

Jaotuse sobitamine

Mudel, tõenäosusjaotus, ühtlane jaotus, binoomjaotus, eksponentjaotus, normaaljaotus, Cauchy jaotus*, lognormaaljaotus, Laplace'i jaotus*, Levy jaotus*, Pareto jaotus*.

Gümnaasium

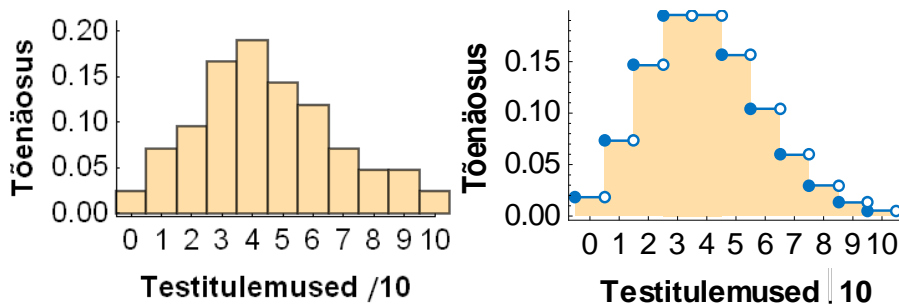
Mudeliks sobiva tõenäosusjaotuse valik

Andmete analüüsimiseks ja nende põhjal järelduste tegemiseks on sageli otstarbekas andmed siduda **mudelig**a, mis neid küllalt hästi kirjeldab. Mudeliks sobib mõni **tõenäosusjaotus**, mille omadused on hästi teada. Kasutades analüüsimisel andmete põhjal moodustatud **empiriilise jaotuse** asemel sobivat tõenäosusjaotust vabaneme me andmetest tulenevast juhuslikkusest ja saame mudelit kasutada analüüsimiseks ja prognoosimiseks.

Selleks on tarvis, et valimisse kuuluvad andmed võiksid põhimõtteliselt olla selle tõenäosusjaotuse üksikväärtused. See tähendab, et kujutleme, et üldkogum ongi sellise tõenäosusjaotusega ja valim pärinebki sellest jaotusest.

Hästi sobib mudel andmetega siis, kui sagedamini esinevatele andmetele vastavad mudeli suurema tõenäosusega väärtused (väärtuste vahemikud). See tähendab, et valimit kirjeldav empiriline jaotus on oma kuju poolest sarnane üldkogumit esindava tõenäosusjaotusega. .

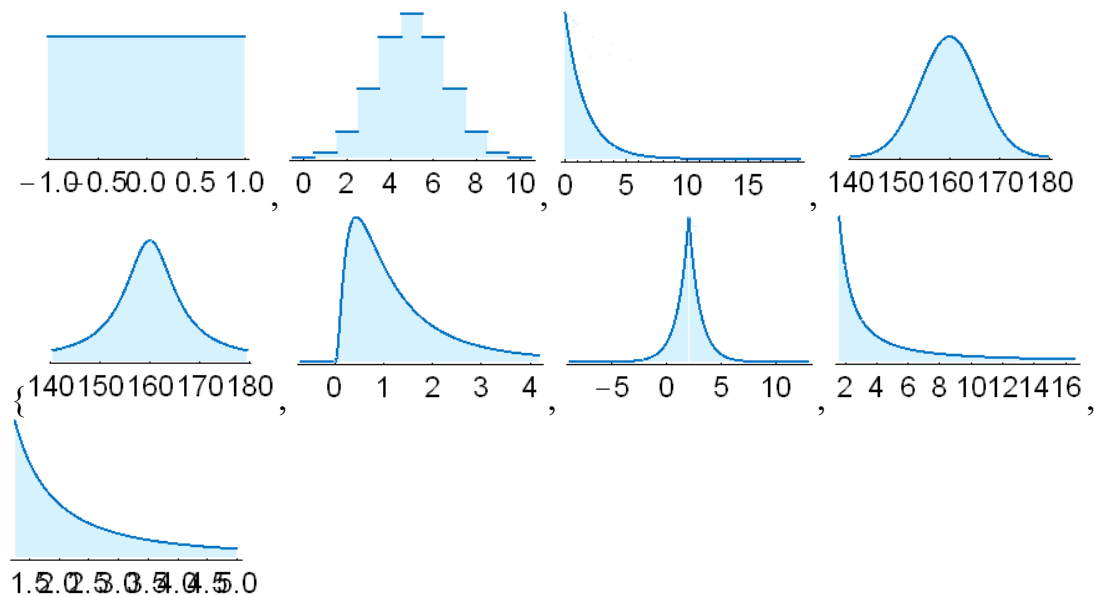
Seda tingimust saab visuaalselt kontrollida. Moodustame andmete põhjal **empiriilise jaotuse histogrammi** ja võrdleme seda erinevate tõenäosusjaotuste histogrammide või tihedusfunktsioonide graafikutega. Sobivaim mudel on see tõenäosusjaotus, mille graafik on oma kuju poolest võimalikult sarnane empirilise jaotuse histogrammiga.



Vasakpoolsel joonisel kujutab histogramm kümne küsimusega testi tulemuste empirilist jaotust, paremal Poissoni jaotust. On näha, et see tõenäosusjaotus sobib kuju poolest empirilise jaotusega üsna hästi.

Mudeli jaoks sobiva tõenäosusjaotuse valimisel on abiks eelteadmised uuritava nähtuse või protsessi kohta. Nii on üldiselt teada, missuguseid nähtusi kirjeldab binoomjaotus või normaaljaotus. Näiteks juhuslikul saabumisel trammipeatusesse kirjeldab ooteaega ühtlane jaotus. Suhteliselt harva esineva nähtuse (nt trükivigade arv raamatu leheküljel) mudeliks sobib Poissoni jaotus. Mõnikord viitab andmete saamise viis sellele, millist mudeljaotust on sobiv kasutada. Siit võivad tulla ka kitsendused, mis välistavad mõne mudeljaotuse kasutamise, näiteks ei sobi mõnikord mudeliks sellised jaotused, mille puhul juhuslik suurus võib omandada ka negatiivseid väärtusi. Kui andmed on saadud loendamise tulemusena, on tegemist diskreetse jaotusega, kuid ka sellise jaotuse mudeliks võib sobida pidev jaotus. Vahel peab mudel kindlasti olema sümmeetriline, teinekord on oluline, et

modeli väärtused on positiivsed. Enamasti on kasulik valida võimalikult lihtne mudel, mida on kerge arvutamisel kasutada ja tõlgendada.



Joonisel on esitatud mõned tõenäosusjaotused, mida jaotuse sobitamisel tihti kasutatakse: [Link tüüpiliste tõenäosusjaotuste videole]. Need on järjest: ühtlane jaotus, binoomjaotus, eksponentjaotus, normaaljaotus, Cauchy jaotus, lognormaaljaotus, Laplace'i jaotus, Levy jaotus, Pareto jaotus.

Enamasti ei ole ülalnimetatud jaotuste puhul tegemist siiski ühe ainsa jaotusega, vaid iga nimetatud tõenäosusejaotus esitab jaotuste peret, millest kõige sobivam tuleb välja valida. Kõige paremini sobib antud empiirilise jaotusega see, mille parameetrid kõige paremini sobivad selle empiirilise jaotusega. Sageli tähendab see, et mudeljaotuse keskväärtuseks valitakse empiirilise jaotuse keskmine. Kui mudeljaotusel on parameetreid rohkem, siis tuleb ka need valimi põhjal hinnata.

Kõige sagedamini kasutatakse mudeljaotusena **normaaljaotust**, see sobib enamasti siis, kui mõõdetava tunnuse väärtus sõltub paljudest juhuslikest teguritest (liidetavatest), nii on olukord paljude bioloogiliste tunnustega. Ka mõõtmisvigade jaotuseks sobib normaaljaotus. Normaaljaotusel on kaks parameetrit – keskväärtus ja standardhälve. Andmetega kõige paremini sobiva mudeli saamiseks kasutatakse normaaljaotuse parameetritena valimi keskmist ja standardhälvet.

Kokkuvõttes, jaotuste sobitamise ülesande lahendamine olemasoleva valimi põhjal koosneb kahest sammust:

1. Sobiva mudeli – tõenäosusjaotuse valimine, mille puhul lähtutakse valimi empiirilisele jaotusele visuaalselt sarnasest jaotusest;
2. Valitud tõenäosusjaotuse parameetrite hindamine valimi põhjal. Sageli sobivad parameetrite hinnanguteks vastavad empiirilise jaotuse parameetrid. (Vt parameetrite hindamine).