-1)}\,.$$

ჩვენ ვნახეთ, რომ $T=\frac{ns^2}{\sigma^2}$ ფუნქცია დამოკიდებულია შერჩევითი ერთობლიობის ელემენტებზე და შესაფასებელ σ^2 პარამეტრზე. რაც შეეხება χ^2-ის განაწილებას, ის დამოკიდებულია მხოლოდ თავისუფლების ხარისხზე ($k=n-1$). T და σ^2 შორის დამოკიდებულება უწყვეტი და მონოტონურია. აქედან ჩანს, რომ T-ს გამოყენება შეიძლება σ^2-ის შესაფასებლად ნდობის ინტერვალების ასაგებად.

მართლაც, ვთქვათ X შემთხვევითი სიდიდე განაწილებულია ისეთი ნორმალური კანონით, რომლის დისპერსიაა σ^2, ხოლო s^2 არის შერჩევითი დისპერსია, რომლის მოცულობაა n. მაშინ σ^2-თვის ნდობის საზღვრები, მაგალითად, 95%-იანი საიმედოობით, ე. ი. როცა $\gamma=0{,}95$, შესაძლებელია აიგოს შემდეგნაირად:

IX ცხრილის საშუალებით $k=n-1$ თავისუფლების ხარისხისათვის ვიპოვით χ^2-ის ორ ისეთ χ_1^2 და χ_2^2 მნიშვნელობებს, რომ

$$P\{\chi^2<\chi_1^2\}=P\{\chi^2>\chi_2^2\}=0{,}025.$$

მაშინ (5.3)-დან გამომდინარეობს, რომ

$$P\left\{\chi_1^2<\frac{ns^2}{\sigma^2}<\chi_2^2\right\}=0{,}95.$$

ან, რაც იგივეა,

$$P\left\{\frac{1}{\chi_1^2}>\frac{\sigma^2}{ns^2}>\frac{1}{\chi_2^2}\right\}=0{,}95.$$

ანდა

$$P\left\{\frac{ns^2}{\chi_1^2}>\sigma^2>\frac{ns^2}{\chi_2^2}\right\}=0{,}95,$$

ანდა საბოლოოდ

$$P\left\{\frac{n\,s^2}{\chi_2^2}<\sigma^2<\frac{ns^2}{\chi_1^2}\right\}=0{,}95.$$

ამგვარად, ორი $\frac{ns^2}{\chi_2^2}$ და $\frac{ns^2}{\chi_1^2}$ რიცხვი განსაზღვრავს σ^2-თვის ნდობის საზღვრებს.

განხილულ შემთხვევაში როგორც ნდობის ინტერვალის ცენტრის მდებარეობა, ისე თვით ამ ინტერვალის სიგრძე შერჩევიდან შერჩევამდე იცვლება.

§ 6. პარამეტრთა შესაფასებლად დასაჯეროების მაქსიმუმის მეთოდის გამოყენება

ვთქვათ, X შემთხვევითი სიდიდეა, მისი შესაძლო მნიშვნელობანი $x_1, x_2, \ldots, x_n$ არიან შერჩევითი ერთობლიობის ელემენტები, რომლის მოცულობაა n. X შემთხვევითი სიდიდის განაწილების სიმკვრივე $\varphi(x, \theta)$ იქნება θ პარამეტრზე დამოკიდებული.

დასაჯერობის ფუნქციას აქვს შემდეგი სახე:

$$L(x_1, x_2, \ldots, x_n; \theta)=\varphi(x_1, \theta)\,\varphi(x_2, \theta) \ldots \varphi(x_n, \theta).$$

დავუშვათ, რომ X დისკრეტული ტიპის შემთხვევითი სიდიდეა, რომელსაც შეუძლია მიიღოს $\xi_1, \xi_2, \ldots, \xi_k$ რიცხვითი მნიშვნელობანი, ხოლო მათი შესაბამისი სიხშირეები შერჩევაში იყოს: $m_1, m_2, \ldots, m_k$, ე. ი. შერჩევითი ერთობლიობის ელემენტთა ის რაოდენობა, რომლებიც ემთხვევიან ξ_1-ს, ξ_2-ს, …, ξ_k-ს სათანადოდ; ცხადია, რომ

$$\sum_{i=1}^{k} m_i=n.$$

დავუშვათ, რომ

$$P(X=\xi_i)=\varphi_i(\theta).$$

ამ შემთხვევაში დასაჯერობის ფუნქციას ექნება სახე:

$$L(x_1, x_2, \ldots, x_n, \theta)=\varphi_1^{m_1}(\theta)\ \varphi_2^{m_2}(\theta) \ldots \varphi_k^{m_k}(\theta). \qquad (6.1)$$

რადგანაც შერჩევითი ერთობლიობის ელემენტები $x_1, x_2, \ldots, x_n$ ჩვენთვის ცნობილია, ამიტომ L-ს განვიხილავთ, როგორც θ პარამეტრის ფუნქციას.

უდიდესი დასაჯერობის მეთოდის არსი იმაში მდგომარეობს, რომ θ პარამეტრის შესაფასებლად იღებენ θ არგუმენტის ისეთ მნიშვნელობას, რომლისთვისაც L ღებულობს მაქსიმუმს. ეს მნიშვნელობა იქნება შერჩევითი ერთობლიობის ელემენტების ფუნქცია და უდიდესი დასაჯერობის შ ე ფ ა ს ე ბ ა ეწოდება.

ამგვარად, უდიდესი დასაჯერობის შეფასების მისაღებად აუცილებელია

$$\frac{\partial L}{\partial \theta}=0 \qquad (6.2)$$

განტოლების ამოხსნა და შერჩევა $\theta=\theta(x_1, x_2, \ldots, x_n)$ ისეთი ამოხსნისა, რომელიც L-ს აძლევს მაქსიმალურ მნიშვნელობას.

(6.2) განტოლების ჩაწერა შემდეგნაირადაც შეიძლება:

$$\frac{\partial L}{\partial \theta}=\frac{1}{L}\,\frac{\partial \ln L}{\partial \theta}=0.$$

იმ შემთხვევაში, როცა შესაფასებელია ორი პარამეტრი, გვექნება შემდეგი განტოლებანი

$$\frac{\partial \ln L}{\partial \theta_1}=0$$

და

$$\frac{\partial \ln L}{\partial \theta_2}=0.$$

მაგალითისათვის განვიხილოთ ურთიერთდამოუკიდებელ ცდათა სქემის შემთხვევა. ვიპოვოთ ცალკეული ცდის დროს ჩვენთვის სასურველი ხდომილობის p ალბათობის შეფასება n ურთიერთდამოუკიდებელ ცდაში მოხდენათა m რიცხვის საშუალებით.

შესაძლებელია p ალბათობის განხილვა, როგორც ისეთი დისკრეტული ტიპის X შემთხვევითი სიდიდის განაწილებაში შემავალი პარამეტრისა, რომელიც ღებულობს ორ მნიშვნელობას $\xi_1=1$ და $\xi_2=0$. ყოველი ცდის დროს ჩვენთვის სასურველ ხდომილობას ან ექნება ადგილი, ან არ ექნება. როცა ექნება, ვაწერთ მნიშვნელობას 1-ს, როცა არა — 0-ს.

(6.1) ფორმულის თანახმად დასაჯერობის ფუნქციას ექნება შემდეგი სახე:

$$L(x_1, x_2; p)=p^m\, q^{n-m}.$$

ვიპოვოთ L-ის კერძო წარმოებული p-თი და გავუტოლოთ ნულს, მაშინ გვექნება:

$$\frac{\partial L}{\partial p}=mp^{m-1}(1-p)^{n-m}-(n-m)\,p^m(1-p)^{n-m-1}=0.$$

ანდა აქედან დავწერთ:

$$p^{m-1}(1-p)^{n-m-1}\,[m(1-p)-(n-m)\,p]=0,$$

საიდანაც

$$m(1-p)-(n-m)p=0,$$

ან რაც იგივეა,

$$m-np=0,$$

ხოლო აქედან

$$p=\frac{m}{n}.$$

მაშასადამე, უდიდესი დასაჯერობის შეფასება ყოფილა ჩვენთვის სასურველი ხდომილობის ფარდობითი სიხშირე. ცხადია, რომ ეს ფარდობითი სიხშირე იქნება აგრეთვე გადაუადგილებელი და ძალდებული შეფასება p-სი.

მაგალითისათვის განვიხილოთ აგრეთვე ისეთი X შემთხვევითი სიდიდე, რომელიც უცნობი a პარამეტრით პუასონის კანონს ემორჩილება. ვთქვათ, შერჩევითი ერთობლიობის ელემენტებია: $x_1, x_2, \ldots, x_n$, შემთხვევითი სიდიდე X-ს შეუძლია მიიღოს ნებისმიერი 0, 1, 2, ... მნიშვნელობანი. ვთქვათ შერჩევაში მოხვედრილ ელემენტთა შორის უდიდესია k, ხოლო $r_0, r_1, \ldots, r_k$-თი აღვნიშნოთ 0, 1, 2... რიცხვთა სიხშირეები შერჩევაში.

როგორც ვიცით,

$$P(X=m)=P_n(m)\cong\frac{a^m}{m!}e^{-a},\ m=0,\ 1,\ 2,\ldots,$$

მაშინ

$$L(x_1,\ x_2,\ \ldots,\ x_n;\ a)=\prod_{m=0}^{k}\left(\frac{a^m e^{-a}}{m!}\right)^{rm},$$

საიდანაც,

$$\ln\ L(x_1,x_2,\ \ldots,\ x_n;a)=\sum_{m=0}^{k} r_m \ln\left(\frac{a^m e^{-a}}{m!}\right)=$$

$$=\sum_{m=0}^{k} r_m\,(m\ln a-a-\ln m!).$$

თუ განვიხილავთ ამ უკანასკნელის კერძო წარმოებულს a-თი და გავუტოლებთ ნულს, გვექნება:

$$\frac{\partial\ln L}{\partial a}=\sum_{m=0}^{k} r_m\left(\frac{m}{a}-1\right)=0,$$

საიდანაც

$$\frac{1}{a}\sum_{m=0}^{k} m r_m = \sum_{m=0}^{k} r_m.$$

აქედან

$$a=\frac{\sum\limits_{m=0}^{k} m\, r_m}{\sum\limits_{m=0}^{k} r_m}.$$

რადგანაც შერჩევის მოცულობა არის n, ამიტომ

$$r_0+r_1+\cdots+r_k=n.$$

ე. ი.

$$\sum_{m=0}^{k} r_m=n,$$

ხოლო, ცხადია

$$\sum_{m=0}^{k} m r_m = \sum_{i=1}^{n} x_i,$$

მაშინ a-თვის საბოლოოდ დავწერთ:

$$a=\frac{1}{n}\sum_{i=1}^{n} x_i=\bar{x}.$$

ამ შემთხვევაში $\bar{x}$ იქნება გადაუადგილებელი და ძალდებული შეფასება a-სი, რომელიც აკმაყოფილებს უდიდესი დასაჯერობის მეთოდის პირობებს.

ახლა შევაფასოთ ორი — a და σ^2 პარამეტრი ნორმალური განაწილების კანონის შემთხვევაში. ვთქვათ X შემთხვევითი სიდიდე ემორჩილება ნორმალური განაწილების კანონს, ხოლო შერჩევითი მოცულობის ელემენტებია $x_1, x_2, \cdots, x_n$.

ამ შემთხვევაში, როგორც ვიცით,

$$L(x_1, x_2, \cdots, x_n;\ a,\ \sigma) =$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right\}.$$

ჩვენს შემთხვევაში L წარმოადგენს უდიდესი დასაჯერობის ფუნქციას და თუ მის ლოგარითმს ავიღებთ, გვექნება:

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2.$$

პარამეტრების შესაფასებელი განტოლებანი იქნებიან:

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0$$

და

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - a)^2 = 0.$$

პირველი განტოლებიდან დავწერთ:

$$\sum_{i=1}^{n}(x_i - a) = 0,$$

ე. ი.

$$\sum_{i=1}^{n}x_i = na,$$

საიდანაც

$$a = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x},$$

ხოლო მეორე განტოლებიდან დავწერთ:

$$-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 = 0,$$

საიდანაც

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2.$$

მაგრამ, რადგანაც $a = \overline{x}$, ამიტომ მივიღებთ:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = s^2.$$

a-ს შეფასება გადაუადგილებელია, ხოლო σ-ს შეფასება, როგორც ეს ადრე ვნახეთ, ცოტათი გადაადგილებულია.

სავარჯიშო

1. გენერალური ერთობლიობიდან ამოღებულია $n=50$ მოცულობის შერჩევითი ერთობლიობა, რომლის სიხშირეთა განაწილებაა

$$X \begin{cases} 1, & 3, & 9, & 11 \\ 14, & 13, & 8, & 15. \end{cases}$$

იპოვეთ გენერალური ერთობლიობის საშუალოს გადაუადგილებადი შეფასება.

პახ. $X = \dfrac{14.1+3.13+9,8+11.15}{50} = 5,8.$

2. გენერალური ერთობლიობიდან აღებულია $n=100$ მოცულობის შერჩევითი ერთობლიობა, რომელიც სიხშირეთა მიხედვით შემდეგნაირადაა განაწილებული:

$$X \begin{cases} 3, & 8, & 10, & 12 \\ 24, & 16, & 35, & 25, \end{cases}$$

იპოვეთ გენერალური საშუალოს გადაუადგილებადი შეფასება.

პას. $\overline{x}=8,5\cdot$

3. $n=41$ მოცულობის მქონე შერჩევის საფუძველზე ნაპოვნია გენერალური დისპერსიის გადაუადგილებელი შეფასება $Dy=3$. მონახეთ გენერალური ერთობლიობის გადაუადგილებელი შეფასება.

პას. $S^2 = \dfrac{n}{n-1} Dy = \dfrac{41}{40} \cdot 3 = 3,075.$

4. $n=101$ მოცულობის მქონე შერჩევის საფუძველზე გენერალური დისპერსიის გადაუადგილებელი შეფასებაა 4. იპოვეთ გენერალური ერთობლიობის გადაუადგილებადი შეფასება.

პას. $-S^2=4,04.$

5. ხელსაწყოს დახმარებით ღეროს სიგრძის 5-ჯერ გაზომვის შედეგად (სისტემატური ცდომილება გამორიცხულია) მიღებულია მნიშვნელობანი მმ-ში: $x_1=92,\ \ x_2=94,\ \ x_3=103,\ \ x_4=105,\ \ x_5=106.$

იპოვეთ ღეროს სიგრძის შერჩევითი საშუალო და ხელსაწყოს ცდომილების შერჩევითი და შესწორებული დისპერსია.

პას. $\overline{x}=100$, $Dy=34$, $S^2=42,5$.

6. რაიმე ფიზიკური სიდიდის 4-ჯერ გაზომვის შედეგად (გამორიცხულია სისტემატური ცდომილებანი) მიღებულია მნიშვნელობანი: $x_1=8$, $x_2=9$, $x_3=11$, $x_4=12$. იპოვეთ გაზომვათა შედეგების შერჩევითი საშუალო და ხელსაწყოს ცდომილების შერჩევითი და შესწორებული დისპერსია.

პას. $\overline{x}=10$, $Dy=2,5$; $S^2=3^1/_3$.

7. ნორმალურად განაწილებული გენერალური ერთობლიობიდან აღებულია შერჩევითი ერთობლიობა, რომლის მოცულობაა $n=49$, გენერალური საშუალო კვადრატული გადახრა უდრის 3-ს, შერჩევითი საშუალო $\overline{x}=16$-ს. იპოვეთ უცნობი გენერალური a საშუალოს შესაფასებლად ნდობის ინტერვალი $\alpha=0,95$-ის ტოლი საიმედოობით.

პას. $106,12 < a < 117,88$.

8. 0,99-ის ტოლი საიმედოობით იპოვეთ a გენერალური საშუალოსათვის ნდობის ინტერვალი, თუ ვიცით, რომ $\overline{x}=10,2$, გენერალური საშუალო კვადრატული გადახრა $\sigma=4$ და შერჩევითი მოცულობა $n=16$.

პას. $7,63 < a < 12,77$.

9. 0.99-ის ტოლი საიმედოობით იპოვეთ a გენერალური საშუალოსათვის ნდობის ინტერვალი, თუ ვიცით, რომ $\overline{x}=16,8$; გენერალური დისპერსია $\sigma^2=25$ და შერჩევის მოცულობა $n=25$.

პას. $14,23 < a < 19,37$.

10. ელექტრონათურათა საკმაოდ დიდი პარტიიდან არჩევენ 100 ცალს. შერჩევის საშუალო (ნათურათა გამძლეობის ხანგრძლიობის თვალსაზრისით) არის 1000 საათი. 0,95-ის ტოლი საიმედოობით იპოვეთ მთელი პარტიისათვის გენერალური a საშუალოს ნდობის ინტერვალი, თუ ცნობილია, რომ ნათურათა გამძლეობის საშუალო კვადრატული გადახრა $\sigma=40$-ს.

პას. $992,16 < a < 1007,84$.

11. გენერალური ერთობლიობიდან ვიღებთ $n=10$ მოცულობის შერჩევას. შერჩევის განაწილება სიხშირის მიხედვით შემდეგია:

$$m\left\{\begin{array}{cccccc} -2, & 1, & 2, & 3, & 4, & 5. \\ 2, & 1, & 2, & 2, & 2, & 1. \end{array}\right.$$

0,95-ის ტოლი საიმედოობით შეაფასეთ ნორმალურად განაწილებული გენერალური ერთობლიობის a მათემატიკური ლოდინი ნდობის ინტერვალის დახმარებით.

პას. $0,3 < a < 3,7$ ინტერვალი 0,95-ის ტოლი საიმედოობით ფარავს უცნობ a მათემატიკური ლოდინის მნიშვნელობას.

12. გენერალური ერთობლიობიდან ვიღებთ $n=12$ მოცულობის შერჩევას, რომლის განაწილება სიხშირეების მიხედვით შემდეგია:

$$m \begin{cases} -0,5; & -0,4; & -0,2; & 0; & 0,2; & 0,6; & 0,8; & 1; & 1,2; & 1,5; \\ 1; & 2; & 1; & 1; & 1; & 1; & 1; & 1; & 2; & 1; \end{cases}$$

0,95-ის საიმედოობით, ნდობის ინტერვალის დახმარებით, შეაფასეთ ნორმალურად განაწილებული გენერალური ერთობლიობის უცნობი a მათემატიკური ლოდინი,

პას. $-0,04 < a < 0,88$.

13. $n=51$ მოცულობის შერჩევის მიხედვით ნაპოვნია, რომ გენერალური ერთობლიობის დისპერსიის გადაადგილებული შეფასება $D_в = 5$.

იპოვეთ გენერალური ერთობლიობის დისპერსიის გადაუადგილებადი შეფასება.

პას. $s^2=5,1$.

14. X შემთხვევითი სიდიდე განაწილებულია ნორმალური კანონის მიხედვით, რომლის მათემატიკური ლოდინი უცნობია და საშუალო კვადრატული გადახრა არის 3.

მონახეთ ნდობის ინტერვალი უცნობი მათემატიკური ლოდინისათვის $\overline{X}$ შერჩევითი საშუალოს დახმარებით, თუ შერჩევის მოცულობა $n=36$ და შეფასების საიმედოობაა $\alpha=0,95$.

პას. $3,12 < a < 5, 08$.

15. X შემთხვევითი სიდიდე განაწილებულია ნორმალური კანონის მიხედვით. $n=16$ მოცულობის შერჩევის საფუძველზე ნაპოვნია, რომ $\overline{X}=20,2$ და შესწორებული საშუალო კვადრატული გადახრა $s=0,8$.

შეაფასეთ უცნობი მათემატიკური ლოდინი ნდობის ინტერვალის დახმარებით. მიიღეთ, რომ საიმედოობაა 0,95.

პას. $19,774 < a < 20, 626$.

16. ვთქვათ X შემთხვევითი სიდიდის გაზომვის შედეგად მივიღეთ შემდეგი შესაძლო მნიშვნელობანი:

5,12; 5,21; 5,13; 5,17; 5,11; 5,31; 5,28; 5,19; 5,16.

იპოვეთ მათი საშუალო არითმეტიკული $\overline{X}$, შესწორებული საშუალო კვადრატული გადახრა $S_{\overline{x}}$. დაუშვით, რომ მნიშვნელოვნების დონეა $\alpha=0,98$ და დაადგინეთ a-თვის ნდობის ინტერვალი.

ჰიპოთეზათა სტატისტიკური შეფასება

§ 1. საკითხის დასმა

სტატისტიკური ჰიპოთეზა ეწოდება ყოველგვარ დაშვებას, რომელიც ეხება განსახილავი შემთხვევითი სიდიდის განაწილების კანონს ან გენერალურ ერთობლიობას. როდესაც რაიმეს ვუშვებთ, ამ შემთხვევაში სხვადასხვა დასკვნებს ვაკეთებთ და ვამოწმებთ, თუ რამდენად ამართლებს ამ დაშვებას სათანადო ცდები. შერჩევითი ერთობლიობის მიხედვით გამოთვლილი რომელიმე სტატისტიკური მახასიათებლის მიმართ დასკვნები იქნება ალბათური მსჯელობის ხასიათისა. ამ მახასიათებლებს შესამოწმებელი კრიტერიუმები ეწოდებათ.

ჰიპოთეზის შემოწმების ხერხი იმაში მდგომარეობს, რომ მოცემული პირობისათვის ისეთი გარკვეული წესის შერჩევა ხდება, რომლის შესრულებისას დაშვებული ჰიპოთეზა არ ეთანხმება ცდებს და ხდება მისი უარყოფა.

იმისათვის, რომ ჩამოვაყალიბოთ ასეთი წესი, პირველ რიგში ვარჩევთ ისეთ სტატისტიკურ მახასიათებელს, რომლის შერჩევითი განაწილება მოცემული ჰიპოთეზისათვის სავსებით ცნობილია.

ვთქვათ, X შემთხვევითი სიდიდეა. თუ X ის განაწილება უცნობია, მაშინ შესაძლებელია რამდენიმე დაკვირვების შედეგად მიღებული მონაცემების ანალიზის საფუძველზე გარკვეული წარმოდგენა შევიქმნათ X-ის განაწილებაზე. ხშირად X შემთხვევითი სიდიდის განაწილების სახე არ არის სავსებით უცნობი, პირიქით, ხშირად შესაძლებელია გვქონდეს არასრული ინფორმაცია განაწილების სახეზე.

მაგალითად, ვთქვათ გვაქვს რომელიმე წარმოებაში დამზადებულ პროდუქციათა N რაოდენობა. პროდუქცია შეიძლება იყოს როგორც ვარგისი, ისე წუნდებული. შემთხვევით ვიღებთ რომელიმე ერთ ცალს და ვამოწმებთ. ვთქვათ, წუნდებულ პროდუქციას აღვნიშნავთ 1-ით, ხოლო ვარგისს—0-ით, მაშინ x შემთხვევით სიდიდეს შეუძლია მიიღოს მხოლოდ ორი რიცხვითი მნიშვნელობა—1 და 0. მოცემულ პარტიაში წუნდებულ პროდუქციათა ფარდობითი რაოდენობა აღვნიშნოთ p-თი, მაშინ $p(X=1)=p$, ხოლო $p(X=0)=1-p$. ამგვარად, თუ p ცნობილია, მაშინ x-ის განაწილება მთლიანად ცნობილი იქნება. ხშირად p-ს მნიშვნელობა უცნობია და ამ შემთხვევაში სწორედ აღებული პარტიიდან ვარჩევთ რამდენიმე ერთეულს, მათ ვამოწმებთ და მიღებული შედეგების საფუძველზე ვაკეთებთ სათა-

ნადო დასკვნებს p-ს მნიშვნელობის შესახებ. თუ p უცნობია, მაშინ ჩვენი ცოდნა X-ის განაწილების შესახებ არასრული იქნება, ჩვენ გვეცოდინება მხოლოდ, რომ მას შეუძლია მიიღოს მნიშვნელობანი 0 ან 1. აღებულ მაგალითში p განიხილება, როგორც უცნობი პარამეტრი, რომელსაც შეუძლია მიიღოს ნებისმიერი მნიშვნელობა 0-დან 1-მდე. ამგვარად, ჩვენს მაგალითში p პარამეტრის მნიშვნელობის სიზუსტით ცნობილი იქნება X შემთხვევითი სიდიდის განაწილება.

განხილულ მაგალითში X შემთხვევითი სიდიდის განაწილების ფუნქციას ექნება შემდეგი სახე:

$$F(\xi)=\begin{cases} 0, & \text{როცა } \xi\leqslant 0 \\ 1-p, & \text{როცა } 0<\xi\leqslant 1 \\ p, & \text{როცა } \xi>1. \end{cases}$$

აგრეთვე, ვთქვათ რაიმე მანძილის გაზომვას ვაწარმოებთ რომელიმე იარაღით, რომლის მიმართ ცნობილია, რომ ცდომილობანი ემორჩილებიან ნორმალური განაწილების კანონს. მაშინ გაზომვის შედეგი x იქნება შემთხვევითი სიდიდე, რომელიც ნორმალური კანონის მიხედვით იქნება განაწილებული. როგორც ცნობილია, ნორმალური განაწილების სიმკვრივეს აქვს შემდეგი სახე:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right].$$

მათემატიკური ლოდინი a და დისპერსია σ^2 არიან ნორმალური განაწილების პარამეტრები, რომლებიც ხშირად უცნობებს წარმოადგენენ. a-ს შეუძლია მიიღოს ნებისმიერი რიცხვითი მნიშვნელობა, ხოლო σ^2-ს—მხოლოდ ნებისმიერი დადებითი მნიშვნელობანი.

ამგვარად, განხილულ მაგალითში განაწილების ფუნქციის სახე ცნობილია a და σ^2 პარამეტრების მნიშვნელობათა სიზუსტით.

როგორც განხილული მაგალითებიდან ჩანს, განაწილების ფუნქციის სახე საზოგადოდ ცნობილია, მაგრამ მასში შემავალი პარამეტრების მნიშვნელობანი უცნობია.

პირველ მაგალითში განაწილების ფუნქცია განისაზღვრება მხოლოდ ერთი p პარამეტრით, რომელიც წუნდებულ პროდუქციათა ფარდობით რაოდენობას გამოსახავს, ხოლო მეორე მაგალითში ასეთ პარამეტრთა რიცხვი ორია: მათემატიკური ლოდინი a და საშუალო კვადრატული გადახრა σ.

ვთქვათ, x შემთხვევითი სიდიდის განაწილების უცნობი პარამეტრებია $\theta_1, \theta_2, \ldots, \theta_k$. თუ დავუშვებთ, რომ $\theta_1, \theta_2, \ldots, \theta_k$ უცნობი პარამეტრები ღებულობენ მხოლოდ ერთადერთ გარკვეულ მნიშვნელო-

ბებს, მაშინ გვექნება უბრალო, მარტივი ჰიპოთეზები; თუ დავუშვებთ, რომ იგივე პარამეტრებს შეუძლიათ მიიღონ რამდენიმე ერთმანეთისაგან განსხვავებული მნიშვნელობა, მაშინ გვექნება რთული ჰიპოთეზები.

მაგალითად, თუ X-ის განაწილებაში შედის სამი θ_1, θ_2 და θ_3 პარამეტრი და დავუშვებთ, რომ $\theta_1=1$, $\theta_2=2$ და $\theta_3=5$, მაშინ საქმე გვექნება მარტივ ჰიპოთეზასთან, ხოლო თუ დავუშვებთ, რომ $\theta_1= =\theta_2=\theta_3$, მაშინ გვექნება რთული ჰიპოთეზა.

თუ დავუშვებთ, რომ ზემოთ განხილულ მაგალითში წუნდებულთა ფარდობითი რაოდენობა p უდრის 0,3-ს, გვექნება მარტივი ჰიპოთეზა, თუ დავუშვებთ, რომ $0,1\leqslant p\leqslant 0,2$, მაშინ გვექნება რთული ჰიპოთეზა. მეორე მაგალითში თუ დავუშვებთ, რომ $a=5$, გვექნება რთული ჰიპოთეზა, რადგან σ^2-ს შეუძლია მიიღოს ნებისმიერი დადებითი რიცხვითი მნიშვნელობა და ამდენად მისი მნიშვნელობა განუზღვრელია.

ვთქვათ, X შემთხვევითი სიდიდეა და გვინდა შევამოწმოთ ჰიპოთეზა მისი განაწილების უცნობი პარამეტრების შესახებ. რომელიმე ჰიპოთეზას ჩვენ მივიღებთ ან უარვყოფთ იმის შემდეგ, როდესაც გავაანალიზებთ დაკვირვების სასრული n რიცხვის შედეგს. როგორც ცნობილია, X შემთხვევით სიდიდეზე დაკვირვებათა სასრულო რაოდენობას შერჩევითი ერთობლიობა ეწოდება, ხოლო შერჩევით ერთობლიობაში შემავალ ელემენტთა რაოდენობას—*შერჩევითი ერთობლიობის მოცულობა*. ასევე, ყველა ელემენტის ერთობლიობას, საიდანაც შერჩევას ვაწარმოებთ, *გენერალური ერთობლიობა* ეწოდება, ხოლო მასში შემავალ ელემენტთა რაოდენობას— *გენერალური ერთობლიობის მოცულობა*.

ვიგულისხმოთ, რომ ცალკეულ დაკვირვებათა შედეგები ერთმანეთისაგან დამოუკიდებელნი არიან. თუ მათ აღვნიშნავთ სათანადოდ $x_1, x_2, \ldots, x_n$-ით, ვიტყვით, რომ ისინი სტატისტიკურად დამოუკიდებელნი არიან, თუკი i-ური დაკვირვების შედეგად მიღებული x_i-ის მნიშვნელობის ალბათობის პირობითი განაწილება დამოუკიდებელია წინა დაკვირვებათა შედეგად მიღებული მნიშვნელობებისაგან. ცხადია, რომ, თუ გვაქვს სასრულო ელემენტებისაგან შემდგარი გენერალური ერთობლიობა და აქედან ვიღებთ შერჩევით ერთობლიობას, მაშინ ზემოთ მოთხოვნილი პირობა, მკაცრად რომ ვთქვათ, არ სრულდება. სახელდობრ, გავარჩიოთ იგივე წუნდებულ პროდუქციასთან დაკავშირებული მაგალითი. ვთქვათ, მთელი პარტიიდან შემთხვევით ვარჩევთ ორ ერთეულს. პირველი აღვნიშნოთ x_1-ით, ხოლო მეორე—x_2-თი.

x_1-ის განაწილების კანონი იქნება:

$$x_1 \begin{cases} 0, & 1 \\ 1-p, & p. \end{cases}$$

განვსაზღვროთ x_2-ის განაწილების კანონი, თუ ცნობილია x_1-ის განაწილების კანონი.

ვთქვათ, გენერალური ერთობლიობის მოცულობა არის N, მაშინ წუნდებულთა რაოდენობა იქნება pN. თუ ცნობილია, რომ $x_1=0$, ე. ი. ის არ არის წუნდებული, მაშინ დარჩენილ ერთობლიობაში, რომლის მოცულობა $N-1$ იქნება, წუნდებულთა რაოდენობა უცვლელი დარჩება და ალბათობა იმისა, რომ შემდეგში შემთხვევით ამოღებული წუნდებული იქნება, ტოლია $\frac{pN}{N-1}$-ის, ხოლო, რომ ვარგისი იქნება, ტოლია $1-\frac{pN}{N-1}$. ამგვარად, x_2-ის განაწილების კანონს იმ პირობით, რომ $x_1=0$ ცნობილია, შემდეგი სახე ექნება:

$$x_2 \begin{cases} 1, & 0 \\ \frac{pN}{N-1}, & 1-\frac{pN}{N-1}. \end{cases}$$

თუ ცნობილია, რომ $x_1=1$, მაშინ x_2-ის განაწილების კანონს შემდეგი სახე ექნება:

$$x_2 \begin{cases} 1, & 0 \\ \frac{pN-1}{N-1}, & 1-\frac{pN-1}{N-1}. \end{cases}$$

ამგვარად, როგორც ამ მაგალითიდან ჩანს, x_2-ის ალბათობის განაწილება დამოკიდებულია პირველი დაკვირვების შედეგზე x_1-ის მნიშვნელობაზე.

საზოგადოდ, როცა გენერალური ერთობლიობა ელემენტებს შეიცავს სასრული რაოდენობით, დაკვირვებათა შედეგების მკაცრად დამოუკიდებლობა შეუძლებელია, მაგრამ თუ გენერალური ერთობლიობის მოცულობა საკმარისად დიდია, მაშინ დაკვირვების შედეგთა შორის დამოკიდებულება იქნება საკმაოდ სუსტი და ხშირად მისი უგულებელყოფა შესაძლებელია.

კიდევ უფრო ნათელი რომ გახდეს დასმული ამოცანის შინაარსი, განვიხილოთ შემდეგი მაგალითიც. ვთქვათ, გვაინტერესებს გავარკვიოთ, თუ რამდენად უკეთესია ელექტრონათურების დამზადების ახალი მეთოდი ადრინდელთან შედარებით. ვიგულისხმოთ, რომ ადრინდელი მეთო-

დით დამზადებული ელექტრონათურების ხანგრძლიობა შეადგენდა 1500 საათს. ახალი მეთოდით დამზადებული ელექტრონათურიდან ვარჩევთ რამდენიმეს შესამოწმებლად და, ვთქვათ, შემოწმებით დადასტურდა, რომ შერჩეული ნათურების საშუალო ხანგრძლიობა უდრის 1675 საათს. ისმება კითხვა: წარმოების ახალი მეთოდი, საზოგადოდ, უკეთესია თუ არა? ამ კითხვაზე და, საზოგადოდ, ამის ანალოგიურ სხვა კითხვებზე დასაბუთებული პასუხი რომ გავცეთ, საჭიროა რაიმე სტატისტიკურ მოდელს ანდა მსგავსი მოვლენის სქემას დავეყრდნოთ. ამ შემთხვევაში ჩვენ განვიხილავთ ნათურათა ხანგრძლიობის მიხედვით განაწილებულ ორ გენერალურ ერთობლიობას. პირველი ერთობლიობის განაწილების საშუალოდ ჩვენ ვღებულობთ $a_1=1500$ საათს და ვგულისხმობთ, რომ ეს საკმარისად ზუსტად ცნობილია. ამოცანის მარტივად ამოხსნა შეიძლება მეორე გენერალური ერთობლიობის a_2 საშუალოს a_1-თან შესაბამისობით, მაგრამ a_2 ჩვენთვის უცნობია და ამიტომ შედარების ეს პირდაპირი გზა არ გამოგვადგება. საშუალო a_2 ჩვენ შეგვიძლია შევაფასოთ მეტი ან ნაკლები სიზუსტით ჩატარებული დაკვირვებების მიხედვით. მიღებული შერჩევითი საშუალო $\overline{x}$ შეიძლება განვიხილოთ როგორც a_2 გენერალური საშუალოს შეფასება, რა თქმა უნდა, $\overline{x}$-ის განაწილების კანონი დამოკიდებულია a_2 უცნობ პარამეტრზე.

იმისათვის, რომ გარკვეული ალბათური ხასიათის დასკვნამდე მაინც მივიდეთ, გენერალური საშუალოების ტოლობაზე ჰიპოთეზურ დაშვებას ვაწარმოებთ $a_2=a_1=a$. ამგვარ დამხმარე ჰიპოთეზას, რომელიც გვეუბნება, რომ გენერალური ერთობლიობების პარამეტრებს შორის ჩვენთვის საინტერესო განსხვავება არ არსებობს, ნულოვანი ჰიპოთეზა ეწოდება. ამგვარად, მეორე გენერალური ერთობლიობიდან შერჩევითი ერთობლიობის შემოწმების საფუძველზე უნდა მივიღოთ ან უკუვაგდოთ ნულოვანი ჰიპოთეზა. ნულოვანი ჰიპოთეზის დაშვების საფუძველზე გავაკეთებთ სათანადო დასკვნებს და მას შევადარებთ იმას, რასაც მივიღებთ შერჩევითი ერთობლიობის შემოწმებისას. ვთქვათ, შერჩევითი ერთობლიობის საშუალო უდრის 1675 საათს. ეს თითქოს მიგვითითებს წარმოების ახალი მეთოდის უპირატესობაზე. ვიგულისხმოთ, აგრეთვე, რომ $\overline{x}$ საშუალო არითმეტიკულის საშუალო კვადრატული გადახრა $\sigma_{\overline{x}}$ არის 140 საათი.

როგორც ცნობილია,

$$P\left\{\ \overline{x}-r\frac{\sigma}{\sqrt{n}}<a<\overline{x}-k\frac{\sigma}{\sqrt{n}}\ \right\}=\gamma, \tag{1.1}$$

სადაც $u=E(x)$, γ არის ნდობის ალბათობა, რომელიც მოიცემა, როგორც მოთხოვნილი (ნდობის) საიმედოობის დონე. (r, k) ის ინტერვალია, რომლის შიგნით არის მოთავსებული

$$u=\frac{\overline{x}-a}{\sigma/\sqrt{n}}$$

ნორმირებული გადახრა.

როგორც (1.1)-დან ჩანს, 95%-იანი ნდობის ინტერვალი მეორე გენერალური ერთობლიობის a_2 საშუალოსათვის იქნება მრგვალ რიცხვებში (1400, 1950). მართლაც, $\sigma/\sqrt{n}=140$ საათს; თუ მივიღებთ, რომ $-r=k$-ს, მაშინ პირობის თანახმად $\gamma=0{,}95$; $1-\gamma=0{,}05$. ე. ი. $q=5\%$. წიგნის ბოლოს მოცემულ III ცხრილში $q=5$-ს შეესაბამება $q=1{,}96$. ამგვარად, $r\cdot\sigma/\sqrt{n}=1{,}96\,.\,140=275$. მაშინ მეორე საშუალოსათვის ინტერვალი იქნება (1675—275, 1675+275)=(1400, 1950). შერჩევითი საშუალო 1675 საათი, რომელიც მიღებულ ინტერვალში ხვდება, აგრეთვე ადვილად მიიღება პირველი გენერალური ერთობლიობიდან, რომლის საშუალოა $a_1=1500$ საათი. ამგვარად, ჩვენ არა გვაქვს სათანადო საფუძველი ნულოვანი ჰიპოთეზის უარყოფისა. თუ სხვა შემთხვევაში გვაქვს, მაგალითად, $\sigma_{\overline{x}}=40$ საათს, მაშინ 95%-იანი ნდობის ინტერვალი იქნება (1600, 1750), რადგანაც აქაც მივიღებთ $-r=k=1{,}96$ და $1{,}96\,.\,40=78{,}5\approx 75$, ამიტომ მეორე საშუალოსათვის ნდობის ინტერვალი იქნება (1675—75, 1675+75)=(1600,1750). ამ შემთხვევაში საკმაოდ დიდი ალბათობით შეგვიძლია უკუვაგდოთ ნულოვანი ჰიპოთეზა და მიღებული სხვაობა ჩავთვალოთ არა შემთხვევითად, რომლის ახსნა შერჩევითი ერთობლიობის შემთხვევითი გარემოებით არ შეიძლება.

განხილული მაგალითის შემთხვევაში ჩვენ შეგვიძლია ნულოვანი ჰიპოთეზის უკუგდება, მაგრამ ეს არ იძლევა არავითარ მითითებას, თუ როგორია თვითონ განსხვავების სიდიდე. ჩვენ შეგვიძლია შემდეგნაირად ვიმოქმედოთ. ისევ ვიგულისხმოთ, რომ ადგილი აქვს ნულოვან ჰიპოთეზას, განვიხილოთ აბსცისათა ღერძზე ($\overline{x}$-ზე) G უსასრულო არე, რომელიც განისაზღვრება

$$\overline{x}>a_1+t\cdot\sigma/\sqrt{n}$$

უტოლობით. ვიგულისხმოთ n საკმარისად დიდად. შეგვიძლია მივიღოთ, რომ ნულოვანი ჰიპოთეზის შემთხვევაში $\overline{x}$-ის მოხვედრის ალბათობა G არეში გამოისახება შემდეგი ინტეგრალით:

$$P\{\bar{x}>a_1+t\cdot\sigma/\sqrt{n}\}=\frac{1}{\sqrt{2\pi}}\int\limits_t^\infty e^{-z^2/2}\,dz=1-\Phi(t). \qquad (1\cdot 2)$$

ჩავსვათ $t=\lambda_{2q}$,
სადაც λ_{2q} არის

$$u_1=\frac{\bar{x}-a_1}{\sigma/\sqrt{n}}$$

ნორმირებული გადახრის 2%-იანი ორმხრივი ზღვარი; მაგრამ რადგანაც

$$P\{|\bar{x}-a_1|>\lambda_{2q}\cdot\sigma/\sqrt{n}\}=2[1-\Phi(\lambda_{2q})]=\frac{2q}{100},$$

ამიტომ გვექნება

$$P\{\bar{x}>a_1+\lambda_{2q}\cdot\sigma/\sqrt{n}\}=1-\Phi(\lambda_{2q})=\frac{q}{100}.$$

ვხედავთ, რომ ჩვენ მიერ აგებული ჰიპოთეზა შესაძლებლობას გვაძლევს $\bar{x}$-ის G არეში მოხვედრის შესახებ სათანადო დასკვნა გავაკეთოთ. თუ მოხვედრის ეს ალბათობა იქნება მცირე, მაშინ საქმე გვექნება პრაქტიკულად შეუძლებელ ხდომილობასთან. თუ უგულებელვყოფთ ასეთი ხდომილობის მოხდენას, დავუშვებთ უმნიშვნელო შეცდომას. ამიტომ, თუ $\bar{x}$ მოხვდა G არეში, ეს იმის მაჩვენებელი იქნება, რომ შესამოწმებელი ჰიპოთეზა უმართებულოა. ამ ჰიპოთეზის უარყოფით შესაძლებელია დავუშვათ საშუალოდ შეცდომა ანალოგიურ პირობებში ამ ხერხის გამოყენების ყველა შემთხვევის $q/100$ ნაწილის ტოლი.

შერჩევა სათანადო მცირე ალბათობისა $q/100$ (ჩვენი ჰიპოთეზის მნიშვნელოვნების დონე) შესაძლებლობას აძლევს მკვლევარს საკმაო სიზუსტით განსაზღვროს შესამოწმებელი ჰიპოთეზის მიღების ან უარყოფის საკითხი, როდესაც სინამდვილეში ეს ჰიპოთეზა სწორია.

ზემოთ განხილულ მაგალითში დავუშვათ, რომ $q=1\%$, მაშინ დართულ ცხრილში (ცხრ. I) ვიპოვით: $\lambda_{2q}=2{,}326$ და G არე შემდეგი უტოლობით განისაზღვრება: $\bar{x}>1500+2{,}326\cdot\sigma/\sqrt{n}$.

ამგვარად აგებული არე შესამოწმებელი ჰიპოთეზისათვის არის „კრიტიკული არის“ მაგალითი.

თუ, როგორც ზემოთ, $\sigma/\sqrt{n}=40$ საათს, მაშინ საშუალო არითმეტიკული 1675 ხვდება კრიტიკულ არეში $\bar{x}>1593$ და დარწმუნებით შეგვიძლია ნულოვანი ჰიპოთეზის უარყოფა (მოსალოდნელია შეცდომა მთელ შემთხვევათა მხოლოდ 1%-ის ტოლი) და ახალი პროცესისათვის უპირატესობის მინიჭება.

ჩვენ შეგვიძლია 95%-იანი გარანტიით მივიღოთ უტოლობა:

$$1675-1,96\cdot\sigma/\sqrt{n}<a_2<1675+1,96\cdot\sigma/\sqrt{n}$$

ანდა, მრგვალ რიცხვებში,

$$1595<a_2<1755$$

და სხვაობა a_2-a_1 აკმაყოფილებს უტოლობას

$$95<a_2-a_1<255.$$

ვთქვათ, X დისკრეტული ტიპის შემთხვევითი სიდიდეა, რომლის განაწილების კანონი ცნობილია

$$X\left\{\begin{array}{l} x_1, x_2, \ldots, x_n \\ p_1, p_2, \ldots, p_n, \end{array}\right.$$

სადაც, როგორც ვიცით,

$$p_i=P(X=x_i).$$

თუ დავუშვებთ, რომ დაკვირვებათა შედეგები ურთიერთდამოუკიდებელია, მაშინ ალბათობა იმისა, რომ დაკვირვებათა ახალი სერია მოგვცემს ადრინდელი სერიის განმეორებას, იქნება

$$p=p_1p_2\cdots p_n$$

ნამრავლის ტოლი, რომელსაც, საზოგადოდ უწოდებენ $x_1, x_2, \ldots, x_n$ შერჩევის ალბათობის ერთობლივ განაწილებას.

თუკი X უწყვეტი ტიპის შემთხვევითი სიდიდეა, რომლის განაწილების სიმკვრივეა $f(x)$, მაშინ $x_1, x_2, \ldots, x_n$ ურთიერთდამოუკიდებელ დაკვირვებათა შედეგების ალბათობათა ერთობლივი განაწილების სიმკვრივე შემდეგი ნამრავლით განისაზღვრება:

$$f(x_1)\,f(x_2)\cdots f(x_n).$$

დავუშვათ, რომ რაიმე ჰიპოთეზის შესამოწმებლად ვატარებთ n დაკვირვებას. ამ n დაკვირვებათა ნებისმიერი ერთობლიობა ქმნის შერჩევით ერთობლიობას, რომლის მოცულობა იქნება n. ჰიპოთეზის შემოწმების მეთოდიკა მდგომარეობს რაიმე ისეთი წესის გამონახვაში, რომელიც ყოველი n მოცულობის შერჩევის შესწავლის შედეგად მოგვცემს შესამოწმებელი ჰიპოთეზის მიღების ან უარყოფის საშუალებას.

ანდა, სხვანაირად, ჰიპოთეზათა შემოწმების მეთოდიკა მდგომარეობს n მოცულობის მქონე ყველა შესაძლებელ შერჩევათა ერთობლიობის ორ ურთიერთარაგადამკვეთ სიმრავლედ დაყოფაში და შემდეგი წესის გამოყენებაში: შესამოწმებელი ჰიპოთეზა იქნება უარსაყოფი, თუ დაკვირვების შედეგად მიღებული შერჩევა მოხვდება I ქვესიმრავლეში, ხოლო თუ ის მოხვდება II ქვესიმრავლეში, მაშინ მიიღება

შესამოწმებელი. ჰიპოთეზა. პირველ ქვესიმრავლეს კრიტიკულ არეს უწოდებენ ამგვარად, შემოწმების ამა თუ იმ მეთოდიკის შერჩევა კრიტიკული არის შერჩევის ეკვივალენტურია.

საილუსტრაციოდ განვიხილოთ რამდენიმე მაგალითი:

ვთქვათ, გვაქვს დამზადებულ პროდუქციათა ერთობლიობა, რომლის მოცულობაა N, და ვაწარმოებთ მთელი პარტიის შემოწმებას. ვიღებთ მთელი პარტიიდან შერჩევით ერთობლიობას, რომლის შემოწმების შედეგად მთელ პარტიას ან მივიღებთ, ან არა. იგულისხმება, რომ მთელი პარტია შედგება მხოლოდ ვარგისი და წუნდებული პროდუქციისაგან. წუნდებულთა ფარდობითი ხვედრი p წინასწარ უცნობია. ვთქვათ p_0, რომელიც მოთავსებულია 0-სა და 1 შორის, ისეთია, რომ, თუ $p \leqslant p_0$, მაშინ მთელი პარტია მიიღება; წინააღმდეგ შემთხვევაში კი, ე. ი. თუ $p > p_0$, მთელი პარტია არ მიიღება.

ვთქვათ, შერჩევითი ერთობლიობის მოცულობა არის n, რომლის შემოწმების შემდეგ ხდება $p \leqslant p_0$ ჰიპოთეზის მიღება ან უარყოფა. ამ შემთხვევაში კრიტიკული არე შემდეგნაირად განისაზღვრება: $p \leqslant p_0$ ჰიპოთეზა არ მიიღება მხოლოდ იმ შემთხვევაში, თუ წუნდებულთა ფარდობითი ხვედრი n მოცულობის მქონე შერჩევაში გადააჭარბებს წინასწარ სათანადოდ შერჩეულ C მუდმივ სიდიდეს.

აგრეთვე, ვთქვათ რაიმე სიდიდის სიგრძეს ვზომავთ იარაღით, რომლის ცდომილებანი გაზომვების დროს ისეთ ნორმალური განაწილების კანონს ემორჩილება, რომლის დისპერსია ერთის ტოლია და მათემატიკური ლოდინი a-სი. ასე რომ, ფაქტიურად გასაზომი სიდიდის ნამდვილი სიგრძე a-ს ტოლი იქნება. დავუშვათ, ჩვენ ვამოწმებთ ჰიპოთეზას, რომელიც გვეუბნება, რომ გასაზომი სიდიდის ნამდვილი სიგრძე არის a_0. ამ ჰიპოთეზის შემოწმება უნდა მოხდეს n რიცხვჯერ, გაზომვის ჩატარების შედეგად მიღებულ x_1, x_2, …, x_n მნიშვნელობათა შესწავლის საფუძველზე. ამ შემთხვევაში კრიტიკული არე შემდეგნაირად განისაზღვრება: $a = a_0$ ჰიპოთეზის უარყოფა მხოლოდ იმ შემთხვევაში მოხდება, როდესაც დაკვირვების შედეგად მიღებული მნიშვნელობანი ისეთია, რომ $|\bar{x} - a_0| \geqslant C$, სადაც $\bar{x}$ არის n დაკვირვების შედეგად მიღებულ მნიშვნელობათა საშუალო არითმეტიკული, ხოლო C სათანადოდ შერჩეული მუდმივია.

საზოგადოდ, კრიტიკული არის შერჩევის უამრავი ხერხი არსებობს. ასე მაგალითად, უკანასკნელ შემთხვევაში ჩვენ შეგვეძლო საშუალო არითმეტიკულის მაგივრად აგვეღო შემთხვევითი სიდიდის სხვა რომელიმე მახასიათებელი. სახელდობრ, მედიანა ან მოდა, ან საშუალო გეომეტრიული, ან საშუალო ჰარმონიული, ანდა დაკვირვებათა შედეგად მიღებული სხვა რომელიმე საშუალო სიდიდე.

ცხადია, რომ სხვადასხვანაირი მიდგომით შერჩეული კრიტიკული არეები შემოწმებისას ერთნაირად კარგ შედეგებს როდი იძლევიან. ამიტომ სტატისტიკურ ჰიპოთეზათა შემოწმებასთან დაკავშირებული ძირითადი პრობლემა მდგომარეობს სწორედ ისეთი პრინციპების დამყარებაში, რომლებიც კრიტიკული არეების სათანადოდ შერჩევის შესაძლებლობას მოგვცემენ. ასეთი პრინციპების წამოყენება ნეიმანსა და პირსონს ეკუთვნით და სწორედ ამ პრინციპებს აქვთ ფუნდამენტური მნიშვნელობა სტატისტიკურ ჰიპოთეზათა შემოწმების თეორიაში.

თეორიულად მეტად საინტერესოა ის კერძო შემთხვევა, როდესაც X შემთხვევითი სიდიდის განაწილება მხოლოდ ერთ ისეთ პარამეტრზეა დამოკიდებული, რომელსაც მარტო ორი θ_0 და θ_1 მნიშვნელობის მიღება შეუძლია.

ვთქვათ, $f(x, \theta)$ არის X შემთხვევითი სიდიდის განაწილება θ პარამეტრის ნებისმიერი მნიშვნელობისათვის. $f_0(x)$-ით აღვნიშნოთ $f(x, \theta_0)$ განაწილება, ხოლო $f_1(x)$-ით $f(x, \theta_1)$ განაწილება. დავუშვათ, ჩვენ ვამოწმებთ $\theta=\theta_0$ ჰიპოთეზას, რომელსაც ნულოვანი ჰიპოთეზა ეწოდება და H_0-ით აღინიშნება. $\theta=\theta_1$ ჰიპოთეზა აღინიშნება H_1-ით და მას θ_0-ის მოქიშპე ჰიპოთეზა ეწოდება. ამგვარად, ჩვენს წინაშე დგას ამოცანა, X შემთხვევით სიდიდეზე n ურთიერთდამოუკიდებელი დაკვირვების შედეგად მიღებული $x_1, x_2, \ldots, x_n$ მნიშვნელობების შესწავლის საფუძველზე შევამოწმოთ H_0 ჰიპოთეზა მოქიშპე H_1 ჰიპოთეზის მიმართ.

როგორც ვიცით, არსებობს კრიტიკულ არეთა სიმრავლე. არსებითია მათგან ისეთის არჩევა, რომელიც აღებულ კონკრეტულ ამოცანას ეთანადება. ამ მიზნით ნეიმანმა და პირსონმა შემდეგი მოსაზრება წამოაყენეს: H_0 ნულოვანი ჰიპოთეზის მიღების ან უარყოფის დროს შესაძლებელია ორგვარი შეცდომის დაშვება. ჩვენ დავუშვებთ პირველი გვარის შეცდომას, თუ უარვყოფთ H_0 ჰიპოთეზას მაშინ, როდესაც ის მართალია, ხოლო დავუშვებთ მეორე გვარის შეცდომას, თუ მივიღებთ H_0-ს, როდესაც მართალი მოქიშპე H_1 ჰიპოთეზაა. პირველი და მეორე გვარის შეცდომათა ალბათობები W კრიტიკული არის შერჩევით ცალსახად განისაზღვრებიან. სინამდვილეში, პირველი გვარის შეცდომის ალბათობა ტოლია დაკვირვების შედეგად მიღებული შერჩევის კრიტიკულ არეში მოხვედრის ალბათობისა, რომელიც გამოთვლილია H_0 ჰიპოთეზის დაშვებით, ხოლო მეორე გვარის შეცდომის ალბათობა ტოლია დაკვირვებათა შერჩევის კრიტიკულ არეში არმოხვედრის ალბათობისა, რომელიც გამოთვლილია H_1 ჰიპოთეზის დაშვებით.

ნებისმიერი W მოცემული კრიტიკული არისათვის პირველი და

შეორე გვარის შეცდომათა ალბათობანი აღინიშნებიან სათანადოდ α და β-თი.

α და β ალბათობებს აქვთ მეტად მნიშვნელოვანი პრაქტიკული ინტერპრეტაცია. ვთქვათ, გვაქვს საკმაოდ დიდი რაოდენობა n მოცულობის მქონე შერჩევებისა. დავუშვათ, რომ ასეთ შერჩევათა რაოდენობა არის m და ვიგულისხმოთ, რომ ყოველი ცალკეული შერჩევის დროს უარვყოფთ H_0 ჰიპოთეზას, თუ შერჩევა ხვდება W არეში, და H_0-ს ვღებულობთ, თუ შერჩევა არ ხვდება W-ში. ამგვარად, ჩვენ ვაკეთებთ m დასკვნას H_0 ჰიპოთეზის მიღების ან უარყოფის შესახებ. საზოგადოდ, ზოგიერთი ამ დასკვნათაგანი შეიძლება იყოს ყალბი. თუ H_0 ჰიპოთეზა მართალია, ხოლო m დიდია, მაშინ თითქმის ერთის ტოლი ალბათობით შეიძლება ჩაითვალოს, რომ ყალბ დასკვნათა წილი დაახლოებით α-ს ტოლი იქნება. თუ მართალია H_1 ჰიპოთეზა, მაშინ თითქმის ერთის ტოლი ალბათობით შეიძლება ყალბ დასკვნათა წილი დაახლოებით β-ს ტოლად ჩაითვალოს. მაშასადამე, შესაძლებელია დავასკვნათ, რომ შერჩევების საკმაოდ დიდი რიცხვის დროს ყალბ დასკვნათა წილი α-ს ტოლია, თუ H_0 ჰიპოთეზა სწორია, და ტოლია β-სი, თუ H_1 ჰიპოთეზაა სწორი.

გასაგებია, რომ, თუ რომელიმე W კრიტიკულ არეს შეესაბამება α და β-ს მცირე მნიშვნელობანი, მაშინ ასეთი არე ჩვენთვის უფრო მისაღებია. კრიტიკული არის ისეთნაირად შერჩევა შეიძლება, რომ ან α, ან β გავხადოთ რაც შეიძლება მცირე, ხოლო ორივე მათგანის n შერჩევითი მოცულობის ფიქსირებული მნიშვნელობისათვის ერთდროულად მცირედ გახდომა შეუძლებელია.

ნათქვამის საილუსტრაციოდ განვიხილოთ ორი ზღვარითი შემთხვევა:

1) W კრიტიკული არე ცარიელია, ე. ი. ჩვენ ყოველთვის ვღებულობთ H_0 ჰიპოთეზას, მიუხედავად იმისა, თუ როგორია ცდის შედეგი. ამ შემთხვევაში $\alpha=0$ და $\beta=1$.

2) W კრიტიკული არე მოიცავს ყველა შესაძლებელ შერჩევათა სიმრავლეს, ე. ი. ჩვენ ყოველთვის უარვყოფთ H_0 ჰიპოთეზას, ამ შემთხვევაში $\alpha=1$ და $\beta=0$.

თუ რაიმე გარემოების გამო გადავწყვეტთ, რომ განვიხილოთ მხოლოდ ისეთი კრიტიკული არე W, რომლისთვისაც α ალბათობა წინასწარ მოცემული სიდიდის ტოლია და ასეთ არეთა სიმრავლე გვაქვს, მაშინ მათგან ისეთი უნდა შევარჩიოთ, რომლისთვისაც β მინიმალურია. ამ შემთხვევაში α-ს კრიტიკული არის დონე ეწოდება, ხოლო $1-\beta$ სიდიდეს—კრიტიკული არის სიმძლავრე.

ერთნაირი დონის მქონე კრიტიკულ არეთა შორის იმას, რომელსაც ყველაზე მეტი სიმძლავრე აქვს, უფრო მეტად მძლავრი არე ეწოდება.

რადგანაც მინიმალურ β-ს შერჩევა 1—β-ს მაქსიმალური მნიშვნელობის შერჩევას ნიშნავს, ამიტომ ნეიმან-პირსონის პრინციპის ჩამოყალიბება, რომელიც კრიტიკული არის შერჩევას ეხება, შემდეგნაირად შეიძლება: თუ შევჩერდებით ერთნაირი დონის მქონე არეების შერჩევაზე, მაშინ მათ შორის კრიტიკულ არედ უნდა შევარჩიოთ ის, რომელიც უფრო მეტად მძლავრია.

თუ შერჩევის ფიქსირებული მოცულობისას ვისარგებლებთ უფრო მეტად მძლავრი კრიტიკული არით, მაშინ β იქნება α-ს ცალსახა ფუნქცია. მაშასადამე, თუ მოცემულია დაკვირვებათა რიცხვი, რომელზედაც დამყარებულია შემოწმება, მაშინ ჩვენ შეგვიძლია ნებისმიერად რომელიმე ერთის ან α-ს, ან β-ს შერჩევა. ნეიმან-პირსონის თეორია ასეთი შერჩევის საკითხს ღიად ტოვებს. საზოგადოდ, ცხადია, რომ, თუ α მცირეა, β იქნება დიდი და, პირიქით, თუ α დიდია, β იქნება მცირე. α-ს ან β-ს სიდიდის შერჩევა ყოველ კერძო შემთხვევაში ძირითადად იმაზეა დამოკიდებული, თუ რამდენად არსებითია ჩვენთვის პირველი და მეორე გვარის შეცდომები. თუ, მაგალითად, ზარალი პირველი გვარის შეცდომით შეადგენს 100 მანეთს, ხოლო მეორე გვარის შეცდომებით—1 მანეთს, მაშინ ცხადია, რომ ჩვენ გვინდა α ალბათობა იყოს რაც შეიძლება პატარა და β ალბათობა იყოს მასზე უფრო დიდი და არავითარ შემთხვევაში პირიქით.

(ნეიმანმა და პირსონმა აჩვენეს, რომ არე, რომელიც შედგება x_1, x_2, ..., x_n ყველა შერჩევისაგან და რომლისთვისაც ადგილი აქვს

$$\frac{f_1(x_1)f_1(x_2)\cdots f_1(x_n)}{f_0(x_1)f_0(x_2)\cdots f_0(x_n)} \geqslant k \tag{1.3}$$

უტოლობას, არის H_0 ჰიპოთეზის შესამოწმებლად H_1 ჰიპოთეზის მიმართ უფრო მეტად მძლავრი არე. k მუდმივის შერჩევა ისეთნაირად ხდება, რომ კრიტიკულ არეს ჰქონდეს განსაზღვრული α დონე.)

ისმება კითხვა, რატომ განისაზღვრება მოცემული შემოწმებისათვის უფრო მეტად კრიტიკული არე (1.3) უტოლობით? სიმარტივისათვის ვიგულისხმოთ, რომ H_0 და H_1 ჰიპოთეზებისას X შემთხვევითი სიდიდის ალბათობის განაწილება დისკრეტულია, ე. ი. $f_i(x_1)\, f_i(x_2) \ldots \ldots f_i(x_n)$ (სადაც $i=0,1$) ნიშნავს ისეთი შერჩევის მიღების ალბათობას, რომელიც დაკვირვების შედეგად მიღებულს ემთხვევა. (1.3) უტოლობით განსაზღვრული კრიტიკული არის აგება შეიძლება დავიწყოთ იმ $E'=(x_1', x_2', \ldots, x_n')$ შერჩევით, რომლისთვისაც

$$\frac{f_1(x_1)f_1(x_2)\dots f_1(x_n)}{f_0(x_1)f_0(x_2)\dots f_0(x_n)} \tag{1.4}$$

ფარდობა მაქსიმალურია. ამის შემდეგ იმავე არეში შეგვყავს $E^2=(x_1^2, x_2^2, \dots, x_n^2)$ შერჩევა, რომლისთვისაც (1.4) ფარდობა მაქსიმალურია ყველა შერჩევათა შორის, საიდანაც ამოღებულია E' შერჩევა. საზოგადოდ, იმის შემდეგ, რაც კრიტიკულ არეში შეყვანილია E^1, E^2, ..., E^r რიცხვით r შერჩევა, ამავე არეში შეიყვანება აგრეთვე E^{r+1} შერჩევა, რომელსაც დარჩენილ ყველა სხვა შერჩევათა შორის აქვს მაქსიმალური (1.4) ფარდობა. ეს პროცესი მანამ გრძელდება, სანამ კრიტიკული არის დონე არ მიაღწევს მოცემულ α სიდიდეს. რადგანაც ამ პროცესის ნებისმიერ მომენტში კრიტიკულ არეში შეყვანილ უკანასკნელ შერჩევას აქვს უფრო მეტი ალბათობათა ფარდობა H_1 და H_0 ჰიპოთეზების შესრულებისას, ვიდრე ნებისმიერ სხვა შერჩევას, რომელიც ჯერ კიდევ არ არის შესული კრიტიკულ არეში. ამიტომ ცხადია, რომ კრიტიკული არის ალბათობა, როცა მართალია H_1 ჰიპოთეზა, მეტია ან ტოლი იმავე დონის მქონე სხვა ნებისმიერი არის სიმძლავრეზე.

მაგალითისათვის განვიხილოთ შემთხვევითი სიდიდე, რომელიც განაწილებულია ისეთი ნორმალური კანონით, რომლის დისპერსია ერთია, ხოლო მათემატიკური ლოდინი $a=a_0$. ვთქვათ, ეს იყოს H_0 ჰიპოთეზა. ვთქვათ აგრეთვე, მოცემულია a_1 და აგრეთვე H_1 ჰიპოთეზა, რომელიც იმაში მდგომარეობს, რომ x განაწილებულია ისეთი ნორმალური კანონით, რომლის დისპერსია ერთია და მათემატიკური ლოდინი $a=a_1$. ვიგულისხმოთ, რომ $a_1 > a_0$. H_0 ჰიპოთეზის H_1-ის მიმართ შემოწმებისას უნდა განვსაზღვროთ

$$\frac{f_1(x_1)f_1(x_2)\dots f_1(x_n)}{f_0(x_1)f_0(x_2)\dots f_0(x_n)}$$

ფარდობა.

რადგანაც

$$f_1(x_1)f_1(x_2)\cdots f_1(x_n)=\frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-a_1)^2\right\}$$

და

$$f_0(x_1)f_0(x_2)\cdots f_0(x_n)=\frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-a_0)^2\right\},$$

ამიტომ (1.3) უტოლობას ექნება შემდეგი სახე:

$$\frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-a_1)^2\right\}}{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-a_0)^2\right\}} \geqslant k$$

ანდა, თუ უტოლობის ორივე მხარეს გავალოგარითმებთ, დავწერთ:

$$\frac{1}{2}\sum_{i=1}^{n}(x_i-a_0)^2 \quad \frac{1}{2}\sum_{i=1}^{n}(x_i-a_1)^2 \geqslant \ln k,$$

ეს უკანასკნელი კი შემდეგნაირად ჩაიწერება

$$(a_1-a_0)\sum_{i=1}^{n}x_i+\frac{1}{2}n(a_0^2-a_1^2)\geqslant \ln k,$$

საიდანაც

$$\sum_{i=1}^{n}x_i \geqslant \frac{\ln k - 1/2\, n(a_0^2-a_1^2)}{a_1-a_0}=k'. \tag{1.5}$$

მიღებულ

$$\sum_{i=1}^{n}x_i \geqslant k' .$$

უტოლობის ორივე მხარეს გამოვაკლოთ na_0 და შემდეგ ორივე მხარე გავყოთ n-ზე, გვექნება

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-a_0) \geqslant \frac{k'-na_0}{n}=k''. \tag{1.6}$$

ახლა k'' შევარჩიოთ ისე, რომ (1.6) უტოლობით განსაზღვრული არე ეთანადებოდეს $\alpha=0{,}05$ დონეს.

ცხადია, რომ H_0 ჰიპოთეზის დროს

$$t=\frac{1}{n}\sum_{i=1}^{n}(x_i-a_0)$$

შემთხვევითი სიდიდის მათემატიკური ლოდინი იქნება ნული, ხოლო დისპერსია $\frac{1}{n}$.

მართლაც,

$$E(t)=\frac{1}{n}\sum_{i=1}^{n}E(x_i-a_0)=\frac{1}{n}\sum_{i=1}^{n}[E(x_i)-a_0]=0,$$

რადგანაც

$$E(x_i)=a_0.$$

$$D(t)=\frac{1}{n^2}\sum_{i=1}^{n}D(x_i-a_0)=\frac{1}{n^2}\sum_{i=1}^{n}D(x_i)=\frac{1}{n},$$

რადგანაც

$$D(x_i)=1.$$

ამიტომ, ნორმალური განაწილების ცხრილის დახმარებით მივიღებთ, რომ

$$k''=\frac{1,64}{\sqrt{n}}.$$

ამგვარად, 0,05 დონის მქონე უფრო მეტად მძლავრი არე არის ყველა იმ შერჩევათა ერთობლიობა, რომელთათვისაც ადგილი აქვს უტოლობას:

$$\frac{1}{n}\sum_{i=1}^{n}(x_t-a_0)\geqslant\frac{1,64}{\sqrt{n}}. \tag{1.7}$$

ამ უკანასკნელი უტოლობით განსაზღვრული არის შესანიშნავი თვისება სწორედ იმაში მდგომარეობს, რომ ის დამოუკიდებელია მოქიშპე a_1 სიდიდისაგან. ამ უტოლობის გამოყვანის დროს ჩვენ ვსარგებლობდით მხოლოდ $a_1>a_0$ უტოლობით. მაშასადამე, კრიტერიუმი, რომელიც განსაზღვრულია (1.7) უტოლობით, არის უფრო მეტად მძლავრი ყველა a_1-თან შედარებით, ე. ი. არის თანაბრად უფრო მეტად მძლავრი კრიტერიუმი იმ შემთხვევაში, როდესაც მოქიშპე მნიშვნელობა პარამეტრისა a_0 სიდიდეს არ აღემატება.

ზემოთ ჩვენ ვგულისხმობდით, რომ შერჩევის მოცულობა n და ალბათობა α მოცემულია და ვეძებდით მათ შესაბამის კრიტიკულ არეს, რომლისთვისაც β ალბათობა მინიმალურია

ახლა საინტერესოა, თუ როგორ განისაზღვრება შერჩევის მოცულობა n, თუ მოცემულია α და β ალბათობანი. ამ შემთხვევაში ამოცანა იმაში მდგომარეობს, რომ ვიპოვოთ n-ის მინიმალური მნიშვნელობა, რომლისთვისაც α დონის მქონე უფრო მეტად მძლავრი არის სიმძლავრე მეტია ან ტოლი $1-\beta$ სიდიდის.

β_n-ით აღვნიშნოთ მეორე გვარის შეცდომის ალბათობა, რომელიც დაკავშირებულია α დონის მქონე უფრო მეტად მძლავრ კრიტიკულ

არესთან, იმ შემთხვევაში, როდესაც ჰიპოთეზის შესამოწმებლად საჭიროა n დაკვირვების ჩატარება. შესაძლებელია იმის ჩვენება, რომ n-ის გადიდებით β_n მცირდება, ყოველ შემთხვევაში ის არ გაიზრდება. საზოგადოდ, $\beta_n \to 0$, როცა $n \to \infty$. აღვნიშნოთ $n(\alpha, \beta)$-თი n-ის უმცირესი მნიშვნელობა, რომლისთვისაც $\beta_n \leqslant \beta$. მაშინ იმ კრიტერიუმის მისაღებად, რომლისთვისაც პირველი გვარის შეცდომის ალბათობაა α, ხოლო მეორე გვარის შეცდომის ალბათობა არ აღემატება β-ს, არსებული თეორიის თანახმად, საჭიროა $h \geqslant n(\alpha, \beta)$ მოცულობის შერჩევის აღება. უფრო მეტად მძლავრი კრიტიკული არის გამოყენების დროს ჩვენ უნდა ავიღოთ $n = n(\alpha, \beta)$ მოცულობის მქონე შერჩევითი ერთობლიობა.

ჰიპოთეზის შემოწმება ნიშნავს ორ მოსალოდნელ მოქმედებათა შორის რომელიმე ერთ-ერთის არჩევას. ვთქვათ, ვირჩევთ რომელიმე მოქმედებას—I ან II, თუ რომელს ავარჩევთ, ეს დამოკიდებულია x შემთხვევითი სიდიდის განაწილების θ პარამეტრზე. ω-თი აღვნიშნოთ θ-ს ყველა მნიშვნელობათა სიმრავლე, რომელთათვისაც I ვანიჭებთ უპირატესობას. ცხადია, რომ θ-ს ყველა სხვა მნიშვნელობისათვის, რომელიც არ შედის ω-ში, უპირატესობა ენიჭება II. თუ შერჩევის მოცულობა არის n, მაშინ დასმული ამოცანა იგივეა, რაც რაიმე H ჰიპოთეზის შემოწმება იმასთან დაკავშირებით, θ-ს მიღებული მნიშვნელობა ეკუთვნის თუ არა ω-ს. თუ შემოწმების შედეგად ვღებულობთ H-ს, ვასრულებთ I მოქმედებას, თუ შემოწმების შედეგად უარვყავით H, მაშინ ვასრულებთ II მოქმედებას.

მაგალითად, ვთქვათ ვამოწმებთ მზა პროდუქციისაგან შედგენილ რაიმე პარტიას. წუნდებულთა ხვედრი p წინასწარ უცნობია. შესაძლებელია ორი მოქმედების ჩატარება: პარტია მთლიანად მიიღება ან არ მიიღება. საზოგადოდ, არსებობს p-ს ისეთი მნიშვნელობა p', რომ თუ $p < p'$, მაშინ მთელი პარტია მიიღება, თუ $p > p'$, მაშინ პარტია არ მიიღება და, თუ $p = p'$, მაშინ სულერთია, პარტიას მივიღებთ თუ არა. ვთქვათ, პარტიის მიღებას ვაწარმოებთ მისგან შემთხვევით შერჩეული n ერთეულის შემოწმების შედეგად. მაშინ ჩვენი ამოცანა იქნება H ჰიპოთეზის შემოწმება იმის შესახებ—ადგილი აქვს თუ არა მოცემული შერჩევისათვის $p \leqslant p'$ უტოლობას. მთელი პარტია მიიღება ან არ მიიღება იმასთან დამოკიდებულებით, მიიღება თუ არა H ჰიპოთეზა.

კრიტიკული არის α დონის არჩევა დიდადაა დამოკიდებული პირველი და მეორე გვარის შეცდომათა ფარდობით მნიშვნელოვნებაზე. თუ ჰიპოთეზის შემოწმება საჭიროა ორი მოქმედებიდან რომელიმე ერთის შესარჩევად, მაშინ პირველი და მეორე გვარის შეცდომათა

ფარდობითი მნიშვნელობა გამომდინარეობს ამა თუ იმ გადაწყვეტილების მიღების პრაქტიკული შედეგების განხილვიდან იმ შემთხვევაში, როდესაც ჩვენთვის საინტერესო პარამეტრის სიდიდე უფრო მეტ უპირატესობას ანიჭებს მიღებული გადაწყვეტილების საწინააღმდეგოს.

§ 2. ჰიპოთეზის შემოწმება საშუალო მნიშვნელობათა ტოლობის შესახებ

დავუშვათ, რომ გენერალური საშუალოების მნიშვნელობები ჩვენთვის უცნობია. დაკვირვების შედეგად მიღებულ მნიშვნელობათა საშუალოები შევადაროთ ერთმანეთს და ვეცადოთ სათანადო დასკვნები გავაკეთოთ გენერალური ერთობლიობების შესახებ.

ვთქვათ, პირველი გენერალური ერთობლიობიდან აღებული შერჩევითი ერთობლიობის საშუალოა $\overline{x}$, ხოლო მეორე გენერალური ერთობლიობიდან აღებული შერჩევითი ერთობლიობის საშუალო $\overline{y}$. დავუშვათ, რომ შერჩევის მოცულობანია სათანადოდ n_x და n_y.

თუ დავუშვებთ, რომ განხილული გენერალური ერთობლიობები ნორმალურად არიან განაწილებული ან შერჩევის მოცულობანი ძალიან დიდია, მაშინ, როგორც ცნობილია, $\overline{x}$ და $\overline{y}$ ნორმალურად იქნებიან განაწილებული ანდა, ყოველ შემთხვევაში, ძალიან ახლოს იქნებიან ნორმალური განაწილების კანონთან. რადგანაც $\overline{x}$ და $\overline{y}$ ემორჩილებიან ნორმალურ კანონს, მათი სხვაობაც დაემორჩილება იმავე განაწილების კანონს და, მაშასადამე, ვიტყვით, რომ $\overline{y}-\overline{x}$ სხვაობა ემორჩილება შემდეგ ნორმალური განაწილების კანონს:

$$N\left(Z;\ a_y-a_x,\ \sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right).$$

ეს გარემოება შესაძლებლობას გვაძლევს $a_y-a_x=0$ ნულოვანი ჰიპოთეზისათვის ავაგოთ კრიტიკული არე, თუკი, გარდა დაშვებულისა, ვიგულისხმებთ, რომ დისპერსიები σ_x^2 და σ_y^2 საკმაოდ ზუსტად არიან შეფასებული. შეფასების კრიტერიუმად მიღებულია შემდეგი ნორმირებული გადახრა:

$$Z=\frac{\overline{y}-\overline{x}}{\sigma_{(\overline{y}-\overline{x})}},$$

ხოლო $q\%$ მნიშვნელოვნების დონით კრიტიკულ არედ მიღებულია

$$|Z|>Z_q$$

აბსოლუტური მნიშვნელობით უდიდესი გადახრის არე.

თუ ნულოვანი ჰიპოთეზის საწინააღმდეგო ჰიპოთეზაა $a_y > a_x$, ანდა პირიქით, $a_x > a_y$, მაშინ q მნიშვნელოვნების დონისათვის ნდობის არე განისაზღვრება სათანადოდ შემდეგი უტოლობებით:

$$Z > Z_{2q}.$$

$$Z < - Z_{2q}.$$

ზემოთ თქმულის საილუსტრაციოდ მოვიყვანოთ შემდეგი მაგალითი:

ვთქვათ, ორ A და B ქარხანაში დამზადებული ელექტრო-ნათურებიდან ვარჩევთ 50 ცალს თითოეულიდან. შემოწმებით დადასტურდა, რომ A ქარხანაში დამზადებული ნათურების ხანგრძლიობათა საშუალო არითმეტიკულია 1282 საათი, ხოლო საშუალო კვადრატული გადახრა —80 საათი; მეორე B ქ რხნისათვის დავუშვათ იგივე მაჩვენებლებია 1208 და 94 საათი.

ჯერ დავადგინოთ — არსებობს თუ არა რეალური ხარისხობრივი განსხვავება A და B ქარხანაში დამზადებულ ელექტრონათურებს შორის.

რადგანაც σ_x^2 და σ_y^2 უცნობია, საჭიროა მათი შეფასება შერჩევითი მოცემულობით. ცხადია, შეფასების დროს გარკვეულ ცდომილებას ექნება ადგილი, მაგრამ რადგანაც შერჩევითი ერთობლიობის მოცულობა უდრის 50-ს, ამიტომ ცდომილება არ იქნება მაინც და მაინც საგრძნობი. ჩვენ საკმაოდ დიდი სიზუსტით დავწერთ:

$$\sigma_{(\overline{y}-\overline{x})} \cong \sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}} = \sqrt{\frac{80^2}{50}+\frac{94^2}{50}} =$$

$$= \sqrt{\frac{15236}{50}} \approx 17{,}5 \quad \text{საათს.}$$

დავუშვათ, რომ კრიტიკული არის მნიშვნელოვნების დონე არის ისევ 5%, მაშინ X ცხრილიდან ვიპოვით, რომ $Z_q = Z_5 = 1{,}96$; ამიტომ საშუალოთა შორის განსხვავება, რომელიც აბსოლუტური მნიშვნელობით აღემატება

$$Z_5 \sigma_{(\overline{y}-\overline{x})} \cong 1{,}96 \cdot 17{,}5 = 34{,}3 \text{ საათს,}$$

უნდა ჩაითვალოს არსებითად.

ჩვენ გვქონდა

$$\overline{y} - \overline{x} = 1282 - 1208 = 74 \text{ საათს.}$$

ეს გადახრა ხვდება კრიტიკულ არეში იმ შემთხვევაშიაც კი, როცა მნიშვნელოვნების დონეს გაცილებით დაბალს ავიღებთ: მაგალითად, როცა $q = 0{,}01\%$. ამგვარად, ჩვენ შეგვიძლია დავასკვნათ, რომ A ქარხნის ნათურათა ხანგრძლიობა მეტია, ვიდრე B ქარხნისა.

ახლა $a_y - a_x$ სხვაობისათვის ნდობის ინტერვალის დადგენა შეიძლება. მაგალითად, ნდობის ალბათობას $p=0{,}95$-ს შეესაბამება $a_y - a_x$ სხვაობისათვის ინტერვალი საზღვრებით:

$$\alpha = \overline{y} - \overline{x} - Z_5 \sigma_{(\overline{y}-\overline{x})} = 74 - 34{,}3 = 39{,}7 \text{ საათს}$$

და

$$\beta = \overline{y} - \overline{x} + Z_5 \sigma_{(\overline{y}-\overline{x})} = 74 + 34{,}3 = 108{,}3 \text{ საათს.}$$

ამგვარად,

$$39{,}7 < a_y - a_x < 108{,}3.$$

§ 3. ჰიპოთეზის შემოწმება საშუალო მნიშვნელობათა ტოლობაზე მცირე რაოდენობის შერჩევის დროს

ამ შემთხვევაში იგულისხმება, რომ გენერალური ერთობლიობები ემორჩილებიან ნორმალური განაწილების კანონს. გარდა ამისა, უნდა დავუშვათ, რომ გენერალურ ერთობლიობათა დისპერსიები ერთმანეთის ტოლია

$$\sigma_x^2 = \sigma_y^2 = \sigma^2.$$

მცირე მოცულობის შერჩევათა დროს დისპერსიათა შეფასება მეტად ძნელია და მოსალოდნელია რიგი ცდომილებანი. ამ შემთხვევაში გამოიყენება სტიუდენტის კრიტერიუმი, რომელიც სარგებლობს შერჩევის დროს მიღებული s_x^2 და s_y^2 დისპერსიებით. ცხადია, გენერალურ ერთობლიობათა დისპერსიების ტოლობის დაშვება ახალი კრიტერიუმის მოქმედების ასპარეზს საგრძნობლად ზღუდავს.

დავუშვათ, რომ შერჩევით ერთობლიობათა მოცულობანია n_x და n_y, ხოლო გენერალურ ერთობლიობათა ცენტრები — a_x და a_y, დისპერსია კი იყოს σ^2.

განვიხილოთ შემდეგი სიდიდე:

$$Z = \frac{\overline{y} - \overline{x} - (a_y - a_x)}{\sigma_{(\overline{y}-\overline{x})}} = \frac{\overline{y} - \overline{x} - (a_y - a_x)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{\overline{y} - \overline{x} - (a_y - a_x)}{\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}.$$

Z-თან ერთად შემოვიყვანოთ კიდევ შემდეგი სიდიდე

$$v^2 = \frac{n_x s_x^2 + n_y s_y^2}{\sigma^2},$$

რომელიც, როგორც ცნობილია, განაწილებულია χ^2 კანონით, რომლის თავისუფლების ხარისხია $n_x + n_y - 2$.

თუ განვიხილავთ Z/v ფარდობას, ის ემორჩილება ცნობილ t განაწილებას. ამგვარად,

$$t=\frac{Z}{v}=\frac{\overline{y}-\overline{x}-(a_y-a_x)}{\sqrt{n_x s_x^2+n_y s_y^2}}\sqrt{\frac{n_x n_y\,(n_x+n_y-2)}{n_x+n_y}}$$

განაწილება დაემორჩილება სტიუდენტის განაწილების კანონს n_x+n_y-2 თავისუფლების ხარისხით. მაშინ, $a_y=a_x$ ნულოვანი ჰიპოთეზის შესამოწმებლად ჩვენ შევარჩევთ q მნიშვნელოვნების დონით კრიტიკულ არეს, რომელიც განისაზღვრება უტოლობით

$$|\,t\,|>t_{q,k},$$

სადაც

$$q=n_x+n_y-2,$$

ხოლო $t_{q,k}$ განისაზღვრება ცხრილების დახმარებით.

ამის შემდეგ ვამოწმებთ, ხვდება თუ არა შერჩევითი მნიშვნელობა

$$t=\frac{\overline{y}-\overline{x}}{\sqrt{n_x s_x^2+n_y s_y^2}}\sqrt{\frac{n_x n_y(n_x+n_y-2)}{n_x+n_y}}$$

ნაჩვენებ კრიტიკულ არეში.

კრიტერიუმი მხოლოდ მაშინ იქნება ძალაში, თუ გენერალურ ერთობლიობათა დისპერსიები ტოლია. თუ ეს პირობა დარღვეულია, მაშინ t კრიტერიუმის გამოყენება არ შეიძლება. ამიტომ ზუსტად უნდა შემოწმდეს დისპერსიების თანაფარდობის საკითხი და მხოლოდ შემდეგ უნდა მოხდეს t კრიტერიუმის გამოყენება. წინააღმდეგ შემთხვევაში ამ კრიტერიუმის გამოყენებით შესაძლებელია ყალბი დასკვნის გაკეთება.

თუ ნაჩვენებია, რომ $a_y=a_x$ ნულოვანი ჰიპოთეზა მიღებულია, მაშინ საინტერესოა ვიცოდეთ, თუ როგორია გენერალურ საშუალოთა შორის მოსალოდნელი განსხვავება. მოცემული p ნდობის ალბათობისათვის ნდობის საზღვრები იძლევიან საშუალოთა შორის განსხვავების მინიმალურ და მაქსიმალურ მნიშვნელობებს.

§ 4. დისპერსიათა ფარდობაზე დამყარებული კრიტერიუმი

იმისათვის, რომ შევამოწმოთ ჰიპოთეზა დისპერსიათა ტოლობის შესახებ, საჭიროა განვიხილოთ შერჩევით დისპერსიათა ფარდობა და შევისწავლოთ ამ ფარდობის განაწილების კანონი იმ დაშვებით, რომ გენერალური ერთობლიობანი, საიდანაც შერჩევას ვაწარმოებთ, ემორჩილებიან ნორმალური განაწილების კანონს.

თუ X და Y შემთხვევითი სიდიდეები ურთიერთდამოუკიდებელნია და განაწილებული არიან χ^2 კანონით, რომლის თავისუფლების ხარისხია სათანადოდ r_1 და r_2, მაშინ, როგორც ცნობილია,

$$F=\frac{\frac{X}{r_1}}{\frac{Y}{r_2}}=\frac{r_2X}{r_1Y}$$

სიდიდის განაწილების სიმკვრივეს შემდეგი სახე ექნება:

$$\varphi_F(f)=\frac{\Gamma\left(\frac{r_1+r_2}{2}\right)r_1^{\frac{r_1}{2}}r_2^{\frac{r_2}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}f^{\frac{r_1-2}{2}}(r_2+r_1f)^{\frac{r_1+r_2}{2}},$$

რომელსაც F-ის განაწილების სიმკვრივე ეწოდება r_1 და r_2 თავისუფლების ხარისხით.

ამ F-ის განაწილების დახმარებით შევამოწმოთ ჰიპოთეზა დისპერსიათა ტოლობის შესახებ.

დავუშვათ, რომ X და Y შემთხვევითი სიდიდეები განაწილებულია ნორმალური კანონით. ვთქვათ ვიღებთ ნორმალურად განაწილებული გენერალური ერთობლიობებიდან ურთიერთდამოუკიდებელ ორ შერჩევას, რომელთა მოცულობანია სათანადოდ n_x და n_y. შერჩევით ერთობლიობათა დისპერსიები აღვნიშნოთ სათანადოდ s_x^2 და s_y^2-ით.

განვიხილოთ შემთხვევითი სიდიდეები

$$\frac{n_xs_x^2}{\sigma_x^2} \text{ და } \frac{n_ys_y^2}{\sigma_y^2},$$

რომლებიც განაწილებული არიან χ^2 კანონით თავისუფლების ხარისხებით სათანადოდ $r_1=n_x-1$; $r_2=n_y-1$. მაშინ

$$\frac{n_x s_x^2}{(n_x-1)\sigma_x^2} \text{ და } \frac{n_y s_y^2}{(n_y-1)\sigma_y^2}$$

სიდიდეთა ფარდობას ექნება $\varphi_F(f)$ განაწილების სიმკვრივე. თუ დავუშვებთ ჰიპოთეზას $\sigma_x=\sigma_y$, ვნახავთ, რომ

$$F=\frac{\left(\frac{n_x}{n_x-1}\right)s_x^2}{\left(\frac{n_y}{n_y-1}\right)s_y^2}=\frac{\bar{s}_x^2}{\bar{s}_y^2}$$

გამოსახულებას, სადაც $\overline{s}_x^2$ და $\overline{s}_y^2$ არიან X და Y შემთხვევით სიდიდეთა დისპერსიების გადაუადგილებელი შეფასებანი, ექნება F განაწილება $r_1 = n_x - 1$ და $r_2 = n_y - 1$ თავისუფლების ხარისხით.

დისპერსიათა ტოლობის შესახებ ჰიპოთეზის შემოწმებისას გენერალურ ერთობლიობათა საშუალო მნიშვნელობებთან დაკავშირებით არავითარი დაშვებები არა გვქონია. როცა ვამოწმებდით საშუალოთა ტოლობაზე ჰიპოთეზას, ვეყრდნობოდით t განაწილების რომელიმე სხვა უცნობი პარამეტრებისაგან დამოუკიდებლობას. ასევე, განხილულ შემთხვევაში ვეყრდნობით F განაწილების გენერალური ერთობლიობის დანარჩენ სხვა უცნობი პარამეტრებისაგან დამოუკიდებლობას.

§ 5. ჰიპოთეზის შემოწმება ორი ურთიერთდამოუკიდებელი შერჩევის ერთსა და იმავე გენერალური ერთობლიობისათვის მიკუთვნებაზე

ვთქვათ, გვაქვს ორი შერჩევა, რომელთაგან ერთის მოცულობაა n_1, ხოლო მეორის—n_2. ვიგულისხმოთ, რომ ეს შერჩევანი ერთმანეთისაგან დამოუკიდებელნი არიან და შეესაბამებიან X_1 და X_2 შემთხვევით სიდიდეებს. დავუშვათ X_1 და X_2 შემთხვევით სიდიდეთა განაწილების კანონები შესაბამის გენერალურ ერთობლიობებში არიან $F_1(x)$ და $F_2(x)$ უწყვეტი ფუნქციები. ამოცანა იმაში მდგომარეობს, რომ მოცემულ დაკვირვებათა მიხედვით შეიძლება თუ არა დავასკვნათ, რომ განაწილების კანონები იგივურად ტოლნი არიან: $F_1(x) \equiv F_2(x)$. ვაჩვენოთ, რომ ის გენერალური ერთობლიობანი, რომელთაგანაც შერჩევას ვაწარმოებთ, იგივური არიან.

ვთქვათ $Sn_1(x)$ და $Sn_2(x)$ არიან ემპირიული განაწილების ფუნქციები, რომლებიც თეორიული განაწილების ფუნქციების $F_1(x)$ და $F_2(x)$-ის სათანადო მიახლოებებს წარმოადგენენ.

ემპირიულ ფუნქციათა შორის განსხვავების ზომად მივიღოთ მათი სხვაობის აბსოლუტური მნიშვნელობის მაქსიმუმი:

$$Dn_1n_2 = \max \mid Sn_2(x) - Sn_1(x) \mid .$$

Dn_1n_2-ის უშუალო გამოთვლა შეიძლება, ან შეიძლება $Sn_1(x)$ და $Sn_2(x)$ ემპირიულ ფუნქციათა შესაბამისი კიბისებური მრუდების აგება.

იმ შემთხვევაში, როდესაც ჰიპოთეზა

$$F(x) = F_1(x) = F_2(x)$$

სამართლიანია x-ის ნებისმიერი მნიშვნელობისათვის, $Dn_1n_2\sqrt{\dfrac{n_1n_2}{n_1+n_2}}$

სიდიდე n_1 და n_2-ის დიდი მნიშვნელობისათვის $F(x)$ ფუნქციის სახისაგან დამოუკიდებლად მიახლოებით ემორჩილება კოლმოგოროვის $K(\lambda)$ განაწილების კანონს

$$P\left\{ D n_1 n_2 < \lambda \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \right\} \underset{\substack{n_1 \to \infty \\ n_2 \to \infty}}{\to} K(\lambda) = 1 - 2 \sum_{\nu=1}^{\infty} (-1)^{\nu-1} e^{-2\nu^2 \lambda^2},$$

ნებისმიერი $\lambda > 0$-თვის.

ეს უკანასკნელი საშუალებას გვაძლევს შევამოწმოთ ჰიპოთეზა თეორიული განაწილების ფუნქციათა ტოლობის შესახებ, როცა შერჩევითი მოცულობანი საკმაოდ დიდია და ისინი 50-ს მაინც აღემატებიან. როცა n_1 და n_2 მცირე სიდიდეებია, მაშინ ზღვარითი თეორემის გამოყენება არ შეიძლება. ამ შემთხვევაში გამოყენებული უნდა იქნეს ფორმულა, რომელიც $D n_1 n_2$-ის ზუსტ განაწილების კანონს შეესაბამება.

α მნიშვნელოვნების დონისათვის კრიტიკული არე ჰიპოთეზის შესამოწმებლად

$$D n_1 n_2 > \lambda_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

უტოლობით განისაზღვრება.

ეს უკანასკნელი იმას ნიშნავს, რომ განსხვავება, რომელიც აღემატება უკანასკნელი უტოლობის მარჯვენა მხარის ზღვარს, შეიძლება α-ზე ნაკლები ცდომილების ალბათობით არსებითად ჩაითვალოს.

§ 6. ნორმალურობის ჰიპოთეზის შემოწმება

როგორც ვიცით, ნორმალური განაწილების ფუნქციას აქვს შემდეგი სახე:

$$\Phi(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{ -\frac{(z-a)^2}{2\sigma^2} \right\} dz.$$

იმ შემთხვევაში, როდესაც გვაქვს ძირითადი ნორმალური განაწილება ($a = 0$, $\sigma = 1$), ნორმალური განაწილების ფუნქცია შემდეგ სახეს ღებულობს:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{ -\frac{z^2}{2} \right\} dz.$$

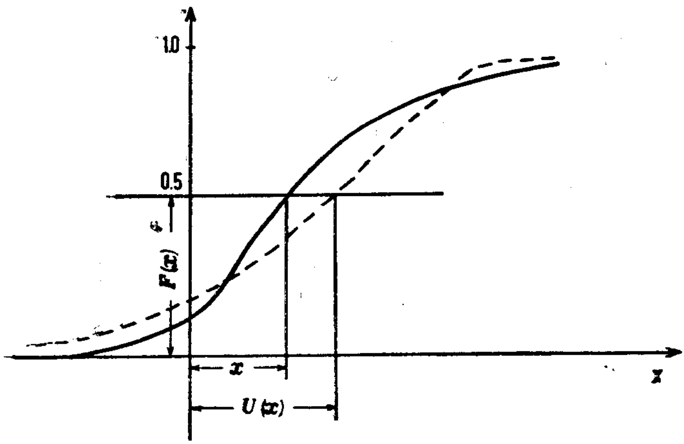
თუ $F(x)$ არის ნებისმიერი უწყვეტი განაწილების ფუნქცია, მაშინ მისი წარმოდგენა ყოველთვის შეიძლება $\Phi(x)$ ნორმალური განაწილების საშუალებით შემდეგნაირად:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u(x)} \exp\left\{-\frac{z^2}{2}\right\} dz = \Phi\,[u(x)],$$

სადაც $u(x)$ არის x-ის მონოტონურად ზრდადი რომელიმე ფუნქცია.

რომ დავრწმუნდეთ ამის სამართლიანობაში, ავაგოთ $F(x)$ და $\Phi(x)$ ფუნქციათა გრაფიკები, რომელთაც შემდეგი სახე ექნებათ (ნახ. 21).

აქ $u(x)$ არის ნორმალური განაწილების მრუდის ისეთი წერტილის აბსცისა, რომლის ორდინატი უდრის განსახილავი $F(x)$ ინტეგრალური მრუდის ორდინატს. ცხადია, რომ ასეთნაირად განხილული $u(x)$ აბსცისა იქნება x-ის ფუნქცია. ამგვარად, თუ ჩვენ შევუსაბამებთ ყოველ x-ს სათანადო $u(x)$-ს, რომლისთვისაც ნორმალური წირის ორდინატი $F(x)$-ის ორდინატის ტოლია, მაშინ ჩვენ მივიღებთ სავსებით განსაზღვრულ $u(x)$ ფუნქციას. ამასთან, ყველგან, სადაც $F(x)=0$, მივიღებთ, რომ $u(x)=-\infty$, ხოლო, სადაც $F(x)=1$, ყველგან მივიღებთ $u(x)=+\infty$.



ნახ. 21.

თუ გვაქვს ნორმალური განაწილება, მაშინ

$$F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-a}{\sigma}} \exp\left\{-\frac{z^2}{2}\right\} dz = \Phi\left(\frac{x-a}{\sigma}\right).$$

მაშასადამე, მხოლოდ და მხოლოდ ამ შემთხვევაში $u(x)$ ფუნქცია იქნება წრფივი ფუნქცია

$$u = \frac{x-a}{\sigma}.$$

(x, u) სიბრტყეზე $u(x)$ ფუნქცია ისე გამოისახება, როგორც წრფე, რომელიც გადის $(a, 0)$ წერტილში და რომლის საკუთხო კოეფიციენტია $\frac{1}{\sigma}$.

ვთქვათ, მოცემულია k წერტილი, რომელთა კოორდინატებია (x_1, F_1); (x_2, F_2), ..., (x_k, F_k), სადაც $F_1, F_2, \ldots, F_k$ არიან $F(x)$ უცნობი განაწილების ფუნქციის მნიშვნელობანი $x_1, x_2, \ldots x_k$ წერტილებში. რომ გავარკვიოთ $F(x)$ განაწილების ფუნქცია ეკუთვნის თუ არა ნორმალური განაწილების ფუნქციათა ტიპს, ჩვენ შეგვიძლია ყოველ F_i-სათვის განვსაზღვროთ სათანადო მნიშვნელობა $u(x_i) = u_i$ ფუნქციისა განტოლებიდან

$$F_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u(x_i)} \exp\left\{-\frac{z^2}{2}\right\} dz,$$

ხოლო შემდეგ (x, u) სიბრტყეზე ავაგოთ (x_i, u_i) წერტილი. თუ მიღებული წერტილები რომელიმე წრფეზე განლაგდებიან, ეს იმას ნიშნავს, რომ ჩვენი ჰიპოთეზა უცნობი $F(x)$ განაწილების ფუნქციის ნორმალობის შესახებ მართებულია.

თუ ჩვენთვის ცნობილია ორი წყვილი მნიშვნელობა (x_1, u_1) და (x_2, u_2), მაშინ a და σ განისაზღვრებიან შემდეგი განტოლებებიდან:

$$u_1 = \frac{x_1 - a}{\sigma} \quad \text{და} \quad u_2 = \frac{x_2 - a}{\sigma},$$

საიდანაც,

$$a = x_1 - u_1\sigma = x_2 - u_2\sigma \quad \text{და} \quad \sigma = \frac{x_1 - x_2}{u_1 - u_2}.$$

იმ შემთხვევაში, როდესაც ჩვენს განკარგულებაშია მხოლოდ შერჩევითი ერთობლიობა, მაშინ ჰიპოთეზის შესამოწმებლად გამოვიყენებთ $S_n(x)$ ემპირიული განაწილების ფუნქციას. როგორც ცნობილია, $S_n(x)$ ემპირიული ფუნქცია წარმოადგენს კიბისებურ მრუდს, რომელიც ყოველ $x = x_i$ წერტილში წარმოადგენს $x_1 < x_2 < \ldots < x_n$ ვარიაციული მწკრივის ერთ-ერთ წევრს. ემპირიული ფუნქცია აკეთებს ნახტომს $\frac{i-1}{n}$ მნიშვნელობიდან $\frac{i}{n}$-ზე. პირობით შეგვიძლია მივიღოთ, რომ

$$S_n(x_i) = \frac{i - 1/2}{n} = \widetilde{S}_i.$$

$\widetilde{u}_i$-ს მნიშვნელობა განვსაზღვროთ

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\widetilde{u}_i} \exp\left\{-\frac{z^2}{2}\right\} dz = \widetilde{S}_i$$

განტოლებიდან და განვიხილოთ $(x_i;\ \widetilde{u}_i)$ $(i=1,\ 2,\ \ldots,\ n)$ წერტილთა სიმრავლე $(x;\ u)$ სიბრტყეზე. n-ის დიდი მნიშვნელობისათვის უფლება გვაქვს $\widetilde{S}_i$ ჩავთვალოთ $F(x_i)$-ის რომელიმე შეფასებად და ამიტომ $(x_i;\ \widetilde{u}_i)$ წერტილები ახლოს უნდა იყოს სათანადო $(x_i;\ u_i)$ წერტილებთან. თუ ეს უკანასკნელი წერტილები მდებარეობენ ერთ $u=\dfrac{x-a}{\sigma}$ წრფეზე, მაშინ $(x_i;\ \widetilde{u}_i)$ წერტილები მხოლოდ შემთხვევით გადასცდებიან ამ წრფეს.

ამგვარად, მიახლოებითი წრფივობა $(u_i;\ x_i)$ წერტილთა განლაგებაში შეგვიძლია ჩავთვალოთ გენერალური ერთობლიობის ნორმალური განაწილების ნიშნად

§ 7. თანადობის კრიტერიუმი

ვთქვათ გენერალური ერთობლიობის განაწილების კანონი უცნობია. შესაძლებელია რაიმე მოსაზრებათა საფუძველზე უცნობი განაწილება მივაკუთვნოთ განაწილებათა რომელიმე ჯგუფს. დავუშვათ არის მოსაზრება, რომ ეს განაწილება ნორმალურია.

უცნობი განაწილების შესახებ ჰიპოთეზის შემოწმება ისევე ხდება, როგორც ამას ვაკეთებდით პარამეტრების შესახებ ჰიპოთეზათა შემოწმების დროს, ე. ი. სპეციალურად შერჩეული შემთხვევითი სიდიდის დახმარებით, რომელსაც თანადობის კრიტერიუმი ეწოდება.

თანადობის კრიტერიუმად იწოდება ნაგულისხმევი უცნობი განაწილების კანონის შესახებ ჰიპოთეზის შემოწმების კრიტერიუმი.

არსებობს თანადობის კრიტერიუმთა სიმრავლე, რომელთა შორის განვიხილავთ პირსონის χ^2 კრიტერიუმს.

პირსონის კრიტერიუმი გამოვიყენოთ ნორმალობის ჰიპოთეზის შესამოწმებლად. ე. ი. უნდა შემოწმდეს ჰიპოთეზა იმის შესახებ, რომ გენერალური ერთობლიობა განაწილებულია ნორმალური კანონით.

რა თქმა უნდა, პირსონის თანადობის კრიტერიუმი სხვა განაწილების შემთხვევაშიაც ანალოგიურად გამოიყენება.

შევადაროთ ერთმანეთს თეორიული და ემპირიული სიხშირეები. ემპირიული სიხშირეები დაკვირვებათა შედეგად მიიღებიან, ხოლო თეორიული კი განაწილების ნორმალობის დაშვებით გამოითვლებიან.

ცხადია, საზოგადოდ თეორიული და ემპირიული სიხშირეები ერთმანეთისაგან განსხვავებულნი იქნებიან.

მაგალითისათვის დავუშვათ, რომ გამოთვლათა შედეგად მივიღეთ შემდეგი სიხშირეები:

ემპირიული—2,7, 17, 22, 19, 10, 2, 1

თეორიული—2,7 16, 22, 19, 1C, 3, 1.

ემპირიული და თეორიული სიხშირეები ერთმანეთისაგან განსხვავდებიან. ისმება კითხვა: შემთხვევითია თუ არა ეს განსხვავება? შესაძლებელია, რომ განსხვავება შემთხვევითია, უმნიშვნელოა და აიხსნება დაკვირვებათა მცირე რაოდენობით, ანდა მათი დაჯგუფების წესით, ანდა სხვა რაიმე მიზეზით. შესაძლებელია მიღებული განსხვავება არ არის შემთხვევითი, მნიშვნელოვანია*და აიხსნება იმით, რომ თეორიული სიხშირეები გამოთვლილია არასწორი ჰიპოთეზის დაშვებით გენერალური ერთობლიობის შესახებ, თითქოს ის იყოს განაწილებული ნორმალური კანონის მიხედვით. დასმულ საკითხებს პასუხობს პირსონის კრიტერიუმი. პირსონის კრიტერიუმი კი არ ამტკიცებს ჰიპოთეზის სამართლიანობას, არამედ ადგენს მნიშვნელოვნების სათანადო დონის შესაბამისად, რამდენად ეთანხმება ან არ ეთანხმება ის დაკვირვების შედეგებს.

დავუშვათ, გვაქვს შერჩევითი ერთობლიობა, რომლის მოცულობაა m. ვთქვათ დაკვირვების შედეგები წარმოადგენენ შემდეგ ცხრილს:

ელემენტები (ვარიანტები) $x_1, x_2, \ldots, x_m$.

ემპირიული სიხშირეები $n_1, n_2, \ldots, n_m$.

ვთქვათ გენერალური ერთობლიობა განაწილებულია ნორმალური კანონით და n_i' თეორიული სიხშირეები გამოთვლილია ამ დაშვების საფუძველზე. α მნიშვნელოვნების დონის შემთხვევაში საჭიროა ნულოვანი ჰიპოთეზის შემოწმება, ე. ი. გენერალური ერთობლიობის ნორმალურად განაწილების შესახებ ჰიპოთეზის შემოწმება.

ნულოვანი ჰიპოთეზის შემოწმების კრიტერიუმად მივიღოთ:

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'} \tag{7.1}$$

შემთხვევითი სიდიდე. ეს სიდიდე შემთხვევითია იმიტომ, რომ სხვადასხვა ცდის დროს ის ღებულობს წინასწარ უცნობ სხვადასხვა მნიშვნელობებს. ცხადია, რამდენადაც ნაკლებად განსხვავდებიან ემპირიული და თეორიული სიხშირეები, იმდენად მცირე იქნება (7.1) კრიტერიუმი

და, მაშასადამე, ის გარკვეული სიზუსტით ახასიათებს ემპირიული და თეორიული განაწილების სიახლოვეს.

ცხადია, სიხშირეთა სხვაობის კვადრატში აყვანა აგვაცილებს უარყოფითი და დადებითი სხვაობების ურთიერთგაბათილების შესაძლებლობებს, ხოლო ამ სხვაობის კვადრატის n_i' თეორიულ სიხშირეზე გაყოფა ყველა შესაკრებს ამცირებს; წინააღმდეგ შემთხვევაში ჯამი იქნებოდა იმდენად დიდი, რომ ის მიგვიყვანდა ნულოვანი ჰიპოთეზის უარყოფამდე მაშინაც კი, როცა ის მართალია.

დამტკიცებულია, რომ როცა m საკმაოდ დიდია, მაშინ (7.1) სიდიდის განაწილების კანონი, იმისდა მიუხედავად, როგორაა განაწილებული გენერალური ერთობლიობა, მიისწრაფვის χ^2 (ხი კვადრატი) განაწილების კანონისაკენ K თავისუფლების ხარისხით. ამიტომ (7.1) შემთხვევითი სიდიდე აღნიშნულია χ^2-ით, ხოლო თვითონ თანადობის კრიტერიუმს უწოდებენ „ხი კვადრატ“ კრიტერიუმს.

თავისუფლების ხარისხის რაოდენობას ითვლიან ფორმულით

$$K = S - 1 - r,$$

სადაც S ჯგუფი (კერძო ინტერვალების) რიცხვია შერჩევაში, r არის ნაგულისხმევ განაწილებაში შესაფასებელ პარამეტრთა რაოდენობა, რომლებიც მოცემული შერჩევით ფასდებიან.

მაგალითად, თუ ნაგულისხმევი განაწილება ნორმალურია, მაშინ აფასებენ ორ პარამეტრს. მათემატიკურ ლოდინს და საშუალო კვადრატულ გადახრას. ამიტომ, ამ შემთხვევაში $r=2$ და თავისუფლების ხარისხი იქნება

$$K = S - 1 - r = S - 3.$$

თუ ნაგულისხმევი განაწილება პუასონის კანონს ემორჩილება, მაშინ გვექნება მხოლოდ ერთი შეფასებული პარამეტრი და თავისუფლების ხარისხი იქნება

$$K = S - 2.$$

იმდენად, რამდენადაც ცალმხრივი კრიტერიუმი უფრო „მკაცრად“ უარყოფს ნულოვან ჰიპოთეზას, ვიდრე ორმხრივი, ავაგოთ მარჯვენა მხარის კრიტიკული არე იმ პირობით, რომ ამ არეში კრიტერიუმის მიხვედრის ალბათობა, როცა იგულისხმება ნულოვანი ჰიპოთეზა მართალია, ტოლი იყოს მიღებული მნიშვნელოვნების α დონის:

$$P\{\chi^2 > \chi^2_{კრ.}(\alpha, k)\} = \alpha.$$

ამგვარად, მარჯვენა მხარის კრიტიკული არე განისაზღვრება

$$\chi^2 > \chi^2_{კრ.}(\alpha, k)$$

უტოლობით, ხოლო ნულოვანი ჰიპოთეზის მიღების არე განისაზღვრება

$$\chi^2 < \chi^2_{კრ.}(\alpha, k)$$

უტოლობით.

ამგვარად, იმისათვის, რომ მოცემული მნიშვნელოვნების დონისათვის შევამოწმოთ ნულოვანი ჰიპოთეზა გენერალური ერთობლიობის ნორმალური კანონით განაწილების შესახებ, საჭიროა პირველ რიგში გამოვთვალოთ თეორიული სიხშირეები, ხოლო შემდეგ დაკვირვების შედეგად მიღებული მნიშვნელობა კრიტერიუმის

$$\chi^2_{დაკვ.} = \sum \frac{(n_i - n_i')^2}{n_i'}, \tag{7.2}$$

და χ^2 განაწილების კრიტიკულ წერტილთა ცხრილიდან, მოცემულ α მნიშვნელოვნების დონისათვის და $K = S - 3$ თავისუფლების ხარისხისათვის, მოვნახოთ $\chi^2_{კრ.}$ (α, k) კრიტიკული წერტილი.

თუ $\chi^2_{დაკვ} < \chi^2_{კრ,}$ — არა გვაქვს ნულოვანი ჰიპოთეზის უარყოფის საფუძველი.

თუ $\chi^2_{დაკვ.} > \chi^2_{კრ.}$ — ნულოვან ჰიპოთეზას უარყოფენ: ცხადია, მიღებული დასკვნები სამართლიანი იქნება დაკვირვებათა დიდი რაოდენობისათვის, სახელდობრ $m > 50$-თვის მაინც. რადგანაც ჰიპოთეზის მიღების ან უარყოფის დროს შესაძლებელია პირველი ან მეორე გვარის შეცდომა, მაშინაც კი, როცა თეორიული და ემპირიული სიხშირეები კარგადაა შეთანხმებული, საჭიროა დიდი სიფრთხილის გამოჩენა. სახელდობრ, საჭიროა ცდათა სერიის განმეორება, დაკვირვებათა რიცხვის გაზრდა, სხვა კრიტერიუმის გამოყენება, განაწილების გრაფიკის აგება, ასიმეტრიის და ექსცესის გამოთვლა.

ნათქვამი გავარჩიოთ მაგალითზე. ემპირიული და თეორიული სიხშირეები ავიღოთ 285-ე გვერდზე მოცემული ცხრილიდან. შევამოწმოთ ნულოვანი ჰიპოთეზა გენერალური ერთობლიობის ნორმალურად განაწილების შესახებ, თუ მნიშვნელოვნების დონეა $\alpha = 0,05$.

პირველ რიგში გამოვთვალოდ $\chi^2_{დაკვ.}$. ამისათვის შევადგინოთ ცხრილი.

ცხრილი 13

1	2	3	4	5	6	7	8
i	n_i	n_i'	n_i-n_i'	$(n_i-n_i')^2$	$\frac{(n_i-n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	2	2	0	0	0	4	2
2	7	7	0	0	0	49	7
3	17	16	1	1	1/16	289	18—1/16
4	22	22	0	0	0	484	22
5	19	19	0	0	0	361	19
6	10	10	0	0	0	100	10
7	2	3	—1	1	1/3	4	1—1/3
8	1	1	0	0	0	1	1
Σ	80	80			19/48		80—19/48

მივიღებთ

$$\chi^2_{\text{დაკვ.}} = 19/48.$$

მიღებული გამოთვლების სისწორის შესამოწმებლად იყენებენ ფორმულას:

$$\sum_{i=1}^{m} \frac{(n_i-n_i')^2}{n_i'} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - m,$$

რომელიც შემდეგნაირად მიიღება:

$$\sum_{i=1}^{m} \frac{(n_i-n_i^1)^2}{n_i^1} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - \sum \frac{2n_i n_i'}{n_i^1} + \sum \frac{n_i'^2}{n_i'} =$$

$$= \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 2\sum n_i + \sum n_i'.$$

მაგრამ რადგანაც, როგორც ვიცით თეორიულ და ემპირიულ სიხშირეთა ჯამი ცალ-ცალკე შერჩევითი მოცულობის ტოლია, ამიტომ დავწერთ:

$$\sum_{i=1}^{m} \frac{(n_i-n_i')^2}{n_i'} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - m.$$

ახლა კი შესამოწმებლად მე-8 სვეტში მიღებულ ჯამს გამოვაკლებთ შერჩევის მოცულობას $m=80$-ს, მივიღებთ:

$$\sum_{i=1}^{m} \frac{n_i^2}{n_i'} - m = 80^{19}/_{48} - 80 = {}^{19}/_{48}.$$

მივიღეთ იგივე, მაშასადამე გამოთვლები სწორადაა ჩატარებული.

ვიპოვოთ თავისუფლების ხარისხი K. შერჩევაში ჯგუფთა რაოდენობა (სხვადასხვა ვარიანტთა) არის 8. ამიტომ

$$K = 8 - 3 = 5.$$

χ^2 განაწილების კრიტიკულ წერტილთა VII ცხრილის დახმარებით, როცა მნიშვნელოვნების დონეა 0,05 და თავისუფლების ხარისხი $K=5$, ვიპოვოთ

$$\chi^2_{კრ.}(0,05;\ 5) = 11,1.$$

როგორც ვხედავთ

$$\chi^2_{დაკვ.} < \chi^2_{კრ.}$$

და არა გვაქვს საფუძველი უარვყოთ ნულოვანი ჰიპოთეზა.

ამგვარად თეორიულ და ემპირიულ განაწილებებს შორის სხვაობა უმნიშვნელოა და შეიძლება დამტკიცებულად ჩაითვალოს ჰიპოთეზა იმის შესახებ, რომ გენერალური ერთობლიობა განაწილებულია ნორმალური კანონების მიხედვით.

სავარჯიშო

1. ისარგებლეთ პირსონის χ^2 კრიტერიუმით, მნიშვნელოვნების დონე იყოს 0,05 და დაადგინეთ, სწორია თუ არა ჰიპოთეზა გენერალური ერთობლიობის ნორმალურად განაწილების შესახებ, თუ შერჩევის მოცულობა $n=200$ და შერჩევის განაწილება სიხშირეთა მიხედვით შემდეგია:

$$m = \begin{cases} 5,\ 7,\ 9,\ 11,\ 13,\ 15,\ 17,\ 19,\ 21 \\ 15,\ 26,\ 25,\ 30,\ 26,\ 21,\ 24,\ 20,\ 13. \end{cases}$$

მითითება: ჯერ იპოვეთ შერჩევითი საშუალო და დისპერსია

$$\overline{X} = 12,63, \quad \sigma_x = 4,695.$$

მივიღოთ $h=2$, $n=200$, $\sigma_x = 4,695$ და გამოვთვალოთ თეორიული სიხშირეები ფორმულით:

$$n_i' = \frac{nh}{\sigma_{\overline{x}}} \varphi(u_i).$$

$\varphi(u_i)$-ს მნიშვნელობებს ვიპოვით დართული ცხრილის დახმარებით. n_i'-ის გამოსათვლელად შევადგინოთ ცხრილი 14.

ცხრილი 14

i	x_i	n_i	$u_i=\frac{x_i-\overline{x}}{\Delta}$	$\varphi(u_i)$	$n'_i=85{,}2\varphi(u_i)$
1	5	15	—1,62	0,3939	36,6
2	7	26	—1,20	0,1942	16,5
3	9	25	—0,77	0,2966	25,3
4	11	30	—0,35	0,3752	32,0
5	13	26	0,08	0,3977	33,9
6	15	21	0,51	0,3503	29,8
7	17	24	0,93	0,2589	22,0
8	19	20	1,36	0,1582	13,5
9	21	13	1,78	0,0818	7,0

გამოვთვალოთ ჯამი

$$\chi_q^2=\sum\frac{(n_i-n'_i)^2}{n'_i}.$$

ამისათვის შევადგინოთ ცხრილი 15.

ცხრილი 15

i	n_i	n'	$n-n'_i$	$(n_c-n'_i)^2$	$\frac{(n_i-n'_i)^2}{n'_i}$
1	15	33,6	—18,6	345,96	10,4
2	26	16,5	9,5	90,25	3,6
3	25	25,3	—0,3	0,09	0,0
4	30	32 0	—0,2	4,00	0,1
5	26	39,9	—7,9	62,41	1,9
6	21	29,8	—8,8	77,44	2,3
7	24	22,0	2,0	4,00	0,2
8	20	13,5	6,5	42,25	3,0
9	13	7,0	6,0	36,00	5,1
					$\chi_q^2\simeq 27$

თავისუფლების ხარისხი იქნება: $K=S-3=9-3=6$.

დართული ცხრილის დახმარებით (ცხრილი VII), როცა $K=6$ და $\chi_q^2=27$, მოვნახავთ

$$P(\chi^2\geqslant\chi_q^2)=0{,}0001.$$

რამდენადაც დაკვირვებით მიღებული ალბათობა ნაკლებია 0,05 მნიშვნელოვნების დონეზე, ჰიპოთეზა გენერალური ერთობლიობის ნორმალურად განაწილების შესახებ არ ეთანხმება მოცემულ შერჩევას.

2. ისარგებლეთ პირსონის χ^2 კრიტერიუმით, მიიღეთ მნიშვნელოვნების დონედ 0,05 და დაადგინეთ, ეთანხმება თუ არა გენერალური

ერთობლიობის ნორმალურად განაწილების შესახებ ჰიპოტეზას. $n=200$ მოცულობის მქონე შერჩევითი ერთობლიობის განაწილება შემდეგია:

$$m = \begin{cases} 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9;\ 1{,}1;\ 1{,}3;\ 1{,}5;\ 1{,}7,\ 19;\ 21;\ 23; \\ 6;\ 9;\ 26;\ 25;\ 30;\ 26;\ 21;\ 24;\ 20;\ 8;\ 5; \end{cases}$$

პას. ეთანხმება [$K=8$; $\chi^2_q=7{,}71$; $P(\chi^2 \geqslant \chi^2_q)=0{,}4664$].

3. პირსონის კრიტერიუმის გამოყენებით, როცა მნიშვნელოვნების დონეა 0,03, დაადგინეთ, შემთხვევითია თუ მნიშვნელოვანია განსხვავება შემდეგ ემპირიულ და თეორიულ (ნორმალურ) განაწილებებს შორის:

$$\left.\begin{array}{ll} \text{ემპ. სიხშირეები—}n_i & \\ \text{თეორიული სიხშირეები } n'_i & \end{array}\right\} \begin{array}{l} 8,\ 16,\ 40,\ 72,\ 36,\ 18,\ 10, \\ 6,\ 18,\ 36,\ 76,\ 39,\ 18,\ \ 7. \end{array}$$

მ ი თ ი თ ე ბ ა: გამოთვალეთ χ^2_q, რისთვისაც შეადგინეთ ცხრილი. იპოვეთ $\chi^2_q=3{,}068$. იანგარიშეთ თავისუფლების ხარისხი $K=4$. დართული ცხრილით იპოვით $P(\chi^2 \geqslant \chi^2_q)=0{,}5578$. ამგვარად განსხვავება ყოფილა შემთხვევითი.

4. პირსონის კრიტერიუმის გამოყენებით, როცა მნიშვნელოვნების დონეა 0,05, დაადგინეთ, შემთხვევითია თუ მნიშვნელოვანი ემპირიულ და თეორიულ სიხშირეებს შორის სხვაობა.

თუ ემპ. სიხშირეებია 6, 8, 13, 15, 20, 16, 10, 7, 5, ხოლო თეორიული სიხშირეები კი— 5, 9, 14, 16, 18, 16, 9, 6, 7, იგულისხმება, რომ გენერალური ერთობლიობა ნორმალურია.

პას. შემთხვევითია: [$K=6$, $\chi^2_q=1{,}52$, $P(\chi^2 \geqslant \chi^2_q)=0.3679$].

5. χ^2 კრიტერიუმის გამოყენებით დაადგინეთ, თუ მნიშვნელოვნების დონეა 0,05, შემთხვევითია თუ მნიშვნელოვანი ემპირიულ და თეორიულ სიხშირეებს შორის განსხვავება, რომლებიც გამოთვლილია იმ ვარაუდით, რომ გენერალური ერთობლიობა ნორმალური კანონითაა განაწილებული.

სიხშირეთა განაწილება შემდეგია:

ემპირიული—5, 10, 20, 8, 7,
თეორიული—6, 14, 18, 7, 5.

პას. შემთხვევითია ($K=2$, $\chi^2_q=2{,}47$; $P(\chi^2 \geqslant \chi^2_q)=0{,}3679$).

6. χ^2 კრიტერიუმის გამოყენებით დაადგინეთ, თუ მნიშვნელოვნების დონეა 0,05, შემთხვევითია თუ მნიშვნელოვანი ემპირიულ და თეორიულ სიხშირეებს შორის განსხვავება, რომლებიც გამოთვლილია იმ ვარაუდით, რომ გენერალური ერთობლიობა ნორმალური კანონით არის განაწილებული?

სიხშირეთა განაწილება შემდეგია:

ემპირიული—14, 18, 32, 70, 20, 36, 10.

თეორიული—10, 24, 34, 80, 18; 22, 12.

პას. მნიშვნელოვანია ($K=4$; $\chi_q^2=13{,}93$; $P(\chi^2 \geqslant \chi_q^2)=0{,}0073$).

7. χ^2 კრიტერიუმის გამოყენებით დაადგინეთ, თუ მნიშვნელოვნების დონეა 0,05, შემთხვევითია თუ მნიშვნელოვანი ემპირიულ და თეორიულ სიხშირეებს შორის განსხვავება, რომლებიც გამოთვლილია იმ ვარაუდით, რომ გენერალური ერთობლიობა ნორმალური კანონით არის განაწილებული?

სიხშირეთა განაწილება შემდეგია:

ემპირიული—5, 7, 15, 14, 21, 16, 9, 7, 6,

თეორიული— 6, 6, 14, 15, 22, 15, 8, 8, 6.

პას. შემთხვევითია ($K=6$; $\chi_q^2=0{,}83$; $P(\chi^2 \geqslant \chi_q^2)= 0{,}9856$).

თავი V

დისპერსიული ანალიზი

§ 1. დისპერსიათა შედარების კრიტერიუმი

სანამ დისპერსიული ანალიზის მეთოდის არსს გავეცნობოდეთ, ვნახოთ, თუ რამდენად არაერთგვაროვანი და ცვალებადია დაკვირვებით მიღებული მონაცემები, და აგრეთვე შევისწავლოთ დისპერსიათა შედარების საკითხი, რომელიც დისპერსიული ანალიზის მეთოდში ძირითადია.

შემთხვევითი სიდიდის მნიშვნელობათა ცვლილება ცალკეული ცდების შემთხვევაში შესაძლებელია დამოკიდებული იყოს მრავალ მიზეზზე, ანუ ფაქტორზე. შემთხვევით სიდიდეზე მოქმედი ფაქტორები შეიძლება ორ ჯგუფად დაიყოს: პირველ ჯგუფში შევლენ ის ფაქტორები, რომლებიც ყოველი ცდის დროს მუდმივად მოქმედებენ, ხოლო მეორე ჯგუფში იქნებიან შემთხვევითი ფაქტორები, რომელთა მოქმედება ცდიდან ცდამდე იცვლება და შემთხვევითი სიდიდის მნიშვნელობანი მათი მოქმედებით ხან მცირდებიან, ხან კი დიდდებიან.

ვთქვათ, მოცემული გვაქვს X შემთხვევითი სიდიდე, რომელიც n-ჯერ დაკვირვების შედეგად ღებულობს შემდეგ მნიშვნელობებს:

$$x_1,\ x_2,\ \ldots,\ x_n.$$

მაშინ დისპერსია, რომელიც იქნება

$$D(x)=\sigma_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2,$$

ყველა ფაქტორის, როგორც მუდმივების, ისე შემთხვევითების, ცვალებადობას, ანუ გაფანტულობას ზომავს.

როგორც ვნახეთ, დისპერსიის თვისებათა დამტკიცებისას, თუ ფაქტორები ურთიერთდამოუკიდებლად მოქმედებენ, მაშინ მათი ერთობლივი მოქმედების შედეგად მიღებულ მნიშვნელობათა დისპერსია ცალცალკე მათი მოქმედებით მიღებულ მნიშვნელობათა დისპერსიების ჯამის ტოლი იქნება.

მაგალითად, თუ X შემთხვევით სიდიდეზე ერთმანეთის დამოუკიდებლად მოქმედებს ორი A და B ფაქტორი, რომლებიც $\overline{x}$ საშუალო მნიშვნელობიდან α და β-ს ტოლ გადახრებს იწვევენ, და აგრეთვე X-ზე მოქმედებენ შემთხვევითი ფაქტორები, რომლებიც x' გადახრებს იწვევენ, მაშინ X-ის რომელიმე შესაძლო მნიშვნელობა x_r შემდეგნაირად წარმოიდგინება:

$$x_r=\overline{x}+\alpha+\beta+x'.$$

X შემთხვევითი სიდიდის σ_x^2 დისპერსია იქნება:

$$\sigma_x^2=\sigma_\alpha^2+\sigma_\beta^2+\sigma_{x'}^2, \tag{1.1}$$

სადაც σ_α^2, σ_β^2, $\sigma_{x'}^2$ არიან A, B და შემთხვევითი ფაქტორების შესაბამისი დისპერსიები.

(1.1) ტოლობით გამოხატული დისპერსიათა თვისება შესაძლებლობას გვაძლევს ერთმანეთს შევადაროთ სხვადასხვა ფაქტორით გამოწვეული დისპერსიები და ამით დავადგინოთ თითოეული ფაქტორის მოქმედების სიდიდე. მაგალითად, σ_x^2 და $\sigma_{x'}^2$ დისპერსიათა შედარება შესაძლებლობას გვაძლევს დავასკვნათ A ფაქტორია უფრო მოქმედი, თუ შემთხვევითი ფაქტორები; ასევე, თუ ერთმანეთს შევადარებთ $\sigma_{x'}^2$ და σ_β^2 დისპერსიებს, გავარკვევთ A და B ფაქტორთაგან რომელი უფრო მოქმედებს აღებულ შემთხვევით სიდიდეზე.

დისპერსიათა შედარებისათვის საჭიროა ცალკეულ ფაქტორთა მოქმედების შედეგად მიღებულ დისპერსიათა ერთმანეთისაგან გამოყოფის ცოდნა.

სანამ დისპერსიათა შედარებას მოვახდენთ, ვნახოთ, თუ როგორ გამოითვლება ჩვენთვის უცნობი გენერალური ერთობლიობის დისპერსია და მოვახდინოთ მისი შეფასება.

გამოვთვლით რა შერჩევით დისპერსიას, რომელსაც S^2-ით აღვნიშნავთ, მისი დახმარებით საჭირო იქნება σ_x^2 გენერალური დისპერსიის პოვნა, ანდა, რაც იგივეა, S შერჩევითი საშუალო კვადრატული გადახრის დახმარებით უნდა ვიპოვოთ σ_x გენერალური ერთობლიობის საშუალო კვადრატული გადახრა.

ვიგულისხმოთ, რომ X შემთხვევითი სიდიდის მიერ მიღებული მნიშვნელობანი (რაც გენერალურ ერთობლიობას შეადგენს) ნორმალური განაწილების კანონს ემორჩილება და შევადგინოთ შემდეგი გამოსახულება:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_x^2},$$

რომელსაც χ^2 განაწილება აქვს. ალბათობა იმისა, რომ უცნობი σ საშუალო კვადრატული გადახრა მოთავსებული იქნება $(S-\varepsilon,\ S+\varepsilon)$ შუალედში, სადაც ε წინასწარ აღებული დადებითი სიდიდეა, $p_1 - p_2$ სხვაობის ტოლი იქნება, სადაც p_1 და p_2 არის $k = n - 1$-თვის და შემდეგი მნიშვნელობათათვის:

$$\chi_1^2 = \frac{(n-1)S^2}{(S+\varepsilon)^2}, \quad \chi_2^2 = \frac{(n-1)S^2}{(S-\varepsilon)^2},$$

VII ცხრილით განსაზღვრული ალბათობანი.

ზემოთ ნათქვამი შემდეგნაირად ჩაიწერება:

$$P(S-\varepsilon < \sigma < S+\varepsilon) = p_1 - p_2.$$

ასეთივე იქნება σ_x^2-ის $[(S-\varepsilon)^2,\ (S+\varepsilon)^2]$ შუალედში მოხვედრის ალბათობაც

$$P[(S-\varepsilon)^2 < \sigma_x^2 < (S+\varepsilon)^2] = p_1 - p_2.$$

მაგალითისათვის დავუშვათ, რომ X შემთხვევით სიდიდეზე, რომელიც ნორმალური განაწილების კანონს ემორჩილება, 15-ჯერ ერთმანეთისაგან დამოუკიდებელი შემთხვევით დაკვირვებათა შედეგად მივიღეთ, რომ $S = 6{,}7\cdot$ რა ალბათობით შეგვიძლია ვამტკიცოთ, რომ უცნობი σ_x საშუალო კვადრატული გადახრა მოთავსებული იქნება (6,5; 6,9) შუალედში?

ამგვარად, გვაინტერესებს, რომ ε სიზუსტე 0,2-ის ტოლი იყოს; გვაქვს: $\varepsilon = 0{,}2$, $k = n - 1 = 14$, მაშინ

$$\chi_1^2 = \frac{14 \cdot (6{,}7)^2}{(6{,}9)^2} = \frac{14.44{,}89}{47{,}61} = 13{,}20,$$

$$\chi_2^2 = \frac{14 \,.\, (6{,}7)^2}{(6{,}5)^2} = \frac{14 \,.\, 44{,}89}{42{,}25} = 14{,}87.$$

ნაპოვნი χ^2 მნიშვნელობისათვის და $k=14$-თვის ცხრილით ვიპოვით:

$$p_1=0{,}5265-0{,}20\cdot(0{,}5265-0{,}4497)=0{,}5111,$$

$$p_2=0{,}4497-0{,}87(0{,}4497-0{,}3782)=0{,}3875,$$

საიდაც

$$P(6{,}5<\sigma_x<6{,}9)=0{,}5111-0{,}3875=0{,}1236.$$

მაშასადამე, 0,2-ის ტოლი სიზუსტით და 0,1236-ის ტოლი სიმტკიცით (საიმედოობით) შეგვიძლია დავწეროთ:

$$\sigma_x\approx 6{,}7.$$

რაც უფრო მცირე იქნება ε სიზუსტე, მით უფრო დიდი იქნება საიმედოობა და, პირიქით.

თუ X შემთხვევითი სიდიდე არ ემორჩილება ნორმალური განაწილების კანონს, მაშინ σ_x უცნობის განსაზღვრა შეიძლება შედარებით ზუსტად, თუ დაკვირვებათა რიცხვი n საკმარისად დიდი იქნება. სახელდობრ, n-ის დიდი მნიშვნელობისათვის საკმაოდ დიდი ალბათობით შეიძლება დარწმუნებული ვიყოთ იმაში, რომ σ_x უცნობი შერჩევითი ერთობლიობის S-გან განსხვავებული იქნება არა უმეტესი $\pm\frac{3S}{\sqrt{2n}}$ სიდიდით. ეს უკანასკნელი არის შერჩევითი ერთობლიობის საშუალო კვადრატული გადახრის (S-ის) საშუალო ცდომილების საზომი.

ახლა განვიხილოთ შერჩევითი ერთობლიობის დისპერსიათა შედარება ორივე შემთხვევისათვის: როცა გენერალური ერთობლიობა ნორმალური განაწილების კანონს ემორჩილება და როდესაც გენერალური ერთობლიობა არ არის ნორმალური კანონის მიხედვით განაწილებული.

ვიგულისხმოთ, რომ გენერალური ერთობლიობა განაწილებულია ნორმალური კანონის მიხედვით. ეს გენერალური ერთობლიობა აღვნიშნოთ S-ით და მისგან ავიღოთ ორი შერჩევითი ერთობლიობა S_1 და S_2, რომელთა შესაბამისი საშუალო კვადრატული გადახრები იყოს s_1 და s_2. აღებულ შერჩევით ერთობლიობათა მოცულობანი შესაბამისად აღვნიშნოთ n_1 და n_2-ით. ვიგულისხმოთ, რომ შერჩევანი ჩატარებულია ერთმანეთის დამოუკიდებლად და ცდებიც ურთიერთდამოუკიდებელია.

s_1^2 და s_2^2 დისპერსიათა შედარება მოვახდინოთ მათი შეფარდების საშუალებით. მათი შეფარდება აღვნიშნოთ T-თი, მაშინ დავწერთ:

$$T=\frac{s_1^2}{s_2^2}.$$

ამ შეფარდებისათვის განაწილების კანონის შედგენა არავითარ სიძნელეს არ წარმოადგენს, რადგანაც S გენერალური ერთობლიობიდან, S_1 და S_2-ის მსგავსად, უამრავ შერჩევათა წარმოება შეიძლება, რომლებიც ურთიერთდამოუკიდებელნი იქნებიან. ამ კანონის მიხედვით შეიძლება შედგენილ იქნეს T-ს მნიშვნელობათა ცხრილი, რომლის საფუძველზე შეიძლება ანგარიში იმისა, რომ მოცემული p ალბათობით შემთხვევითი $\frac{s_1^2}{s_2^2}$ ფარდობა T-ზე ნაკლები არ იქნება.

პრაქტიკული მიზნებისათვის სავსებით საკმარისია ცხრილის შედგენა $p=0{,}05$ და $p=0{,}01$-თვის. შესაბამისი ცხრილები წიგნის ბოლოს არის მოყვანილი (V, VI), რომელთა შინაარსი შემდეგში მდგომარეობს: V ცხრილში მოცემულია $k_1=n_1-1$-ს და $k_2=n_2-1$-ს მნიშვნელობისათვის T-ს მნიშვნელობანი, რომელთაც ის თვისება აქვთ, რომ ცნობილი k_1 და k_2-თვის $\frac{s_1^2}{s_2^2}$ ფარდობის შემთხვევითი მნიშვნელობა $p=0{,}05$-ის ტოლი ალბათობით არანაკლები იქნება, ვიდრე სათანადო k_1 და k_2 სიდიდეებით მოცემული T-ს მნიშვნელობა V ცხრილიდან.

სხვანაირად რომ ვთქვათ, თუ $\frac{s_1^2}{s_2^2}$ ფარდობის შემთხვევით მნიშვნელობას აღვნიშნავთ T'-ით, მაშინ

$$P(T' \geqslant T)=0{,}05,$$

სადაც T-ს მნიშვნელობა, სათანადო k_1 და k_2-ის მოცემით, V ცხრილიდან აიღება.

მაგალითად, თუ ვიხილავთ ნორმალური ერთობლიობიდან ორ ურთიერთდამოუკიდებელ შერჩევით ერთობლიობებს, რომელთა მოცულობანია შესაბამისად 13 და 24, ე. ი. $n_1=13$, $n_2=24$ და, ამგვარად, $k_1=12$, $k_2=23$, მაშინ V ცხრილის დახმარებით ვიპოვით T-ს მნიშვნელობას

$$T=2{,}20.$$

მაშასადამე, 0,05-ის ტოლი ალბათობით გარანტირებული ვართ, რომ დისპერსიათა შემთხვევითი $T'=\frac{s_1^2}{s_2^2}$ შეფარდება იქნება არანაკლები 2,20-ზე.

$$P(T' \geqslant 2{,}20)=0{,}05.$$

ანალოგიურად არის შედგენილი VI ცხრილიც. 0,05 და 0,01 სიდიდეებს, რომელთათვის არის შედგენილი V და VI ცხრილები, ნდობის ალბათობებს უწოდებენ.

V და VI ცხრილები შედგენილია $T>1$-თვის, მაგრამ ეს ცხრილები ზოგად შემთხვევაშიც გამოდგება, რადგანაც შერჩევითი დისპერსიების გამოთვლის შემდეგ s_1^2-ად ყოველთვის მათ შორის უფრო დიდის მიღება შეგვიძლია.

მიღებული შეფასებანი გამომდინარეობენ მცირეალბათობიან ხდომილობათა პრაქტიკულად შეუძლებლობის პრინციპიდან. ამ პრინციპის თანახმად, თუ მოცემული k_1 და k_2-თვის დაკვირვების შედეგად მიღებული T-ს მნიშვნელობა, რომელიც $T_ϱ$-ით აღინიშნება, იქნება არანაკლები, ვიდრე V ცხრილში მოცემული T-ს მნიშვნელობა, მაშინ მისი ალბათობა, როგორც შემთხვევითი ხდომილობისა, 0,05-ზე მეტი არ იქნება და, თუ ამ ალბათობას საკმარისად მცირედ ჩავთვლით, მაშინ მიღებული მნიშვნელობა არაშემთხვევითად უნდა მივიჩნიოთ და, მაშასადამე, s_1^2 და s_2^2-ს შორის განსხვავება არსებითად უნდა ჩაითვალოს. თუ ვნახავთ, რომ $T_ϱ$ ნაკლებია ცხრილში მოცემულ T-ზე, მაშინ ის შეიძლება შემთხვევითად ჩაითვალოს, ხოლო დისპერსიათა შორის განსხვავება უმნიშვნელოდ ჩაითვლება.

ამგვარად, s_1^2 და s_2^2 დისპერსიათა შედარების შესახებ შემდეგ დასკვნამდე მივდივართ:

გამოვთვლით

$$T_ϱ = \frac{s_1^2}{s_2^2}$$

გამოსახულებას, სადაც s_1^2-ს ვგულისხმობთ უფრო მეტს, ვიდრე s_2^2-ია. შემდეგ $k_1 = n_1 - 1$ და $k_2 = n_2 - 1$ მნიშვნელობების შესაბამისად V და VI ცხრილებიდან ვიპოვით T-ს მნიშვნელობას. თუ აღმოჩნდა, რომ

$$T_ϱ \geqslant T,$$

მაშინ დისპერსიათა შორის განსხვავება არსებითად ჩაითვლება, თუ

$$T_ϱ < T,$$

მაშინ მათ შორის განსხვავება შეიძლება უმნიშვნელოდ ჩაითვალოს.

თუ შერჩევით დისპერსიათა შორის სხვაობა უმნიშვნელოა, ეს იმის მაჩვენებელი იქნება, რომ ისინი აღებულია ერთი და იგივე გენერალური ერთობლიობიდან, რომელიც ნორმალური განაწილების კანონს ემორჩილება; თუ შერჩევით დისპერსიათა შორის სხვაობა არსებითია, ეს იმის მაჩვენებელი იქნება, რომ შერჩევანი არ არის დამოუკიდებლად ჩატარებული, ანდა უმართებულოა ჩვენი დაშვება იმის შესახებ, რომ შერჩევითი ერთობლიობანი აღებულია ერთი და იგივე გენერალური ერთობლიობიდან, რომელიც ნორმალური განაწილების კანონს ემორჩილება.

უნდა შევნიშნოთ, რომ k_1 და k_2 რიცხვები ყოველთვის არ არიან n_1-1 და n_2-1 სიდიდეთა ტოლი. ზოგად შემთხვევაში ისინი განისაზღვრებიან, როგორც იმ სიდიდის ერთმანეთისაგან დამოუკიდებელ მნიშვნელობათა რიცხვი, რომლისთვისაც გამოითვლებიან შესადარებელი დისპერსიები. ამ რიცხვებს ეწოდებათ სწორედ თავისუფლების ხარისხები.

თუ X შემთხვევითი სიდიდის მიერ მიღებული მნიშვნელობანი ერთმანეთთან დაკავშირებულია r პირობით, მაშინ შერჩევითი ერთობლიობის დისპერსია განისაზღვრება ფორმულით:

$$s^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n_1-r},$$

სადაც n_1 არის შერჩევითი ერთობლიობის მოცულობა, ხოლო k_1 იქნება n_1-r.

როცა ვიხილავდით X შემთხვევით სიდიდეზე დაკვირვებით მიღებულ მნიშვნელობებს, მაშინ ისინი ერთმანეთთან დაკავშირებული იყვნენ მხოლოდ ერთი პირობით: სახელდობრ, X-ზე დაკვირვების შედეგად მიღებულ მნიშვნელობათა ჯამი შერჩევითი საშუალოსა და მისი მოცულობის ნამრავლის ტოლია. ამიტომ შერჩევითი დისპერსიების გამოსათვლელად ვიღებდით ფორმულებს:

$$s_1^2=\frac{\sum_{i=1}^{n_1}(x_i-\overline{x})^2}{n_1-1} \quad \text{და} \quad s_2^2=\frac{\sum_{i=1}^{n_2}(x_i-\overline{x})^2}{n_2-1}.$$

იმ შემთხვევაში, როდესაც T-ს მნიშვნელობათა ცხრილი ხელთ არ გვაქვს, შესაძლებელია ვიხმაროთ სხვა კრიტერიუმი დისპერსიათა განსხვავების შესაფასებლად. ეს კრიტერიუმი შემდეგი ტოლობით განისაზღვრება:

$$\theta=\frac{k_1-2}{k_1}\cdot\frac{s_2^2}{s_1^2}.$$

θ-ს თეორიული მნიშვნელობა ერთის ტოლია, ხოლო დისპერსია, როდესაც შერჩევას ვაწარმოებთ ნორმალური ერთობლიობიდან, შემდეგნაირად გამოისახება:

$$\sigma_\theta{}^2=\frac{2(k_1+k_2-2)}{k_2(k_1-4)}.$$

აქედან კი σ_θ-თვის დავწერთ:

$$\sigma_\theta = \sqrt{\frac{2(k_1+k_2-2)}{k_2(k_1-4)}}.$$

ახლა შევადგინოთ შემდეგნაირი გამოსახულება:

$$\frac{|\theta-1|}{\sigma_\theta}.$$

თუ ეს გამოსახულება $\geqslant 3$-ზე, მაშინ დისპერსიათა განსხვავება არსებითია, თუკი < 3-ზე, მაშინ დისპერსიათა შორის განსხვავება არ იქნება არსებითი. k_1 და k_2 რიცხვები ისევე განისაზღვრებიან, როგორც წინა T კრიტერიუმით სარგებლობის დროს. ამ მეორე კრიტერიუმს θ კრიტერიუმი ეწოდება და ის გამოიყენება მაშინ, როდესაც ერთ-ერთი k_1 და k_2 სიდიდეთაგან 4-ს აღემატება. განხილულ შემთხვევაში k_1 უნდა იყოს 4-ზე მეტი.

განვიხილოთ შემთხვევა, როდესაც შერჩევას ვაწარმოებთ ისეთი გენერალური ერთობლიობიდან, რომელიც ნორმალური განაწილების კანონს არ ემორჩილება. თუ აღებული გენერალური ერთობლიობა მცირედ განსხვავდება ნორმალური განაწილებისაგან, მაშინ ზემოთ მიღებული შედეგების გადატანა შეიძლება ამ შემთხვევაშიაც: თუ აღებული გენერალური ერთობლიობის განაწილება საგრძნობლად განსხვავდება ნორმალური განაწილებისაგან, მაშინ შეგვიძლია მხოლოდ გარკვეული გარანტიით შერჩევითი დისპერსიები ერთმანეთს შევადაროთ იმ შემთხვევაში, როდესაც შერჩევითი ერთობლიობის მოცულობანი საკმაოდ დიდი რიცხვებია.

ასეთ შემთხვევაში დისპერსიათა შედარების კრიტერიუმად შეიძლება შემდეგი შეფარდება ავიღოთ:

$$\frac{|s_1-s_2|}{\sqrt{\frac{s_1^2}{2n_1}+\frac{s_2^2}{2n_2}}}.$$

თუ ეს შეფარდება $\leqslant 3$-ზე, მაშინ დისპერსიათა შორის განსხვავება არსებითად ჩაითვლება, თუ ეს შეფარდება > 3-ზე, მაშინ დისპერსიათა შორის განსხვავება არაარსებითი იქნება.

§ 2. დისპერსიული ანალიზის მეთოდი

სხვადასხვა სტატისტიკური გამოკვლევისათვის უდიდესი მნიშვნელობა აქვს დისპერსიული ანალიზის მეთოდის გამოყენებას, რომელშიაც ძირითად როლს დისპერსიათა შედარება ასრულებს.

დისპერსიული ანალიზის მეთოდის არსი შემდეგში მდგომარეობს: ვთქვათ, მოცემული გვაქვს x შემთხვევითი სიდიდე, რომელზედაც დაკ-

ვირვებას ვაწარმოებთ. ცხადია, X შემთხვევით სიდიდეზე მოქმედებს სხვადასხვა ფაქტორი; ვთქვათ, ისინი არიან:

$$F_1,\ F_2,\ \ldots,\ F_i. \tag{2.1}$$

მაგალითად, ვაკვირდებით რაიმე კულტურის (პურის, სიმინდის, ბრინჯის და სხვ.) მოსავლიანობას. ცხადია, მოსავლიანობაზე გარკვეული გავლენის მოხდენა შეუძლია სხვადასხვა სასუქის გამოყენებას, მაგრამ მოსავლიანობისათვის აგრეთვე დიდი მნიშვნელობა აქვს მთელ რიგ სხვა ფაქტორებსაც, მაგალითად ამინდს, ჰაერს, დროულ მოვლას, სტიქიურ შემთხვევებს და სხვ. ვიგულისხმოთ, რომ ყოველი ფაქტორის მოქმედებას ვამოწმებთ j რიცხვჯერ; ამგვარად, გვექნება შემთხვევით სიდიდეთა ij რაოდენობა.

შემთხვევით სიდიდეზე დაკვირვების შედეგად მიღებული საერთო დისპერსია s^2-ით აღვნიშნოთ. დისპერსია დავყოთ ორ ნაწილად: ერთი ნაწილი იყოს (2.1) ფაქტორების მოქმედებით მიღებული დისპერსია, ხოლო მეორე ნაწილი შემთხვევითი ფაქტორების მოქმედებით მიღებული დისპერსია; პირველი აღვნიშნოთ $s^2{}_f$-ით, ხოლო მეორე s_e^2-ით. s_e^2 დისპერსიას უწოდებენ აგრეთვე ნარჩენ დისპერსიას. s_f^2 და s_e^2 დისპერსიათა შედარებით შეგვიძლია განვსაზღვროთ X შემთხვევით სიდიდეზე მოქმედი ფაქტორები შემთხვევით ფაქტორებთან შედარებით არსებითია თუ არა და ამით განვსაზღვროთ მთლიანად აღებულ ფაქტორთა მოქმედება X შემთხვევითი სიდიდის მიერ მიღებული ყველა მნიშვნელობის ერთობლიობაზე.

X შემთხვევით სიდიდეზე მოქმედი ფაქტორების დეტალურად შესწავლის მიზნით, თითოეული ფაქტორის მოქმედებაზე ცდები უნდა ვაწარმოოთ ცალ-ცალკე და შემდეგ X შემთხვევითი სიდიდის საშუალო მნიშვნელობათა განსხვავების შეფასება მოვახდინოთ.

პირველ რიგში ვნახოთ, თუ როგორ გამოითვლება s^2, s_f^2, s_e^2 დისპერსიები. ჯერ უნდა ვიპოვოთ X ის ყოველგვარ შესაძლო მნიშვნელობათა $\overline{x}$ საშუალო, ამისათვის საჭიროა X-ის ყველა შესაძლო მნიშვნელობათა ჯამი გავყოთ მათ რიცხვზე ij-ზე. შემდეგ ვიპოვოთ X შემთხვევით სიდიდეზე F_1 ფაქტორის მოქმედების შედეგად მიღებულ მნიშვნელობათა საშუალო; ის აღვნიშნოთ a_1-ით, შემდეგ ვიპოვოთ X შემთხვევით სიდიდეზე F_2 ფაქტორის მოქმედების შედეგად მიღებულ მნიშვნელობათა საშუალო, რომელსაც a_2-ით აღვნიშნავთ და ა. შ. ამგვარად, გვექნება შემდეგი საშუალოები:

$$\overline{x}\,,\ a_1,\ a_2,\ \ldots,\ a_i.$$

მაგალითად, a_1-ის მისაღებად, რადგანაც თითოეული ფაქტორის

მოქმედებაზე ცდას j-ჯერ ვაწარმოებთ, ყველა მიღებული მნიშვნელობა სწორედ j-ზე უნდა გაიყოს.

ამ საშუალოების პოვნის შემდეგ გამოვთვლით გადახრათა კვადრატების ჯამს როგორც მთელი ერთობლიობისათვის, ისევე შემთხვევითი ფაქტორების მოქმედების შედეგად მიღებული სიდიდეებისათვის.

გადახრათა კვადრატების საერთო ჯამი აღვნიშნოთ S-ით, ის იქნება:

$$S = \Sigma(x - \overline{x})^2,$$

სადაც x ღებულობს ij მნიშვნელობას.

ფაქტორების მოქმედების შედეგად მიღებულ სიდიდეთათვის გადახრათა კვადრატების ჯამი S_f-ით აღვნიშნოთ:

$$S_f = j[(a_1 - \overline{x})^2 + (a_2 - \overline{x})^2 + \dots + (a_i - \overline{x})^2] = j\sum_{r=1}^{i}(a_r - \overline{x})^2.$$

ნარჩენი დისპერსია, ანუ შემთხვევითი ფაქტორების მოქმედების შედეგად მიღებულ სიდიდეთათვის გადახრათა კვადრატების ჯამი, რომელიც აღინიშნება S_e-თი, გამოითვლება შემდეგი სხვაობის საშუალებით:

$$S_e = S - S_f.$$

სხვანაირად, S_e არის X შემთხვევით სიდიდეზე დაკვირვების შედეგად მიღებულ მნიშვნელობათა ფაქტორების მიხედვით საშუალოებიდან გადახრათა კვადრატების ჯამი, ე. ი. S_e-ს საპოვნელად შემდეგნაირად იქცევიან: ჯერ პოულობენ X შემთხვევით სიდიდეზე F_1 ფაქტორის მოქმედების შედეგად მიღებულ მნიშვნელობათა a_1 საშუალოდან გადახრათა კვადრატების ჯამს, შემდეგ კი—X-ზე F_2 ფაქტორის მოქმედების შედეგად მიღებულ გადახრათა კვადრატების ჯამს და ა. შ. ყველა ფაქტორის მიხედვით. ყველა მიღებულ გადახრათა კვადრატების ჯამს შევაერთებთ და მივიღებთ სწორედ S_e-ს.

როგორც კი ვიპოვით S, S_f და S_e სიდიდეებს, s^2, s_f^2 და s_e^2 სიდიდეთა პოვნა ხდება შემდეგი ფორმულებით:

$$s^2 = \frac{S}{ij - 1}, \quad s_f^2 = \frac{S_f}{i - 1}, \quad s_e^2 = \frac{S_e}{i(j - 1)}.$$

მნიშვნელში მდგომი სიდიდეები გამოსახავენ სათანადოდ აღებული დისპერსიისათვის თავისუფლების ხარისხების რიცხვებს.

ამგვარად, თუ X შემთხვევით სიდიდეზე დაკვირვების შედეგად მიღებული მნიშვნელობების ერთობლიობას ჩავთვლით შერჩევით ერთობლიობად, რომელიც გენერალური ერთობლიობიდან არის აღებული, მაშინ დისპერსიათა შესადარებლად გამოგვადგება როგორც T, ისე θ კრიტერიუმი.

ვიგულისხმოთ, რომ $F_1, F_2, \ldots, F_i$ ფაქტორები X შემთხვევით სიდიდეზე ვერ ახდენენ მნიშვნელოვან გავლენას (ასეთ დაშვებას ნულოვან ჰიპოთეზას უწოდებენ). დისპერსიული ანალიზის თეორიის თანახმად, საკმაოდ დიდი ალბათობით უნდა მივიღოთ ყველა s^2, s_f^2 და s_e^2 დისპერსიათა შორის მხოლოდ არაარსებითი განსხვავება. ამიტომ, თუ T ან θ კრიტერიუმის გამოყენება გვიჩვენებს, რომ s_f^2 და s_e^2 არსებითად განსხვავდებიან, ეს იმის მაჩვენებელი იქნება, რომ ნულოვანი ჰიპოთეზის დაშვება სამართლიანი არ არის.

თუ

$$s_f^2 > s_e^2,$$

მაშინ მივიღებთ:

$$T = \frac{s_f^2}{s_e^2}, \quad k_1 = i-1, \quad k_2 = i(j-1);$$

თუ

$$s_f^2 < s_e^2,$$

მაშინ საჭიროა ავიღოთ:

$$T = \frac{s_e^2}{s_f^2}, \quad k_1 = i(j-1), \quad k_2 = i-1.$$

T-ს შეფასება ზემოთ გვქონდა აღწერილი.

თუ გვინდა θ კრიტერიუმის გამოყენება, მაშინ უნდა გამოვთვალოთ θ და σ_θ შემდეგი ფორმულებით:

$$\theta = \frac{i(j-1)-2}{i-1}\,\frac{S_f}{S_e},$$

$$\sigma_\theta = \sqrt{\frac{2ij-6}{(i-1)(ij-i-4)}};$$

დისპერსიული ანალიზის გამოყენება შეიძლება პროდუქციის ხარისხის სტატისტიკური კონტროლით შემოწმების შემთხვევაშიც.

ვთქვათ, პროდუქციის ხარისხზე სტატისტიკური მეთოდების გამოყენებით დაკვირვების შედეგად გამოირკვა, რომ წარმოების პროცესი არ არის წონასწორობაში, მაგრამ იმ მიზეზის დადგენა, რომლითაც გამოწვეულია ეს გარემოება, უშუალოდ შეუძლებელია.

ასეთ შემთხვევაში შეიძლება გამოყენებულ იქნეს რამდენიმე ჰიპოთეზა, რომლებიც მიგვითითებენ იმ ფაქტორებზე, რომელთაც შეეძლოთ წარმოების პროცესის ნორმალური მდგომარეობიდან გამოყვანა.

იმის გასარკვევად, თუ რომელი ფაქტორი როგორ მოქმედებს და რამდენად შეუძლია მას წარმოების პროცესზე იმოქმედოს, მიმართავენ ცდებს, რომელთა ჩატარებისას ცვლიან იმ ფაქტორის მოქმედების

სიდიდეს, რომელსაც შეეძლო წარმოების პროცესის წონასწორობის მდგომარეობიდან გამოყვანა.

დისპერსიული ანალიზის გამოყენებით, ცდათა რაც შეიძლება მცირე რაოდენობით, შეგვიძლია დავადგინოთ გამოსაცდელ ფაქტორთა მოქმედების რეალური სიდიდე.

ეს ამოცანა, როგორც ვხედავთ, უშუალოდ დაკავშირებულია პროდუქციის ხარისხის კონტროლთან, რადგანაც საკონტროლო დიაგრამების საშუალებით ხდება წარმოების პროცესზე დაკვირვება და მისი ნორმალური მდგომარეობიდან გადახრის გამოვლინება, ხოლო დისპერსიული ანალიზის საშუალებით ხდება ამ გადახრათა გამომწვევი მიზეზების შეფასება. ამგვარად, ორივე ეს მეთოდი პროდუქციის ხარისხის სტატისტიკური კონტროლის სხვადასხვა სახეს წარმოადგენს.

დისპერსიული ანალიზის გამოყენება შეიძლება იმ შემთხვევაშიც, როცა გვინდა პროდუქციის ხარისხის გაუმჯობესება.

მაგალითად, პროდუქციის ხარისხის გასაუმჯობესებლად, ვთქვათ, არსებული რეჟიმის შეცვლით, ვაწარმოებთ რაიმე ახალ ღონისძიებებს. იმისათვის, რომ მოვნახოთ გამოსაცდელ ფაქტორთა როგორც ცალ-ცალკე, ისე მათი ერთობლივი მოქმედების საუკეთესო ვარიანტები, ამისათვის საჭიროა ერთმანეთს შევადაროთ ყოველგვარი შესაძლო კომბინაციისათვის გამოსაცდელ ფაქტორთა მოქმედების შედეგები. რა თქმა უნდა, ამ ღონისძიების ჩასატარებლად დისპერსიული ანალიზი დიდად დაგვეხმარება.

ამგვარად, როგორც ვხედავთ, დისპერსიული ანალიზის მეთოდის გამოყენებას მთელი რიგი ტექნიკური ამოცანების ამოსახსნელად არსებითი მნიშვნელობა აქვს. მისი დახმარებით შეგვიძლია წარმოების პროცესზე მოქმედი ამა თუ იმ ფაქტორის დადგენა, მაგრამ ამ ფაქტორის მოქმედების ვერც ხარისხის და ვერც ფორმის დადგენის შესაძლებლობას ის ვერ იძლევა. ამ უკანასკნელი ამოცანის ამოსახსნელად უნდა გამოვიყენოთ კორელაციური ანალიზი, რომელსაც შემდეგ თავში ვიხილავთ.

სავარჯიშო

დისპერსიული ანალიზის მეთოდის გამოყენების ცხადყოფის მიზნით, განვიხილოთ შემდეგი მაგალითი, რომელიც ნ. ფ. დერევიცკის მიერ იყო შედგენილი.

მაგალითი 1. ქვემომოყვანილ ცხრილში მოცემულია სათესი კულტურის ერთ-ერთი სახის მოსავლიანობაზე სხვადასხვა სასუქის მოქმედების შედეგები; ერთი და იგივე ცდა განმეორებულია 4-ჯერ, ცდათა

ჩასატარებლად 4 სახის სასუქია გამოყენებული, ერთი დაკვირვება უსასუქოდ არის ჩატარებული.

საცდელი საშუალებანი	მოსავალი				ჯამი	საშუალო
უსასუქოდ	67	67	55	42	231	57,75
$K_2O + N$	98	96	91	66	351	87,75
$K_2O+P_2O_5$	60	69	50	35	214	53,50
$N+P_2O_5$	79	64	81	70	294	73,50
$K_2O+P_2O_5+N$	90	70	79	88	327	81,75
					1417	70,85

საჭიროა მოსავლიანობაზე ამ სასუქთა მოქმედების შეფასება.

ა მ ო ხ ს ნ ა: პირველ რიგში ვიპოვოთ $\overline{x}$ საშუალო

$$\overline{x} = \frac{1417}{4 \cdot 5} = 70{,}85;$$

შემდეგ გამოვთვლით ცალკეული ფაქტორების მიხედვით საშუალოებს:

$$a_1 = \frac{231}{4} = 57{,}75; \quad a_2 = \frac{351}{4} = 87{,}75; \quad a_3 = \frac{214}{4} = 53{,}50;$$

$$a_4 = \frac{294}{4} = 73{,}50; \quad a_5 = \frac{327}{4} = 81{,}75.$$

ამ გამოთვლათა შედეგებიც შეტანილია ცხრილში. ახლა გამოვთვალოთ S, S_f და S_e გადახრათა კვადრატების ჯამი. როგორც ვიცით,

$$S = \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

გამოთვლების გამარტივების მიზნით ეს ჯამი შემდეგნაირად წარმოვადგინოთ:

$$S = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 .$$

მოცემულობის თანახმად,

$$S = 67^2 + 67^2 + 55^2 + \ldots + 88^2 - \frac{1}{20}(1417)^2 = 5698{,}55.$$

ამგვარად,

$$S = 5698{,}55.$$

ახლა ვიანგარიშოთ S_f. გვაქვს ფორმულა:

$$S_f = j \sum_{r=1}^{i} (a_r - \overline{x})^2.$$

ეს შემდეგნაირად წარმოვიდგინოთ:

$$S_f = j \sum_{r=1}^{i} (a_r - \overline{x})^2 = j \left[\sum_{r=1}^{i} a_r^2 - \frac{1}{i} \left(\sum_{r=1}^{i} a_r \right)^2 \right].$$

ჩვენს შემთხვევაში $j = 4$; $i = 5$-ს, a_1, a_2, a_3, a_4 და a_5 ცხრილშია მოცემული; ამიტომ შეგვიძლია დავწეროთ:

$$S_f = 4[(57{,}75)^2 + (87{,}75)^2 + (53{,}50)^2 + (73{,}50)^2 + (81{,}75)^2 - \frac{1}{5}(57{,}75 + $$

$$+ 87{,}75 + 53{,}50 + 73{,}50 + 81{,}75)^2] = 3536{,}3,$$

ე. ი. მივიღეთ:

$$S_f = 3536{,}3.$$

რადგანაც

$$S_e = S - S_f,$$

ამიტომ გვექნება:

$$S_e = 5698{,}55 - 3536{,}30 = 2162{,}25.$$

ახლა შეგვიძლია s^2, s_f^2 და s_e^2 დისპერსიათა ანგარიში.

გვაქვს:

$$s_f^2 = \frac{S_f}{i-1} = \frac{3536{,}30}{4} = 884{,}075.$$

$$s_e^2 = \frac{S_e}{i(j-1)} = \frac{2162{,}25}{5\,.\,3} = 144{,}150.$$

მიღებულ დისპერსიათა შედარება გვაძლევს, რომ

$$s_f^2 > s_e^2,$$

ამიტომ T_ϱ-ს ვიპოვით ფორმულით:

$$T_\varrho = \frac{s_f^2}{s_e^2} = \frac{884{,}075}{144{,}150} = 6{,}13.$$

განხილულ მაგალითში:

$$k_1 = i - 1 = 4, \quad k_2 = i(j-1) = 15,$$

რომელთა მეშვეობით VI ცხრილიდან ვიპოვით სათანადო T-ს.

$$T = 3{,}06.$$

თუ ვისარგებლებთ VI ცხრილით, ვნახავთ, რომ

$$T = 4,89.$$

ორივე შემთხვევაში T_{ϱ} მეტია ცხრილით მიღებულ T-ს მნიშვნელობებზე; ეს იმის მაჩვენებელია, რომ მოსავალზე სასუქთა მოქმედების ფაქტი დამტკიცებულია.

ამრიგად, იმ გარემოებიდან, რომ

$$T_{\varrho} = 6,13 > 4,89,$$

დავასკვნით, რომ მხოლოდ 0,01-ზე ნაკლები ალბათობით შეიძლება ველოდოთ s_f^2 და s_e^2 შორის შემთხვევით განსხვავებას, რომელიც დაკვირვებით მიღებულზე მეტი ან ტოლი იქნება.

ახლა აღებული მაგალითისათვის გამოვიყენოთ θ კრიტერიუმი. გვქონდა:

$$\theta = \frac{i(j-1)-2}{i-1}\frac{S_f}{S_e} = \frac{13}{4}\cdot\frac{3536,30}{2162,25} = 5,32;$$

$$\sigma_\theta = \sqrt{\frac{2ij-6}{(i-1)(ij-i-4)}} = \sqrt{\frac{34}{44}} = 0,879.$$

შევადგინოთ გამოსახულება

$$\frac{|\theta-1|}{\sigma_\theta} = \frac{4,32}{0,879} = 4,91 > 3.$$

ამგვარად, θ კრიტერიუმმაც იგივე დასკვნამდე მიგვიყვანა, რომელიც მიღებული გვქონდა T კრიტერიუმის გამოყენების შედეგად. მოსავლიანობაზე სასუქების მოქმედება დადასტურებულია.

შესაძლებელია აგრეთვე დავადგინოთ მოსავლიანობაზე ცალკეულ სასუქთა მოქმედების შედეგებიც, და ამით ვნახავთ, თუ რამდენად ეფექტურია ამა თუ იმ სასუქის გამოყენება.

შევადგინოთ დისპერსიული ანალიზის ზოგადი სქემა. აღებული მაგალითის შემთხვევაში მას შემდეგი სახე ექნება:

დისპერსიის სახე	კვადრატების ჯამი	თავისუფლების ხარისხთა რაოდენობა	დისპერსია
საერთო	5698,55	19	299,924
ფაქტორების მოქმედებით მიღებული	3536,30	4	884,075
ნარჩენი (s_e^2)	2162,25	15	144,150

დისპერსიული ანალიზის სქემა კი შემდეგნაირად წარმოიდგინება:

დისპერსიის სახე	კვადრატების ჯამი	თავისუფლების ხარისხთა რაოდენობა	დისპერსია
საერთო	$S=\Sigma(x-\overline{x})^2$	$ij-1$	$s^2=\dfrac{S}{ij-1}$
ფაქტორების მოქმედებით მიღებული	$S_f=j\sum\limits_{r=1}^{i}(a_r-\overline{x})^2$	$i-1$	$s_f^2=\dfrac{S_f}{i-1}$
ნარჩენი	$S_e=S-S_f$	$i(j-1)$	$s_e^2=\dfrac{S_e}{i(j-1)}$

როგორც ვნახეთ, დისპერსიული ანალიზის გამოყენებისას ერთ-ერთი მნიშვნელოვანი ადგილი უჭირავს თავისუფლების ხარისხთა რიცხვის განსაზღვრას. გავარკვიოთ, თუ რაში მდგომარეობს თავისუფლების ხარისხთა აზრი.

საერთო დისპერსიის გამოსათვლელად გვქონდა ფორმულა:

$$s^2=\frac{S}{ij-1},$$

სადაც

$$S=\Sigma(x-\overline{x})^2$$

არის ყველა ij დაკვირვებათა შედეგად მიღებულ სიდიდეთა მათ საშუალოდან გადახრათა კვადრატების ჯამი. s^2-თვის თავისუფლების ხარისხთა რიცხვი იქნება $ij-1$.

s_f^2 გამოითვლება ფორმულით:

$$s_f^2=\frac{S_f}{i-1},$$

სადაც

$$S_f=i\sum_{r=1}^{i}(a_r-\overline{x})^2,$$

ეს გამოსახულება შედგენილია $a_1, a_2, \ldots, a_i$ საშუალოების (რომლებიც მიღებულია $F_1, F_2, \ldots, F_i$ ფაქტორების მიხედვით) $\overline{x}$ საშუალოდან გადახრათა კვადრატების ჯამისაგან; ამ შემთხვევაში საერთო დაკვირვებათა მიხედვით აღებული $\overline{x}$ წარმოადგენს $a_1, a_2, \ldots, a_i$ საშუალოთა საშუალოს:

$$\bar{x} = \frac{a_1 + a_2 + \cdots + a_i}{i}.$$

ამის თანახმად, s_f^2-თვის თავისუფლების ხარისხთა რიცხვი იქნება $i-1$. როგორც ზემოთ ვნახეთ,

$$s_e^2 = \frac{S_e}{i(j-1)},$$

სადაც S_e ყველა დაკვირვების შედეგად მიღებულ სიდიდეთა $a_1, a_2, \cdots a_i$ საშუალოებიდან გადახრათა კვადრატების ჯამს წარმოადგენს. s_e^2-თვის თავისუფლების ხარისხთა რიცხვი იქნება:

$$ij - i = i(j-1),$$

ამგვარად, ნებისმიერი დისპერსიისათვის თავისუფლების ხარისხთა რიცხვი ამ დისპერსიის მისაღებად შედგენილ გადახრათა კვადრატების ჯამში შემავალ ცვლადთა რაოდენობისა და იმ სიდიდეთა რიცხვის სხვაობას წარმოადგენს, რომლიდანაც გადახრებს იღებენ.

s^2-თვის თავისუფლების ხარისხთა რიცხვი, როგორც ვხედავთ, s_f^2 და s_e^2-თვის თავისუფლების ხარისხთა რიცხვის ჯამის ტოლია. მართლაც,

$$i - 1 + i(j-1) = ij - 1.$$

თავისუფლების ხარისხთა რიცხვი ფორმალურად განვსაზღვრეთ. მისი შინაგანი არსი მდგომარეობს: თუ ნულოვანი ჰიპოთეზა სამართლიანია, მაშინ s^2, s_f^2 და s_e^2 დისპერსიები, რომლებიც მიიღებიან S, S_f და S_e გადახრათა კვადრატების ჯამის სათანადოდ თავისუფლების ხარისხთა რიცხვზე გაყოფით, წარმოადგენენ ერთი საერთო σ_x^2 დისპერსის მიახლოებით მნიშვნელობებს. ისინი ისეთი გენერალური ერთობლიობიდან არიან აღებული, რომელიც ნორმალური განაწილების კანონს ემორჩილება. ამიტომ შეიძლება დავასკვნათ, რომ ნულოვანი ჰიპოთეზის სამართლიანობისათვის საჭიროა s^2, s_f^2 და s_e^2 დისპერსიათა შორის განსხვავება მხოლოდ შემთხვევითი იყოს.

მაგალითი 2. წინა მაგალითში ჩვენ განვიხილეთ დისპერსიული ანალიზი ისეთი შემთხვევებისათვის, როდესაც სხვადასხვა ფაქტორთა მოქმედებაზე დაკვირვებათა ერთი და იგივე რაოდენობა გვქონდა, მაგრამ არსებობს ისეთი შემთხვევებიც, როდესაც ცალკეული ფაქტორების მიხედვით დაკვირვებათა რიცხვი სხვადასხვაა. ასეთ შემთხვევაში დისპერსიული ანალიზის არსის გასარკვევად განვიხილოთ შემდეგი მაგალითი:

ეს ცხრილი წარმოადგენს 1937 წ. ტაშკენტის ახლოს კუი-აულის ბრინჯის საცდელ სადგურში სამ მწკრივად დაკვირვების შედეგს. პირველი მწკრივი (S_1) საკონტროლოა, რომელშიაც ყოველგვარი სასუქის გარეშე ბრინჯის მოსავალი მოცემულია 7 დანაყოფზე, მეორე მწკრივში (S_2) ბრინჯის მოსავალი მოცემულია 3 დანაყოფზე, როდესაც დათესვის წინ იხმარეს P_2O_5 სასუქი, ხოლო შემდეგ 3 ვადაში N სასუქი. მესამე მწკრივში (S_3) ბრინჯის მოსავალი მოცემულია 4 დანაყოფზე, როდესაც თესვის წინ იხმარეს N და P_2O_5 სასუქი. იგულისხმება, რომ მოცემული მწკრივები ერთმანეთისაგან დამოუკიდებელია, მოსავალი მოცემულია თითოეულ ჰექტარზე ცენტნერებში.

S_1	S_2	S_3
35,50	43,44	36,38
32,66	47,51	31,36
30,56	53,80	42,20
36.63		36,20
42,28		
34,78		
40.20		
252,61	144,75	146,14

ჩვენი მიზანია ამ მონაცემების საფუძველზე ჩავატაროთ დისპერსიული ანალიზი.

განოყიერების ხერხები აღვნიშნოთ სათანადოდ F_1, F_2 და F_3-ით. ცალ-ცალკე ვიანგარიშოთ მათთვის მოსავალთა ჯამი, მათი საშუალოები, და შევადგინოთ შესაბამისი ცხრილი.

F_1-ის მიხედვით მოსავალთა ჯამი იქნება 252,61 ც, F_2-ის მიხედვით—144,75 ც, ხოლო F_3-ის მიხედვით—146,14 ც. F_1-თვის a_1 საშუალო იქნება:

$$a_1 = \frac{252,61}{7} = 36,09;$$

F_2-თვის a_2 იქნება:

$$a_2 = \frac{144,75}{3} = 48,26;$$

ხოლო F_3-თვის a_3 იქნება:

$$a_3 = \frac{146,14}{4} = 36,53.$$

მაშინ, ცხრილს შემდეგი სახე ექნება:

განოყიერების ხერხები	მოსავალთა ჯამი	განმეორებათა რიცხვი	საშუალო
F_1	252,61	7	36,09
F_2	144,75	3	48,26
F_3	146,14	4	36,53

ამ მონაცემებით შევადგინოთ დისპერსიული ანალიზის ცხრილი. წინა შემთხვევის ანალოგიურად გამოითვლება გადახრათა კვადრატების ჯამი. დაკვირვებათა რიცხვი იქნება $7+3+4=14$.

$$S=(35,50)^2+(32,66)^2+\ldots+(40,20)^2+\ldots+(36,20)^2-$$
$$-\frac{(35,50+32,66+\ldots+36,20)^2}{14}=553,3442.$$

ფაქტორების მიხედვით გადახრათა კვადრატების ჯამი გამოითვლება სხვანაირად, ვიდრე წინა შემთხვევის დროს გვქონდა. გამოვთვალოთ ჯერ S_f:

$$S_f=\frac{(252,61)^2}{7}+\frac{(144,75)^2}{3}+\frac{(146,14)^2}{4}-$$
$$-\frac{(252,61+144,75+146,14)^2}{14}=339,9391.$$

S_e ჯამი იქნება:

$$S_e=S-S_f=553,3442-339,9391=213,4051.$$

თავისუფლების ხარისხთა რიცხვი შემდეგნაირად იანგარიშება:

საერთო დისპერსიის თავისუფლების ხარისხთა რიცხვი იქნება ყველა დაკვირვების რიცხვი ერთით შემცირებული, ე. ი. $14-1=13$. ფაქტორების მიხედვით გამოთვლილი დისპერსიათა რიცხვი იქნება ყველა მწკრივის რიცხვს გამოკლებული ერთი, ე. ი. $3-1=2$; ხოლო ნარჩენი დისპერსიისათვის თავისუფლების ხარისხის რიცხვი იქნება ზემოგამოთვლილ თავისუფლების ხარისხთა რიცხვების სხვაობა, ე. ი. $13-2=11$.

გამოვთვლით s^2, s_f^2 და s_e^2 დისპერსიებს ფორმულებით:

$$s^2=\frac{S}{13},\quad s_f^2=\frac{S_f}{2},\quad s_e^2=\frac{S_e}{11}.$$

მაშასადამე,

$$s^2=\frac{553,4464}{13}=42,5728,$$

$$s_f^2=\frac{339,9391}{2}=169,96955;$$

$$s_e^2=\frac{213,4051}{11}=19,4005.$$

ახლა ვნახოთ, თუ რამდენად ეფექტურია ამა თუ იმ სასუქის გამოყენება. ამისათვის მივმართოთ ჯერ T კრიტერიუმს.

ამ კრიტერიუმის თანახმად დავწეროთ:

$$T_{\varrho} = \frac{s_f^2}{s_e^2} = \frac{169{,}970}{19{,}400} = 8{,}76.$$

ვიპოვოთ T-ს მნიშვნელობა VI ცხრილის დახმარებით; აქ $k_1 = 2$, $k_2 = 11$, მივიღებთ:

$$T = 7{,}20,$$

ე. ი. 0,01-ზე ნაკლები ალბათობით შეიძლება შევხვდეთ შემთხვევით მნიშვნელობას $T \geqslant T_{\varrho} = 8{,}76$.

ცხრილით ნაპოვნი T ნაკლებია დაკვირვებით მიღებულ T_{ϱ}-ზე და, ამგვარად, სასუქის მოქმედება ეფექტურად შეიძლება ჩაითვალოს.

ახლა გამოვიყენოთ θ კრიტერიუმი. ამისათვის უნდა ვიანგარიშოთ θ და σ_θ; s_1^2 და s_2^2-ის მაგივრად მივიღოთ s_f^2 და s_e^2, განხილულ მაგალითში $k_1 = 11$, $k_2 = 2$; მაშასადამე, დავწეროთ:

$$\theta = \frac{k_1 - 2}{k_1} \frac{s_f^2}{s_e^2} = \frac{9}{11} \cdot \frac{169{,}97}{19{,}40} = 7{,}17;$$

$$\sigma_\theta = \sqrt{\frac{2(k_1 + k_2 - 2)}{(k_1 - 4)k_2}} = \sqrt{\frac{22}{2 \cdot 7}} = \sqrt{\frac{11}{7}} \simeq 1{,}25.$$

ახლა ვიანგარიშოთ შემდეგი გამოსახულება:

$$\frac{|\,\theta - 1\,|}{\sigma_\theta} = \frac{6{,}17}{1{,}25} = 4{,}93 > 3.$$

მაშასადამე, მივიღეთ იგივე შედეგი—სასუქის გამოყენება ეფექტური ყოფილა.

თავი VI

კორელაციის თეორიის ელემენტები

§ 1. კორელაციური და ფუნქციონალური დამოკიდებულება

შემთხვევით სიდიდეზე შესაძლებელია მრავალი ფაქტორი მოქმედებდეს. სხვადასხვა ფაქტორი სხვადასხვანაირად მოქმედებს და მათი მოქმედების შედეგიც განსაკუთრებული იქნება. საკმარისი არ არის

მარტო მოქმედ ფაქტორთა დადგენა, არამედ საჭიროა ამ ფაქტორთა მოქმედების როგორც სიდიდის ისე ფორმის განსაზღვრა.

შემთხვევით სიდიდეთა შორის ურთიერთდამოკიდებულების ცნება რამდენადმე განსხვავდება საზოგადოდ მათემატიკაში ცნობილ სიდიდეთა შორის ურთიერთდამოკიდებულების ცნებისაგან. მათემატიკაში საზოგადოდ „დამოკიდებულება“ გულისხმობს მკაცრ ფუნქციონალურ დამოკიდებულებას. X და Y ორი სიდიდე ფუნქციონალურად დამოკიდებულია, თუ ერთის მნიშვნელობის ცოდნით შეიძლება ზუსტად ვაჩვენოთ მეორის მნიშვნელობა.

ალბათობის თეორიაში ვიხილავთ დამოკიდებულების უფრო ზოგად ცნებას, რომელსაც ალბათური, ანუ სტოქასტური დამოკიდებულება ეწოდება. თუ X სიდიდე Y-თან დაკავშირებულია ალბათური დამოკიდებულებით, მაშინ X-ის მნიშვნელობის ცოდნით არ შეიძლება Y-ის მნიშვნელობის ზუსტად ჩვენება; ამ შემთხვევაში შეიძლება მხოლოდ მისი განაწილების კანონის ჩვენება, რომელიც დამოკიდებული იქნება იმაზე, თუ რა მნიშვნელობა მიიღო X სიდიდემ.

შემთხვევით სიდიდეთა შორის რამდენადაც უფრო მჭიდროა ალბათური დამოკიდებულება, იმდენად ის ფუნქციონალურ დამოკიდებულებას უახლოვდება.

განვიხილოთ ფუნქციონალური და ალბათური დამოკიდებულების მაგალითები.

როგორც ცნობილია, გავლილი გზა, როცა საწყისი სიჩქარე ნულის ტოლია, დროის ფუნქციას წარმოადგენს და ამ შემთხვევაში ვიტყვით, რომ t დროსა და s მანძილს შორის არსებობს გარკვეული ფუნქციონალური დამოკიდებულება

$$s=a\frac{t^2}{2}\ .$$

აქ t-ს გარკვეულ მნიშვნელობას სათანადო s შეესაბამება.

ვთქვათ, X შემთხვევითი სიდიდე შემთხვევით აღებული ადამიანის სიგრძეს გამოხატავს და Y შემთხვევითი სიდიდე კი იმავე ადამიანის სხეულის წონას. ცხადია, X და Y იმყოფება ერთმანეთის მიმართ გარკვეულ ალბათურ დამოკიდებულებაში; ეს დამოკიდებულება იმით გამოიხატება, რომ ადამიანის სხეული რაც უფრო გრძელია, საზოგადოდ, მით უფრო მეტს იწონის. შესაძლებელია ისეთი ემპირიული ფორმულის შედგენა, რომელიც მიახლოებით ამ ალბათურ დამოკიდებულებას ფუნქციონალურით შეცვლის. ასეთია, მაგალითად, საზოგადოდ ცნობილი ფორმულა, რომელიც მიახლოებით გამოხატავს ადამიანის სიგრძესა და წონას შორის დამოკიდებულებას: თუ ადამიანის

სხეულის წონას Y კილოგრამებით გამოვსახავთ, ხოლო სიგრძეს—X სანტიმეტრებით, მაშინ

$$Y = X - 100.$$

ეს ფორმალური ჩანაწერი გვეუბნება, რომ ადამიანის სხეულის წონა დაახლოებით იმდენი კილოგრამია, რამდენი სანტიმეტრიცაა მისი სიგრძე 100-ის გამოკლებით.

უკანასკნელი ფორმულა, რა თქმა უნდა, არაზუსტია და ხშირად შესაძლებელია მისგან გადახვევას ჰქონდეს ადგილი.

ახლა, ვთქვათ, X არის ისევ შემთხვევით აღებული ადამიანის სხეულის სიგრძე, ხოლო Y კი—მისი ასაკი. ცხადია, თუ მხედველობაში მივიღებთ დაახლოებით 25 წელზე მეტი ასაკის ადამიანებს, მაშინ X და Y შორის თითქმის პრაქტიკულად არავითარი დამოკიდებულება არ არსებობს; პირიქით, უფრო მცირე ასაკის ადამიანების შემთხვევაში X და Y ურთიერთდამოკიდებულია.

შემთხვევით სიდიდეთა შორის ალბათური დამოკიდებულების ან, რაც იგივეა, კორელაციური დამოკიდებულების სიდიდის საზომად კორელაციის კოეფიციენტს იყენებენ, რომლის არსიც შემდგომ იქნება გარჩეული.

§ 2. წრფივი კორელაციური კავშირი

ვთქვათ, მოცემული გვაქვს X და Y შემთხვევითი სიდიდეები, რომლებიც სხვადასხვა ობიექტს ეკუთვნის, მაშინ X-ის ყოველ მნიშვნელობას Y-ის მნიშვნელობები შეესაბამება. სიდიდეთა შორის ამგვარ დამოკიდებულებას **სტატისტიკურ კავშირს** უწოდებენ. სიდიდეებს შორის სტატისტიკური კავშირის მთლიანი შესწავლა გულისხმობს X-ის თითოეული მნიშვნელობისათვის Y-ის მნიშვნელობათა განაწილების და X-ის ცვლილებასთან ერთად ამ განაწილების ცვლილების განხილვას.

შემთხვევით სიდიდეთა შორის სტატისტიკური კავშირის წარმოდგენა შესაძლებელია ისეთი ცხრილების საშუალებით, სადაც ნაჩვენები იქნება (X, Y) წყვილ მნიშვნელობათა სიხშირეები (ცხრილი 16).

გავეცნოთ ცხრილი 16-ის აგებულებას. აქ X და Y-ის 100 წყვილი მნიშვნელობა გვაქვს. X და Y დაყოფილია გარკვეულ ჯგუფებად და აღებული გვაქვს ამ ჯგუფთა შუა მნიშვნელობანი: მაგალითად, X-თვის: 20—30, 30—40, 40—50, ..., 120—130, ხოლო Y-თვის 10,5—15,5; 15,5—20,5, ..., 65,5—70,5.

X \ Y	13	18	23	28	33	38	43	48	53	58	63	68	n_x
25	3	2	—	—	—	—	—	—	—	—	—	—	5
35	—	6	4	—	—	—	—	1	—	—	—	—	11
45	—	1	13	5	—	—	—	—	—	—	—	—	19
55	—	1	2	4	8	1	—	—	—	—	—	—	16
65	—	—	1	—	4	4	2	—	—	—	—	—	11
75	—	—	—	—	2	6	6	1	—	—	—	—	15
85	—	—	—	—	—	—	1	5	—	—	—	—	6
95	—	—	—	—	—	—	—	1	4	1	—	—	6
105	—	—	—	—	—	—	—	—	2	4	1	1	8
115	—	—	—	—	—	—	—	—	—	1	—	1	2
125	—	—	—	—	—	—	—	—	—	—	—	1	1
m_y	3	10	20	9	14	11	9	8	6	6	1	3	100

ცხრილში მოცემულია სიხშირეები· მაგალითად, მეორე სტრიქონის და მესამე სვეტის გადაკვეთაში მოცემული რიცხვი 4 გვიჩვენებს, რომ (X, Y)-ის 100 წყვილი მნიშვნელობიდან 4 არის ისეთი, რომელთათვისაც $X=35$-ს შეესაბამება Y-ის მნიშვნელობა 23. უფრო ზუსტად, ეს იმას ნიშნავს, რომ 4 ისეთი შემთხვევა გვაქვს, როდესაც X-ის გარკვეულ მნიშვნელობას 30—40 შუალედში Y-ის მნიშვნელობანი შეესაბამება 20,5—25,5 შუალედში. მესამე სტრიქონში რიცხვი 5 ნიშნავს, რომ ხუთი შემთხვევა გვაქვს, როდესაც $X=45$ და $Y=28$ და ა. შ.

ბოლო სვეტში მოცემულია X-ის მნიშვნელობათა სიხშირეები; 5 შემთხვევა გვაქვს, როცა $X=25$-ს, 11 შემთხვევა, როცა $X=35$-ს და ა. შ. სულ სიხშირეთა ჯამი უდრის 100-ს. უკანასკნელ სტრიქონში მოთავსებულია Y-ის მნიშვნელობათა სიხშირეები. X-ის სიხშირეები აღნიშნულია n_x-ით, Y-ის კი m_y-ით.

განხილული ცხრილი ტიპიურ კორელაციურ ცხრილს წარმოადგენს, რომლის მიხედვით გავარკვევთ, თუ როგორი ხასიათის დამოკიდებულება არსებობს X და Y სიდიდეებს შორის. როგორც ცხრილიდან ჩანს, X-ის ცალკეულ მნიშვნელობას Y-ის რიგი მნიშვნელობანი შეესაბამება და X-ის გადიდებით საშუალოდ Y-იც იზრდება.

ნათელი რომ გახდეს აღებულ სიდიდეებს შორის კავშირის ნამდვილი აზრი, X-ის მნიშვნელობებს შევუსაბამოთ Y-ის კერძო საშუალოები, რომლებიც იანგარიშება შემდეგნაირად: X-ის მნიშვნელობა—25-ს შეესაბამება Y-ის სამი მნიშვნელობა 13 და ორი მნიშვნელობა 18, რომელთა საშუალო, თუ მას აღვნიშნავთ $\overline{y}_1$-ით, იქნება:

$$\overline{y}_1=\frac{3 \cdot 13+2 \cdot 18}{5}=15.$$

ასევე ვიანგარიშებთ Y-ის დანარჩენ კერძო საშუალოებს

$$\overline{y_2}=\frac{6\cdot 18+4.23+1\cdot 48}{11}=22,6.$$

$\overline{y_2}$ შეესაბამება $X=35$ მნიშვნელობას.

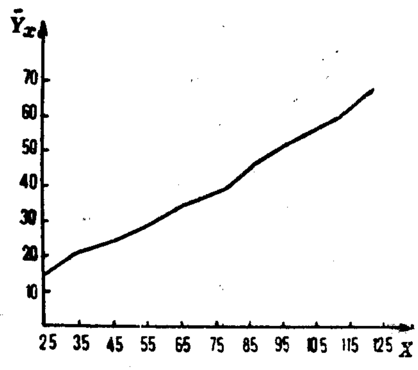
თუ ანალოგიურად ვიანგარიშებთ დანარჩენ კერძო საშუალოებს, მივიღებთ შემდეგს (ცხრილი 17).

როგორც მე-17 ცხრილიდან ჩანს, X-ის ზრდით $\overline{y_x}$-იც იზრდება და ეს ზრდა თითქმის თანაბარია.

ცხრილი 17

X	25	35	45	55	65	75	85	95	105	115	125
$\overline{y_x}$	15,0	22,6	24,1	29,9	35,3	40,0	47,2	53,0	58,6	63,0	68,0

X და Y_x-ის ცვალებადობის გრაფიკს ექნება ასეთი სახე (ნახ. 22). გრაფიკის აგებამ მოგვცა თითქმის წრფე. ასეთ შემთხვევაში ვიტყვით, რომ X და y_x-ს შორის არსებობს წრფივი კორელაციური კავშირი.



ნახ. 22.

ამგვარად, კორელაციური კავშირი შესაძლებელია იყოს როგორც წრფივი, ისე არაწრფივიც.

სტატისტიკური კავშირების გარკვევისას არსებითია კავშირის ძალისა და ფორმის განსაზღვრა.

თუ ერთ-ერთი რომელიმე შემთხვევითი სიდიდის სხვადასხვა მნიშვნელობისათვის მეორე შემთხვევითი სიდიდის შესაბამისი მნიშვნელობანი უცვლელია, მაშინ მათ შორის სტატისტიკურ კავშირზე ლაპარაკი

არ შეიძლება. წინააღმდეგ შემთხვევაში აღებულ შემთხვევით სიდიდეებს შორის გარკვეული სტატისტიკური კავშირი არსებობს სტატისტიკური კავშირის ფორმის გარკვევა შესაძლებელია იმ წრფის საშუალებით, რომლის მახლობლობაში განლაგდება აგებული გრაფიკის წერტილები. იმისდა მიუხედავად, თუ როგორი კორელაციური კავშირი გვექნება— არაწრფივი თუ წრფივი, საჭირო იქნება კავშირის ძალის სხვადასხვა საზომთა განხილვა. ასეთი საზომი წრფივი კორელაციის შემთხვევაში კორელაციის კოეფიციენტი იქნება, ხოლო არაწრფივი კორელაციის დროს—კორელაციური ფარდობა.

პირველ რიგში განვიხილოთ წრფივი კორელაციური კავშირი და შევისწავლოთ კორელაციის კოეფიციენტთა გამოთვლა. დავუბრუნდეთ ისევ მე-16 ცხრილს

n_{xy}-ით აღვნიშნოთ ამ ცხრილში X და Y-ის რომელიმე ერთი წყვილი მნიშვნელობის სიხშირე. შევადგინოთ შემდეგი ჯამი:

$$C_{xy}=\frac{\Sigma n_{xy}(x-\overline{x})\,(y-\overline{y})}{n}\ .$$

C_{xy}-ს ეწოდება X და Y-ის კოვარიაცია. თუ X და Y შემთხვევით სიდიდეთა საშუალო კვადრატულ გადახრებს სათანადოდ σ_x და σ_y-ით აღვნიშნავთ, მაშინ კოვარიაცია σ_x და σ_y საშუალო კვადრატულ გადახრათა ნამრავლზე გაყოფილი გვაძლევს კორელაციის კოეფიციენტს, რომელიც R-ით აღინიშნება. ამგვარად, დავწერთ:

$$R=\frac{C_{xy}}{\sigma_x\,\sigma_y}\ .$$

გამოთვლების გამარტივების მიზნით C_{xy} კოვარიაციის წარმოდგენა შეიძლება შემდეგნაირად:

$$C_{xy}=\frac{\Sigma n_{xy}xy}{n}-\overline{x}\,\overline{y}\ ,$$

ხოლო საშუალო კვადრატული გადახრები გამოითვლება ფორმულებით:

$$\sigma_x=\sqrt{\frac{\Sigma\, n_x x}{n}-\overline{x}^2},$$

$$\sigma_y=\sqrt{\frac{\Sigma\, m_y y}{n}-\overline{y}^2},$$

სადაც n_x და m_y არის X და Y-ის სათანადო სიხშირეები. გამოვთვალოთ კორელაციის კოეფიციენტი მე-16 ცხრილის მონაცემების მიხედვით.

როგორც მე-16 ცხრილიდან ჩანს, X-ის მნიშვნელობათა შორის სხვაობა

მუდმივი სიდიდეა და 10-ის ტოლია, ხოლო Y-ის მნიშვნელობათა შორის სხვაობა 5-ის ტოლია.

გამოთვლების გამარტივების მიზნით შემოვიყვანოთ ახალი ცვლადები:

$$u=\frac{x-15}{10}, \qquad v=\frac{y-8}{5}.$$

როგორც ვხედავთ, u და v სიდიდეები ღებულობს მხოლოდ მთელ მნიშვნელობებს, დაწყებულს 1-დან და ა. შ.

X და Y სიდიდეები u და v-თი გამოისახება შემდეგნაირად:

$$x=10u+15, \qquad y=5v+8.$$

აქედან ცხადია, რომ საშუალოებისათვის დავწერთ:

$$\overline{x}=10\overline{u}+15, \quad \overline{y}=5\overline{v}+8.$$

ეს მნიშვნელობანი შევიტანოთ C_{xy}-ის გამოსახულებაში, მივიღებთ:

$$C_{xy}=\frac{\Sigma n_{uv}(10u+15-10\overline{u}-15)(5v+8-5\overline{v}-8)}{n}=$$

$$=\frac{\Sigma n_{uv}10(u-\overline{u})\ 5(v-\overline{v})}{n}=50\frac{\Sigma n_{uv}(u-\overline{u})(v-\overline{v})}{n}=50\,C_{uv}.$$

ამგვარად, გვექნება:

$$C_{xy}=50\,C_{uv}.$$

აქ n_{uv} არის u და v-ს რომელიმე წყვილი მნიშვნელობის შესაბამისი სიხშირე.

ახლა დავამყაროთ კავშირი σ_x, σ_y, σ_u და σ_v სიდიდეთა შორის:

$$\sigma_x=\frac{1}{n}\sqrt{n_x(x-\overline{x}^2)}=\frac{1}{n}\sqrt{n_u(10u+15-10\overline{u}-15)^2}=$$

$$=\frac{1}{n}\sqrt{n_u 10^2(u-\overline{u})^2}=10\frac{1}{n}\sqrt{n_u(u-\overline{u})^2}=10\sigma_u,$$

ე. ი. მივიღეთ, რომ

$$\sigma_x=10\,\sigma_u,$$

ანალოგიურად, გვექნება:

$$\sigma_y=10\,\sigma_v.$$

კორელაციის კოეფიციენტი R გამოისახება შემდეგნაირად:

$$R=\frac{C_{xy}}{\sigma_x\sigma_u}=\frac{50\,C_{uv}}{10\sigma_u\cdot 5\sigma_v}=\frac{C_{uv}}{\sigma_u\sigma_v}.$$

ამგვარად, როგორც ვხედავთ, X და Y-თვის კორელაციის კოეფიციენტი იგივეა, რაც u და v-თვის, ამიტომ სავსებით საკმარისი იქნება უკანასკნელი კორელაციის კოეფიციენტის გამოთვლა. აგრეთვე ცხადია, რომ n_{xy}, n_x, m_y სიხშირეები იგივეა, რაც n_{uv}, n_u, m_v სიხშირეები.

მე-16 ცხრილის შესაბამისი ცხრილი u და v-თვის იქნება მე-18 ცხრილი.

ცხრილი 18

u \ v	1	2	3	4	5	6	7	8	9	10	11	12	n_u
1	3	2											5
2		6	4					1					11
3		1	13	5									19
4		1	2	4	8	1							16
5			1		4	4	2						11
6					2	6	6	1					15
7							1	5					6
8								1	4	1			6
9									2	4	1	1	8
10										1		1	2
11												1	1
m_v	3	10	20	9	14	11	9	8	6	6	1	3	100

u-თვის სიხშირეები ემთხვევა X-ის სიხშირეებს, v-ს სიხშირეები კი Y-ის სიხშირეებს, ე. ი.

$$n_x = n_u, \quad m_y = m_v.$$

კორელაციური ცხრილის ყოველ სვეტს, რომელიც Y-ის ამა თუ იმ მნიშვნელობას შეესაბამება, ამ Y-ის რიგი ეწოდება, სათანადოდ კორელაციური ცხრილის სტრიქონებს—X-ის რიგები.

Y-ის რიგების მიმართ ჯამები Y-ის სათანადო რიგების სიხშირეთა X-ის სათანადო მნიშვნელობებზე ნამრავლთა ჯამისაგან შედგება. ამგვარადვე იანგარიშება ჯამები u-ს რიგის მიმართ. მაგალითად,

$$1 \cdot 3 + 2 \cdot 2 = 7; \quad 2 \cdot 6 + 3 \cdot 4 + 8 \cdot 1 = 32,$$

$$2 \cdot 1 + 3 \cdot 13 + 4 \cdot 5 = 61, \quad 2 \cdot 1 + 3.2 + 4 \cdot 4 + 5 \cdot 8 + 6 \cdot 1 = 70$$

და ა. შ.

ვიანგარიშოთ ჯამები v-ს რიგის მიმართ:

$$3 \cdot 1 = 3; \quad 2 \cdot 1 + 6 \cdot 2 + 1 \cdot 3 + 1 \cdot 4 = 21;$$

$$4 \cdot 2 + 13 \cdot 3 + 2 \cdot 4 + 1 \cdot 5 = 60; \quad 5 \cdot 3 + 4 \cdot 4 = 31$$

და ა. შ.

შევადგინოთ v-ს რიგის მიმართ ჯამების სათანადო v-ზე ნამრავლთა ჯამები, გვექნება:

$$7 \cdot 1=7; \quad 32 \cdot 2=64; \quad 61 \cdot 3=183.$$

და ა. შ.

გამოთვლათა შედეგები მოცემულია მე-19-ე ცხრილში.

ცხრილი 19

u	$n_v=m_x$	v	$m_v=m_y$	v რიგის მიმართ ჯამები	$\sum\limits_{n} n_{uv}uv$	v რიგის მიმართ ჯამები	$\sum\limits_{v} u_{uv}uv$
1	5	1	3	7	7	3	3
2	11	2	10	32	64	21	42
3	19	3	20	61	183	60	183
4	16	4	9	70	170	31	124
5	11	5	14	61	305	64	320
6	15	6	11	96	576	60	360
7	6	7	9	47	329	53	371
8	6	8	8	54	432	51	408
9	8	9	6	81	724	50	450
10	2	10	6	22	229	54	540
11	1	11	1	12	152	9	94
		12	3			20	360
	$n=100$		$n=100$	$v=553$	$\sum\limits_{u,v} n_{uv}uv = 3257$	$u=486$	$\sum\limits_{v,u} n_{uv}uv = 3257$

მე-19 ცხრილიდან აგრეთვე ვიანგარიშებთ:

$$\overline{u} = \frac{u}{n} = \frac{486}{100} = 4,86,$$

$$\overline{u} = \frac{v}{n} = \frac{543}{100} = 5,43.$$

$$C_{uv} = \frac{\Sigma n_{uv}uv}{n} - \overline{uv} = \frac{3257}{100} - 26,3898 = 6,1802.$$

ახლა გამოვთვალოთ u და v-ს საშუალო კვადრატული გადახრები, გვექნება:

$$\sigma_u^2 = \frac{\Sigma\, n_u n^2}{n} - \overline{u}^2 = \frac{2938}{100} - 23,6196 = 5,7604,$$

ე. ი.

$$\sigma_u = 2,4.$$

ანალოგიურად,

$$\sigma_y^2 = \frac{\Sigma\, m_v v^2}{u} - \overline{v}^2 = \frac{3705}{100} - 29,4849 = 7,5651.$$

ე. ი.

$$\sigma_y = 2,75.$$

ახლა კი შეგვიძლია კორელაციის კოეფიციენტის ანგარიში:

$$R = \frac{C_{uv}}{\sigma_u \sigma_v} = \frac{6.1802}{2,4 \cdot 2,75} = 0,936.$$

ვიანგარიშოთ სათანადო ელემენტები X და Y-თვის. გვქონდა

$$X = 10\,u + 15,$$

საიდანაც დავწერთ

$$\overline{x} = 10\,\overline{u} + 15 = 63,5,$$

ე. ი.

$$\overline{x} = 63,5.$$

ასევე

$$Y = 5v + 8, \ \overline{y} = 5\,\overline{v} + 8 = 35,5.$$

ე. ი.

$$\overline{y} = 35,5.$$

გვქონდა

$$\sigma_x = 10\,\sigma_u, \quad \sigma_y = 5\sigma_v,$$

ამიტომ

$$\sigma_x = 24, \quad \sigma_y = 13,75.$$

გვქონდა $C_{xy} = 50\,C_{uv}$, ამიტომ გვექნება:

$$C_{xy} = 309,01.$$

§ 3. წრფივი კორელაციური კავშირის და რეგრესიის წირის განტოლება

ვთქვათ, მოცემული გვაქვს ორი შემთხვევითი სიდიდე X და Y, რომელთა შორის არსებობს წრფივი კორელაციური კავშირი. მათი საშუალო კვადრატული გადახრებია შესაბამისად σ_x და σ_y. კორელაციის კოეფიციენტის და საშუალო კვადრატულ გადახრათა დახმარებით შესაძლებელია X და Y შორის წრფივი კორელაციური კავშირის გამომსახველი განტოლების შედგენა.

X და Y სიდიდეთა საშუალოები აღვნიშნოთ $\overline{x}$ და $\overline{y}$-ით, ხოლო მათი კორელაციის კოეფიციენტი r-ით. აგრეთვე, $\overline{y}_x$-ით აღვნიშნოთ X-ის რომელიმე მოცემული მნიშვნელობისათვის Y-ის კერძო საშუალო. მაშინ Y-ის X-თან წრფივი კორელაციური კავშირის განტოლება შეიძლება დაიწეროს შემდეგნაირად:

$$\overline{y_x}-\overline{y} = R\frac{\sigma_y}{\sigma_x}(x-\overline{x}). \qquad (3.1)$$

ეს განტოლება გვიჩვენებს, რომ Y-ის კერძო საშუალოს საერთო საშუალო სიდიდიდან გადახრა X-ის საშუალოდან გადახრის პროპორციულია; პროპორციულობის კოეფიციენტია

$$\rho_y = R\frac{\sigma_y}{\sigma_x} \qquad (3.2)$$

სიდიდე, რომელსაც X-ის მიმართ Y-ის რეგრესიის კოეფიციენტი ეწოდება.

ეს უკანასკნელი ფორმულა არის Y-ის რეგრესიის განტოლება X-ის მიმართ. თუ X და Y შორის სტატისტიკური კავშირი არ არის ზუსტად წრფივი, მაშინ (3.1) განტოლებით განსაზღვრული კერძო საშუალოები x-ის მოცემული შესაბამისი მნიშვნელობისათვის კერძო საშუალოს მხოლოდ მიახლოებითი მნიშვნელობანი იქნება.

ამ მიახლოებით კერძო საშუალოებს ის თვისება აქვს, რომ მათი ნამდვილი $\overline{y_x}$ კერძო საშუალოდან საშუალო კვადრატული გადახრა მინიმალურ მნიშვნელობას აღწევს, იმ მიახლოებით კერძო საშუალოებთან შედარებით, რომელთა მიღება შეგვიძლია $\overline{y_x}$ და X-ის სხვა წრფივი კორელაციური კავშირების დახმარებით.

სხვა წრფივი კორელაციური კავშირები შესაძლებელია უბრალოდ შემდეგნაირად გამოისახოს:

$$\overline{y_x} = aX + b,$$

სადაც a და b მუდმივი რიცხვებია.

ყოველივე ზემოთქმულის ნათელსაყოფად განვიხილოთ ისევ მე-18 ცხრილი.

მე-18 ცხრილისათვის გვქონდა:

$$\overline{x} = 63{,}5,$$

$$\overline{y} = 35{,}15,$$

$$\sigma_x = 24.$$

$$\sigma_y = 13{,}75,$$

$$R = 0{,}936.$$

ამიტომ Y-ის რეგრესიის კოეფიციენტი X-ის მიმართ იქნება:

$$\rho_y = R\frac{\sigma_y}{\sigma_x} = 0{,}936 \cdot \frac{13{,}75}{24} = 0{,}536.$$

რეგრესიის (3.1) განტოლება დაიწერება შემდეგნაირად:

$$\overline{y_x}-35,15=0,536(X-63,5),$$

საიდანაც

$$\overline{y_x}=0,536\,X+1,11. \qquad (3.3)$$

(3.3) განტოლება გამოსახავს კავშირს კერძო საშუალო $\overline{y_x}$-სა და სათანადო X-ს შორის, რომელიც აღებულია მე-16 ცხრილიდან. თუ (3.3) განტოლებაში ჩავსვამთ X-ის სხვადასხვა მნიშვნელობას მე-16 ცხრილიდან, მივიღებთ სათანადოდ $\overline{y_x}$-ის მნიშვნელობებს. მაგალითად, $X=x_1=25$-თვის.

$$\overline{y_1}=0,536\,.\,25+1,11\simeq 14,5$$

ან

$$\overline{y_1}=14,5.$$

როცა $X=x_2=35$, მაშინ

$$\overline{y_2}=0,536\,.\,35+1,11=19,9.$$

თუ ანალოგიურად გავაგრძელებთ გამოთვლებს, მივიღებთ (ცხრილი 20):

ცხრილი 20

x	$\overline{y'_x}$	$\overline{y_x}$	$\overline{y'_x}-\overline{y_x}$
25	14,5	15,0	—0,5
35	19,9	22,6	—2,7
45	25,2	24,1	1,1
55	30,6	29,9	0,7
65	36,0	35,3	0,7
75	41,3	40,0	1,3
85	46,7	47,2	—0,5
95	52,0	53,0	—1,0
105	57,4	58,6	—1,2
115	62,8	68,0	—0,2
125	68,1	68,0	0,1

აქ (3.1) განტოლებით გამოთვლილი $\overline{y_x}$ კერძო საშუალოები აღნიშნულია $\overline{y'_x}$-ით, ხოლო მათი რეალური მნიშვნელობანი, რომელიც მიღებული გვქონდა წინა ცხრილში (იხ. ცხრილი 17), უბრალოდ $\overline{y_x}$-ით არის აღნიშნული.

როგორც მე-20 ცხრილიდან ჩანს, გამოთვლილ კერძო საშუალოებს და ნამდვილ კერძო საშუალოებს შორის სხვაობა უმნიშვნელო სიდიდეებს წარმოადგენს.

§ 4. კორელაციის კოეფიციენტის თვისებები

ახლა გავიხსენოთ კორელაციის კოეფიციენტის ზოგიერთი თვისება და ზოგიერთი ახალიც დავუმატოთ.

1. კორელაციის კოეფიციენტის მნიშვნელობა ყოველთვის მოთავსებულია —1 და +1 შორის:

$$-1 \leqslant R \leqslant 1.$$

2. როდესაც R კორელაციის კოეფიციენტი უდრის —1 ან +1-ს, მაშინ X და Y ერთმანეთთან დაკავშირებულია ზუსტი წრფივი კავშირით, რომელსაც ექნება შემდეგი სახე:

$$Y = aX + b$$

ან

$$X = cY + d.$$

თუ პირველს X-ის მიმართ ამოვხსნით, მივიღებთ მეორე განტოლებას.

პირიქით, თუ X და Y ერთმანეთთან დაკავშირებულია წრფივი კავშირით, მაშინ მათთვის კორელაციის კოეფიციენტი იქნება —1 ან +1. როდესაც კორელაციის კოეფიციენტი $R = +1$-ს, მაშინ a და c კოეფიციენტები უნდა იყოს დადებითი. თუ კორელაციის კოეფიციენტი $R = -1$-ს, მაშინ a და c კოეფიციენტი იქნება უარყოფითი.

3. როდესაც კორელაციის კოეფიციენტი $R = 0$-ს, მაშინ X და Y სიდიდეთა შორის წრფივი კორელაციური კავშირი არ არსებობს, მაგრამ მათ შორის შესაძლებელია მრუდწირული კორელაციური კავშირის არსებობა.

4. რაც უფრო ახლოს იქნება R ის მნიშვნელობა +1 ან —1-თან, მით უფრო ზუსტი წრფივი კორელაციური კავშირი გვექნება X და Y სიდიდეებს შორის და, პირიქით, ეს კავშირი მით უფრო სუსტი იქნება, რაც უფრო კორელაციის კოეფიციენტი უახლოვდება 0-ს.

5. Y-ის, X-ის მიმართ რეგრესიის განტოლებით გამოთვლილ $\overline{y}_x$ კერძო საშუალოების მახლობლობაში დაკვირვების შედეგად მიღებულ Y მნიშვნელობათა დისპერსია (მას Σ_y^2-ით აღვნიშნავთ) გამოისახება შემდეგნაირად:

$$\Sigma_y^2 = \sigma_y^2 (1 - R^2).$$

ეს გამოსახულება მით უფრო მცირე იქნება, რაც R უფრო ახლოს იქნება +1-თან, ან —1-თან, ე. ი. უფრო ახლოს იქნება $\overline{y}_x$ კერძო საშუალოების მნიშვნელობანი დაკვირვების შედეგად მიღებულ Y-ის მნიშვნელობებთან.

როდესაც R უახლოვდება ნულს, მაშინ Σ_y^2 უახლოვდება Y-ის მნიშვნელობის σ_y^2 საერთო დისპერსიას.

$\overline{x_y}$ კერძო საშუალოს მახლობლად X-ის მნიშვნელობათა დისპერსია კი გამოისახება შემდეგნაირად

$$\Sigma_x^2 = \sigma_x^2 (1 - R^2).$$

Σ_x^2-თვის იგივე შენიშვნაა სამართლიანი, რაც იყო Σ_y^2-ის მიმართ. Σ_y^2 და Σ_x^2-ს შესაბამისად ეწოდება Y-ის X-ის მიმართ და X-ის Y-ის მიმართ რეგრესიის დისპერსიები.

§ 5. არაწრფივი კორელაციური კავშირი და კორელაციური ფარდობა

შემთხვევით სიდიდეთა შორის, გარდა წრფივი კორელაციური კავშირისა, არსებობს აგრეთვე არაწრფივი კორელაციური კავშირი.

განვიხილოთ მე-16 ცხრილი, რომელიც შემთხვევით სიდიდეთა შორის კორელაციური კავშირის ტიპიურ სურათს გვაძლევს. მოცემული დაკვირვების შედეგად მიღებულ შემთხვევით სიდიდეთა შორის არაწრფივი კორელაციური კავშირის ძალის საერთო საზომად იღებენ კორელაციურ შეფარდებას. თუ განვიხილავთ Y-ის კორელაციურ კავშირს X-თან და X-ის კორელაციურ კავშირს Y-თან და მათთვის ავიღებთ კორელაციურ შეფარდებებს, ვნახავთ, რომ ისინი ერთმანეთისაგან განსხვავებული იქნება.

განვიხილოთ $\overline{y_x}$ კერძო საშუალოთა X-ის შესაბამის მნიშვნელობებთან დამოკიდებულება, მაშინ Y-ის კორელაციური კავშირი x-თან განისაზღვრება შემდეგი ტოლობით:

$$\eta_y = \frac{\sigma(\overline{y_x})}{\sigma_y},$$

სადაც σ_y არის y შემთხვევითი სიდიდის საშუალო კვადრატული გადახრა, ხოლო $\sigma(\overline{y_x})$ არის კვადრატული ფესვი Y-ის მნიშვნელობათა საერთო $\overline{y}$ საშუალოს მახლობლად $\overline{y_x}$ კერძო საშუალოთა დისპერსიიდან, რომელიც გამოისახება შემდეგნაირად:

$$\sigma^2(\overline{y_x}) = \frac{1}{n} \Sigma n_x(\overline{y_x} - \overline{y})^2, \tag{5.1}$$

საიდანაც

$$\sigma(\overline{y_x}) = \sqrt{\frac{1}{n} \Sigma n_x(\overline{y_x} - \overline{y})^2}. \tag{5.2}$$

გამოთვლათა გამარტივების მიზნით (5.1) ტოლობა გარდავქმნათ შემდეგნაირად:

$$\sigma^2(\overline{y_x}) = \frac{1}{n}\sum n_x \overline{y}_x^2 - 2\frac{1}{n}\sum n_x \overline{y_x}\,\overline{y} + \frac{1}{n}\sum n_x \overline{y}^2 =$$

$$= \frac{1}{n}\sum n_x \overline{y}_x^2 - 2\overline{y}\,\frac{1}{n}\sum n_x \overline{y_x} + \overline{y}^2\frac{1}{n}\sum n_x, \tag{5.3}$$

მაგრამ, როგორც ვიცით, სიხშირეთა ჯამი არის დაკვირვებათა რაოდენობის ტოლი, ე. ი.

$$\sum n_x = n. \tag{5.4}$$

ხოლო Y-ის მნიშვნელობათა საერთო საშუალო $\overline{y}$ კერძო საშუალოებით გამოისახება შემდეგნაირად:

$$\overline{y} = \frac{1}{n}\sum n_x \overline{y_x} \tag{5.5}$$

(5.4) და (5.5) ტოლობების მხედველობაში მიღებით (5.3) ტოლობა გადაიწერება შემდეგნაირად:

$$\sigma^2(\overline{y_x}) = \frac{1}{n}\sum n_x \overline{y}_x^2 - 2\overline{y}\,\overline{y} + \overline{y}^2 = \frac{1}{n}\sum n_x \overline{y_x}^2 - \overline{y}^2 .$$

ამგვარად მივიღებთ:

$$\sigma^2(\overline{y_x}) = \frac{1}{n}\sum n_x \overline{y}_x^2 - \overline{y}^2 . \tag{5.6}$$

გავარჩიოთ ისევ მე-16 ცხრილი. $\overline{y_x}$ კერძო საშუალოებისათვის მიღებული გვქონდა შემდეგი მნიშვნელობანი (ცხრილი 20):

15,0; 22,6; 42,1; 29,9; 35,3; 40,0; 47,2; 53,0; 58,6; 63,0; 68,0,

რომელთა შესაბამისი n_x სიხშირეებია:

5, 11, 19, 16, 11, 15, 6, 6, 8, 2, 1.

გამოთვლილი გვქონდა საერთო საშუალო

$$\overline{y} = 35{,}15.$$

ამ მონაცემებით (5.6) ფორმულის დახმარებით გამოვთვალოთ $\sigma^2(\overline{y_x})$ დისპერსია:

$$\sigma^2(\overline{y_x}) = \frac{1}{100}\,[5\cdot 15^2 + 11(22{,}6)^2 + 19(24{,}1)^2 + 16(29{,}9)^2 +$$

$$+ 11(35{,}3)^2 + 15\cdot 40^2 + 6(47{,}2)^2 + 6\cdot 53^2 + 8(58{,}6)^2 + 2\cdot 63^2 +$$

$$+ 1.68^2 - (35{,}15)^2 = 1400{,}4462 - 1235{,}5225 = 164{,}9237,$$

საიდანაც საშუალო კვადრატული გადახრისათვის მივიღებთ:

$$\sigma(\overline{y_x}) = \sqrt{164{,}9237} \simeq 12{,}84.$$

Y შემთხვევითი სიდიდის საშუალო კვადრატული გადახრისათვის მიღებული გვქონდა:

$$\sigma_y = 13{,}75.$$

ახლა შეგვიძლია ვიანგარიშოთ კორელაციური ფარდობა

$$\eta_y = \frac{\sigma(\overline{y_x})}{\sigma_y} = \frac{12{,}84}{13{,}75} = 0{,}934.$$

გავარჩიოთ მიღებული კორელაციური ფარდობის ძირითადი თვისებები:

1. კორელაციური ფარდობა ყოველთვის დადებითია და მოთავსებულია 0 და 1 შორის;

2. კორელაციური ფარდობა ყოველთვის აღემატება სათანადო კორელაციის კოეფიციენტის რიცხობრივ მნიშვნელობას;

3. როდესაც კორელაციური ფარდობა კორელაციის კოეფიციენტის აბსოლუტური მნიშვნელობის ტოლია

$$\eta_y = |\, R \,|,$$

მაშინ Y ის რეგრესია X-ის მიმართ ზუსტად წრფივია და, პირიქით.

4. თუ Y და X შემთხვევით სიდიდეთა შორის კორელაციური კავშირი არ არის, მაშინ

$$\eta_y = 0$$

და, პირიქით, თუ Y არის X-თან ცალსახად დაკავშირებული, ე. ი. როდესაც X-ის ყოველ მნიშვნელობას Y-ის ერთი გარკვეული მნიშვნელობა შეესაბამება, მაშინ

$$\eta_y = 1$$

და, პირიქით.

5. რაც უფრო ახლოს არის η_y-ის მნიშვნელობა 0-თან, მით უფრო სუსტი იქნება X და Y-ს შორის კორელაციური კავშირი და, რაც უფრო ახლოს იქნება η_y-ის მნიშვნელობა 1-თან, კორელაციური კავშირი მით უფრო ძლიერი იქნება.

ამრიგად, განვსაზღვრეთ Y-ის X-თან კავშირისათვის η_y კორელაციური ფარდობა. ანალოგიურად დავწერთ:

$$\eta_x = \frac{\sigma(\overline{x_y})}{\sigma_x},$$

სადაც

$$\sigma^2\,(\overline{x}_y)=\frac{1}{n}\sum m_y\,\overline{x}_y^2-\overline{x}^2,$$

ხოლო

$$\sigma\,(\overline{x}_y)=\sqrt{\frac{1}{n}\sum m_y\,\overline{x}_y^2-\overline{x}^2}\,.$$

ახლა შევაფასოთ კორელაციური ფარდობა.

როდესაც დაკვირვებათა რიცხვი საკმაოდ დიდია, კორელაციური ფარდობის შესაფასებლად იყენებენ მის შესაბამის საშუალო კვადრატულ გადახრას, რომელიც გამოითვლება შემდეგი მიახლოებითი ფორმულით:

$$\sigma_{\eta_y}\simeq\frac{1-\eta_y^2}{\sqrt{n}}\,.$$

მაგრამ n-ის საკმაოდ დიდი მნიშვნელობისათვის ეს ფორმულა დიდად საიმედო არ არის, ამიტომ კორელაციური ფარდობის შესაფასებლად, ე. ი. არაწრფივი კორელაციური კავშირის დასადგენად, უკეთესია კორელაციური ფარდობის შედარება კორელაციის კოეფიციენტთან.

თუ ვიგულისხმებთ, რომ დაკვირვებები წარმოადგენს ისეთ შერჩევით ერთობლიობებს, რომლებიც ნორმალური გენერალური ერთობლიობიდან არის აღებული, მაშინ დაკვირვებათა ნებისმიერი რიცხვისათვის კორელაციური ფარდობისა და კორელაციის კოეფიციენტის შედარება ზუსტ შედეგს მოგვცემს.

ასეთი შეფარდებისათვის ითვლიან ე. წ. T კრიტერიუმს, რომელიც განისაზღვრება შემდეგი ტოლობით:

$$T_\eta=\frac{(n-s)(\eta_y^2-R^2)}{(s-2)\,(1-\eta_y^2)}\,,$$

სადაც s აღნიშნავს იმ კორელაციური ცხრილის X-ის მწკრივთა რაოდენობას, რომელთათვისაც R და η_y არის გამოთვლილი.

თუ მივიღებთ, რომ $k_1=s-2$, $k_2=n-s$, მაშინ T_η-ს მნიშვნელობა V და VI ცხრილების დახმარებით მოიძებნება. თუ აღმოჩნდა, რომ დაკვირვების შედეგად მიღებული T_η ცხრილში ნაპოვნ T-ს მნიშვნელობაზე არანაკლები იქნება, ეს იმას გვიჩვენებს, რომ η_y და R მნიშვნელოვნად განსხვავდება ერთმანეთისაგან და, მაშასადამე, Y და X სიდიდეთა შორის არაწრფივი კორელაციური კავშირი არსებობს. თუკი T_η ნაკლები აღმოჩნდება ცხრილით ნაპოვნ T-ს მნიშვნელობაზე, მაშინ T_η და R შორის განსხვავება შემთხვევითად შეიძლება ჩაითვალოს და Y-ის რეგრესია X-ის მიმართ წრფივი იქნება.

§ 6. მრავლობითი კორელაცია

სტატისტიკური კავშირი შესაძლებელია ორზე მეტ შემთხვევით სიდიდეთა შორის არსებობდეს. მაგალითად, სატელეფონო მავთულის გამძლეობა დამოკიდებულია: ტემპერატურაზე, ნალექთა რაოდენობაზე, ფაბრიკა-ქარხნების ტერიტორიულ სიახლოვეზე და სხვ. აგრეთვე, ქვემეხიდან მიზანში სროლისას მოხვედრა დამოკიდებულია: ყუმბარის სიდიდეზე, ატმოსფერულ წნევაზე, დამიზნების სიზუსტეზე და სხვ. როგორც ვხედავთ, სტატისტიკური კავშირი არსებობს რამდენიმე შემთხვევით ფაქტორს შორის და ჩვენთვის საინტერესოა ამ კავშირის მათემატიკურად გამოსახვა.

სიმარტივისათვის განვიხილოთ სამი შემთხვევითი სიდიდე X, Y, და Z და მათ შორის დავამყაროთ სტატისტიკური კავშირი. პირველ რიგში განვიხილოთ შემთხვევა, როდესაც მათ შორის წრფივი სტატისტიკური კავშირი არსებობს. თუ ვიგულისხმებთ, რომ Z დამოკიდებულია X და Y-ზე, მაშინ აღნიშნული წრფივი სტატისტიკური კავშირი შეიძლება ჩაიწეროს შემდეგნაირად:

$$Z = aX + bY + c, \tag{6.1}$$

სადაც a, b და c მუდმივი კოეფიციენტებია. X, Y და Z სიდიდეთა ნაცვლად უფრო მოხერხებულია მათი საშუალოსაგან გადახრათა განხილვა. მაშინ მათ შორის წრფივი კორელაციური კავშირი გამოისახება შემდეგნაირად:

$$Z - \overline{z} = A(X - \overline{x}) + B(Y - \overline{y}). \tag{6.2}$$

ამ განტოლებაში A და B არის რეგრესიის კოეფიციენტები, რომლებიც კორელაციის კოეფიციენტების და საშუალო კვადრატულ გადახრათა საშუალებით გამოისახება შემდეგნაირად:

$$A = \frac{R_{xz} - R_{yz}R_{xy}}{1 - R^2}\,\frac{\sigma_z}{\sigma_x},$$

$$B = \frac{R_{yz} - R_{xz}R_{xy}}{1 - R^2}\,\frac{\sigma_z}{\sigma_y}.$$

აქ R_{xy}, R_{xz} და R_{yz} გამოხატავს სათანადოდ კორელაციურ კავშირს X და Y, X და Z და Y და Z სიდიდეთა შორის.

წრფივი კავშირის ძალის საზომად ე. წ. კორელაციის ერთობლივი კოეფიციენტი იხმარება, რომელსაც აქვს ასეთი სახე:

$$r = + \sqrt{\frac{R_{xz}^2 - 2R_{xy}R_{xz}R_{yz} + R_{yz}^2}{1 - R_{xy}^2}}.$$

r ყოველთვის დადებითია და მოთავსებულია 0 და 1 შორის. თუ $r=0$, ეს იმის მაჩვენებელი იქნება, რომ, ერთი მხრივ, Z-ს და მეორე მხრივ, X და Y-ს შორის არავითარი წრფივი კორელაციური კავშირი არ არსებობს.

თუ $r=1$, მაშინ X, Y და Z-ს შორის ზუსტი წრფივი კორელაციური კავშირი არსებობს და ის გამოისახება (6.2) ფორმულის სახით.

როდესაც ვარჩევთ მრავალ შემთხვევით სიდიდეთა შორის კორელაციურ კავშირს, საჭიროა აგრეთვე ვიცოდეთ გამოსაკვლევ სიდიდეზე ცალკეული შემთხვევითი სიდიდის გავლენა და მისი შეფასება. ამის მიღწევა შეიძლება კორელაციის კერძო კოეფიციენტების საშუალებით, რომელთაც სამი შემთხვევითი სიდიდის დროს აქვს ასეთი სახე:

$$R_{xz(y)} = \frac{R_{xz} - R_{xy}R_{yz}}{\sqrt{(1-R_{xy}^2)\,(1-R_{yz}^2)}},$$

$$R_{yz(x)} = \frac{R_{yz} - R_{xy}R_{xz}}{\sqrt{(1-R_{xy}^2)\,(1-R_{xz}^2)}}.$$

ამ კერძო კოეფიციენტთა მნიშვნელები ყოველთვის დადებითი ნიშნით აიღება. $R_{xz(y)}$ კერძო კოეფიციენტის აზრი იმაში მდგომარეობს, რომ ის არის X და Z-ს შორის წრფივი კავშირის გამომსახველი იმ შემთხვევაში, როდესაც Y ღებულობს მუდმივ მნიშვნელობას და Z-ზე მხოლოდ X მოქმედებს. ასევე $R_{yz(x)}$ გამოხატავს კერძო კოეფიციენტს, როდესაც X ღებულობს მუდმივ მნიშვნელობას, ხოლო Z-ზე მხოლოდ Y მოქმედებს. ორივე ეს კერძო კოეფიციენტი, ჩვეულებრივ, კორელაციის კოეფიციენტს ჰგავს. მათი მნიშვნელობანი მოთავსებული იქნება $+1$ და -1-ს შორის. როდესაც რომელიმე მათგანი ნულის ტოლია, ეს იმას ნიშნავს, რომ შესაბამის X ან Y და Z-ს შორის წრფივი კორელაციური კავშირი არ არსებობს, ხოლო როცა კერძო კოეფიციენტთა მნიშვნელობა $+1$ ან -1-ს გაუტოლდება, მაშინ ზუსტი წრფივი კორელაციური კავშირი გვექნება.

ცხრილები

ცხრილი I

ნორმალური განაწილების ფუნქცია და ნორმალური განაწილების სიმკვრივე

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} c^{-\frac{1}{2}t^2} dt$$

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0,00	0,3989	0,5000	45	3605	6736	0,90	0,2661	0,8159
01	3989	5040	46	3589	6772	91	2637	8186
02	3989	5080	47	3572	6808	92	2613	8212
03	3988	5120	48	3555	6844	93	2589	8238
04	3986	5160	49	3538	6879	94	2565	8264
05	3984	5199	0,50	0,3521	0,6915	95	2541	8289
06	3982	5239	51	3503	6950	96	2516	8315
07	3980	5279	52	3485	6985	97	2492	8340
08	3977	5319	53	3467	7019	98	2468	8365
09	3973	5359	54	3448	7054	99	2444	8389
0,10	0,3970	0,5398	55	3429	7088	1,00	0,2420	0,8413
11	3965	5438	56	3410	7123	01	2396	8438
12	3961	5478	57	3391	7157	02	2371	8461
13	3956	5517	58	3372	7190	03	2347	8485
14	3951	5557	59	3352	7224	04	2323	8508
15	3945	5596	0,60	0,3332	0,7257	05	2299	8531
16	3939	5636	61	3312	7291	06	2275	8554
17	3932	5675	62	3292	7324	07	2251	8577
18	3925	5714	63	3271	7357	08	2227	8599
19	3918	5753	64	3251	7389	09	2203	8621
0,20	0,3910	0,5793	65	3230	7422	0,10	0,2179	0,8643
21	3902	5832	66	3209	7454	11	2155	8665
22	3894	5871	67	3187	7486	12	2131	8686
23	3885	5910	68	3166	7517	13	2107	8708
24	3876	5948	69	3144	7549	14	2083	8729
25	3867	5987	0,70	0,3123	0,7580	15	2059	8749
26	3857	6026	71	3101	7611	16	2036	8770
27	3847	6064	72	3079	7642	17	2012	8790
28	3836	6103	73	3056	7673	18	1989	8810
29	3825	6141	74	3034	7703	19	1965	8830
0,30	0,3814	0,6179	75	3011	7734	1,20	0,1942	0,8849
31	3802	6217	76	2989	7764	21	1919	8869
32	3790	6265	77	2966	7794	22	1895	8888
33	3778	6293	78	2943	7823	23	1872	8907
34	3765	6331	79	2920	7852	24	1849	8925
35	3752	6368	0,80	0,2897	0,7881	25	1826	8944
36	3739	6406	81	2874	7910	26	1804	8962
37	3725	6443	82	2850	7939	27	1881	8980
38	3712	6480	83	2827	7967	28	1858	8997
39	3697	6517	84	2803	7995	29	1836	9015
0,40	0,3683	0,6557	85	2780	8023	1,30	0,1714	0,9032
41	3668	6591	86	2756	8051	31	1691	9049
42	3653	6628	87	2732	8078	32	1669	8066
43	3637	6664	88	2709	8106	33	1647	9082
44	3621	6700	89	2685	8133	34	1626	9099

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
35	1604	9115	1,80	0,0790	0,9641	50	0175	9938
36	1582	9131	81	0775	9649	52	0167	9941
37	1561	9147	82	0761	9656	54	0158	9945
38	1539	9162	83	0748	9664	56	0151	9948
39	1518	9177	84	0734	9671	58	0143	9951
1,40	0,1497	0,9192	85	0721	9678	2,60	0,0136	0,9953
41	1476	9207	86	0707	9686	62	0129	9956
42	1456	9222	87	0694	9693	64	0122	9959
43	1435	9236	88	0681	9699	66	0116	9961
44	1415	9251	89	0669	9706	68	0110	9963
45	1394	9265	1,90	0,0656	0,9713	70	0104	9965
46	1374	9279	91	0644	9719	72	0099	9967
47	1354	9292	92	0632	9729	74	0093	9969
48	1334	9306	93	0620	9732	76	0088	9971
49	1315	9319	94	0608	9738	78	0084	9973
1,50	0,1295	0,9332	95	0596	9744	2,80	0,0079	0,9974
51	1276	9345	96	0584	9750	82	0075	9976
52	1257	9357	97	0573	9756	84	0071	9977
53	1238	9370	98	0562	9761	86	0067	9979
54	1219	9382	99	0551	9767	88	0063	9980
55	1200	9394	2,00	0,0540	0,9772	90	0,0060	0,9981
56	1182	9406	02	0519	9783	92	0056	9982
57	1163	9418	04	0498	9793	94	0053	9984
58	1145	9429	06	0478	9803	96	0050	9985
59	1127	9441	08	0459	9812	98	0047	9986
1,60	0,1109	0,9452	10	0440	9821	3,00	00443	0,99865
61	1092	9463	12	0422	9830	3,10	00327	99903
62	1074	9474	14	0404	9838	3,20	00238	99931
63	1057	9484	16	0387	9846	3,30	00172	99951
64	1040	9495	18	0371	9854	3,40	00123	99966
65	1023	9505	2,20	0,0355	0,9861	3,50	00087	99976
66	1006	9515	22	0339	9868	3,60	00061	69984
67	0989	9525	24	0325	9875	3,70	00042	99989
68	0973	9535	26	0310	9881	3,80	00029	99993
69	0957	9545	28	0297	9887	3,80	00020	99995
1,70	0,0940	5,9554	30	0283	9893	4,00	0,0001338	0,999968
71	0925	9564	32	0270	9898	4,50	0000160	999997
72	0909	9573	34	0258	9904	5,00	0000015	99999997
73	0893	9583	36	0246	9909			
74	0878	9591	38	0235	9913			
75	0863	9599	2,40	0,0224	0,9918			
76	0848	9608	42	0213	9922			
77	0833	9616	44	0203	9927			
78	0818	9625	46	0194	9931			
79	0804	9633	48	0184	9934			

$$\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$$ ფუნქცია

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,00000	0,45	0,34729	0,90	0,63188
01	00798	46	35448	91	63718
02	01596	47	36164	92	64243
03	02593	48	36877	93	64763
04	03191	49	37587	94	65278
05	039[illegible]8	0,50	0,38292	95	65789
06	04784	51	38995	96	66294
07	05581	52	39694	97	66795
08	06576	53	40389	98	67291
09	07171	54	41080	99	67783
0,10	0,07966	55	41768	1,00	0,68269
11	08759	56	42452	01	68750
12	09552	57	43132	02	69227
13	10343	58	43809	03	69699
14	11134	59	44481	04	70166
15	11924	0,60	0,45149	05	70628
16	12712	61	45814	06	71086
17	13499	62	46474	07	71538
18	14285	63	47131	08	71986
19	15069	64	47783	09	72429
0,20	0,15852	65	48431	1,10	0,72867
21	16633	66	49075	11	73300
22	17413	67	49714	12	73729
23	18191	68	50350	13	74152
24	18967	69	50981	14	74571
25	19741	0,70	0,51607	15	74986
26	20514	71	52230	16	75395
27	21284	72	52348	17	75800
28	22052	73	53461	18	76200
29	22818	74	54070	19	76595
0,30	0,23582	75	54675	1,20	0,76986
31	24344	76	55275	21	77372
32	25103	77	55870	22	77754
33	25860	78	56461	23	78130
34	26614	79	57047	24	78502
35	27366	0,80	0,57629	25	78870
36	28115	81	58206	26	79233
37	28862	82	58778	27	79592
38	29605	83	59346	28	79945
39	30346	84	59909	29	80295
0,40	0,31084	85	[illegible]0468	1,30	0,80640
41	31819	86	61[illegible]21	31	80980
42	32552	87	61570	32	81316
43	33280	88	62114	33	81642
44	34006	89	62653	34	81975

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,35	0,82298	1,85	0,93569	2,35	0,98123
36	82617	86	93711	36	98172
37	82931	87	93852	37	98221
38	83241	88	93989	38	98269
39	83547	89	94124	39	98315
1,40	0,83849	1,90	0,94257	2,40	0,98360
41	84146	91	94387	41	98405
42	84439	92	94514	42	98448
43	84728	93	94639	43	98490
44	85013	94	94762	44	98531
45	85294	95	94882	45	98571
46	85571	96	95000	46	98611
47	85844	97	95116	47	98649
48	86113	98	95230	48	98686
49	86378	99	95341	49	98723
1,50	0,86639	2,00	0,95450	2,50	0,08758
51	86696	01	95557	51	98793
52	87149	02	95662	52	98826
53	87398	03	95764	53	98859
54	87644	04	95865	54	98891
55	87886	05	95964	55	98923
56	88124	06	96060	56	98953
57	88358	07	96155	57	98983
58	88589	08	96247	58	99012
59	88817	09	96338	59	99040
1,60	0,89040	2,10	0,96427	2,60	0,99068
61	89260	11	96514	61	99095
62	89477	12	96599	62	99121
63	89690	13	96683	63	99146
64	89899	14	96765	64	99171
65	90106	15	96844	65	99195
66	90309	16	96923	66	99219
67	90508	17	96999	67	99241
68	90704	18	97074	68	99263
69	90897	19	97148	69	99285
1,70	0,91087	2,20	0,97219	2,70	0,99307
71	91273	21	97289	71	99327
72	91452	22	97358	72	99347
73	91637	23	97425	73	99367
74	91814	24	97491	74	99386
75	91988	25	97555	75	99404
76	92159	26	97618	76	99422
77	92327	27	97679	77	99439
78	92492	28	97739	78	99456
79	92655	29	97798	79	99473
1,80	0,92814	2,30	0,97855	2,80	0,99489
81	92970	31	97911	81	99505
82	93124	32	97966	82	99520
83	93275	33	98019	83	99535
84	93423	34	98072	84	99549

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,85	0,99563	3,25	0,99885	3,65	0,99974
86	99576	26	99889	66	99975
87	99590	27	99892	67	99976
88	99602	28	99896	68	99977
89	99615	29	99900	69	99978
2.90	0,99627	3,30	0,99903	3,70	0,99978
91	99639	31	99907	71	99979
92	99650	32	99610	72	99980
93	99661	33	99913	73	99981
94	99672	34	99916	74	99982
95	99682	35	99919	75	99982
96	99692	36	99922	76	99983
97	99702	37	99925	77	99984
98	99712	38	99928	78	99984
99	99721	39	99930	79	99985
3,00	0,99730	3,40	0,99933	3,80	0,99986
01	99739	41	99935	81	99986
02	99747	42	99937	82	99987
03	99755	43	99940	83	99987
04	99763	44	99942	84	99988
05	99771	45	99944	85	99988
06	99779	46	99946	86	99989
07	99786	47	99948	87	99989
08	99793	48	99950	88	99990
09	99800	49	99952	89	99990
3,10	0,99806	3.50	0,99953	3,90	0,99990
11	99813	51	99955	91	99991
12	99819	52	99957	92	99991
13	99825	53	99958	93	99992
14	99831	54	99960	94	99992
15	99837	55	99961	95	99992
16	99842	56	99963	96	99992
17	99848	57	99964	97	99993
18	99853	58	99966	98	99993
19	99858	59	99967	99	99993
3,20	0,99863	3,60	0,99968		
21	99867	61	99969		
22	99872	62	99971		
23	99876	63	99972		
24	99880	64	99973		

პუასონის ფუნქცია

$$P_n(m) \simeq \frac{a^m e^{-a}}{m!}$$

m \ a	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003

m \ a	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367879	0,270671	0,149361
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090224	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8		0,000002	0,000004	0,000009	0,000859	0,008101
9				0,000001	0,000191	0,002701
10					0,000038	0,000810
11					0.000007	0,000221
12					0,000001	0,000055
13						0,000013
14						0,000003
15						0,000001

m \ a	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111
2	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998
3	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994
4	0,195367	0,175467	0,133853	0,091226	0,057252	0,033737
5	0,156293	0,175467	0,160623	0,027717	0,091604	0,060727
6	0,104194	0,146223	0,160623	0,149003	0,122138	0,091090
7	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116
8	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756
9	0,013231	0,036266	0,068838	0,101405	0,124077	0,131756
10	0,005292	0,018133	0,041303	0,070983	0,099262	0,118085
11	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020
12	0,000642	0,003434	0,011262	0,026350	0,048127	0,072765
13	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376
14	0,000056	0,000472	0,002228	0,007094	0,016924	0,032384
15	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786
18		0,000004	0,000039	0,000232	0,000944	0,002893
19		0,000001	0,000012	0,000085	0,000397	0,001370
20			0,000004	0,000030	0,000159	0,000617
21			0,000001	0,000010	0,000061	0,000264
22				0,000003	0,000022	0,000108
23				0,000001	0,000008	0,000042
24					0,000003	0,000016
25					0,000001	0,000006
26						0,000002
27						0,000001

სტიუდენტის განაწილების მიხედვით $P(|t|\geqslant t_1)$ ალბათობანი

t_1 \ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	∞
0,0	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,0000
0,1	0,937	0,929	0,927	0,925	0,924	0,924	0,923	0,923	0,923	0,922	0,922	0,922	0,922	0,922	0,922	0,922	0,922	0,921	0,921	0,921	0,9203
0,2	874	860	854	851	849	848	847	846	846	845	845	845	845	844	844	844	844	844	844	844	8415
0,3	814	792	784	779	776	774	773	772	771	770	770	769	769	769	768	768	768	768	767	767	7642
0,4	758	728	716	710	706	703	701	700	698	698	697	696	696	695	695	694	694	694	694	693	6892
0,5	705	667	651	643	628	635	632	631	629	628	627	626	625	625	624	624	623	623	623	623	6171
0,6	656	609	591	581	575	570	567	565	563	562	561	560	559	558	557	557	556	556	556	555	5485
0,7	611	556	534	523	515	510	507	504	502	500	498	497	496	495	495	494	493	493	492	492	4839
0,8	570	508	482	469	460	454	450	447	444	442	441	439	438	437	436	435	435	434	434	433	4237
0,9	533	463	434	419	409	403	398	394	392	389	387	384	384	383	382	381	381	380	379	379	3681
1,0	500	423	391	374	363	356	351	347	343	341	359	337	336	334	333	332	331	331	330	329	3173
1,1	470	386	352	333	321	313	308	303	300	297	295	293	291	290	289	288	287	286	285	284	2713
1,2	442	353	316	296	284	275	269	264	261	258	255	253	252	250	249	248	247	246	245	244	2301
1,3	417	323	284	263	250	241	235	230	226	223	220	218	216	215	213	212	211	210	209	208	1936
1,4	395	296	256	234	220	221	204	199	195	192	189	187	185	183	182	181	179	179	178	177	1615
1,5	374	272	231	208	194	184	177	172	168	165	162	159	158	156	154	153	152	151	150	149	1336
1,6	356	251	208	185	170	161	154	148	144	141	138	136	134	132	130	129	128	127	126	125	1096
1,7	339	231	188	164	150	140	133	128	123	120	117	115	113	111	110	108	107	106	105	105	0891
1,8	323	214	170	146	132	122	115	110	105	102	099	097	095	093	092	091	090	089	088	087	0719
1,9	308	198	154	130	116	106	099	094	090	087	084	082	080	078	077	076	075	074	073	072	0574
2,0	295	184	139	116	102	092	086	081	077	073	071	069	067	065	064	063	062	061	060	059	0455
2,1	283	171	127	104	090	080	074	069	065	062	060	058	056	054	053	052	051	050	049	049	0357
2,2	272	159	115	093	079	070	064	059	055	052	050	048	046	045	044	043	042	041	040	040	0278
2,3	261	148	105	083	070	061	055	050	047	044	042	040	039	037	036	035	034	034	033	032	0214
2,4	251	138	096	074	062	053	047	043	040	037	035	034	032	031	030	029	028	027	027	026	0164
2,5	242	130	088	067	054	047	041	037	034	031	030	028	027	025	024	024	023	022	022	021	0124
2,6	234	122	080	060	048	041	035	032	029	026	025	023	022	021	020	019	019	018	018	017	0093
2,7	226	114	074	054	043	036	031	027	024	022	021	019	018	017	016	016	015	015	014	014	0069
2,8	218	107	068	049	038	031	027	023	021	019	017	016	015	014	013	013	012	012	011	011	0051
2,9	211	101	063	044	034	027	023	020	018	016	014	013	012	012	011	010	010	010	009	009	0037

t \ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	∞
3,0	205	095	058	040	030	024	020	017	015	013	012	011	010	010	009	008	008	008	007	007	0027
3,1	199	090	053	036	027	021	017	015	013	011	010	009	008	008	007	007	006	006	006	005	0019
3,2	193	085	048	033	024	019	015	013	011	009	008	008	007	006	006	006	005	005	005	004	0014
3,3	187	081	045	030	021	016	013	011	009	008	007	006	006	005	005	005	004	004	004	004	0010
3,4	182	077	042	027	019	014	011	009	008	007	006	005	005	004	004	004	003	003	003	003	0007
3,5	177	073	039	025	017	013	010	008	007	006	005	004	004	004	003	003	003	003	002	002	0004
3,6	172	069	037	023	016	011	009	007	006	005	004	004	003	003	002	002	002	002	002	002	0003
3,7	168	066	034	021	014	010	008	006	005	004	004	003	003	002	002	002	002	002	002	001	0002
3,8	164	063	032	019	013	009	007	005	004	003	003	003	002	002	002	002	001	001	001	001	0001
3,9	160	060	030	018	011	008	006	005	004	003	002	002	002	002	001	001	001	001	001	001	0001
4,0	156	057	028	016	010	007	005	004	003	003	002	002	002	001	001	001	001	001	001	001	0001
4,1	152	055	026	015	009	006	005	003	003	002	002	001	001	001	001	001	001	001	001	001	0000
4,2	149	052	025	014	008	006	004	003	002	002	001	001	001	001	001	001	001	001	000	000	
4,3	145	050	023	013	008	005	004	003	002	002	001	001	001	001	001	001	000	000			
4,4	142	048	022	012	007	005	003	002	002	001	001	001	001	001	001	000					
4,5	139	046	020	011	006	004	003	002	001	001	001	001	001	000	000						
4,6	136	044	019	010	006	004	002	002	001	001	001	001	000								
4,7	133	042	018	009	005	003	002	002	001	001	001	001									
4,8	131	041	017	009	005	003	002	001	001	001	001	000									
4,9	128	039	016	008	004	003	002	001	001	001	000										
5,0	126	038	015	007	004	002	002	001	001	001											
5,2	121	035	014	007	003	002	001	001	001	000											
5,4	117	033	012	006	003	002	001	001	000												
5,6	112	030	011	005	003	001	001	001													
5,8	109	028	010	004	002	001	001	000													
6,0	105	027	009	004	002	001	001														

ცხრილი V

$P=0,05$ ნდობის ალბათობისათვის T-ს მნიშვნელობანი

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	3,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,52
45	4,06	3,21	3,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	

ცხრილი VI

P = 0,01 ნდობის ალბათობისათვის T-ს მნიშვნელობანი

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	4052,10	4999,03	5403,49	5625,14	5764,08	5859,39	5981,34	6105,83	6234,16	6366,48
2	98,49	99,01	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,82	4,31	3,99	3,75	3,45	3,12	2,75	2,30
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,37	1,90
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,82
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,23	1,75
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,18	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,78	2,45	2,07	1,53
80	6,96	4,88	4,04	3,56	3,26	3,04	2,74	2,42	2,03	1,47
90	6,92	4,85	4,01	3,53	3,23	3,01	2,72	2,39	2,00	1,43
100	6,90	4,82	3,98	3,51	3,21	2,99	2,69	2,37	1,98	1,39
125	6,84	4,78	3,94	3,47	3,17	2,95	2,66	2,33	1,94	1,32
150	6,81	4,75	3,91	3,45	3,14	2,92	2,63	2,31	1,92	1,27
200	6,76	4,71	3,88	3,41	3,11	2,89	2,60	2,28	1,88	1,21
300	6,72	4,68	3,85	3,38	3,08	2,86	2,57	2,24	1,85	1,14
400	6,70	4,66	3,83	3,37	3,06	2,85	2,56	2,23	1,84	1,11
500	6,69	4,65	3,82	3,36	3,05	2,84	2,55	2,22	1,83	1,08
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,53	2,20	1,81	1,04
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	

χ^2 კრიტერიუმისათვის p ალბათობანი

χ^2 \ k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,3173	0,6065	0,8013	0,9098	0,9626	0,9856	0,9948	0,9982	0,9994	0,9998	0,9999	1,0000	1,0000	1,0000	1,0000
2	1574	3679	5724	7358	8491	9197	9598	9810	9915	9963	9985	0,9994	0,9998	0,9999	1,0000
3	0833	2231	3916	5578	7000	8088	8850	9344	9643	9814	9907	9955	9979	9991	0,9996
4	0455	1353	2615	4060	5494	6767	7798	8571	9114	9473	9699	9834	9912	9955	9977
5	0254	0821	1718	2873	4159	5438	6600	7576	8343	8912	9312	9580	9752	9858	9921
6	0143	0498	1116	1991	3062	4232	5398	6472	7399	8153	8734	9161	9462	9665	9797
7	0081	0302	0719	1359	2206	3208	4289	5366	6371	7254	7991	8576	9022	9347	9576
8	0047	0183	0460	0916	1562	2381	3326	4335	5341	6288	7133	7851	8436	8893	9238
9	0027	0111	0293	0611	1091	1736	2527	3423	4373	5321	6219	7029	7729	8311	8775
10	0016	0067	0186	0404	0752	1247	1886	2650	3505	4405	5304	6160	6939	7622	8197
11	0009	0041	0117	0266	0514	0884	1386	2017	2757	3575	4433	5289	6108	6860	7526
12	0005	0025	0074	0174	0348	0620	1006	1512	2133	2851	3626	4457	5276	6063	6790
13	0003	0015	0046	0113	0234	0430	0721	1119	1626	2237	2933	3690	4478	5265	6023
14	0002	0009	0029	0073	0156	0296	0512	0818	1223	1730	2330	3007	3738	4497	5255
15	0001	0006	0018	0047	0104	0203	0360	0591	0909	1321	1825	2414	3074	3782	4514
16	0001	0003	0011	0030	0068	0138	0251	0424	0669	0996	1411	1912	2491	3134	3821
17	0000	0002	0007	0019	0045	0093	0174	0301	0487	0744	1079	1496	1993	2562	3189
18		0001	0004	0012	0029	0062	0120	0212	0352	0550	0816	1157	1575	2068	2627
19		0001	0003	0008	0019	0042	0082	0149	0252	0403	0611	0885	1231	1649	2137
20		0000	0002	0005	0013	0028	0056	0103	0179	0293	0453	0671	0952	1301	1719
21		0000	0001	0003	0008	0018	0038	0071	0126	0211	0334	0504	0729	1016	1368
22		0000	0001	0002	0005	0012	0025	0049	0089	0151	0244	0375	0554	0786	1078
23		0000	0000	0001	0003	0008	0017	0034	0062	0107	0177	0277	0417	0603	0841
24		0000	0000	0001	0002	0005	0011	0023	0043	0076	0127	0203	0311	0458	0651
25		0000	0000	0001	0001	0003	0008	0016	0030	0053	0091	0148	0231	0346	0499
26		0000	0000	0000	0001	0002	0005	0010	0020	0037	0065	0107	0170	0259	0380
27		0000	0000	0000	0001	0001	0003	0007	0014	0026	0046	0077	0124	0193	0287
28		0000	0000	0000	0000	0001	0002	0005	0010	0018	0032	0053	0090	0142	0216
29		0000	0000	0000	0000	0001	0001	0003	0006	0012	0023	0039	0065	0104	0161
30		0000	0000	0000	0000	0000	0001	0002	0004	0009	0016	0028	0047	0076	0119

გაგრძელება

k / χ^2	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1,0000	1,0000	1.0000	1,0000	1.0000	1,0000	1,0000	1.0000	1.0000	1.0000	1.0000	1.0000	1,0000	1,0000
2	1,0000	1,0000	1.0000	1.0000	1.0000	1,0000	1,0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0,9998	0,9999	1.0000	1.0000	1.0000	1,0000	1,0000	1.0000	1,0000	1,0000	1.0000	1.0000	1.0000	1.0000
4	9989	9995	0,9998	0,9999	1.0000	1,0000	1,0000	1.0000	1.0000	1.0000	1.0000	1.0000	1,0000	1.0000
5	9958	9978	9989	9994	0.9997	0,9999	0,9999	1.0000	1.0000	1.0000	1,0000	1.0000	1.0000	1.0000
6	9881	9932	9962	9979	9989	9994	9997	0.9999	0.9999	1,0000	1.0000	1,0000	1,0000	1.0000
7	9733	9835	9901	9942	9967	9981	9990	9995	9997	0.9999	0.9999	1.0000	1.0000	1.0000
8	9489	9665	9786	9867	9919	9951	9972	9984	9991	9995	9997	0.9999	0.9999	1,0000
9	9134	9403	9597	9735	9829	9892	9933	9960	9976	9986	9992	9995	9997	0,9999
10	8666	9036	9319	9539	9682	9789	9863	9913	9945	9967	9980	9988	9993	9996
11	8095	8566	8944	9238	9462	9628	9747	9832	9890	9929	9955	9972	9983	9990
12	7440	8001	8472	8856	9161	9396	9574	9705	9799	9866	9912	9943	9964	9977
13	6728	7362	7916	8386	8774	9086	9332	9520	9661	9765	9840	9892	9929	9954
14	5987	6671	7291	7837	8305	8696	9015	9269	9466	9617	9730	9813	9872	9914
15	5246	5955	6620	7226	7764	8230	8622	8946	9208	9414	9573	9694	9784	9850
16	4530	5238	5925	6573	7166	7696	8159	8553	8881	9148	9362	9529	9658	9755
17	3856	4544	5231	5899	6530	7111	7634	8093	8487	8818	9091	9311	9486	9622
18	3239	3888	4557	5224	5874	6490	7060	7575	8030	8424	8758	9035	9261	9443
19	2687	3285	3918	4568	5218	5851	6453	7012	7520	7971	8364	8700	8981	9213
20	2202	2742	3328	3946	4579	5213	5830	6419	6968	7468	7916	8308	8645	8929
21	1785	2263	2794	3368	3971	4589	5207	5811	6387	6926	7420	7863	8253	8591
22	1432	1847	2320	2843	3405	3995	4599	5203	5793	6357	6887	7374	7813	8202
33	1137	1493	1906	2373	2888	3440	4017	4608	5198	5776	6329	6850	7330	7765
24	0895	1194	1550	1962	2424	2931	3472	4038	4616	5194	5760	6303	6815	7289
25	0698	0947	1249	1605	2014	2472	2971	3503	4058	4624	5190	5745	6278	6782
26	0540	0745	0998	1302	1658	2064	2517	3009	3532	4076	4631	5186	5730	6255
27	0415	0581	0790	1047	1353	1709	2112	2560	3045	3559	4093	4638	5182	5717
28	0316	0449	0621	0834	1094	1402	1757	2158	2600	3079	3585	4110	4644	5179
29	0239	0345	0484	0660	0878	1140	1449	1803	2201	2639	3111	3609	4125	4651
30	0180	0263	0374	0518	0699	0920	1185	1494	1848	2243	2676	3142	3632	4140

t_α-ს ის მნიშვნელობანი, რომელთათვისაც $p(-t_\alpha<t<t_\alpha)=\alpha$

k \ α	0,9	0,95	0,98	0,99	0,999
1	6,31	12,71	31,82	63,66	636,2
2	2,92	4,30	6,97	9,93	31,60
3	2,35	3,18	4,54	5,84	12,94
4	2,13	2,78	3,75	4,60	8,61
5	2,02	2,57	3,37	4,03	6,86
6	1,94	2,45	3,14	3,70	5,96
7	1,90	2,37	3,00	3,50	5,40
8	1,86	2,30	2,90	3,36	5,04
9	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,49
12	1,78	2,18	2,68	3,06	4,32
13	1,77	2,18	2,65	3,06	4,12
14	1,76	2,14	2,62	2,98	4,14
15	1,75	2,13	2,60	2,95	4,07
16	1,75	2,12	2,58	2,92	4,02
17	1,74	2,11	2,57	2,90	3,97
18	1,73	2,10	2,55	2,88	3,92
19	1,73	2,09	2,54	2,86	3,88
20	1,72	2,09	2,53	2,85	3,85
21	1,72	2,08	2,52	2,83	3,82
22	1,72	2,07	2,51	2,82	3,79
23	1,71	2,07	2,50	2,81	3,77
24	1,71	2,06	2,49	2,80	3,75
25	1,71	2,06	2,49	2,79	3,72
26	1,71	2,06	2,48	2,78	3,71
27	1,70	2,05	2,47	2,77	3,69
28	1,70	2,05	2,47	2,76	3,67
29	1,70	2,05	2,46	2,76	3,66
30	1,70	2,04	2,46	2,75	3,65
40	1,68	2,02	2,42	2,70	3,55
60	1,67	2,00	2,39	2,66	3,36
120	1,66	1,98	2,36	2,62	3,37
∞	1,65	1,96	2,33	2,58	3,29

$$P(\chi^2>\chi_q^2)=\frac{1}{\Gamma\left(\frac{k}{2}\right)\cdot 2^{k/2}}\int\limits_{\chi_q^2}^{\infty} x^{(k/2)-1}\, e^{-x/2}\, dx$$

ალბათობასთან დამოკიდებულებით χ_q^2 მნიშვნელობათა და χ^2-განაწილების თავისუფლების ხარისხთა k რიცხვი

თავისუფლების ხარისხთა რიცხვი	$P(\chi^2>\chi_q)$ ალბათობა															
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	0,55	0,75	1,14	1,61	2,35	3,00	4,35	6,1	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7	23,9	27,2	30,1	33,7	36,2	48,5	41,5	43,8
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	36,7	40,3	42,5	46,0	48,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	52,5	56,0	58,3
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

z_q ნორმალური გადახრის q-პროცენტიანი მნიშვნელობანი და

$$q=100P(|x-a|>z_q\sigma)=100\left(1-\frac{1}{\sqrt{2\pi}}\int_{-z_q}^{z_q}\exp\{-t^2/2\}\,dt\right)$$

ნორმალური განაწილების ალბათობა

z_q როგორც q-ს ფუნქცია		q როგორც z_q-ს ფუნქცია	
$q=100\,Q$	z_q	z_q	q
100	0,0000	0 0	100,000
95	0,0627	0,2	84,148
90	0,1257	0,4	68,916
85	0,1891	0,6	54,851
80	0,2533	0,8	42,371
75	0,3186	1,0	31,731
70	0,3853	1,2	23,014
65	0,4538	1,4	16,151
60	0,5244	1,6	10,960
55	0,5978	1,8	7,186
50	0,6745	2,0	4,550
45	0,7554	2,2	2,781
40	0,8416	2,4	1,640
35	0,9346	2,6	6,932
30	1,0364	2,8	0.511
25	1,1503	3,0	0,270
20	1,2816	3,2	0,137
15	1,4395	3,4	0,067
10	1,6449	3,6	0,032
5	1,9600	3,8	0,014
1	2,5758	4,0	0,006
0,27	3,0000		
0,1	3,2905		
0,01	3,8906		

ლიტერატურა

1. А. Я. Боярский, Математика для экономистов, Госстатиздат, М. 1961.
2. Б. В. Гнеденко, Курс теории вероятностей, ГТТИ. 1954.
3. И. В. Дунин-Борковский и Н. В. Смирнов, Теория вероятностей и математическая статистика в технике. ГТТИ, 1955.
4. პ. ზერაგია და გ. მანია, უმაღლესი მათემატიკის კურსი, თბილისის უნივერსიტეტის გამომცემლობა, 1967 წ.
5. ჰ. კრინსკი, მათემატიკა ეკონომისტებისათვის, თბილისის უნივერსიტეტის გამომცემლობა, 1973.
6. გ. მანია, მათემატიკური სტატისტიკა ტექნიკაში, გამომცემლობა „საბჭოთა საქართველო", 1958 წ.
7. გ. მანია, ალბათობის თეორიის კურსი, თსუ გამომცემლობა, 1962 წ.
8. გ. მანია, მათემატიკური სტატისტიკის ზოგიერთი მეთოდი, საქართველოს სსრ მეცნიერებათა აკადემიის გამომცემლობა, 1963 წ.

საძიებელი

შინაარსი

V თ ა ვ ი. შემთხვევითი სიდიდის მახასიათებელი

VI თ ა ვ ი. ნორმალური განაწილების კანონი

VII თ ა ვ ი. დიდ რიცხვთა კანონი

ნაწილი მეორე

მათემატიკური სტატისტიკა

I თავი. უმცირეს კვადრატთა ხერხი

II თავი. შერჩევითი მეთოდი

III თავი. განაწილების პარამეტრების სტატისტიკური შეფასება

IV თავი. ჰიპოთეზათა სტატისტიკური შეფასება

V თავი. დისპერსიული ანალიზი

რედაქტორი ა. ტორონჯაძე
გამომცემლობის რედაქტორი ა. სტურუა
ტექრედაქტორი ი. ხუციშვილი
კორექტორი ც. მოლოდინი

გადაეცა წარმოებას 30/IV-76
ხელმოწერილია დასაბეჭდად 30/IV-76
ქაღალდის ფორმატი 60×90/16
ნაბეჭდი თაბახი 22,25
სააღრიცხვო-საგამომცემლო თაბახი 18,49

შეკვეთა 812 უე 06566 ტირაჟი 2000

ფასი 88 კაპ.

თბილისის უნივერსიტეტის გამომცემლობა,
თბილისი, 380028, ი. ჭავჭავაძის პროსპექტი, 14.
Издательство Тбилисского университета,
Тбилиси, 380028, пр. И. Чавчавадзе, 14.
თბილისის უნივერსიტეტის სტამბა,
თბილისი, 380028, ი. ჭავჭავაძის პროსპექტი, 1.
Типография Тбилисского университета.
Тбилиси, 380028, пр. И. Чавчавадзе, 1.

Гванджи Михайлович Мания

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

(на грузинском языке)

Издательство Тбилисского университета
Тбилиси 1976