

1 Sissejuhatus tõenäosusteooriasse

1.1 Juhuslik sündmus ja selle tõenäosus

1.1.1 Naturaalarvu faktoriaal

Definitsioon. Naturaalarvu n faktoriaaliks nimetatakse kõikide naturaalarvude korrutist 1-st n -ni, st

$$n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1)n$$

Seega

$$1! = 1$$

$$2! = 1 \cdot 2 = 2$$

$$3! = 1 \cdot 2 \cdot 3 = 6$$

$$4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$$

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

$$6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$$

$$7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$$

$$8! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 = 40320$$

$$9! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 = 362880$$

$$10! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 3628800$$

Eraldi juurde defineeritud suuruseks on

$$0! = 1$$

mis teooriaga on igati kooskõlas.

Faktoriaalidega arvutamisel on sageli kasulik mõned tegurid eraldi välja kirjutada nagu

$$10! = 10 \cdot 9! = 10 \cdot 9 \cdot 8! = 10 \cdot 9 \cdot 8 \cdot 7!$$

jne.

Näide 1. Arvutame $8! - 7! = 8 \cdot 7! - 7! = 7!(8 - 1) = 7 \cdot 7!$

Näide 2. Arvutame $\frac{21!}{19!} = \frac{21 \cdot 20 \cdot 19!}{19!} = 21 \cdot 20 = 420$

1.1.2 Permutatsioonid

Definitsioon. Hulga permutatsioonideks nimetatakse hulga elementide kõikvõimalikke erinevaid järjestusi.

n -elemendilise hulga permutatsioonide hulka tähistatakse P_n .

Vaatleme kolmeelemendilist hulka $A = \{a; b; c\}$. Selle hulga permutatsioonid on a, b, c a, c, b b, a, c b, c, a c, a, b c, b, a . Neid on kokku 6 ehk

$$P_3 = 6 = 3!$$

Lisades hulka neljanda elemendi d , st vaadeldes hulka $B = \{a; b; c; d\}$, saame kolmeelemendilise hulga elementide igale permutatsioonile neljanda elemendi d paigutada neljale erinevale kohale. Kolmeelemendilise hulga esimese permutatsiooni korral d, a, b, c a, d, b, c a, b, d, c a, b, c, d . Samamoodi saab neljanda elemendi lisada neljale erinevale kohale ülejäänud viies kolmeelemendilise hulga permutatsioonis. Neljaelemendilise hulga permutatsioone on seega $P_4 = 4 \cdot 3! = 4!$.

Lisades hulka viienda elemendi, saame viienda elemendi lisada igas neljaelemendilise hulga permutatsioonis viiele erinevale kohale, st viieelemendilise hulga permutatsioone on $P_5 = 5 \cdot 4! = 5!$. Arutlust jätkates saame, et n -elemendilise hulga permutatsioone on

$$P_n = n!$$

Näide 3. Mitu erinevat formaalset sõna saab moodustada tähtedest A, A, A, B, S ja S?

Formaalse sõna all mõistame tähtade suvalist järjestust, näiteks AA-ABSS. Kokku on kuus tähte, mille järjestamiseks on P_6 võimalust. Need ei ole siiski kõik erinevad, sest kolme A omavahelise järjestuse muutmine ei tekita uut formaalset sõna. Samuti ei teki uut formaalset S-de omavahelise järjestuse muutmisel.

A-de järjestusi on P_3 ja S-de järjestusi P_2 . Seega on kuue tähe kõikvõimalike järjestuste seas $P_3 P_2$ ühesugust ja erinevaid formaalseid sõnu saab moodustada

$$\frac{P_6}{P_3 P_2} = \frac{720}{6 \cdot 2} = 60$$

Nende hulgas on ka üks sisulise tähendusega sõna SAABAS.

Näide 4. Mitu viiekohalist arvu saab moodustada numbritest 0, 1, 2, 3 ja 4?

Antud viit numbrit on võimalik järjestada P_5 erineval viisil. Nende hulgas järjestused, milles esimeseks numbriks on 0 nagu 01234, ei ole viiekohalsed arvud. Neid on parajasti nii palju, kui on võimalusi nulli taga nelja ülejäänud numbrit järjestada, st P_4 . Seega viiekohalisi arve on võimalik moodustada

$$P_5 - P_4 = 5! - 4! = 120 - 24 = 96$$

Võib arutleda ka teisiti. Viiekohalise arvu moodustamiseks on esimesele kohale võimalik valida nelja numbrit 1, 2, 3 või 4 hulgast, st neli võimalust. Üle jääb 4 numbrit, mida on esimese valitud numbrit järel võimalik järjestada P_4 erineval viisil ja kokku saame

$$4 \cdot P_4 = 4 \cdot 24 = 96$$

erinevat viiekohalist arvu.

1.1.3 Kombinatsioonid

Vaatleme viieelemendilist hulka $\{a, b, c, d, e\}$. Selle hulga elementide kombinatsioonid ühe kaupa on üheelemendilised alamhulgad $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$ ja $\{e\}$.

Kombinatsioonid kahe kaupa on $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{a, e\}$, $\{b, c\}$, $\{b, d\}$, $\{b, e\}$, $\{c, d\}$, $\{c, e\}$ ja $\{d, e\}$, st kaheelemendilised alamhulgad, mis erinevad üksteisest vähemalt ühe elemendi poolest.

Kombinatsioonid kolme kaupa on $\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, c, d\}$, $\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$, $\{b, d, e\}$ ja $\{c, d, e\}$, st kolmelemendilised alamhulgad, mis erinevad üksteisest vähemalt ühe elemendi poolest.

Kombinatsioonid nelja kaupa on $\{a, b, c, d\}$, $\{a, b, c, e\}$, $\{a, b, d, e\}$, $\{a, c, d, e\}$ ja $\{b, c, d, e\}$, st neljaelemendilised alamhulgad, mis erinevad üksteisest vähemalt ühe elemendi poolest.

Ainsaks kombinatsiooniks viie kaupa on hulk ise.

Kui hulgas on n elementi, on sellest võimalik moodustada ühe-, kahe-, kolme-, ..., n -elemendilisi alamhulki.

Definitsioon. Kombinatsioonideks n elemendist k kaupa ($1 \leq k \leq n$) nimetatakse n -elemendilise hulga k -elemendilisi alamhulki, mis erinevad üksteisest vähemalt ühe elemendi poolest.

Eespool veendusime, et kombinatsioonide viiest elemendist ühekaupa on 5, kahekaupa ja kolmekaupa on 10, neljakaupa 5 ja viiekaupa 1. Kombinatsioonide arvu n elemendist k kaupa tähistatakse C_n^k ja see leitakse valemist

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

Kontrolliks leiame

$$C_5^1 = \frac{5!}{1!4!} = \frac{5 \cdot 4!}{4!} = 5$$

$$C_5^2 = \frac{5!}{2!3!} = \frac{5 \cdot 4 \cdot 3!}{2 \cdot 3!} = 10$$

ja

$$C_5^5 = \frac{5!}{5!0!} = \frac{5!}{5!} = 1$$

Näide 5. Maleturniiril osaleb 8 mängijat. Leiame, mitu partiid mängitakse turniiril, eeldusel, et iga osaleja mängib iga vastasega ühe partii.

Partiisid mängitakse turniiril nii palju, kui on võimalik moodustada erinevaid paare 8 mängija hulgast, st kui on kombinatsioone 8 elemendist (mängijast) kahekaupa. Arvutame

$$C_8^2 = \frac{8!}{2!6!} = \frac{8 \cdot 7 \cdot 6!}{2 \cdot 6!} = 4 \cdot 7 = 28$$

Seega mängitakse sellel turniiril 28 partiid.

Näide 6. Leiame, mitmel viisil on malelauale võimalik paigutada kolme etturit.

Malelaua on 64 ruutu. Erineva paigutuse saame, kui vähemalt ühe etturi paigutame uuele ruudule. Kokku on erinevaid paigutusi niipalju kui on kombinatsioone 64 elemendist (ruudust) kolmekaupaga. Arvutame

$$C_{64}^3 = \frac{64!}{3!61!} = \frac{64 \cdot 63 \cdot 62 \cdot 61!}{6 \cdot 61!} = 64 \cdot 21 \cdot 31 = 41664$$

Seega on kolme etturit võimalik paigutada malelauale 41664 erineval viisil.

1.1.4 Variatsioonid

Vaatleme viieelemendilist hulka $\{a, b, c, d, e\}$. Selle hulga elementide variatsioonid kahekaupa on $\{a, b\}, \{b, a\}, \{a, c\}, \{c, a\}, \{a, d\}, \{d, a\}, \{a, e\}, \{e, a\}, \{b, c\}, \{c, b\}, \{b, d\}, \{d, b\}, \{b, e\}, \{e, b\}, \{c, d\}, \{d, c\}, \{c, e\}, \{e, c\}, \{d, e\}$ ja $\{e, d\}$, st 20 kaheelemendilist alamhulka, mis erinevad üksteisest elementide endi või nende järjestuse poolest.

Definitsioon. Variatsioonideks n elemendist k kaupaga ($1 \leq k \leq n$) nimetatakse n -elemendilise hulga k -elemendilisi alamhulki, mis erinevad üksteisest vähemalt ühe elemendi või elementide järjestuse poolest.

Variatsioonide arvu n elemendist k kaupaga tähistatakse V_n^k ja see leitakse valemist

$$V_n^k = \frac{n!}{(n-k)!}$$

Kinnituseks leiame

$$V_5^2 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{5 \cdot 4 \cdot 3!}{3!} = 20$$

Näide 7. Leiame, mitmel viisil on võimalik malelauale paigutada kuningat, lippu ja etturit.

Kui oleme kolm ruutu välja valinud ja kolm malendit neile paigutada, siis näiteks kuninga ja lipu vahetamine annab meile uue paigutuse, st tulemus sõltub sellest, millises järjekorras malendid kolmele välja valitud ruudule paigutada. Seega on võimalusi kuninga, lipu ja etturi paigutamiseks malelauale

$$V_{64}^3 = \frac{64!}{(64-3)!} = \frac{64!}{61!} = \frac{64 \cdot 63 \cdot 62 \cdot 61!}{61!} = 64 \cdot 63 \cdot 62 = 249984$$

Kuningat, lippu ja etturit on võimalik paigutada malelauale 249984 erineval viisil.

Näide 8. Leiame, mitu erinevat neljakohalist arvu on võimalik moodustada numbrimärkidest 0, 1, 2, 3, 4, 5 ja 6.

Numbrimärke on 7, st saame valida seitsmest nelja kaupa. Erinev numbrimärkide järjestus tekitab erineva arvu. Neid arve on

$$V_7^4 = \frac{7!}{(7-4)!} = \frac{7!}{3!} = \frac{5040}{6} = 840$$

Nende 840 hulgas on ka numbrit järjestused $0xxx$, mis ei ole neljakohalised arvud. 0-le järgnevale kolmele kohale on võimalik valida 6 erineva numbrimärgi hulgast ja neid valikuid on

$$V_6^3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{720}{6} = 120$$

Järelikult on numbrimärkidest 0, 1, 2, 3, 4, 5 ja 6 võimalik moodustada $840 - 120 = 720$ erinevat neljakohalist arvu.

1.1.5 Katse ja sündmus

Katse ja sündmuse mõisteid selgitame näidete varal.

1. Katse: viskame münti ühe korra; sündmus: viskamise (katse) tagajärjel tuleb vapp.
2. Katse: viskame täringut ühe korra; sündmus: viskamise (katse) tagajärjel tuleb kuus silma.
3. Katse: viskame täringut ühe korra; sündmus: viskamise (katse) tagajärjel tuleb paarisarv silmi.
4. Katse: Valime õpperühmast juhuslikult kaks üliõpilast; sündmus: mõlemad valitud üliõpilased on meessoost.
5. Katse: tõmbame kaardipakist juhuslikult kolm kaarti; sündmus: tõmmatud kaartide hulgas on üks risti, üks poti ja üks ruutu.

Kõik nendes näidetes kirjeldatud sündmused on *juhuslikud sündmused*, st katse tagajärjel need võivad toimuda, aga võivad ka mitte toimuda. Kui visata münti üks kord, siis vapp võib tulla, aga võib ka mitte tulla. Kui visata täringut ühe korra, siis kuus silma võib tulla, aga võib ka mitte tulla. Kui valida õpperühmast juhuslikult kaks üliõpilast, siis valituks võivad osutuda mehed, kuid valitud võivad olla ka naised ja valituks võivad osutuda üks mees ja üks naine.

Juhuslikke sündmusi hakkame tähistama suurte tähtedega (nagu hulki) $A, B, C, \dots$. Vajaduse korral indekseeritult $A_1, A_2, A_3, \dots$. Sündmusi defineerime järmsel viisil:

A - mündi viskamisel tuleb vapp;

B - täringu viskamisel tuleb kuus silma;

C - täringu viskamisel tuleb paarisarv silmi.

Kaks äärmuslikku sündmust on *võimatu sündmus* ja *kindel sündmus*.

Sündmus - täringu viskamisel tuleb seitse silma - on võimatu sündmus, sest täringul puudub seitse silma. Võimatut sündmust tähistatakse

$$\emptyset$$

Sündmus - täringu viskamisel tuleb üks kuni kuus silma - on kindel sündmus, sest üks nendest silmade arvust täringu viskamisel kindlasti tuleb. Kindlat sündmust tähistatakse

$$\Omega$$

Sündmuse saab lahutada *elementaarsündmusteks*. Näites sündmuse "täringu viskamisel tuleb paarisarv silmi" saab lahutada kolmeks elementaarsündmuseks: täringu viskamisel tuleb kaks silma, neli silma või kuus silma.

Kui eeldada, et kaardipakis on 36 kaarti, siis sündmuse - kolme kaardi juhuslikul tõmbamisel tuleb üks risti, üks poti ja üks ruutu - saab lahutada elementaarsündmusteks: tuleb risti kuus, poti kuus ja ruutu kuus, tuleb risti kuus, poti kuus ja ruutu seitse jne. Neid elementaarsündmusi on kokku $9 \cdot 9 \cdot 9 = 729$.

Sündmused - mündi viskamisel tuleb vapp - ja - täringu viskamisel tuleb kuus silma - on ise elementaarsündmused.

1.1.6 Sündmuse tõenäosus

Olgu sündmus A - täringu viskamisel saadakse paarisarv silmi.

Kõiki elementaarsündmusi, mis saavad täringu viskamisel toimuda on 6, st täringu viskamisel tuleb üks, kaks, kolm, neli, viis või kuus silma. Paarisarv silmade tulekuks *soodsad* elementaarsündmused on: tuleb kaks, neli või kuus silma, st sündmuse toimumiseks soodsaid elementaarsündmusi on kolm.

Sündmuse toimumise tõenäosus on sündmuse toimumiseks soodsate elementaarsündmuste arv jagatud kõikide elementaarsündmuste arvuga. Seega sündmuse - täringu viskamisel saadakse paarisarv silmi - toimumise tõenäosus on

$$\frac{3}{6} = \frac{1}{2}$$

Sündmuse A toimumise tõenäosust tähistatakse $P(A)$. Kui kõiki katse tagajärjel toimuvaid elementaarsündmusi on n ja sündmuse A toimumiseks soodsaid elementaarsündmusi on m , siis

$$P(A) = \frac{m}{n}$$

Võimatu sündmuse toimumiseks soodsaid elementaarsündmusi ei ole, st nende arv on $m = 0$ ja võimatu sündmuse tõenäosus

$$P(\emptyset) = \frac{0}{n} = 0$$

Kindla sündmuse korral on kõik elemetaarsündmused ühtlasi sündmuse toimumiseks soodsad, st soodsate elementaarsündmuste arv $m = n$ ja kindla sündmuse tõenäosus

$$P(\Omega) = \frac{n}{n} = 1$$

Üldiselt on sündmuse toimumiseks soodsaid elementaarsündmusi mitte vähem kui 0, ja mitte rohkem, kui on kõiki elementaarsündmusi, st

$$0 \leq m \leq n$$

Jagades viimaseid võrratusi kõikide elementaarsündmusete arvuga n , saame

$$0 \leq \frac{m}{n} \leq 1$$

st igasuguse juhusliku sündmuse A toimumise tõenäosus on 0 ja 1 vahel paiknev suurus, st

$$0 \leq P(A) \leq 1$$

Näide 1. Leiame tõenäosuse, et Tähtede A , A , K aj K juhulikul rittaladumisel saadakse sõna $KAAK$.

Olgu sündmus B - saadakse sõna $KAAK$. Kõiki elementaarsündmusi on nii palju, kui on võimalusi nelja tähe järjestamiseks, st $n = 4! = 24$. Arvestades sellega, et sõnas $KAAK$ kahe A ja kahe K ümberjärjestamine sõna ei muuda, on soodsaid elementaarsündmusi selle sõna tekkeks $m = 2! \cdot 2! = 4$. Sündmuse B tõenäosus

$$P(B) = \frac{4}{24} = \frac{1}{6}$$

Näide 2. Karbis on 12 kollast ja 10 valget tennisepalli. Karbist võetakse juhuslikult kaks palli. Leiame tõenäosuse, et mõlemad võetud pallid on kollased.

Olgu sündmus C - mõlemad võetud pallid on kollased. Et karbis on kokku 22 palli on kahe palli võtmiseks (st kõiki elementaarsündmusi) $n = C_{22}^2 = 231$ võimalust. Soodsaid elementaarsündmusi selleks, et mõlemad oleksid kollased, on nii palju, kui on võimalik valida 12 kollase tennisepalli hulgast kahte, st $m = C_{12}^2 = 66$. Seega sündmuse C - mõlemad võetud pallid on kollased - tõenäosus on

$$P(C) = \frac{66}{231} = \frac{2}{7} = 0,286$$

Vastus on ümardatud kolme kohani peale koma.

Näide 3. Raamaturiiulil on 15 eestikeelset ja 5 inglisekeelset raamatut. Riiulilt võetakse huupi kolm raamatut. Leiame tõenäosuse, et võetud kolme raamatu hulgas on 2 eestikeelset ja 1 inglisekeelne raamat.

Olgu sündmus A - võetud kolme raamatu hulgas on 2 eestikeelset ja 1 inglisekeelne raamat.

Kokku on riiulil 20 raamatut. Kõiki elementaarsündmusi on nii palju, kui on võimalik valida 20 raamatu hulgast kolme, st $n = C_{20}^3 = 1140$. Soodsaid elementaarsündmusi selleks, et võetud kolme raamatu hulgas oleks 2 eestikeelset raamatut ja 1 inglisekeelne raamat on $m = C_{15}^2 \cdot C_5^1 = 105 \cdot 5 = 525$. Järelikult sündmuse A tõenäosus

$$P(A) = \frac{525}{1140} = \frac{35}{76} = 0,461$$

Definitsioon. Sündmuse A *vastandsündmuseks* nimetatakse sündmust, mis seisneb sündmuse A mittetoimumises.

Sündmuse A vastandsündmust tähistatakse $\bar{A}$.

Kui sündmus A - täringu viskamisel tuleb 6 silma, siis sündmus $\bar{A}$ - täringu viskamisel ei tule 6 silma, st tuleb kas 1, 2, 3, 4 või 5 silma.

Kui sündmus B - õpperühmast juhuslikult valitud kaks üliõpilast on mõlemad naissoost, siis $\bar{B}$ - vähemalt üks valitute on meessoost, st valitute hulgas on üks meessoost või on mõlemad meessoost.

Kui sündmus C - korvpalluri sooritatud vabavisetest vähemalt üks tabab, siis $\bar{C}$ - ükski vabavise ei taba.

Olgu A juhuslik sündmus. Olgu kõikide elementaarsündmuste arv n ja sündmuse A toimumiseks soodsate elementaarsündmuste arv m . Sellisel juhul sündmuse $\bar{A}$ toimumiseks (st A mittetoimumiseks) soodsate elementaarsündmuste arv on $n - m$. Vastandsündmuse tõenäosus

$$P(\bar{A}) = \frac{n - m}{n} = \frac{n}{n} - \frac{m}{n} = 1 - P(A)$$

Siit saame, et

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

st **sündmuse ja selle vastandsündmuse tõenäosuste summa võrdub 1-ga.**

Kui saadud võrdus kirjutada

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

siis näeme, et sündmuse tõenäosuse arvutamiseks saab kasutada vastandsündmuse tõenäosust.

Näide 4. Vaagnal on 15 lillat ja 10 kollast ploomi. Vaagnalt võetakse juhuslikult neli ploomi. Leiame tõenäosuse, et nende hulgas on vähemalt üks lilla ploom.

Tähistame sündmuse A - võetud nelja ploomi hulgas on vähemalt üks lilla. Selle sündmuse tõenäosuse otse arvutamist oleks selleks vaja teha palju arvutusi. Saab lihtsamalt, kui võtame appi vastandsündmuse $\bar{A}$ - võetud ploomide hulgas ei ole ühtegi lillat, st kõik neli on kollased.

Kõiki elementaarsündmusi ehk kõiki võimalusi 25 ploomi hulgast nelja valimiseks on $n = C_{25}^4 = 12650$. Sündmuse $\bar{A}$ toimumiseks soodsaid elementaarsündmusi on nii palju, kui on võimalusi valida 10 kollase ploomi hulgast nelja, st $m = C_{10}^4 = 210$. Saame $\bar{A}$ tõenäosuse

$$P(\bar{A}) = \frac{210}{12650} = 0,017$$

ja selle abil sündmuse A tõenäosuse

$$P(A) = 1 - P(\bar{A}) = 1 - 0,017 = 0,983$$

1.1.7 Sündmuste algebra

Vaatleme kahte sündmust A ja B .

Definitsioon 1. Sündmuste A ja B *summaks* nimetatakse sündmust, mis seisneb sündmuse A või sündmuse B toimumises (sh nende koostoimumises).

Sündmuste A ja B summat tähistatakse $A + B$. Kõnekeeles - toimub A või toimub B .

Kui A on sündmus - märklaua tulistamisel tabatakse kümnesse ja B on sündmus - märklaua tulistamisel tabatakse üheksasse, siis $A + B$ - märklaua tulistamisel tabatakse kümnesse või üheksasse.

Kui A on sündmus - täringu viskamisel tuleb paarisarv silmi ja B - täringu viskamisel tuleb kolmega jaguv arv silmi, siis $A + B$ on sündmus - täringu viskamisel tuleb 2, 3, 4 või 6 silma.

Kui A on sündmus - kaardipakist juhuslikult tõmmatud kaart on ärtu ja B - kaardipakist juhuslikult tõmmatud kaart on pilt, siis $A + B$ on sündmus - kaardipakist juhuslikult tõmmatud kaart on ärtu või pilt (sh ärtu pilt).

Definitsioon 2. Sündmuste A ja B korrutiseks nimetatakse sündmust, mis seisneb sündmuste A ja B samaaegses toimumises.

Sündmuste A ja B korrutist tähistatakse AB . Kõnekeeles - toimub A ja toimub B .

Kui A on sündmus - täringu viskamisel tuleb paarisarv silmi ja B - täringu viskamisel tuleb kolmega jaguv arv silmi, siis AB on sündmus - täringu viskamisel tuleb 6 silma.

Kui A on sündmus - kaardipakist juhuslikult tõmmatud kaart on ärtu ja B - kaardipakist juhuslikult tõmmatud kaart on pilt, siis AB on sündmus - kaardipakist juhuslikult tõmmatud kaart on ärtu pilt.

Kehtivad väited

1. $A + \Omega = \Omega$

2. $A\Omega = A$
3. $A + \emptyset = A$
4. $A\emptyset = \emptyset$
5. $A + \overline{A} = \Omega$
6. $A\overline{A} = \emptyset$

1.2 Tõenäosuste arvutamine

1.2.1 Sündmuste korrutise tõenäosus

Definitsioon 1. Kahte sündmust nimetatakse sõltumatuteks, kui teise sündmuse tõenäosus ei sõltu sellest, kas esimene sündmus toimus või ei toimunud.

Definitsioon 2. Kahte sündmust nimetatakse sõltuvateks, kui teise sündmuse tõenäosus sõltub sellest, kas esimene sündmus toimus või ei toimunud.

Näide 1. Esimeses kastis on 4 kollast ja 6 valget tennisepalli, teises kastis 10 kollast ja 2 valget tennisepalli. Kummasti kastist võetakse juhuslikult üks pall. Defineerime sündmused

A - esimesest kastist võetakse kollane pall;

B - teisest kastist võetakse kollane pall.

Need kaks sündmust on sõltumatud, sest esimesest kastist kollase palli võtmise tõenäosus $P(A) = \frac{4}{10} = \frac{2}{5}$ ei sõltu sellest, kas teisest kastist võeti kollane pall või mitte. Samuti ei sõltu teisest kastist kollase palli võtmise tõenäosus $P(B) = \frac{10}{12} = \frac{5}{6}$ sellest, kas esimesest kastist võeti kollane pall või ei võetud.

Näide 2. Kastis on 4 kollast ja 6 valget tennisepalli. Kastist võetakse järgemööda kaks palli. Defineerime sündmused

A - esimene kastist võetud pall on kollane;

B - teine kastist võetud pall on kollane.

Need kaks sündmust on sõltuvad, sest kui sündmus A toimus, jäi kasti järele 9 palli, millest 3 on kollased ning tõenäosus selleks, et ka teine pall on kollane $P(B) = \frac{3}{9} = \frac{1}{3}$. Kui aga sündmus A ei toimunud, st esimene võetud pall oli valge, jäi kasti alles 9 palli, millest 4 on kollased ning tõenäosus selleks, et teine pall on kollane on $P(B) = \frac{4}{9}$. Teise sündmuse B tõenäosus sõltub sellest kas sündmus A toimus või ei toimunud.

Teoreem 1 (Sõltumatute sündmuste korrutise tõenäosus). Kui A ja B on sõltumatud sündmused, siis nende korrutise tõenäosus võrdub tõenäosuste korrutisega:

$$P(AB) = P(A)P(B)$$

Näide 3. Esimeses kastis on 4 kollast ja 6 valget tennisepalli, teises kastis 10 kollast ja 2 valget tennisepalli. Kummasti kastist võetakse juhuslikult üks pall. Leiame tõenäosuse, et mõlemad võetud pallid on kollased.

Defineerime sündmused

A - esimesest kastist võetakse kollane pall;

B - teisest kastist võetakse kollane pall.

Sel juhul sündmus - mõlemad võetud pallid on kollased - on AB (so toimub A ja toimub B). Teoreemi järgi

$$P(AB) = P(A)P(B) = \frac{2}{5} \cdot \frac{5}{6} = \frac{1}{3}$$

st tõenäosus selleks, et mõlemast kastist võetud pallid on kollased, on $\frac{1}{3}$.

Näide 4. Münti visatakse järgemööda kaks korda. Leiame tõenäosuse, et vähemalt ühel korral tuleb vapp.

Sündmus - vähemalt ühel korral tuleb vapp - on kolme sündmuse summa (esimesel korral tuleb vapp ja teisel ei tule või esimesel korral ei tule vapp ja teisel tuleb või mõlemal korral tuleb vapp). Seepärast on lihtsam sündmuse tõenäosust arvutada, kasutades vastandsündmuse tõenäosust.

Olgu sündmus A - vähemalt ühel viskel tuleb vapp, siis vastandsündmus $\bar{A}$ - mõlemal viskel vappi ei tule (mõlemal viskel tuleb kiri).

Tähistame sündmused B - esimesel viskel tuleb kiri - ja C - teisel viskel tuleb kiri. Nende tõenäosused $P(B) = \frac{1}{2}$ ja $P(C) = \frac{1}{2}$. Sündmus $\bar{A} = BC$ ja teoreemi järgi

$$P(\bar{A}) = P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

ning

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{1}{4} = \frac{3}{4}$$

Tõenäosus, et vähemalt ühel viskel tuleb vapp, on $\frac{3}{4}$

Teoreem 2 (Sõltuvate sündmuste korrutise tõenäosus). Kui A ja B on sõltuvad sündmused, siis nende korrutise tõenäosus

$$P(AB) = P(A)P(B/A)$$

kus $P(B/A)$ on *tinglik tõenäosus*, st sündmuse B tõenäosus tingimusel, et sündmus A toimus.

Näide 5. (vt Näide 2) Kastis on 4 kollast ja 6 valget tennisepalli. Kastist võetakse järgemööda kaks palli. Arvutame tõenäosuse, et mõlemad pallid on kollased.

Tähistame sündmused

A - esimesena kastist võetud pall on kollane;

B - teisena kastist võetud pall on kollane.

Siis AB on sündmus - mõlemad pallid on kollased, so sündmus, mille tõenäosus tuleb arvutada.

Sündmuse A tõenäosus $P(A) = \frac{2}{5}$ ja tinglik tõenäosus $P(B/A) = \frac{1}{3}$. Seega Teoreemi 2 põhjal

$$P(AB) = \frac{2}{5} \cdot \frac{1}{3} = \frac{2}{15}$$

Teoreemi 2 väite saab kirjutada ka

$$\boxed{P(AB) = P(A)P(B/A) = P(B)P(A/B)} \quad (1.1)$$

1.2.2 Teineteist välistavate sündmuste summa tõenäosus

Definitsioon. Kahte sündmust A ja B nimetatakse teineteist välistavateks, kui ühe sündmuse toimumine välistab teise toimumise ja vastupidi.

Võib öelda ka teisiti: kui ükski sündmuse A toimumiseks soodne elementaarsündmus ei ole soodne B toimumiseks ja vastupidi.

Järeldus.

Näide 1. Kaardipakist (36 kaarti) tõmmatakse juhuslikult üks kaart. Olgu sündmus A - see kaart on ruutu - ja B - see kaart on ärtu. Sündmuse A toimumiseks on soodasaid elementaarsündmusi 9 (tõmmatud kaart on ruutu 6, ..., ruutu äss). Sündmuse B toimumiseks on ka 9 elementaarsündmust (tõmmatud kaart on ärtu 6, ..., ärtu äss). Ükski A toimumiseks soodne elementaarsündmus ei ole soodne B toimumiseks ja vastupidi. Sündmused A ja B on teineteist välistavad.

Olgu A ja B teineteist välistavad sündmused. Olgu katse tagajärjel kõiki toimuvaid elementaarsündmusi n , sündmuse A toimumiseks soodsaid elementaarsündmusi m_1 ja sündmuse B toimumiseks soodsaid m_2 . Sel juhul sündmuse $A + B$ (toimub A või toimub B) toimumiseks soodsaid elementaarsündmusi on $m_1 + m_2$ ning $A + B$ toimumise tõenäosus on

$$P(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

Järelikult kehtib.

Teoreem. Teineteist välistavate sündmuste summa tõenäosus on võrdne nende sündmuste tõenäosuste summaga

$$\boxed{P(A + B) = P(A) + P(B)} \quad (1.2)$$

Näide 2. Kaardipakist (36 kaarti) tõmmatakse juhuslikult üks kaart. Olgu sündmus A - see kaart on ruutu - ja B - see kaart on ärtu. Leiame tõenäosuse, et kaardipakist juhuslikult tõmmatud kaart on punane.

Sündmus - kaardipakist juhuslikult tõmmatud kaart on punane - on $A + B$, st kaart on ruutu või kaart on ärtu. Kõiki elementaarsündmusi on 36,

mõlema sündmuse A ja B toimumiseks soodsaid elementarsündmusi on 9. Sündmuste A ja B tõenäosused on $P(A) = \frac{9}{36} = \frac{1}{4}$ ning $P(B) = \frac{9}{36} = \frac{1}{4}$. Seega sündmuse $A + B$ tõenäosus on

$$P(A + B) = P(A) + P(B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

1.2.3 Teineteist mittevälitavate sündmuste summa tõenäosus

Definitsioon. Sündmusi nimetatakse teineteist mittevälitavateks, kui ühe sündmuse toimumine ei välista teise toimumist, st need võivad toimuda ka samaaegselt.

Näide 1. Kaardipakist, milles on 36 kaarti, tõmmatakse juhuslikult üks kaart. Defineerime sündmused A - see kaart on ärtu, B - see kaart on pilt. Sündmused A ja B ei ole teineteist välistavad, sest kaart saab samaaegselt olla nii ärtu kui ka pilt, st ärtu pilt.

Oletame, et katse tagajärjel toimuvate kõikide elementaarsündmuste arv on n , sündmuse A toimumiseks soodsate elementaarsündmuste arv m_1 , sündmuse B toimumiseks soodsate elementaarsündmuste arv m_2 ning sündmuste A ja B samaaegseks toimumiseks (sündmuse AB toimumiseks) soodsate elementaarsündmuste arv m_3 . Sellisel juhul on sündmuse $A + B$ toimumiseks soodsate elementaarsündmuste arv $m_1 + m_2 - m_3$.

Toodud näites 1 on kõiki elementaarsündmusi 36, sündmuse A toimumiseks soodsaid elementaarsündmusi 9, sündmuse B toimumiseks soodsaid elementaarsündmusi 16 ja sündmuste samaaegseks toimumiseks (sündmuse AB - kaart on ärtu pilt) soodsaid elementaarsündmusi 4. Sündmuse $A + B$ toimumiseks soodsaid elementaarsündmusi on $9 + 16 - 4 = 21$.

Pöördume üldjuhu juurde tagasi. Sündmuse $A + B$ tõenäosus

$$P(A + B) = \frac{m_1 + m_2 - m_3}{n} = \frac{m_1}{n} + \frac{m_2}{n} - \frac{m_3}{n}$$

Arvestades sellega, et $\frac{m_1}{n} = P(A)$, $\frac{m_2}{n} = P(B)$ ja $\frac{m_3}{n} = P(AB)$, saame teoreemi.

Teoreem. Teineteist mittevälitavate sündmuste summa tõenäosus on nende sündmuste tõenäosuste summa miinus nende korrutise tõenäosus, st

$$P(A + B) = P(A) + P(B) - P(AB)$$

Kui sündmused A ja B on sõltumatud, siis

$$\boxed{P(A + B) = P(A) + P(B) - P(A)P(B)} \quad (1.3)$$

Et toodud näites 1 sündmused A ja B on sõltumatud, siis $A + B$ - kaart on ärtu või pilt - tõenäosus on

$$P(A + B) = \frac{9}{36} + \frac{16}{36} - \frac{9}{36} \cdot \frac{16}{36} = \frac{21}{36} = \frac{7}{12}$$

Sündmuste A ja B summa $A + B$ on peale selle, et toimub sündmus A või sündmus B (mis ei välista nende samaaegset toimumist), sõnastatav ka: toimub vähemalt üks sündmustest A või B . Toimub vähemalt üks neist tähendab ju samuti, et toimub A või toimub B või toimub AB .

Näide 2. Üliõpilased Piibeleht ja Vestmann lahendavad teineteisest sõltumatult tõenäosusteooria ülesannet. Tõenäosus, et Piibeleht lahendab ülesande õigesti, on 0,7 ja tõenäosus, et Vestmann lahendab ülesande õigesti, on 0,8. Leiame tõenäosuse, et vähemalt üks neist lahendab ülesande õigesti.

Ülesande lahendame kahel viisil. Esiteks, olgu

sündmus A - Piibeleht lahendab ülesande õigesti

ja sündmus B - Vestmann lahendab ülesande õigesti.

Siis on sündmus $A + B$ - vähemalt üks neist lahendab ülesande õigesti. Selle tõenäosus on valemi (1.3) järgi

$$P(A + B) = P(A) + P(B) - P(A)P(B) = 0,7 + 0,8 - 0,7 \cdot 0,8 = 0,94$$

Teine lahendus. Jääme esimese lahenduse tähistuste juurde. $A + B$ vastandsündmus, kumbki ei lahenda ülesannet õigesti, st Piibeleht ei lahenda ülesannet õigesti ja Vestmann ei lahenda ülesannet õigesti, on $\overline{A} \cdot \overline{B}$ ning selle tõenäosus

$$P(\overline{A} \cdot \overline{B}) = P(\overline{A})P(\overline{B}) = 0,3 \cdot 0,2 = 0,06$$

Seega antud sündmuse tõenäosus

$$P(A + B) = 1 - P(\overline{A} \cdot \overline{B}) = 1 - 0,06 = 0,94$$

1.2.4 Täistõenäosuse valem

Oletame, et meil on n sündmust $A_1, A_2, \dots, A_n$, mis on paarikaupa teineteist välistavad ja millest üks toimub kindlasti, st

$$A_i A_j = \emptyset, \text{ kui } i \neq j$$

ja

$$\sum_{i=1}^n A_i = \Omega$$

Sellist sündmuste süsteemi nimetatakse *täielikuks sündmuste süsteemiks*.

Oletame, et sündmus B toimub ühe sündmustest $A_1, A_2, \dots, A_n$ toimumise tagajärjel. Siis

$$P(B) = P(B\Omega) = P\left(B \sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(A_i)P(B/A_i)$$

sest kui sündmused $A_1, A_2, \dots, A_n$ on paarikaupa teineteist välistavad, siis on seda ka sündmused $BA_1, BA_2, \dots, BA_n$.

Seega oleme saanud valemi

$$P(B) = \sum_{i=1}^n P(A_i)P(B/A_i) \quad (1.4)$$

mida nimetatakse *täistõenäosuse valemiks*.

Näide. Firma autopargis on 5 BMW-d, 10 Audi ja 20 Opelit. Tõenäosus, et BMW läbib tehnõlevaatuse, on 0,98, tõenäosus, et Audi läbib tehnõlevaatuse, on 0,95 ja tõenäosus, et Opel läbib tehnõlevaatuse, on 0,92. Arvutame tõenäosuse, et firma autopargist juhusliult valitud auto läbib tehnõlevaatuse.

Olgu sündmused

A_1 juhuslikult valitud auto on BMW;

A_2 juhuslikult valitud auto on Audi;

A_3 juhuslikult valitud auto on Opel.

Need moodustavad täieliku sündmuste süsteemi, kusjuures $P(A_1) = \frac{5}{35} = \frac{1}{7}$, $P(A_2) = \frac{2}{7}$ ja $P(A_3) = \frac{4}{7}$.

Olgu sündmus B - auto läbib tehnõlevaatuse.

Edasi tinglikud tõenäosused. Tõenäosus, et auto läbib tehnõlevaatuse, eeldusel et see auto on BMW $P(B/A_1) = 0,98$. Samuti $P(B/A_2) = 0,95$ ja $P(B/A_3) = 0,92$. Seega täistõenäosuse valemi (1.4) järgi

$$\begin{aligned} P(B) &= P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3) = \\ &= \frac{1}{7} \cdot 0,98 + \frac{2}{7} \cdot 0,95 + \frac{4}{7} \cdot 0,92 = 0,937 \end{aligned}$$

1.2.5 Bayesi valem

Eeldame jälle, et sündmused $A_1, A_2, \dots, A_n$ moodustavad täieliku sündmuste süsteemi ja et sündmus B toimub ühe nendest toimumise tagajärjel. Tingliku tõenäosuse valemi (1.1) järgi iga $i = 1, 2, \dots, n$ korral

$$P(B)P(A_i/B) = P(A_i)P(B/A_i)$$

millest tinglik tõenäosus

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$$

Asendades $P(B)$ täistõenäosuse valemist (1.4), saame nn *Bayesi valemi*

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^n P(A_i)P(B/A_i)} \quad (1.5)$$

Bayesi valemi kasutamise puhul nimetataske sündmusi $A_1, A_2, \dots, A_n$ hüpoteesideks. Seega Bayesi valemi abil arvutatakse hüpoteesi A_i tõenäosust, eeldusel, et sündmus B on toimunud.

Näide. Laual on kaks ühesugust hoiukarpi. Ühes on 10 2-eurost münti, teises on 5 2-eurost ja 5 üheeurost münti. Juhuslikust karbist võeti juhuslik münt, see osutus 2-euroseks. Leiame tõenäosuse, et münt võeti karbist, kus kõik kümme münti on 2-eurosed.

Tähistame hüpoteesid:

A_1 - valiti karp, kus kõik mündid on 2-eurosed;

A_2 - valiti karp, kus on 5 kahe- ja 5 üheeurost münti.

Sündmus B - münt osutus 2-euroseks.

Antud tähistustes peame arvutama $P(A_1/B)$, st 2-euroseks osutunud münt valiti karbist, kus kõik on 2-eurosed. Et kardid on ühesugused, siis $P(A_1) = \frac{1}{2}$ ja $P(A_2) = \frac{1}{2}$. Kaheeurose münti võtmine karbist, kus kõik on 2-eurosed, on kindel sündmus, st $P(B/A_1) = 1$ ja kaheeurose münti võtmise tõenäosus karbist, kus pooled on 2-eurosed, on $P(B/A_2) = \frac{1}{2}$. Bayesi valemi järgi

$$\begin{aligned} P(A_1/B) &= \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2)} = \\ &= \frac{\frac{1}{2} \cdot 1}{\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3} \end{aligned}$$

1.2.6 Bernoulli valem

Eelmistes punktides vaatlesime sündmust, mis toimus või ei toimunud ühe katse tagajärjel. Käesolevas punktis eeldame, et tehakse n järjestikust sõltumatut katset ja kõikidel katsetel on sündmuse toimumise tõenäosus võrdne.

Oletame, et tehakse 5 katset, igal katsel on sündmuse A toimumise tõenäosus p ja mittetoimumise tõenäosus $q = 1 - p$, st $P(A) = p$ ja $P(\bar{A}) = q$. Sündmus A toimub 5-st katsest kahel korral, kui esimesel kahel korral sündmus toimub ja ülejäänud kolmel korral ei toimu, st toimub sündmus

$$AA\bar{A}\bar{A}\bar{A}$$

Samuti toimub sündmus A viiest katsest kahel korral, kui toimuvad sündmused

$$\begin{aligned} &A\bar{A}A\bar{A}\bar{A}, \quad A\bar{A}\bar{A}A\bar{A}, \quad A\bar{A}\bar{A}\bar{A}A, \\ &\bar{A}A\bar{A}\bar{A}\bar{A}, \quad \bar{A}A\bar{A}\bar{A}A, \quad \bar{A}\bar{A}\bar{A}\bar{A}A, \\ &\bar{A}\bar{A}A\bar{A}\bar{A}, \quad \bar{A}\bar{A}\bar{A}\bar{A}A \text{ ja } \bar{A}\bar{A}\bar{A}A\bar{A} \end{aligned}$$

Katsed on sõltumatud, seega tõenäosuste korrutamise teoreemi põhjal

$$P(AA\bar{A}\bar{A}\bar{A}) = P(A)P(A)P(\bar{A})P(\bar{A})P(\bar{A}) = ppqqq = p^2q^3$$

Ilmselt on ka ülejäänud 9 sündmuse tõenäosused p^2q^3 . Et kõik need 10 sündmust on paarikaupa teineteist välistavad, siis tõenäosus, et sündmus toimub viiest katsest kahel korral on

$$P(AA\bar{A}\bar{A}\bar{A} + \dots + \bar{A}\bar{A}\bar{A}AA) = P(AA\bar{A}\bar{A}\bar{A}) + \dots + P(\bar{A}\bar{A}\bar{A}AA) = 10p^2q^3$$

Kordaja 10 saaksime ka järgmiselt arutledes. Viiest katsest valida neid kahte, mille korral sündmus toimub, on parajasti nii palju kui $C_5^2 = 10$. Tähistades tõenäosuse, et sündmus toimub 5-st katsest kahel korral $P_{5,2}$, saame

$$P_{5,2} = C_5^2 p^2 q^3$$

Analoogilisel viisil arutledes jõuame tulemuseni, et tõenäosus selleks, et sündmus toimuks n katsest k -l korral ($0 \leq k \leq n$), on arvutatav Bernoulli valemi

$$P_{n,k} = C_n^k p^k q^{n-k} \quad (1.6)$$

abil.

Näide. Korvpallur tabab vabaviske tõenäosusega 0,8. Arvutame tõenäosuse, et korvpallur tabab 6-st vabaviskest viiel korral.

Antud ülesandes $n = 6$, $k = 5$, $p = 0,8$ ja $q = 0,2$. Seega tõenäosus

$$P_{6,5} = C_6^5 \cdot 0,8^5 \cdot 0,2^{6-5} = 6 \cdot 0,8^5 \cdot 0,2 = 0,393$$

Lisaks arvutame tõenäosuse, et korvpallur tabab kuuest vabaviskest neljal korral. Sellisel juhul muud andmed on endised, ainult et $k = 4$. Arvutame

$$P_{6,4} = C_6^4 \cdot 0,8^4 \cdot 0,2^{6-4} = 15 \cdot 0,8^4 \cdot 0,2^2 = 0,246$$

1.2.7 Tõenäoseim sündmuse toimumiste arv

Eelmise punkti näitest selgus muuhulgas, et tõenäosus selleks, et sündmus toimuks taetud katsete arvu korral erinev arv kordi, on üldjuhul erinev. Käesolevas punktis püstitame küsimuse: "Milline sündmuse toimumiste arv on suurima tõenäosusega?"

Oletame, et suurima tõenäosusega sündmuse toimumste arv on k , so nõuame, et

$$P_{n,k+1} \leq P_{n,k}$$

ja et

$$P_{n,k-1} \leq P_{n,k}$$

Esimene tingimustest tähendab (1.6) järgi, et

$$\frac{n!}{(k+1)!(n-k-1)!} p^{k+1} q^{n-k-1} \leq \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Jagades selle võrratuse mõlemaid pooli korrutisega $n!p^kq^{n-k-1}$, saame

$$\frac{p}{(k+1)!(n-k-1)!} \leq \frac{q}{k!(n-k)!}$$

Korrutades selle võrratuse mõlemaid pooli korrutisega $(k+1)!(n-k)!$, saame

$$p(n-k) \leq q(k+1)$$

ehk

$$np - q \leq k(p+q)$$

Et aga $p+q=1$, siis saame tingimuse

$$np - q \leq k$$

Täpselt samal viisil arutledes, saaksime võrratusest $P_{n,k-1} \leq P_{n,k}$ tingimuse

$$k \leq np + p$$

Seega tõenäoiseima sündmuse toimumiste arvu k saame tingimustest

$$np - q \leq k \leq np + p \quad (1.7)$$

Näide. Leiame eelmise punkti näites vaadeldud olukorras korpalluri suurima tõenäosusega tabavate vabavisete arvu.

Selles näites $n=6$, $p=0,8$ ja $q=0,2$. Seega suurima tõenäosusega tabavate vabavisete arvu saame (1.7) järgi

$$6 \cdot 0,8 - 0,2 \leq k \leq 6 \cdot 0,8 + 0,8$$

ehk

$$4,6 \leq k \leq 5,6$$

st $k=5$. Seega antud tingimustel viie vabaviske tabamine ongi suurima tõenäosusega.

1.3 Statistiline tõenäosus

Uurime mingi sündmuse A toimumist katse tagajärjel. Kordame katsed n korda ja oletame, et sündmus toimub k katsel. Sündmuse toimumise *suhteliseks sageduseks* nimetatakse suhet

$$P^*(A) = \frac{k}{n}$$

Täringut visati 10 korda. Kuus silma ei tulnud kordagi. Kuue silma esinemise suhteline sagedus on $\frac{0}{n} = 0$. Täringut visati 100 korda. kuus silma tuli 10 korda, kuue silma esinemise suhteline sagedus on $\frac{10}{100} = 0,1$. Täringut

visati 10000 korda. kuus silma tuli 1661 korda, kuue silma esinemise suhteline sagedus on $\frac{1661}{10000} = 0,1661$.

Kui sündmus toimub kõikidel katsetel, siis on $k = n$ ja suhteline sagedus on 1. Kui sündmus ei toimu ühegi katse tagajärjel, siis $k = 0$ ja suhteline sagedus on 0. Üldiselt on suhteline sagedus ratsionaalarv, mis kuulub lõiku $[0; 1]$.

Kindla sündmuse suhteline sagedus on 1, sest kindel sündmus toimub igal katsel. Kuid sellest, et sündmuse suhteline sagedus on 1, ei järeldu, et tegemist on kindla sündmusega.

Samuti on võimatu sündmuse suhteline sagedus 0. Sellest aga, et sündmuse suhteline sagedus on 0, ei järeldu, et sündmus on võimatu.

Suhtelisel sagedusel on omadus: küllalt pika katseseeria korral on sündmuse toimumise suhteline sagedus ligikaudu võrdne sündmuse toimumise tõenäosusega. Seepärast küllalt pika katseseeria tulemusel saadud suhtelist sagedust nimetatakse sündmuse toimumise *statistiliseks tõenäosuseks*.

Inglise matemaatik Carl Pearson viskas münti 24000 korda ja vapp tuli 12012 korda. Vapi tuleku suhteline sagedus on

$$\frac{12012}{24000} = 0,5005$$

On teada, et vapi tuleku tõenäosus on 0,5, seega nii pika katseseeria puhul erineb suhteline sagedus sündmuse tõenäosusest väga väikese suure võrra ja seda suhtelist sagedust võib võtta statistilise tõenäosusena.

Sageli esitatakse statistilist tõenäosust protsentides, seega Pearsoni poolt saadud vapi tuleku statistiline tõenäosus on 50,05%.

1.4 Juhuslik suurus

1.4.1 Juhusliku suuruse jaotus

Katse tagajärjel võivad toimuda mitmed erinevad sündmused, st katse tagajärg on *juhuslik suurus*. Kui täringut veeretada ühe korra, siis silmade arv on juhuslik suurus, mis võib omandada väärtusi 1-st 6-ni. Juhuslik suurus omandab konkreetse väärtuse on juhuslik sündmus. Täringu veeretamisel juhuslik suurus - silmade arv - omandab väärtuse 2 on juhuslik sündmus - täringu veeretamisel saadakse kaks silma.

Juhuslikke suurusi tähistatakse suurte tähestiku lõpu tähtedega X , Y , Z , U , V jne.

Defineerime juhusliku suuruse X - täringu viskamisel saadav silmade arv. Siis $X = 2$ tähendab, et täringu viskamisel saadav silmade arv on 2, so sündmus $X = 2$ on juhuslik sündmus - täringu veeretamisel tuleb 2 silma. Selle tõenäosus

$$P(X = 2) = \frac{1}{6}$$

Juhuslikku suurust X , mis võib omandada lõpliku hulga eraldiseisvaid väärtusi $x_1, x_2, \dots, x_n$, nimetatakse *diskreetseks juhuslikuks suuruseks*. Täringu veeretamisel saadav silmade arv on tüüpiline diskreetne juhuslik suurus.

Juhusliku suurust, mis võib omandada niisuguseid väärtusi, et iga kahe mis tahes võimaliku väärtuse vahel võib see juhuslik suurus omandada veel mingi väärtuse, nimetatakse *pidevaks juhuslikuks suuruseks*. Pidevaks juhuslikuks suuruseks on kõikvõimalikud mõõtmistulemused. Näiteks, erinevad üliõpilased mõõdavad sama silindrikujulise detaili läbimõõtu. Mõõtmistulemus on juhuslik suurus. Üks üliõpilane saab tulemuseks 2,50 cm, teine 2,60 cm, kolmas aga võib saada mõõtmistulemuseks 2,54 cm. Pole välistatud, et neljas üliõpilane saab mõõtmistulemuseks 2,56 cm jne.

Definitsioon 1. Juhusliku suuruse *jaotuseks ehk jaotusseaduseks* nimetatakse eeskirja, mis igale juhusliku suuruse väärtusele seab vastavusse selle väärtuse omandamise tõenäosuse.

Diskreetse juhusliku suuruse jaotust saab esitada tabelina või graafikuna, mida nimetatakse *jaotuspolügooniks*. Pideva juhusliku suuruse jaotust esitatakse reeglina kas *jaotusfunktsioonina* või *jaotustihedusena*.

Vaatleme ühte diskreetse juhusliku suuruse jaotusseaduse koostamist.

Näide 1. Riiulil on 18 pooleliitrist veepudelit, neist 10 Bonaqua ja 8 Vichy pudelit. Riiulilt võetakse huupi 4 pudelit. Koostame nende hulka sattuvate Bonaqua pudelite arvu kui juhusliku suuruse jaotusseaduse.

Olgu juhuslikuks suuruseks X nelja pudeli hulgas olevate Bonaqua pudelite arv. See diskreetne juhuslik suurus võib omandada väärtused 0, 1, 2, 3 ja 4. Arvutame nende väärtuste omandamise tõenäosused. Tõenäosus selleks, et juhuslik suurus omandab väärtuse 0, st võetud 4 pudeli hulgas pole ühtegi Bonaqua pudelit:

$$P(X = 0) = \frac{C_8^4}{C_{18}^4} = \frac{70}{3060} = \frac{7}{306}$$

Tõenäosus selleks, et juhuslik suurus omandab väärtuse 1, so võetud nelja pudeli hulgas on üks Bonaqua pudel:

$$P(X = 1) = \frac{C_{10}^1 C_8^3}{C_{18}^4} = \frac{10 \cdot 56}{3060} = \frac{28}{153}$$

Tõenäosus selleks, et juhuslik suurus omandab väärtuse 2, so võetud nelja pudeli hulgas on kaks Bonaqua pudelit:

$$P(X = 2) = \frac{C_{10}^2 C_8^2}{C_{18}^4} = \frac{45 \cdot 28}{3060} = \frac{7}{17}$$

Tõenäosus selleks, et juhuslik suurus omandab väärtuse 3, so võetud nelja pudeli hulgas on kolm Bonaqua pudelit:

$$P(X = 3) = \frac{C_{10}^3 C_8^1}{C_{18}^4} = \frac{120 \cdot 8}{3060} = \frac{16}{51}$$

Tõenäosus selleks, et juhuslik suurus omandab väärtuse 4, so kõik võetud neli on Bonaqua pudelid:

$$P(X = 4) = \frac{C_{10}^4}{C_{18}^4} = \frac{210}{3060} = \frac{7}{102}$$

Tulemused koondame tabelisse, mille ülemisse ritta kirjutame juhusliku suuruse võimalikud väärtused ja alumisse ritta nende väärtuste omandamise tõenäosused.

X	0	1	2	3	4
P	$\frac{7}{306}$	$\frac{28}{153}$	$\frac{7}{17}$	$\frac{16}{51}$	$\frac{7}{102}$

Saadud tabel kujutab endast juhusliku suuruse jaotusseadust, st tabel korraldab vastavuse juhusliku suuruse võimalike väärtuste ja nende omandamise tõenäosuse vahel. Kui liita alumises reas olevad tõenäosused kokku, on tulemuseks 1. See on loomulik, sest kõik sündmused on paarikaupa teineteist välistavad ja üks neist toimub kindlasti.

Teiseks selle juhusliku suuruse jaotusseaduse esitusviisiks on graafiline esitus ehk jaotuspolügoon. Abstsissiteljele kanname juhusliku suuruse X võimalikud väärtused ja ordinaatteljele vastavad tõenäosused.

Üldjuhul eeldades, et diskreetsel juhuslikul suurusel on n võimalikku väärtust $x_1, x_2, \dots, x_n$ ja p_k on tõenäosus, et X omandab väärtuse x_k , kus $k = 1, 2, \dots, n$, so

$$p_k = P(X = x_k)$$

siis juhusliku suuruse jaotusseadus on

X	x_1	x_2	$\dots$	x_n
P	p_1	p_2	$\dots$	p_n

ja selles tabelis kõikide tõenäosuste summa võrdub ühega, so

$$p_1 + p_2 + \dots + p_n = 1$$

Viimane on kõikide diskreetsete juhuslike suuruste jaotuste ühine omadus ja selle abil on võimalik kontrollida jaotusseaduse koostamise õigsust.

Diskreetse juhusliku suuruse võimalike väärtuste arv ei pruugi olla lõplik, vaid saab olla ka loenduv. Vaatleme järgmist näidet.

Näide 2. Laskur tulistab märklauda, kusjuures igal lasul on märgi tabamise tõenäosus 0,7. Laskur tulistab kuni esimese möödalasuni. Koostame tabavate laskude arvu kui juhusliku suuruse X jaotuse.

Tõenäosus, et tabamuste arv on 0, st laskur laseb esimese lasu märgist mööda

$$P(X = 0) = 0,3$$

Tõenäosus, et tabamuste arv on 1, st laskur esimese lasuga tabab märklauda ja teise laseb mööda

$$P(X = 1) = 0,7 \cdot 0,3$$

Tõenäosus, et tabamuste arv on 2, st laskur esimese ja teise lasu tabab ning kolmanda laseb mööda

$$P(X = 2) = 0,7 \cdot 0,7 \cdot 0,3 = 0,7^2 \cdot 0,3$$

Sellisel jätkates saame, et tõenäosus selleks, et tabamuste arv on k , st laskur esimese k lasuga tabab märklauda ja $k + 1$ korral laseb mööda, on

$$P(X = k) = 0,7^k \cdot 0,3$$

Selle juhusliku suuruse jaotustabel on

X	0	1	2	...	k	...
P	0,3	$0,7 \cdot 0,3$	$0,7^2 \cdot 0,3$	...	$0,7^k \cdot 0,3$	...

Alumise rea summa

$$\sum_{k=0}^{\infty} 0,7^k \cdot 0,3$$

on geomeetiline rida, mis on koonduv, sest $0,7 < 1$ ja selle summa

$$\sum_{k=0}^{\infty} 0,7^k \cdot 0,3 = \frac{0,3}{1 - 0,7} = 1$$

1.4.2 Juhusliku suuruse jaotusfunktsioon

Järgmiseks juhusliku suuruse jaotuse esitamise viisiks on selle juhusliku suuruse *jaotusfunktsioon*.

Definitsioon. Juhusliku suuruse X jaotusfunktsiooniks $F(x)$ nimetatakse tõenäosust, et juhuslik suurus X omandab väärtuse, mis on väiksem kui x , st

$$F(x) = P(X < x) \quad (1.8)$$

Näide 1. Koostame eelmise alampunkti Näites 1 vaadeldud juhusliku suuruse jaotusfunktsiooni. Esiteks, kui $x \leq 0$, siis

$$F(x) = P(X < x) = 0$$

Teiseks, kui $0 < x \leq 1$, siis

$$F(x) = \frac{7}{306}$$

Kolmandaks, kui $1 < x \leq 2$, siis

$$F(x) = \frac{7}{306} + \frac{28}{153} = \frac{7}{34}$$

Neljandaks, kui $2 < x \leq 3$, siis

$$F(x) = \frac{7}{34} + \frac{7}{17} = \frac{21}{34}$$

Viiendaks, kui $3 < x \leq 4$, siis

$$F(x) = \frac{21}{34} + \frac{16}{51} = \frac{95}{102}$$

Lõpuks kui $x > 4$, siis

$$F(x) = \frac{95}{102} + \frac{7}{102} = 1$$

Viimase väärtuse 1 oleksime saanud ka arutlusega, et mis tahes 4-st suurema x väärtuse korral on $X < x$ kindel sündmus. Seega vaadeldava juhusliku suuruse jaotusfunktsioon on

$$F(x) = \begin{cases} 0, & \text{kui } x \leq 0 \\ \frac{7}{306}, & \text{kui } 0 < x \leq 1 \\ \frac{7}{34}, & \text{kui } 1 < x \leq 2 \\ \frac{21}{34}, & \text{kui } 2 < x \leq 3 \\ \frac{95}{102}, & \text{kui } 3 < x \leq 4 \\ 1, & \text{kui } x > 4 \end{cases}$$

Uurime jaotusfunktsiooni omadusi.

Jaotusfunktsiooni definitsiooni tõttu kuuluvad kõik selle väärtused lõiku $[0; 1]$, st

Omadus 1. $0 \leq F(x) \leq 1$

Valides kaks reaalarvu x_1 ja x_2 selliselt, et $x_1 < x_2$, saame sündmuse $(X < x_2)$ kirjutada kahe teineteist välistava sündmuse summana

$$(X < x_2) = (X < x_1) + (x_1 \leq X < x_2)$$

ja valemi (1.2) põhjal

$$F(x_2) = P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2) = F(x_1) + P(x_1 \leq X < x_2) \quad (1.9)$$

Et $P(x_1 \leq X < x_2) \geq 0$, siis

$$F(x_1) \leq F(x_2)$$

st

Omadus 2. Jaotusfunktsioon on monotoonselt kasvav funktsioon.

Omaduste 2 ja 1 järgi on jaotusfunktsioon monotoonselt kasvav tõkestatud funktsioon. Monotoonselt kasvaval ja tõkestatud funktsioonil on piirprotsessides $x \rightarrow \infty$ ja $x \rightarrow -\infty$ olemas lõplik piirväärtus, st kehtib.

Omadus 3.

$$\lim_{x \rightarrow \infty} F(x) = 1$$

ja

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

Võrdusest (1.9) saame

Järeldus. Tõenäosus selleks, et juhusliku suuruse X väärtused kuuluvad poollõiku $[x_1; x_2)$

$$P(x_1 \leq X < x_2) = F(x_2) - F(x_1) \quad (1.10)$$

Nüüd on võimalik korrektselt defineerida pidev juhuslik suurus.

Definitsioon 3. Pidevaks juhuslikuks suuruseks nimetatakse juhuslikku suurust X , mille jaotusfunktsioon $F(x)$ on pidev funktsioon.

Näide 3. Funktsioon

$$F(x) = \begin{cases} 0 & \text{kui } x \leq 1 \\ \frac{x-1}{2} & \text{kui } 1 < x \leq 3 \\ 1, & \text{kui } x > 3 \end{cases}$$

on ilmselt mingi pideva juhusliku suuruse X jaotusfunktsioon, sest kõik jaotusfunktsiooni omadused 1. - 3. on täidetud ja funktsioon on pidev kogu arvteljel.

Kui $F(x)$ on pidev funktsioon, siis (1.10) põhjal

$$P(X = x_1) = \lim_{x \rightarrow x_1} P(x_1 \leq X < x) = \lim_{x \rightarrow x_1} (F(x) - F(x_1)) = 0$$

Niisiis, pideva juhusliku suuruse mis tahes üksikväärtuse omandamise tõenäosus on 0. See aga ei tähenda seda, et sündmus $(X = x_1)$ on võimatu.

Järelikult **pideva** juhusliku suuruse X korral saame arvutuseeskirja (1.10) asemel kirjutada

$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

aga ka

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$$

st pideva juhusliku suuruse tõenäosuse arvutamisel pole oluline jälgida, kas otspunktid kuuluvad hulka või mitte.

1.4.3 Pideva juhusliku suuruse jaotustihedus

Jaotusfunktsioon on üks pideva juhusliku suuruse jaotuse defineerimise võimalus. Kui jaotusfunktsioon $F(x)$ on diferentseeruv, va lõplikus arvus punktides, siis jaotusfunktsiooni tuletist

$$f(x) = F'(x)$$

nimetatakse pideva juhusliku suuruse *jaotustiheduseks*.

Näide 1. Eelmise alampunkti Näites 3 vaadeldud jaotusfunktsioon ei ole diferentseeruv ainult kahes punktis $x = 1$ ja $x = 3$. Selle tuletis, so vaadeldava juhusliku suuruse jaotustihedus on

$$f(x) = \begin{cases} 0 & \text{kui } x \leq 1 \\ \frac{1}{2} & \text{kui } 1 < x \leq 3 \\ 0, & \text{kui } x > 3 \end{cases}$$

Vaatleme jaotustiheduse omadusi. Et $F(x)$ on monotoonselt kasvav funktsioon, siis

Omadus 1. Jaotustihedus on mittenegatiivne, $f(x) \geq 0$.

Omadus 2. Jaotusfunktsioon

$$F(x) = \int_{-\infty}^x f(x)dx \quad (1.11)$$

Tõestus. Päratu integraali definitsioonist ja sellest, et $F(x)$ on jaotustiheduse $f(x)$ algfunktsioon

$$\int_{-\infty}^x f(x)dx = \lim_{M \rightarrow -\infty} \int_M^x f(x)dx = \lim_{M \rightarrow -\infty} F(x) \Big|_M^x = \lim_{M \rightarrow -\infty} (F(x) - F(M)) = F(x)$$

Viimane võrdus järeldub jaotusfunktsiooni 3. omadusest.

Omadus 3. Sündmuse $(x_1 < X < x_2)$ tõenäosus

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx \quad (1.12)$$

Viimane võrdus järeldub vahetult sellest, et vasak pool on $F(x_2) - F(x_1)$ ning Newton-Leibnizi valemist.

Omadus 4.

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

See võrdus järeldub omadusest 2, sest

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{N \rightarrow \infty} \int_{-\infty}^N f(x) dx = \lim_{N \rightarrow \infty} F(N) = 1$$

Viimane võrdus järeldub jälle jaotusfunktsiooni 3. omadusest.

Geomeetriliselt tähendab omadus 4 seda, et jaotustiheduse graafiku ja x -telje vahele jääva kõvertrapetsi pindala on alati võrdne 1-ga.

Omadus 3 geomeetriline sisu: tõenäosus $P(x_1 < X < x_2)$ on võrdne sellise kõvertrapetsi pindalaga, mis on alt piiratud x -teljega, ülalt jaotustiheduse $f(x)$ graafikuga, vasakult sirgega $x = x_1$ ja paremalt sirgega $x = x_2$.

Näide 2. Olgu juhusliku suuruse X jaotustihedus

$$f(x) = \begin{cases} a \sin x, & \text{kui } 0 \leq x \leq \pi \\ 0 & \text{kui } x < 0 \text{ või } x > \pi \end{cases}$$

Määrame konstandi a ning arvutame tõenäosuse, et juhusliku suuruse väärtus paikneks vahemikus $\left(\frac{\pi}{3}; \frac{\pi}{2}\right)$.

Kordaja a määrame jaotustiheduse 4. omadusest

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\pi} a \sin x dx = 1$$

ehk

$$-a \cos x \Big|_0^{\pi} = 1$$

millest $2a = 1$ ehk $a = \frac{1}{2}$

Tõenäosuse $P\left(\frac{\pi}{3} < X < \frac{\pi}{2}\right)$ leiame (1.12) abil

$$P\left(\frac{\pi}{3} < X < \frac{\pi}{2}\right) = \int_{\frac{\pi}{3}}^{\frac{\pi}{2}} \frac{1}{2} \sin x dx = -\frac{1}{2} \cos x \Big|_{\frac{\pi}{3}}^{\frac{\pi}{2}} = -\frac{1}{2} \left(0 - \frac{1}{2}\right) = \frac{1}{4}$$

1.5 Juhusliku suuruse arvkarakteristikud

Juhusliku suuruse jaotus, st diskreetse juhusliku suuruse jaotustabel või jaotusfunktsioon ning pideva juhusliku suuruse jaotusfunktsioon või jaotustihedus iseloomustavad seda juhuslikku suurust täielikult.

Taetud vaatevinklist lähtudes aga on jaotuseduse näol informatsiooni liiga palju. Seepärast kasutatakse jaotuse iseloomustamiseks juhuslikku suurust iseloomustavaid arvulisi näitajaid - juhusliku suuruse arvkarakteristikuid. Neid on kahte tüüpi: paiknemist iseloomustavad arvkarakteristikud (nt juhusliku suuruse keskväärts) ja hajuvust iseloomustavad arvkarakteristikud (nt juhusliku suuruse dispersioon)

1.5.1 Juhusliku suuruse keskväärts

Olgu diskreetse juhusliku suuruse X jaotus antud tabelina

X	x_1	x_2	$\dots$	x_n
P	p_1	p_2	$\dots$	p_n

Selle keskväärts leitakse valemist

$$EX = \sum_{k=1}^n x_k p_k \quad (1.13)$$

st keskväärts kujutab endast juhusliku suuruse väärtuste kaalutud keskmist, kusjuures iga väärtuse kaaluks on selle omandamise tõenäosus.

Juhul kui diskreetsel juhuslikul suurusel on loenduv arv väärtusi, leitakse selle keskväärts valemiga

$$EX = \sum_{k=1}^{\infty} x_k p_k \quad (1.14)$$

Siin tuleb eeldada, et rida on absoluutselt koonduv, st rida

$$\sum_{k=1}^{\infty} |x_k| p_k$$

on koonduv. Vastasel korral juhuslikul suurusel keskväärts ei eksisteeri.

Kui pideva juhusliku suuruse X jaotustihedus on $f(x)$, siis selle keskväärts leitakse valemiga

$$EX = \int_{-\infty}^{\infty} x f(x) dx \quad (1.15)$$

Ka siin tuleb eeldada, et päratu integraal on absoluutselt koonduv.

Näide 1. Leiame alampunkti 1.7.1 Näites 1 toodud juhusliku suuruse keskvaartuse.

$$EX = 0 \cdot \frac{7}{306} + 1 \cdot \frac{28}{153} + 2 \cdot \frac{7}{17} + 3 \cdot \frac{16}{51} + 4 \cdot \frac{7}{102} = 2,222$$

Keskvaartus on 2,222 vaatamata sellele, et sellist väärtust antud juhuslik suurus omandada ei saa.

Näide 2. Leiame juhusliku suuruse, mille jaotustihedus on

$$f(x) = \begin{cases} \frac{1}{2} \sin x, & \text{kui } 0 \leq x \leq \pi \\ 0 & \text{kui } x < 0 \text{ või } x > \pi \end{cases}$$

keskvaartuse.

Valemi (1.15) järgi

$$EX = \int_0^{\pi} x \cdot \frac{1}{2} \sin x dx = \frac{1}{2} \int_0^{\pi} x \sin x dx$$

Integreerime ositi, valides $u = x$ ja $dv = \sin x dx$. Siis $du = dx$ ja $v = -\cos x$ ning

$$EX = \frac{1}{2} \left[-x \cos x \Big|_0^{\pi} + \int_0^{\pi} \cos x dx \right] = \frac{1}{2} \left(\pi + \sin x \Big|_0^{\pi} \right) = \frac{\pi}{2}$$

Vaatleme keskvaartuse omadusi.

Omadus 1. Konstandi keskvaartus on võrdne selle konstandiga

$$Ec = c$$

Konstantne juhuslik suurus omandab ainult ühe väärtuse c tõenäosusega 1, so selle keskvaartus on tõepoolest $c \cdot 1 = c$.

Omadus 2. Kahe juhusliku suuruse X ja Y summa keskvaartus on võrdne nende keskvaartuste summaga

$$E(X + Y) = EX + EY$$

Omadus 3. Konstantse teguri saab tuua keskvaartuse ette

$$E(cX) = cEX$$

Tõestuseks märgime, et kui juhuslik X suurus omandab väärtuse x , siis juhuslik suurus cX omandab väärtuse cx ja valemi (1.15) abil

$$E(cX) = \int_{-\infty}^{\infty} cx f(x) dx = c \int_{-\infty}^{\infty} x f(x) dx = cEX$$

Omadus 4. Kui X ja Y on sõltumatud juhuslikud suurused, siis

$$E(XY) = EX \cdot EY$$

1.5.2 Juhusliku suuruse dispersioon ja standardhälve

Näide 1. Olgu antud juhuslikud suurused X ja Y jaotusseadustega

X	1	4	5	8
P	0,1	0,4	0,4	0,1

Y	1	4	5	8
P	0,4	0,1	0,1	0,4

Juhuslikel suurustel on samad võimalikud väärtused. Arvutame mõlema juhusliku suuruse keskväärtuse

$$EX = 1 \cdot 0,1 + 4 \cdot 0,4 + 5 \cdot 0,4 + 8 \cdot 0,1 = 4,5$$

$$EY = 1 \cdot 0,4 + 4 \cdot 0,1 + 5 \cdot 0,1 + 8 \cdot 0,4 = 4,5$$

Selgus, et nende juhuslike suuruste keskväärtused on ka võrdsed. Jaotustest on aga näha, et X keskväärtusele lähemal olevate väärtuste tõenäosused on suured ja keskväärtusest kaugemal asuvate väärtuste tõenäosused väikesed. Juhusliku suuruse Y puhul on olukord vastupidine. Juhusliku suuruse Y väärtused on rohkem hajutatud keskväärtuse ümber kui juhusliku suuruse X väärtused.

Definitsioon. Juhusliku suuruse X *hällbeks* nimetatakse juhuslikku suurust $X - EX$.

Koostame näites 2 antud juhusliku suuruse X hällbe jaotusseaduse. Hällbe võimalikud väärtused on $1 - 4,5 = -3,5$, $4 - 4,5 = -0,5$, $5 - 0,5 = 0,5$ ja $8 - 0,5 = 3,5$. Kui juhuslik suurus X omandab väärtuse 1 tõenäosusega 0,1, siis hällve $X - EX$ omandab väärtuse $-3,5$ sama tõenäosusega. Jätkates arutlust samal viisil, saame hällbe jaotuse

$X - EX$	$-3,5$	$-0,5$	$0,5$	$3,5$
P	0,1	0,4	0,4	0,1

Hällbe kui juhusliku suuruse keskväärtus on

$$E(X - EX) = -3,5 \cdot 0,1 - 0,5 \cdot 0,4 + 0,5 \cdot 0,4 + 3,5 \cdot 0,1 = 0$$

Hällbe keskväärtus on 0. See pole juhus, sest mis tahes juhusliku suuruse X hällbe keskväärtus on keskväärtuse omaduste (eelmise alampunti omadused 2 ja 1) põhjal

$$E(X - EX) = E(X + (-EX)) = EX + E(-EX) = EX - EX = 0$$

mille tõttu hällbe keskväärtus ei kõlba juhusliku suuruse hajuvuse iseloomustamiseks. Juhusliku suuruse väärtuste hajutatust keskväärtuse ümber iseloomustab juhusliku suuruse *dispersioon*.

Definitsioon. Juhusliku suuruse dispersiooniks nimetatakse hällbe ruudu keskväärtust.

$$DX = E(X - EX)^2$$

Dispersiooni omadused.

Omadus 1. Konstantse suuruse dispersioon $Dc = 0$.

Tõepoolest, keskvaartuse 1. omaduse põhjal

$$E(c^2 - (Ec)^2) = E(c^2 - c^2) = E(0) = 0$$

Omadus 2. $D(cX) = c^2DX$.

Tõestuseks kasutame keskvaartuse 3. omadust, mille järgi

$$D(cX) = E((cX)^2 - (E(cX))^2) = E(c^2X^2 - c^2(EX)^2) = E(c^2(X^2 - (EX)^2)) = c^2DX$$

Omadus 3.

$$DX = EX^2 - (EX)^2 \quad (1.16)$$

Selle tõestamiseks kasutame keskvaartuse omadusi 1. - 3.

$$\begin{aligned} E(X - EX)^2 &= E(X^2 - 2X \cdot EX + (EX)^2) = E(X^2) - 2EX \cdot EX + (EX)^2 = \\ &= E(X^2) - 2(EX)^2 + (EX)^2 = E(X^2) - (EX)^2 \end{aligned}$$

Dispersiooni saab arvutada definitsiooni abil, kuid üldjuhul arvutatakse dispersioon valemi (1.16) järgi.

Omadus 4. Kui juhuslikud suurused X ja Y on sõltumatud, siis

$$D(X + Y) = DX + DY$$

Näide 2. Selleks, et leida näites 1 antud juhuslike suuruste X ja Y dispersioone, koostame esmalt juhuslike suuruste X ja Y ruutude jaotuse. Nende koostamisel arutleme järgmiselt. Kui juhuslik suurus X omandab väärtuse 1 tõenäosusega 0,1, siis X^2 omandab väärtuse 1 samuti tõenäosusega 0,1. Kui X omandab väärtuse 4 tõenäosusega 0,4, siis X^2 omandab väärtuse 16 samuti tõenäosusega 0,4 jne.

X^2	1	16	25	64
P	0,1	0,4	0,4	0,1

Y^2	1	16	25	64
P	0,4	0,1	0,1	0,4

Leiame

$$EX^2 = 1 \cdot 0,1 + 16 \cdot 0,4 + 25 \cdot 0,4 + 64 \cdot 0,1 = 22,9 \text{ ja}$$

$$DX = 22,9 - 4,5^2 = 2,65$$

$$\text{Teiseks leiame } EY^2 = 1 \cdot 0,4 + 16 \cdot 0,1 + 25 \cdot 0,1 + 64 \cdot 0,4 = 30,1 \text{ ja}$$

$$DY = 30,1 - 4,5^2 = 9,85$$

Juhusliku suuruse Y dispersioon on oluliselt suurem juhusliku suuruse X dispersioonist.

Definitsioon. Juhusliku suuruse X *standardhälbeks* nimetatakse ruutjuurt dispersioonist

$$\sigma_X = \sqrt{DX}$$

Näide 3. Leiame näites 2 antud juhuslike suuruste standardhälbed

$$\sigma_X = \sqrt{2,65} = 1,63$$

ja

$$\sigma_Y = \sqrt{9,85} = 3,14$$

Näide 4. Eeldades, et näites 1 toodud juhuslikud suurused X ja Y on sõltumatud, leiame nende summa $X + Y$ dispersiooni. Selles koostame kõigepealt summa jaotuse

$X + Y$	2	5	6	8	9	10	12	13	16
P	0,04	0,17	0,17	0,04	0,16	0,04	0,17	0,17	0,04

Selgitame kahe näite varal, kuidas on saadud selle tabeli alumises reas olevad tõenäosused.

Summa $X + Y = 2$ kui $X = 1$ ja $Y = 1$ ning et X ja Y on sõltumatud, siis

$$P((X = 1)(Y = 1)) = P(X = 1)P(Y = 1) = 0,1 \cdot 0,4 = 0,04$$

Summa $X + Y = 9$ kui $X = 1$ ja $Y = 8$ või $X = 4$ ja $Y = 5$ või $X = 5$ ja $Y = 4$ või $X = 8$ ja $Y = 1$. Need paarid on paarikaupa teineteist välistavad sündmused. Kasutades teineteist välistavate sündmuste tõenäosuste liitmisteoreemi ja taas kord sõltumatute sündmuste tõenäosuste korrutamisteoreemi, saame

$$\begin{aligned} P(X + Y = 9) &= P(X = 1)P(Y = 8) + P(X = 4)P(Y = 5) + \\ &+ P(X = 5)P(Y = 4) + P(X = 8)P(Y = 1) = \\ &= 0,1 \cdot 0,4 + 0,4 \cdot 0,1 + 0,4 \cdot 0,1 + 0,1 \cdot 0,4 = 0,16 \end{aligned}$$

Edasi leiame summa ruudu keskväärtuse

$$\begin{aligned} E(X + Y)^2 &= 4 \cdot 0,04 + 25 \cdot 0,17 + 36 \cdot 0,17 + 64 \cdot 0,04 + 81 \cdot 0,16 + \\ &+ 100 \cdot 0,04 + 144 \cdot 0,17 + 169 \cdot 0,17 + 256 \cdot 0,04 = 93,5 \end{aligned}$$

Keskväärtuse 2. omaduse põhjal

$$E(X + Y) = EX + EY = 4,5 + 4,5 = 9$$

ning (1.16) järgi

$$D(X + Y) = 93,5 - 81 = 12,5 = 2,65 + 9,85 = DX + DY$$

Loomulikult ei saa Näidet 4 võtta dispersiooni Omadus 4 tõestusena, see on siiski pelgalt ainult üks näide.

Näide 5. Leiame juhusliku suuruse, mille jaotustihedus on

$$f(x) = \begin{cases} \frac{1}{2} \sin x, & \text{kui } 0 \leq x \leq \pi \\ 0 & \text{kui } x < 0 \text{ või } x > \pi \end{cases}$$

dispersiooni ja standardhälbe.

Leiame valemi (1.15) abil X^2 keskvaartuse kasutades kaks korda ositi integreerimise valemit

$$\begin{aligned}
 EX^2 &= \int_0^\pi x^2 \cdot \frac{1}{2} \sin x dx = \\
 &= \frac{1}{2} \left(-x^2 \cos x \Big|_0^\pi + \int_0^\pi 2x \cos x dx \right) = \\
 &= \frac{1}{2} \left(\pi^2 + 2 \int_0^\pi x \cos x dx \right) = \frac{\pi^2}{2} + \int_0^\pi x \cos x dx = \\
 &= \frac{\pi^2}{2} + x \sin x \Big|_0^\pi - \int_0^\pi \sin x dx = \frac{\pi^2}{2} - 2
 \end{aligned}$$

Nüüd valemist (1.16) saame, arvestades eelmise alampunkti Näites 2 saadud sama juhusliku suuruse keskvaartust,

$$DX = \frac{\pi^2}{2} - 2 - \left(\frac{\pi}{2}\right)^2 = \frac{\pi^2 - 8}{4}$$

ning lõpuks standardhälve

$$\sigma_X = \sqrt{DX} = \frac{1}{2} \sqrt{\pi^2 - 8}$$

1.5.3 Juhusliku suuruse mood ja mediaan

Lisaks keskvaartusele kasutatakse tõenäosusteoorias veel kahte juhusliku suuruse väärtuste paiknemist iseloomustavat arvkarakteristikut moodi ja mediaani.

Definitsioon 1. Diskreetse juhusliku suuruse X moodiks $M_o X$ nimetatakse selle juhusliku suuruse väärtust, mille omandamise tõenäosus on suurim.

Definitsioon 2. Pideva juhusliku suuruse X moodiks nimetatakse selle juhusliku suuruse väärtust, mille korral jaotustihedus omab maksimumi.

Võib juhtuda, et moodisid on rohkem kui üks. Samuti on võimalik, et mood puudub.

Definitsioon 3. Juhusliku suuruse X mediaaniks $M_e X$ nimetatakse selle juhusliku suuruse väärtust x_0 , mille korral $P(X < x_0) = P(X > x_0)$.

Sellest definitsioonist jäeldub, et pideva juhusliku suuruse X korral

$$P(X < M_e X) = P(X > M_e X) = 0,5$$

st sirege $x = M_e X$ jaotab jaotustiheduse graafiku ja x -telje vahelise piirkonna kaheks pindvõrdseks osaks, millede pindalad on 0,5.

Näide. Leiame juhusliku suuruse, mille jaotustihedus on

$$f(x) = \begin{cases} \frac{1}{2} \sin x, & \text{kui } 0 \leq x \leq \pi \\ 0 & \text{kui } x < 0 \text{ või } x > \pi \end{cases}$$

moodi ja mediaani.

Leiame x väärtuse, mille korral jaotustihedus $f(x) = \frac{1}{2} \sin x$ omab vahemikus $(0; \pi)$ maksimumi. Võrrandil, jaotustiheduse tuletis $f'(x) = \frac{1}{2} \cos x$ võrdub 0-ga, st

$$\frac{1}{2} \cos x = 0$$

on vahemikus $(0; \pi)$ täpselt üks lahend $x = \frac{\pi}{2}$, seega antud juhusliku suuruse mood

$$M_o X = \frac{\pi}{2}$$

Mediaani leiame tingimusest $P(X < M_e X) = 0,5$ ehk

$$\int_0^{M_e X} \frac{1}{2} \sin x dx = 0,5$$

Pärast võrduse korrutamist 2-ga saame

$$\int_0^{M_e X} \sin x dx = 1$$

ehk

$$-\cos x \Big|_0^{M_e X} = 1$$

st

$$-\cos M_e X + 1 = 1$$

ning võrrandil

$$\cos M_e X = 0$$

on vahemikus $(0; \pi)$ jälle ainsaks lahendiks

$$M_e X = \frac{\pi}{2}$$

Antud juhusliku suuruse jaoks langevad mood ja mediaan kokku. Üldjuhul ei pruugi see nii olla.

1.6 Klassikalised jaotused

Kõigepealt vaatleme diskreetsete juhuslike suuruste jaotusi.

1.6.1 Binoomjaotus

Oletame, et sooritatakse n sõltumatut katset, kus igal katsel on sündmuse toimumise tõenäosus P ja mittetoimumise tõenäosus $q = 1 - p$.

Definitsioon. Binoomjaotusega juhuslikuks suuruseks X nimetatakse juhuslikku suurust, mis omandab täisarvulisi väärtusi $0, 1, 2, \dots, n$, mille omandamise tõenäosused leitakse Bernoulli valemi abil, st

$$P(X = k) = P_{n,k} = C_n^k p^k q^{n-k}$$

Näide. Korvpallur tabab vabaviske tõenäosusega $0,8$. Koostame 6 -st vabaviskest tabavate visete arvu kui juhusliku suuruse X jaotusseaduse.

Punkti 1.2.6 näites oleme juba leidnud $P_{6,5} = 0,393$ ja $P_{6,4} = 0,246$. Lisaks leiame

$$P_{6,6} = C_6^6 \cdot 0,8^6 \cdot 0,2^{6-6} = 0,8^6 = 0,262$$

$$P_{6,3} = C_6^3 \cdot 0,8^3 \cdot 0,2^{6-3} = 20 \cdot 0,512 \cdot 0,008 = 0,082$$

$$P_{6,2} = C_6^2 \cdot 0,8^2 \cdot 0,2^{6-2} = 15 \cdot 0,64 \cdot 0,0016 = 0,015$$

$$P_{6,1} = C_6^1 \cdot 0,8^1 \cdot 0,2^{6-1} = 6 \cdot 0,8 \cdot 0,00032 = 0,00154$$

$$P_{6,0} = C_6^0 \cdot 0,8^0 \cdot 0,2^{6-0} = 0,2^6 = 0,00006$$

Antud juhuslik suurus on binoomjaotusega, sest võimalike väärtuste tõenäosused on arvutatud Bernoulli valemi abil, ja selle jaotus on

X	0	1	2	3	4	5	6
P	0,00006	0,00154	0,015	0,082	0,246	0,393	0,262

Alumise rea summa ei ole täpselt 1 ümardamisvigade tõttu.

Allpool kasutame Newtoni binoomvalemit: iga naturaalarvulise n väärtuse korral

$$(p + q)^n = \sum_{k=0}^n C_n^k p^k q^{n-k} \quad (1.17)$$

Leiame binoomjaotusega juhusliku suuruse X keskvaartuse. Binoomjaotusega juhusliku suuruse X korral $p_k = C_n^k p^k q^{n-k}$, seega valemi (1.13) järgi

$$EX = \sum_{k=0}^n k \cdot C_n^k p^k q^{n-k}$$

Selles summas 1. liige on 0, seega

$$EX = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

ehk pärast k -ga taandamist ja summerimisindeksi nihutamist

$$\begin{aligned} EX &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} = \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} p^{k+1} q^{n-k-1} = \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} = \end{aligned}$$

ehk valemi (1.17) tõttu

$$EX = np(p+q)^{n-1} = np$$

sest $p+q=1$.

Dispersiooni leidmiseks avaldame X^2 keskvärtuse, kasutades eelnevaga sarnaseid teisendusi

$$\begin{aligned} EX^2 &= \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k} = \sum_{k=1}^n k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k} = \\ &= \sum_{k=1}^n k \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} = \sum_{k=0}^{n-1} (k+1) \frac{n!}{k!(n-k-1)!} p^{k+1} q^{n-k-1} = \\ &= \sum_{k=0}^{n-1} \frac{k \cdot n!}{k!(n-k-1)!} p^{k+1} q^{n-k-1} + np \sum_{k=0}^n \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} = \\ &= \sum_{k=1}^{n-1} \frac{n!}{(k-1)!(n-k-1)!} p^{k+1} q^{n-k-1} + np(p+q)^{n-1} = \\ &= \sum_{k=0}^{n-2} \frac{n!}{k!(n-k-2)!} p^{k+2} q^{n-k-2} + np = \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-k-2)!} p^k q^{n-k-2} + np = \\ &= (n^2 p^2 - np^2)(p+q)^{n-2} + np = n^2 p^2 - np^2 + np \end{aligned}$$

ja valemi (1.16) põhjal

$$DX = n^2 p^2 + np(1-p) - n^2 p^2 = npq$$

Binoomjaotusega juhusliku suuruse standardhälve on

$$\sigma_X = \sqrt{npq}$$

Leiame punkti alguses oleva näite juhusliku suuruse, 6-st vabaviskest tabavate visete arv, keskvärtuse, dispersiooni ja standardhälbe.

Selles näites $n=6$, $p=0,8$ ja $q=0,2$. Seega

$$EX = 6 \cdot 0,8 = 4,8$$

$$DX = 6 \cdot 0,8 \cdot 0,2 = 0,96$$

ja

$$\sigma_X = \sqrt{0,96} = 0,98$$

1.6.2 Poissoni jaotus

Poissoni jaotus on diskreetse juhusliku suuruse jaotus, mida kasutatakse binoomjatuse asemel, kui katsete arv on oluliselt suurem kui sündmuse toimumiste arv antud katseseerias ehk kui binoomjaotuses n on suur ja p väike. Poissoni jaotust kasutatakse eeldustel:

- Juhuslik suurus X on sündmuse toimumiste arv mingis ajavahemikus ja omandab väärtusi $0, 1, 2, \dots$
- Katsed on sõltumatud, so sündmuse toimumine ei mõjuta järgmise sündmuse toimumise tõenäosust.
- Sündmuse toimumiste arv ajaühikus ei sõltu ühiku asukohast, vaid ainult selle pikkusest.
- Kaks sündmust ei saa samaaegselt toimuda samal ajahetkel, so igas lõpmatult väikeses ajaühikus kas sündmus toimub või ei toimu.
- Sündmuse toimumise tõenäosus mingis ajaühikus on võrdeline ajaühiku pikkusega.

Üks kuulus ajalooline Poissoni jaotuse praktiline kasutamine toimus Preisi armees. Nimelt, Poissoni jaotuse abil hinnati, kui palju ratsaarmee sõdurid hukkub aastas hobuste kabjalöökide tagajärjel.

Poissoni jaotusega juhuslikud suurused on veel näiteks

- Maad tabavate rohkem kui 1-meetrise läbimõõduga meteoriitide arv aastas.
- Erakorralise meditsiini osakonda saabuvate patsientide arv 22:00 ja 23:00 vahel.

Eeldame, et n on suur ja tõenäosus p väike ning keskvärtus $EX = np = \lambda$ on konstant. Preisi armee ajaloolise näite puhul on sõdurite arv suur, tõenäosus hobuse kabjalöögist surma saada väike ja keskmine kabjalöögist aasta jooksul hukkunud sõdurite arv konstantne.

Suurst λ nimetatakse ka *Poissoni jaotuse parameetrik*s. Tõenäosus selleks, et antud ajavahemikus sündmus toimuks k -l korral on

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (1.18)$$

Näide. Teatud jõel toimuvad ulatuslikud üleujutused keskmiselt üks kord 100 aasta jooksul. Arvutame tõenäosuse, et 100 aastase ajavahemiku jooksul ulatuslikud üleujutused toimuvad 0, 1, 2 või 3 korda.

Et keskmine üleujutuste arv 100 aasta jooksul on 1, siis $\lambda = 1$. Valemi (1.18) järgi

$$P(X = k) = \frac{1^k}{k!} e^{-1} = \frac{e^{-1}}{k!}$$

Siit tõenäosus, et 100 aasta jooksul üleujutusi ei toimu

$$P(X = 0) = \frac{e^{-1}}{0!} = e^{-1} = 0,368$$

Tõenäosus, et 100 aasta jooksul toimub üleujutus ühe korra

$$P(X = 1) = \frac{e^{-1}}{1!} = e^{-1} = 0,368$$

Tõenäosus, et 100 aasta jooksul toimub üleujutus kaks korda

$$P(X = 2) = \frac{e^{-1}}{2!} = 0,184$$

Tõenäosus, et 100 aasta jooksul toimub üleujutus kolm korda

$$P(X = 3) = \frac{e^{-1}}{3!} = 0,061$$

Poissoni jaotusega juhusliku suuruse keskväärtuse ja dispersiooni leidmisel kasutame funktsiooni e^x arendust astmereaks (Maclaurini reaks)

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (1.19)$$

Poissoni jaotuse puhul on $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$, seega keskväärtuse arvutamise valemi (1.14) järgi

$$\begin{aligned} EX &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \end{aligned}$$

Nüüd valemi (1.19) järgi

$$EX = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Dispersiooni leidmiseks leiame esmalt jälle X^2 keskväärtuse

$$\begin{aligned}
EX^2 &= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} = \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{k=0}^{\infty} (k+1) \frac{\lambda^{k+1}}{k!} = \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k+1}}{k!} + \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k-1)!} + \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \\
&= \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}
\end{aligned}$$

Kasutades valemit (1.19), saame

$$EX^2 = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda$$

ning dispersioon

$$DX = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Niisiis, Poissoni jaotuse kaks põhikarakteristikut, keskväärtus ja dispersioon, on võrdsed.

1.6.3 Ühtlane jaotus

Ühtlast jaotust vaatleme nii diskreetse kui ka pideva juhusliku suuruse korral.

Diskreene juhuslik suurus X on ühtlase jaotusega, kui kõigi n väärtuse omandamise tõenäosused on võrdsed, so

$$P(X = x_k) = \frac{1}{n}, \text{ kui } k = 1, 2, \dots, n$$

Näide 1. Täringu veeretamisel saadav silmade arv on ühtlase jaotusega juhuslik suurus, selle jaotus on

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Pidev juhuslik suurus X on ühtlase jaotusega, kui selle jaotustihedus defineeritakse:

$$f(x) = \begin{cases} 0, & \text{kui } x < a \\ \frac{1}{b-a}, & \text{kui } a \leq x \leq b \\ 0, & \text{kui } x > b \end{cases} \quad (1.20)$$

Ühtlase jaotusega juhusliku suuruse jaotusfunktsiooni leiame valemi (1.11) abil. Kui $a \leq x \leq b$, siis

$$F(x) = \int_{-\infty}^x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^x dx = \frac{x-a}{b-a}$$

Seega ühtlase jaotusega juhusliku suuruse jaotusfunktsioon avaldub

$$F(x) = \begin{cases} 0, & \text{kui } x < a \\ \frac{x-a}{b-a}, & \text{kui } a \leq x \leq b \\ 1, & \text{kui } x > b \end{cases} \quad (1.21)$$

Tõenäosus, et ühtlase jaotusega juhuslik suuruse on vahemikus $(x_1; x_2) \subset [a; b]$, on

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \frac{x_2 - x_1}{b - a}$$

Ühtlase jaotusega juhusliku suuruse keskväärus

$$EX = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{b+a}{2}$$

Edasi leiame X^2 keskvääruse

$$EX^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^3}{3} \Big|_a^b = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{b^2 + ab + a^2}{3}$$

Seega ühtlase jaotusega juhusliku suuruse dispersioon on

$$DX = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} = \frac{(b-a)^2}{12}$$

ning standardhälve

$$\sigma_X = \frac{b-a}{2\sqrt{3}}$$

Näide 2. Trammide intervall teatud trammiliinil on 4 minutit. Reisija, kes trammiliikluse graafikut ei tea, satub ooteplatformile juhuslikul ajahetkel. Olgu juhuslik suurus X aeg, kui kaua reisija peab trammi tulekut ootama. Leiame ooteaja keskvääruse ja dispersiooni ning tõenäosuse, et reisija ei pea trammi ootama üle 1 minuti.

Tegemist on ühtlase jaotusega juhusliku suurusega, sest kõikidel ajahetkedel on reisija jõudmine ooteplatformile võrdse tõenäosusega. Selle juhusliku suuruse jaotustihedus on

$$f(x) = \frac{1}{4} \quad \text{kui } 0 < x < 4$$

Ooteaja keskväärtus on

$$EX = \frac{4+0}{2} = 2$$

minutit, dispersioon

$$DX = \frac{(4-0)^2}{12} = \frac{4}{3}$$

ja tõenäosus, et riesijal ei tule oodata üle ühe minuti

$$P(0 < X \leq 1) = \frac{1-0}{4} = \frac{1}{4}$$

1.6.4 Normaaljaotus

Normaaljaotus on loodus- ja sotsiaalteadustes enim kasutatav jaotus. Juhuslik suurus on normaaljaotusega, kui seda mõjutavad paljud tegurid, kuid kõikide tegurite mõju on seljuures väike ning domineerivad tegurid puuduvad.

Normaaljaotusega on igasugused mõõtmistulemused, mõõtmisvead jne.

Normaaljaotus on kaheparameetriline jaotus. Tähistades parameetrid a ja σ , kus a on mis tahes reaalarv ja $\sigma > 0$, on normaaljaotusega juhusliku suuruse X jaotustihedus

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Kui $a = 0$ ja $\sigma = 1$, siis nimetatakse juhuslikku suurust X *normeeritud normaaljaotusega* juhuslikuks suuruseks. Normeeritud normaaljaotusega juhusliku suuruse jaotustihedus

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Edasistes arutlustes võtame ilma tõestuseta teadmiseks, et

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \quad (1.22)$$

Selle valemi abil saab näidata, et normeeritud normaaljaotusega juhusliku suuruse jaotustihedus on tõepoolest mingi juhusliku suuruse jaotustihedus, sest tehes päratus integraalis

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

muutuja vahetuse $t = \frac{x}{\sqrt{2}}$, saame $dt = \frac{dx}{\sqrt{2}}$ ehk $dx = \sqrt{2}dt$ tõttu

$$\int_{-\infty}^{\infty} \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \sqrt{2} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1$$

Samamoodi, tehes päratus integraalis

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

muutuja vahetuse $t = \frac{x-a}{\sigma\sqrt{2}}$, saame, et

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Normaaljaotusega juhusliku suuruse X jaotustiheduse tuletis on

$$f'(x) = -\frac{x-a}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Tuletis võrdub 0-ga, kui $x = a$. Kui $x < a$, siis $f'(x) > 0$ ja kui $x > a$, siis $f'(x) < 0$. Seega on jaotustiheduse graafikul punktis $\left(a; \frac{1}{\sigma\sqrt{2\pi}}\right)$ maksimum. Sealjuures σ suurenedes maksimum väheneb ja vastupidi.

Normaaljaotusega juhusliku suuruse graafikut nimetatakse *Gaussi kõveraks*. Selle maksimumpunkti leidsime, lisaks on see kõver sümmeetriline sirge $x = a$ suhtes.

Normeeritud normaaljaotuse jaotustiheduse graafiku maksimumpunkt on $\left(0; \frac{1}{\sqrt{2\pi}}\right)$ ja graafik (nagu iga paarisfunktsiooni graafik) on sümmeetriline y -telje suhtes. Seega y -telg jaotab selle jaotustiheduse graafiku ja x -telje vahelise piirkonna pindala kaheks võrdseks osaks, st

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \quad (1.23)$$

Leiame normaaljaotusega juhusliku suuruse X keskvaartuse

$$\begin{aligned} EX &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a+a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} a e^{-\frac{(x-a)^2}{2\sigma^2}} dx \end{aligned}$$

Esimese liidetava jagame punkti a kaheks

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_a^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

ja teeme mõlemas muutuja vahetuse $t = \frac{(x-a)^2}{2\sigma^2}$. Sel juhul $dt = \frac{2(x-a)dx}{2\sigma^2}$ ehk $(x-a)dx = \sigma^2 dt$. Leiame rajad: kui $x = a$, siis $t = 0$, ja kui $x \rightarrow \pm\infty$, siis $t \rightarrow \infty$. Seega

$$\begin{aligned} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\infty}^0 e^{-t} \sigma^2 dt + \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\infty} e^{-t} \sigma^2 dt = \\ &= -\frac{\sigma}{\sqrt{2\pi}} \int_{\infty}^0 e^{-t} dt + \frac{\sigma}{\sqrt{2\pi}} \int_0^{\infty} e^{-t} dt = 0 \end{aligned}$$

Seega keskvaartus

$$EX = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} a e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

Selles pärratus integraalis teeme muutuja vahetuse $t = \frac{x-a}{\sigma\sqrt{2}}$, millest $dx = \sigma\sqrt{2}dt$ ning

$$EX = \frac{a}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \sigma\sqrt{2} dt = \frac{a}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt$$

millest (1.22) tõttu

$$EX = a$$

Dispersiooni

$$DX = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

arvutamiseks kasutame jälle muutuja vahetust $t = \frac{x-a}{\sigma\sqrt{2}}$, millest $x-a = \sigma\sqrt{2}t$ ning $dx = \sigma\sqrt{2}dt$, millest

$$DX = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} 2\sigma^2 t^2 e^{-t^2} \sigma\sqrt{2} dt = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt$$

Saadud integraali integreerime ositi, valides

$$u = t \quad \text{ja} \quad dv = te^{-t^2}$$

Siis

$$du = dt \quad \text{ja} \quad v = -\frac{1}{2}e^{-t^2}$$

ja

$$DX = \frac{2\sigma^2}{\sqrt{\pi}} \left[-\frac{t}{2}e^{-t^2} \Big|_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2} dt \right]$$

Et

$$\lim_{t \rightarrow \pm\infty} \frac{t}{2}e^{-t^2} = \frac{1}{2} \lim_{t \rightarrow \pm\infty} \frac{t}{et^2} = 0$$

siis valemit (1.22) rakendades

$$DX = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2}\sqrt{\pi} = \sigma^2$$

Järelikult normaaljaotusega juhusliku suuruse standardhälve

$$\sigma_X = \sigma$$

Seega tähendavad normaaljaotusega juhusliku suuruse parameetrid a ja σ vastavalt selle juhusliku suuruse keskvaärtust ja standardhälvet.

1.6.5 Normaaljaotusega juhusliku suuruse vahemikku kuulumise tõenäosus. Laplace'i funktsioon

Normeeritud normaaljaotusega juhusliku suuruse jaotusfunktsioon on

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

Kasutades võrdust (1.23) saame viimase kirjutada

$$F(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

Funktsiooni

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

nimetatakse *Laplace'i funktsiooniks*. Laplace'i funktsioon on paaritu funktsioon, sest kasutades integraalis

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

muutuja vahetust $u = -t$, saame $dt = -du$ tõttu

$$\Phi(-x) = -\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du = -\Phi(x)$$

Laplace'i funktsiooni väärtused leitakse tabelist (vt Lisa 1).

Normeeritud normaaljaotusega juhusliku suuruse jatosfunktsioon on nüüd kirjutatav

$$F(x) = \frac{1}{2} + \Phi(x)$$

Tõenäosus selleks, et normeeritud normaaljaotusega juhusliku suuruse väärtused kuuluvad vahemikku $(x_1; x_2)$ on arvutatav Laplace'i funktsiooni väärtuste abil, sest

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \frac{1}{2} + \Phi(x_2) - (\frac{1}{2} + \Phi(x_1))$$

ehk

$$P(x_1 < X < x_2) = \Phi(x_2) - \Phi(x_1)$$

Normaaljaotusega juhusliku suuruse, mille parameetrid on a ja σ , jaotusfunktsioon on

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt$$

Kui selles integraalis teha muutuja vahetus $u = \frac{t-a}{\sigma}$, saame $du = \frac{dt}{\sigma}$ ehk $dt = \sigma du$. Vastav ülemine raja on $\frac{x-a}{\sigma}$, järelikult

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}} du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}} du = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right)$$

Seega tõenäosus selleks, et sellise normaaljaotusega juhusliku suuruse väärtused kuuluvad vahemikku $(x_1; x_2)$ on

$$P(x_1 < X < x_2) = \frac{1}{2} + \Phi\left(\frac{x_2 - a}{\sigma}\right) - \frac{1}{2} - \Phi\left(\frac{x_1 - a}{\sigma}\right)$$

ehk

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right) \quad (1.24)$$

Leiame tõenäosuse, et normaaljaotusega juhuslik suuruse X väärtus erineb keskväärtusest a vähem kui ε võrra

$$P(|X - a| < \varepsilon) = P(-\varepsilon < X - a < \varepsilon) = P(a - \varepsilon < X < a + \varepsilon)$$

Valemi (1.24) põhjal

$$P(|X - a| < \varepsilon) = \Phi\left(\frac{a + \varepsilon - a}{\sigma}\right) - \Phi\left(\frac{a - \varepsilon - a}{\sigma}\right)$$

ehk, arvestades sellega, et Laplace'i funktsioon on paaritu,

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) \quad (1.25)$$

Näide 1. Firms toodetakse kuullaagreid, kuuli nominaalne diameeter on 10 mm. Kuulide diameeter on juhuslik suurus X , mis allub normaaljaotusele keskväärtusega 10 mm, standardhällbega 0,2 mm. Kuul loetakse praaktooteks, kui selle diameeter ületab 10,3 mm või on alla 9,7 mm. Leiame tõenäosuse, et kuul osutub praaktooteks.

Tõenäosus, et kuul osutub praaktooteks on

$$P(|X - a| > 0,3) = 1 - P(|X - a| < 0,3)$$

Valemi (1.25) põhjal

$$P(|X - a| < 0,3) = 2\Phi\left(\frac{0,3}{0,2}\right) = 2\Phi(1,5) \approx 2 \cdot 0,433 = 0,866$$

Seega tõenäosus, et kuul osutub praaktooteks on

$$P(|X - a| > 0,3) = 1 - 0,866 = 0,134$$

Normaaljaotusega juhusliku suuruse *keskmiseks veaks* nimetatakse sellist suurust E , et

$$P(|X - a| < E) = 0,5$$

st $P(-E < X - a < E) = 0,5$ ehk

$$P(a - E < X < a + E) = 0,5$$

Geomeetriliselt tähendab see, et jaotustiheduse graafiku, x -telje ja sirgete $x = a - E$ ning $x = a + E$ vahele jääva piirkonna pindala on 0,5, st moodustab poole jaotustiheduse graafiku ja x -telje vahele jääva piirkonna pindalast.

Valemi (1.25) põhjal

$$2\Phi\left(\frac{E}{\sigma}\right) = 0,5$$

ehk

$$\Phi\left(\frac{E}{\sigma}\right) = 0,25$$

Laplace'i funktsiooni tabelis on väärtused $\Phi(0,67) = 0,24857$ ja $\Phi(0,68) = 0,25175$. Seega

$$\frac{E}{\sigma} = 0,675$$

ehk keskmine viga avaldub standardhälbe kaudu

$$E = 0,675\sigma \quad (1.26)$$

Näide 2. Lennuk lendab 20 km laiuses õhukoridoris. Lennutrajektoori säilitamisel tehtav süstemaatiline viga on 0,5 km ja keskmine viga 2 km. Kui suur on tõenäosus, et lennuk satub väljaspoole õhukoridori?

Juhuslikuks suuruseks X on lennutrajektoori hälbed, mis alluvad normaaljaotusele, mille keskväärtuseks on süstemaatiline viga $a = 0,5$. Valemist (1.26) saame

$$2 = 0,675\sigma$$

millest

$$\sigma \approx 2,96$$

Sündmus - lennuk satub väljaspoole õhukoridori - on sündmuse - lennuk püsib õhukoridoris - vastandsündmus. Tõenäosus, et lennuk püsib õhukoridoris on

$$\begin{aligned} P(-10 < X < 10) &= \Phi\left(\frac{10 - 0,5}{2,96}\right) - \Phi\left(\frac{-10 - 0,5}{2,96}\right) = \\ &= \Phi(3,21) + \Phi(3,55) = 0,49933 + 0,49980 \approx 0,99913 \end{aligned}$$

Seega tõenäosus, et lennuk satub väljaspoole õhukoridori, on

$$1 - 0,9991 = 0,00087$$

1.7 Tõenäosusteooria piirteoreemid

Tõenäosusteooria piirteoreemid seovad massiliselt esinevate juhuslike nähtuste teoreetilisi ja eksperimentaalseid karakteristikuid. Piisavalt pikkade katsete seeriatega korral saab hakata tegema küllaltki tõepäraseid oletusi järgnevate katsete tulemuste kohta.

1.7.1 Tšebõševi teoreem

Vaatleme juhuslike suuruste jada

$$X_1, X_2, \dots, X_n, \dots \quad (1.27)$$

Definitsioon. Juhuslike suuruste jada (1.27) koondub tõenäosuse järgi reaalarvuks c , kui iga $\varepsilon > 0$ korral

$$\lim_{n \rightarrow \infty} P(|X_n - c| < \varepsilon) = 1$$

Viimane tingimus tähendab sisuliselt seda, et küllalt suure n väärtuse korral on sündmuse $|X_n - c| < \varepsilon$ mittetoimumine väga väikese tõenäosusega.

Tšebõševi teoreem. Kui juhuslikud suurused X_k ($k = 1, 2, \dots$) on sõltumatud ja tõkestatud dispersioonidega $DX_k \leq d$ ($k = 1, 2, \dots$), siis iga $\varepsilon > 0$ korral

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n EX_k\right| < \varepsilon\right) = 1$$

st juhuslike suuruste aritmeetiline keskmine koondub tõenäosuse järgi nende juhuslike suuruste keskväärtuste keskmiseks.

Kui kõikide juhuslike suuruste keskväärtused on võrdsed, st $EX_k = EX$, siis Tšebõševi teoreem väidab, et

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - EX\right| < \varepsilon\right) = 1$$

1.7.2 Bernoulli teoreem

Teoreem. Oletame, et on sooritatud n sõltumatut katset ja sündmus A toimub selles katseseerias m korda. Sündmuse A tõenäosus igal katsel olgu p . Siis iga $\varepsilon > 0$ korral

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1$$

st sündmuse toimumise suhteline sagedus koondub tõenäosuse järgi selle sündmuse toimumise tõenäosuseks.

Teoreemi väite võib kirjutada ka

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| > \varepsilon\right) = 0$$

st mida pikem on katsaeseeria, seda väiksmaks muutub tõenäosus, et sündmuse toimumise suhteline sagedus erineb sündmuse tõenäosusest ükskõik kui väikese suuruse võrra.

1.7.3 Tsentraalne piirteoreem

Kui juhuslikud suurused X_k ($k = 1, 2, \dots, n$) on sõltumatud ja ühesuguse jaotusega, siis on nende juhuslike suuruste summa

$$\sum_{k=1}^n X_k$$

samuti mingi juhuslik suurus. Selle keskväärtus on

$$E \sum_{k=1}^n X_k$$

standardhälve

$$\sqrt{D \sum_{k=1}^n X_k}$$

Juhuslik suurus

$$\frac{\sum_{k=1}^n X_k - E \sum_{k=1}^n X_k}{\sqrt{D \sum_{k=1}^n X_k}} \quad (1.28)$$

on antud juhuslike suuruste normeeritud summa.

Teoreem. Kui juhuslikud suurused X_k ($k = 1, 2, \dots, n$) on sõltumatud ja ühesuguse jaotusega ning iga $k = 1, 2, \dots, n$ korral eksisteerivad EX_k , DX_k ja $E|X_k - EX_k|^3$ (kolmandat järku absoluutne keskmoment), siis

$$\lim_{n \rightarrow \infty} P\left(-\infty < \frac{\sum_{k=1}^n X_k - E \sum_{k=1}^n X_k}{\sqrt{D \sum_{k=1}^n X_k}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (1.29)$$

Võrdusmärgist vasakul pool olev tõenäosus on normeeritud summa (1.28) jaotusfunktsioon. Võrdusmärgist paremal pool on aga normeeritud normaaljaotusega juhusliku suuruse jaotusfunktsioon.

Seega väidab tsentraalne piirteoreem, et juhuslike suuruste normeeritud summa jaotusfunktsioon koondub normeeritud normaaljaotusega juhusliku suuruse jaotusfunktsiooniks.

Tsentraalne piirteoreem annab teoreetilise põhjenduse normaaljaotuse kasutamiseks, kui juhuslik suurus on mõjutatud paljude erinevate tegurite poolt ($n \rightarrow \infty$), kusjuures iga iga üksiku teguri mõju juhuslikule suurusele on tühine.

1.7.4 Moivre-Laplace'i integraalne piirteoreem

Tsentraalses piirteoreemis on juhusliud suurused $X_1, X_2, \dots, X_n$ üldjuhul mis tahes jaotusega. Kui aga eeldada, et need juhuslikud surused on ühesuguse jaotusega ja omandavad ainult kaht võimalikku väärtust 0 ja 1, saame tsentraalse piirteoreemi erijuhu, nn moivre-Laplace'i integraalse teoreemi.

Moivre-Laplace'i integraalne piirteoreem. Kui k on sündmuse toimumiste arv n sõltumatu katse korral ja igal katsel on sündmuse toimumise tõenäosus p , siis

$$\lim_{n \rightarrow \infty} P\left(\alpha < \frac{k - np}{\sqrt{npq}} < \beta\right) = \Phi(\beta) - \Phi(\alpha)$$

Teisendame võrratusi

$$\alpha < \frac{k - np}{\sqrt{npq}} < \beta$$

$$\alpha\sqrt{npq} < k - np < \beta\sqrt{npq}$$

ehk

$$np + \alpha\sqrt{npq} < k < np + \beta\sqrt{npq}$$

Tähistades $np + \alpha\sqrt{npq} = k_1$ ja $np + \beta\sqrt{npq} = k_2$, saame, et

$$\alpha = \frac{k_1 - np}{\sqrt{npq}}$$

ja

$$\beta = \frac{k_2 - np}{\sqrt{npq}}$$

ning Moivre-Laplace'i integraalne piirteoreemi väitest saame

$$\lim_{n \rightarrow \infty} P(k_1 < k < k_2) = \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) \quad (1.30)$$

Märkus. Sündmuse toimumiste arv n katse korral on binoomjaotusega juhuslik suurus. Järelikult võimaldab Moivre-Laplace'i integraalne piirteoreem arvutada binoomjaotusega juhuslikum suuruse antud vahemikku kuulumise tõenäosust, kasutades selleks Laplace'i funktsiooni ehk sisuliselt normeeritud normaaljaotusega juhuslikku suurust. Piisavalt pika katseseeria, st

piisavalt suure n väärtuse korral

$$P(k_1 < k < k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) \quad (1.31)$$

Näide. Korvpallur tabab vabaviske tõenäosusega 0,7. Nädala jooksul sooritab ta 2000 vabaviset. Arvutame tõenäosuse, et ta tabab rohkem kui 1000 ja vähem kui 1420 vabaviset.

Kasutades valemit (1.31) on $n = 2000$, $p = 0,7$, $k_1 = 1000$ ja $k_2 = 1420$. Seega $np = 1400$, $npq = 420$ ning

$$\begin{aligned} P(1000 < k < 1420) &\approx \Phi\left(\frac{1420 - 1400}{\sqrt{420}}\right) - \Phi\left(\frac{1000 - 1400}{\sqrt{420}}\right) = \\ &= \Phi(0,98) - \Phi(-19,52) = 0,33646 + 0,5 \approx 0,84 \end{aligned}$$

1.7.5 Laplace'i lokaalne piirteoreem

Bernoulli valemiga

$$P_{n,k} = C_n^k p^k q^{n-k}$$

arvutatakse tõenäosus, et juhuslik sündmus toimub n katsest k -l korral eeldusel, et igal katsel on sündmuse toimumise tõenäosus p . Kui n on suur, on tegemist väga mahukate arvutustega. Neid on võimalik vältida kasutades Laplace'i lokaalset piirteoreemi.

Laplace'i lokaalne piirteoreem.

$$\lim_{n \rightarrow \infty} \frac{P_{n,k}}{\frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}} = 1$$

Siit saame tõenäosuse arvutamiseks ligikaudse valemi

$$P_{n,k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

Võrreldes seda normaaljaotuse jaotustihedusega

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

näeme, et ligikaudse võrduse paremal pool on normaaljaotuse jaotustihedus parameetritega $a = np$ ja $\sigma = \sqrt{npq}$.

Ligikaudset valemit on mugavam kasutada kujul

$$P_{n,k} \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \text{kus } x = \frac{k - np}{\sqrt{npq}} \quad (1.32)$$

Näide. Tõenäosus, et supermarketisse saabunud maasikakastis on riknenud marju, on 0,1. Kui suur on tõenäosus, et 300-st saabunud kastist 20-s esineb riknenud marju?

Antud näites, $n = 300$, $k = 20$, $p = 0,1$ ja $q = 0,9$. Arvutame $np = 30$, $\sqrt{npq} = \sqrt{30 \cdot 0,9} = 3\sqrt{3}$,

$$x = \frac{20 - 30}{3\sqrt{3}} = -\frac{10}{3\sqrt{3}}$$

ning valemi (1.32) järgi arvutatav tõenäosus

$$P_{300,20} \approx \frac{1}{3\sqrt{6\pi}} e^{-\frac{100}{2 \cdot 27}} = 0,012$$

Märkus. Arvestades sellega, et antud näites n on suur ja p on väike, saaksime tõenäosuse arvutamiseks kasutada ka Poissoni jaotust. Valemis (1.18) $\lambda = 30$ ning

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda} = \frac{30^{20}}{20!} e^{-30} = 0,013$$

1.8 Juhuslik vektor

Juhuslike nähtuste uurimisel ei ole paljudel juhtudel piisav vaadelda ainult ühte juhuslikku suurust, vaid tuleb kasutada kahte, kolme või enam juhuslikku suurust. Näiteks punkt, kuhu maandub pall tenniseväljakul on kahe juhusliku koordinaadiga - abstsissi ja ordinaadiga.

Mitme juhusliku suuruse samaaegsel käsitlemisel kasutatakse juhuslikku vektorit. Juhuslikuks vektoriks nimetatakse vektorit, mille komponendid on juhuslikud suurused. Juhusliku vektori uurimisel ei piisa ainult selle komponentide kui juhuslike suuruste uurimisest, vaid on vaja ka nende koordinaatide vahelisi sõltuvusi.

Kahemõõtmeline juhuslik vektor (X, Y) on vaadeldav ka tasandi juhusliku punktina, mille kohavektoriks on (X, Y) .

Kolmemõõtmeline juhuslik vektor (X, Y, Z) on vaadeldav juhusliku punktina, mille kohavektoriks on (X, Y, Z) , ruumis.

Edasises piirdume kahemõõtmeliste juhuslike vektorite uurimisega.

1.8.1 Diskreetse juhusliku vektori jaotus

Juhusliku vektori koordinaadid omandavad väärtuse on sündmus, millel on tõenäosus. Kui vektori juhuslik komponent X omandab väärtuse x_i ja komponent Y omandab väärtuse y_j , siis vektorit (x_i, y_j) nimetatakse juhusliku vektori *realisatsiooniks*.

Definitsioon 1. Diskreetseks juhuslikuks vektoriks nimetatakse juhuslikku vektorit, mille realisatsioonide hulk on lõplik või äärmisel juhul loenduv.

Kui tegemist on diskreetse juhusliku vektoriga, mille realisatsioonide hulk on lõplik, siis komponent X omadab väärtusi x_i , $i = 1, 2, \dots, m$, ja komponent Y väärtusi y_j , $j = 1, 2, \dots, n$. Tähistades sündmuse $(X = x_i, Y = y_j)$

tõenäosuse

$$P(X = x_i, Y = y_j) = p_{ij}$$

on diskreetse juhusliku vektori jaotusseadus esitatav tabelina

$X \backslash Y$	y_1	y_2	$\dots$	y_n
x_1	p_{11}	p_{12}	$\dots$	p_{1n}
x_2	p_{21}	p_{22}	$\dots$	p_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mn}

Sellel tabelil on omadus

$$\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1$$

sest kõik sündmused $(X = x_i, Y = y_j)$ $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ on paarikaupa teineteist välistavad ja üks neist realiseerub kindlasti. Sealjuures

$$P(X = x_i) = \sum_{j=1}^n P(X = x_i, Y = y_j) = \sum_{j=1}^n p_{ij}$$

ja

$$P(Y = y_j) = \sum_{i=1}^m P(X = x_i, Y = y_j) = \sum_{i=1}^m p_{ij}$$

1.8.2 Juhusliku vektori jaotusfunktsioon

Definitsioon 1. Juhusliku vektori jaotusfunktsiooniks nimetatakse tõenäosust

$$F(x, y) = P(X < x, Y < y)$$

Geomeetriliselt tähendab juhusliku vektori jaotusfunktsioon juhusliku punkti (X, Y) sattumise tõenäosust xy -tasandi punktihulka, mis jäävad III veerandisse tipuga (x, y)

Jaotusfunktsiooni omadused.

- $0 \leq F(x, y) \leq 1$
- Jaotusfunktsioon on monotoonselt kasvav mõlema argumendi järgi.

Tõepoolest, valides näiteks kaks X väärtust x_1 ja x_2 selliselt, et $x_1 < x_2$, saame sündmuse $(X < x_2, Y < y)$ kirjutada teineteist välistavate sündmuste summana

$$(X < x_2, Y < y) = (X < x_1, Y < y) + (x_1 \leq X < x_2, Y < y)$$

Seega

$$P(X < x_2, Y < y) = P(X < x_1, Y < y) + P(x_1 \leq X < x_2, Y < y)$$

ehk

$$F(x_2, y) = F(x_1, y) + P(x_1 \leq X < x_2, Y < y)$$

ja et $P(x_1 \leq X < x_2, Y < y) \geq 0$, siis $F(x_2, y) \geq F(x_1, y)$.

- Tähistades komponendi X jaotusfunktsiooni $F_1(x)$ ja komponendi Y jaotusfunktsiooni $F_2(y)$, kehtivad võrdused

$$\lim_{y \rightarrow \infty} F(x, y) = F_1(x)$$

ja

$$\lim_{x \rightarrow \infty} F(x, y) = F_2(y)$$

st kui juhusliku vektori üks argument kasvab tõkestamatult, siis juhusliku vektori jaotusfunktsioon läheneb teise komponendi jaotusfunktsioonile.

- $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0$
- $\lim_{(x, y) \rightarrow (\infty, \infty)} F(x, y) = 1$

Definitsioon 2. Juhusliku vektorit (X, Y) nimetatakse pidevaks, kui selle jaotusfunktsioon $F(x, y)$ on pidev mõlema argumendi järgi.

1.8.3 Juhusliku vektori jaotustihedus

Jaotusfunktsiooni saab kasutada nii diskreetse kui ka pideva juhusliku vektori jaotuse kirjeldamiseks. Pidevate juhuslike vektorite jaotuse iseloomustamiseks sobib paremini juhusliku vektori jaotustihedus.

Definitsioon. Kui pideva juhusliku vektori jaotusfunktsioonil on olemas teist järku segatuletis $\frac{\partial^2 F}{\partial x \partial y}$, siis seda nimetatakse juhusliku vektori (X, Y) jaotustiheduseks ja tähistatakse

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}$$

Funktsiooni $z = f(x, y)$ graafikuks on pind ruumis, mida nimetatakse *jaotuspinnaks*. Teades jaotustihedust, saame arvutada tõenäosuse selleks, et juhuslik punkt asuks piirkonnas D valemiga

$$P((X, Y) \in D) = \iint_D f(x, y) dx dy$$

Geomeetriliselt tähendab see sellise kõversilindri ruumala, mis pealt on piiratud jaotuspinnaga $z = f(x, y)$, alt xy -tasandi piirkonnaga D ja mille moodustaja on paralleelne z -teljega.

Jaotusfunktsioon $F(x, y)$ on tõenäosus, et juhuslik punkt asub xy -tasandi III veerandis tipuga (x, y) , st

$$F(x, y) = \int_{-\infty}^x dx \int_{-\infty}^y f(x, y) dy \quad (1.33)$$

Juhusliku vektori jaotustiheduse omadused.

Omadus 1. Jaotustihedus on mittenegatiivne funktsioon

$$f(x, y) \geq 0$$

See järeldub jaotustiheduse definitsioonist ja sellest, et juhusliku vektori jaotusfunktsioon on monotoonselt kasvav mõlema argumendi järgi.

Omadus 2.

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x, y) dy = 1$$

See järeldub võrdusest (1.33) ja juhusliku vektori jaotusfunktsiooni omadusest $\lim_{(x, y) \rightarrow (\infty, \infty)} F(x, y) = 1$

Geomeetriliselt tähendab omadus 2 seda, et jaotuspinna ja xy -tasandi vahelise piirkonna ruumala võrdub 1-ga.

Näide. Juhusliku vektori (X, Y) jaotustihedus ruudus $0 \leq x \leq \frac{\pi}{2}$, $0 \leq y \leq \frac{\pi}{2}$ on $f(x, y) = a \sin(x+y)$ ja väljaspool seda ruutu $f(x, y) = 0$. Määrame kordaja a ja leiame jaotusfunktsiooni.

Väljaspool ruutu on $f(x, y) = 0$, seega Omadus 2 järgi

$$a \int_0^{\frac{\pi}{2}} dx \int_0^{\frac{\pi}{2}} \sin(x+y) dy = 1$$

Arvutame seesmise integraali

$$\int_0^{\frac{\pi}{2}} \sin(x+y) dy = -\cos(x+y) \Big|_0^{\frac{\pi}{2}} = -\cos\left(x + \frac{\pi}{2}\right) + \cos x = \sin x + \cos x$$

ning välise integraali

$$\int_0^{\frac{\pi}{2}} (\sin x + \cos x) dx = (-\cos x + \sin x) \Big|_0^{\frac{\pi}{2}} = 1 + 1 = 2$$

Seega $2a = 1$, millest $a = \frac{1}{2}$

Kui $x < 0$ või kui $y < 0$, siis ilmselt $F(x, y) = 0$. Samuti kui $x > 1$ või $y > 1$, siis $F(x, y) = 1$. Kui $0 \leq x \leq \frac{\pi}{2}$ ja $0 \leq y \leq \frac{\pi}{2}$, siis $f(x, y) = \frac{1}{2} \sin(x+y)$ ning

$$F(x, y) = \frac{1}{2} \int_0^x dx \int_0^y \sin(x+y) dy$$

Integreerimisel tuleb arvestada, et ülemine raja y on konstant, aga samas on seesmist integraali arvutades on y muutuja.

$$\int_0^y \sin(x+y) dy = -\cos(x+y) \Big|_0^y = \cos x - \cos(x+y)$$

Viimases on nüüd y konstant ja

$$F(x, y) = \frac{1}{2} \int_0^x [\cos x - \cos(x+y)] dx = \frac{1}{2} [\sin x - \sin(x+y)] \Big|_0^x = \frac{1}{2} [\sin x + \sin y - \sin(x+y)]$$

Kui $0 \leq x \leq \frac{\pi}{2}$ ja $y > \frac{\pi}{2}$, saame komponendi X jaotusfunktsiooni

$$F_1(x) = \frac{1}{2} [\sin x + \sin \frac{\pi}{2} - \sin(x + \frac{\pi}{2})] = 1 + \sin x - \cos x$$

ja kui $x > \frac{\pi}{2}$ ja $0 \leq y \leq \frac{\pi}{2}$, saame komponendi Y jaotusfunktsiooni

$$F_2(y) = 1 + \sin y - \cos y$$

1.8.4 Juhusliku vektori komponentide jaotustihedused. Tinglikud jaotusseadused

Olgu $f(x, y)$ juhusliku vektori (X, Y) jaotustihedus. Jaotusfunktsiooni omaduste põhjal saame komponendi X jaotusfunktsiooni

$$F_1(x) = \lim_{y \rightarrow \infty} F(x, y) = \int_{-\infty}^x dx \int_{-\infty}^{\infty} f(x, y) dy$$

ja komponendi Y jaotusfunktsiooni

$$F_2(y) = \lim_{x \rightarrow \infty} F(x, y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^y f(x, y) dy$$

Diferentseerides esimest võrdust muutuja x järgi, saame komponendi X jaotustiheduse

$$f_1(x) = \frac{dF_1}{dx} = \int_{-\infty}^{\infty} f(x, y) dy \quad (1.34)$$

ja diferentseerides teist võrdust muutuja y järgi, saame komponendi Y jaotustiheduse

$$f_2(y) = \frac{dF_2}{dy} = \int_{-\infty}^{\infty} f(x, y) dx \quad (1.35)$$

Viimaste valemite abil saab leida juhusliku vektori komponentide jaotustihedusi, teades juhusliku vektori jaotustihedust.

Vastupidine ülesanne ei ole üldjuhul üheselt lahenduv. Juhusliku vektori komponentide jaotustihedused iseloomustavad kumbagi juhusliku suurust üksikult, kuid ei iseloomusta nende omavahelisi seoseid.

Seega, kui juhusliku vektori komponendid on sõltuvad, ei saa juhusliku vektori jaotusseadust iseloomustada ainult vektori komponentide jaotusseaduste kaudu. See toob meid tingliku jaotusseaduse mõiste juurde.

Definitsioon. Vektori ühe koordinaadi *tinglikuks jaotusseaduseks* nimetatakse selle koordinaadi jaotust tingimusel, et teine koordinaat omandab kindla väärtuse.

Komponendi X tinglikku jaotusfunktsiooni tingimusel, et $Y = y$ tähistatakse $F(x/y)$ ja tingliku jaotustihedust $f(x/y)$. Komponendi Y tinglikku jaotusfunktsiooni tingimusel, et $X = x$ tähistatakse $F(y/x)$ ja tingliku jaotustihedust $f(y/x)$.

Komponendi X tinglik jaotustihedus leitakse seosest

$$f(x/y) = \frac{f(x, y)}{f_2(y)} \quad (1.36)$$

Komponendi Y tinglik jaotustihedus leitakse seosest

$$f(y/x) = \frac{f(x, y)}{f_1(x)} \quad (1.37)$$

Viimased võrdused saab kirjutada

$$f(x, y) = f_1(x)f(y/x) = f_2(y)f(x/y) \quad (1.38)$$

st juhusliku vektori jaotustiheduse leidmiseks on vaja teada ühe komponendi jaotustihedust ja teise komponendi tinglikku jaotustihedust.

Kasutades valemeid (1.34) ja (1.35), saame võrdustest (1.36) ja (1.37)

$$f(x/y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \quad (1.39)$$

ja

$$f(y/x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy} \quad (1.40)$$

Tinglikel jaotustihedustel on kõik jaotustiheduse omadused, sealhulgas

$$\int_{-\infty}^{\infty} f(x/y) dx = 1$$

ja

$$\int_{-\infty}^{\infty} f(y/x) dy = 1$$

1.8.5 Sõltumatud ja sõltuvad juhuslikud suurused

Definitsioon. Juhuslikku suurust X nimetatakse sõltumatuks juhuslikust suurusest Y , kui juhusliku suuruse X jaotusseadus ei sõltu sellest, millise väärtuse omandab juhuslik suurus Y .

Samamoodi saab defineerida juhusliku suuruse Y sõltumatuse juhuslikust suurusest X .

Kui sõltumatud juhuslikud suurused X ja Y on pidevad, siis juhusliku suuruse X tinglik jaotustihedus ei sõltu Y väärtustest, st tinglik jaotustihedus on võrdne komponendi X jaotustihedusega

$$f(x/y) = f_1(x)$$

Kui aga juhuslik suurus X sõltub juhuslikust suurusest Y , siis

$$f(x/y) \neq f_1(x)$$

Lause. Kui juhuslik suurus X on sõltumatu juhuslikust suurusest Y , siis ka juhuslik suurus Y on sõltumatu juhuslikust suurusest X .

Tõepoolest võrdusest (1.38) saame

$$f_1(x)f(y/x) = f_2(y)f_1(x)$$

millest järeldub, et

$$f(y/x) = f_2(y)$$

Teoreem. Pidevate juhuslike suuruste X ja Y sõltumatuseks on tarvilik ja piisav, et juhusliku vektori (X, Y) jaotustihedus avalduks komponentide jaotustiheduste korrutisena

$$f(x, y) = f_1(x)f_2(y)$$

Näide. Olgu juhusliku vektori (X, Y) jaotustihedus

$$f(x, y) = \frac{1}{\pi^2(1 + x^2 + y^2 + x^2y^2)}$$

Kas X ja Y on sõltumatud või sõltuvad juhuslikud suurused.

Et veenduda juhuslike suuruste X ja Y sõltumatuses, piisab vektori jaotustiheduse teisendamisest kujule

$$f(x, y) = \frac{1}{\pi^2(1 + x^2)(1 + y^2)} = \frac{1}{\pi(1 + x^2)} \cdot \frac{1}{\pi(1 + y^2)}$$

Vektori komponentide jaotustihedused

$$f_1(x) = \frac{1}{\pi(1 + x^2)}$$

ja

$$f_2(y) = \frac{1}{\pi(1 + y^2)}$$

Sellise jaotustihedusega juhuslike suurusi nimetatakse *Cauchy jaotusega* juhuslikeks suurusteks.

1.8.6 Juhusliku vektori arvkarakteristikud

Juhusliku vektori jaotusseadus iseloomustab juhuslikku vektorit täielikult. Katseandmete järgi aga ei pruugi jaotusseadus alati olla määratav. Seejärel kasutatakse sageli juhusliku vektori arvkarakteristikuid, mis teatud määral annavad ka ettekujutuse jaotuseadusest.

Diskreetse kahemõõtmelise vektori $(X; Y)$ jaotusseadusega

$X \backslash Y$	y_1	y_2	$\dots$	y_n
x_1	p_{11}	p_{12}	$\dots$	p_{1n}
x_2	p_{21}	p_{22}	$\dots$	p_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mn}

komponentide X ja Y keskvaartused arvutatakse valemitega

$$EX = \sum_{i=1}^m x_i \sum_{j=1}^n p_{ij} \quad (1.41)$$

ja

$$EY = \sum_{j=1}^n y_j \sum_{i=1}^m p_{ij} \quad (1.42)$$

Pideva kahemõõtmelise vektori $(X; Y)$ komponentide X ja Y keskvaartused arvutatakse valemitega

$$EX = \int_{-\infty}^{\infty} x f_1(x) dx = \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} f(x, y) dy \quad (1.43)$$

ja

$$EY = \int_{-\infty}^{\infty} y f_2(y) dy = \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} f(x, y) dx \quad (1.44)$$

Näide 1. Diskreetse juhusliku vektori $(X; Y)$ jaotusseadus on

$X \backslash Y$	2	3	4
1	0,1	0	0,2
3	0,1	0,2	0,1
5	0	0,2	0,1

Arvutame komponentide X ja Y keskväärtused ja dispersioonid.

Kõigepealt leiame valemi (1.41) järgi

$$EX = 1 \cdot 0,3 + 3 \cdot 0,4 + 5 \cdot 0,3 = 3,0$$

siis sama valemi järgi

$$EX^2 = 1^2 \cdot 0,3 + 3^2 \cdot 0,4 + 5^2 \cdot 0,3 = 11,4$$

Seega

$$DX = 11,4 - 3,0^2 = 2,4$$

Edasi valemi (1.42) järgi

$$EY = 2 \cdot 0,2 + 3 \cdot 0,4 + 4 \cdot 0,4 = 3,2$$

Sama valemi abil

$$EY^2 = 2^2 \cdot 0,2 + 3^2 \cdot 0,4 + 4^2 \cdot 0,4 = 10,8$$

ja lõpuks dispersiooni

$$DY = 10,8 - 3,2^2 = 0,56$$

Näide 2. Juhusliku vektori $(X; Y)$ jaotustihedus kolmnurgas $x \geq 0$, $y \leq 1$ ja $y \geq x$ on $f(x, y) = 8xy$ ja väljaspool seda kolmnurka $f(x, y) = 0$. Leiame komponendi X keskväärtuse ja dispersiooni.

Esiteks leiame komponendi X jaotustiheduse

$$f_1(x) = \int_x^1 8xy dy = 4xy^2 \Big|_x^1 = 4x - 4x^3$$

seejärel valemi (1.43) abil X keskväärtuse

$$EX = \int_0^1 x(4x - 4x^3) dx = \left(\frac{4x^3}{3} - \frac{4x^5}{5} \right) \Big|_0^1 = \frac{8}{15}$$

ja X^2 keskväärtuse

$$EX^2 = \int_0^1 x^2(4x - 4x^3)dx = \left(x^4 - \frac{4x^6}{6}\right)\Big|_0^1 = \frac{1}{3}$$

Komponendi X dispersioon on seega

$$DX = EX^2 - (EX)^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}$$

Definitsioon 1. Juhusliku vektori (X, Y) *kovariatsiooniks* nimetatakse komponentide X ja Y hälvete korrutise keskväärtust

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)] \quad (1.45)$$

Diskreetsete juhuslike suuruste X ja Y kovariatsioon arvutatakse valemiga

$$\text{cov}(X, Y) = \sum_{i=1}^m \sum_{j=1}^n (x_i - EX)(y_j - EY)p_{ij} \quad (1.46)$$

ning pidevate juhuslike suuruste kovariatsioon valemiga

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY)f(x, y)dxdy \quad (1.47)$$

Keskväärtuse omadusi kasutades saame

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - EX)(Y - EY)] = E(XY - X \cdot EY - Y \cdot EX + EX \cdot EY) = \\ &= E(XY) - EX \cdot EY - EY \cdot EX + EX \cdot EY \end{aligned}$$

ehk

$$\text{cov}(X, Y) = E(XY) - EX \cdot EY \quad (1.48)$$

Juhusliku vektori kovariatsioon iseloomustab peale juhuslike komponentide hajuvuse ka nende vastastikust mõju. Selle mõju tugevuse hindamiseks kasutatakse dimensioonita suhet.

Definitsioon 2. Juhuslike suuruste X ja Y *korrelatsioonikordajaks* nimetatakse suhet

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad (1.49)$$

Definitsioon 3. Juhuslikke suurusi X ja Y nimetatakse mittekorreleeruvaks, kui nende kovariatsioon (st ka korrelatsioonikordaja) võrdub 0-ga.

Kehtib väide.

Lause. Sõltumatud juhuslikud suurused X ja Y on mittekorreleeruvad.

Selle lause väide järeldub vahetult sõltumatute juhuslike suuruste kesk-
väärtuse omadusest $E(XY) = EX \cdot EY$.

Osutub, et juhuslike suuruste mittekorreleeruvus on tarvilik nende juhus-
like suuruste sõltumatuseks, kuid mitte piisav. Juhuslikud suurused X ja Y
võivad olla mittekorreleeruvad ja samal ajal sõltuvad. Vaatleme selle kohta
näidet.

Näide. Olgu juhuslik vektor (X, Y) ühtlase jaotusega ringis raadiusega
 r , keskpunktiga koordinaatide alguses, st

$$f(x, y) = \begin{cases} \frac{1}{\pi r^2}, & \text{kui } x^2 + y^2 \leq r^2 \\ 0, & \text{kui } x^2 + y^2 > r^2 \end{cases} \quad (1.50)$$

Leiame kõigepealt komponentide X ja Y jaotustihedused. Valemi (1.34)
järgi

$$f_1(x) = \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2} dy = \frac{2}{\pi r^2} \sqrt{r^2 - x^2}$$

kui $|x| \leq r$ ja $f_1(x) = 0$, kui $|x| > r$. Valemi (1.35) järgi

$$f_2(y) = \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{1}{\pi r^2} dx = \frac{2}{\pi r^2} \sqrt{r^2 - y^2}$$

kui $|y| \leq r$ ja $f_2(y) = 0$, kui $|y| > r$.

Valemi (1.36) järgi leiame komponendi X tingliku jaotustiheduse

$$f(x/y) = \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2} \sqrt{r^2 - y^2}} = \frac{1}{2\sqrt{r^2 - y^2}}$$

kui $|x| \leq \sqrt{r^2 - y^2}$ ja $f(x/y) = 0$, kui $|x| > \sqrt{r^2 - y^2}$. Valemi (1.37) järgi
leiame komponendi Y tingliku jaotustiheduse

$$f(y/x) = \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2} \sqrt{r^2 - x^2}} = \frac{1}{2\sqrt{r^2 - x^2}}$$

kui $|y| \leq \sqrt{r^2 - x^2}$ ja $f(y/x) = 0$, kui $|y| > \sqrt{r^2 - x^2}$. Ilmselt

$$f(x/y) \neq f_1(x)$$

ega ka

$$f(y/x) \neq f_2(y)$$

st juhuslikuud suurused X ja Y on sõltuvad.

Komponentide X ja Y kovariatsiooni arvutamiseks kasutame valemit (1.47). Selleks arvutame esiteks komponendi X keskväärtuse

$$\int_{-\infty}^{\infty} x f_1(x) dx = \frac{2}{\pi r^2} \int_{-r}^r x \sqrt{r^2 - x^2} dx$$

Kui selles integraalis teha muutuja vahetus $t = r^2 - x^2$, saame $dt = -2x dx$ ehk $x dx = -\frac{dt}{2}$. Kui $x = -r$, siis $t = 0$ ning kui $x = r$, siis samuti $t = 0$. Pärast muutuja vahetust

$$\int_{-\infty}^{\infty} x f_1(x) dx = -\frac{1}{\pi r^2} \int_0^0 \sqrt{t} dt = 0$$

Arvutades komponendi Y keskväärtuse, saame samuti

$$\int_{-\infty}^{\infty} y f_2(y) dy = 0$$

Järelikult valemi (1.47) põhjal

$$\text{cov}(X, Y) = \iint_D xy \frac{1}{\pi r^2} dx dy = \frac{1}{\pi r^2} \iint_D xy dx dy$$

kus D on ring $x^2 + y^2 \leq r^2$. Kahekordse integraali arvutamiseks teisendame selle polaarkoordinaatidesse $x = \rho \cos \varphi$, $y = \rho \sin \varphi$ ja $|J| = \rho$. Siis

$$\text{cov}(X, Y) = \frac{1}{\pi r^2} \int_0^{2\pi} d\varphi \int_0^r \rho^3 \cos \varphi \sin \varphi d\rho$$

Seesmine integraal

$$\cos \varphi \sin \varphi \int_0^r \rho^3 d\rho = \frac{r^4}{4} \cos \varphi \sin \varphi$$

Kui välises integraalis

$$\frac{1}{\pi r^2} \frac{r^4}{4} \int_0^{2\pi} \cos \varphi \sin \varphi d\varphi$$

teha muutuja vahetus $t = \sin \varphi$, saame $dt = \cos \varphi d\varphi$ tõttu

$$\text{cov}(X, Y) = \frac{r^2}{4\pi} \int_0^0 t dt = 0$$

st juhusliku vektori komponendid on mittekorreleeruvad.

2 Matemaatilise statistika elemendid

2.1 Matemaatilise statistika põhimõisted

Matemaatiliseks statistikaks nimetatakse matemaatika haru, mis uurib statistiliste andmete põhjal järelduste tegemise meetodeid. Nimetame mõningad matemaatilise statistika suunad: katseplaneerimine, statistiliste hüpoteeside kontrollimine, statistilised hinnangud, statistilised otsustused jne. Tõenäosusteooria on matemaatilise statistika teoreetiline alus. Matemaatilises statistikas on kasutusel tõenäosusteooria terminoloogia ja meetodid.

2.1.1 Matemaatilise statistika põhimõisted

Üldkogumiks nimetatakse kõikide uuritavate objektide hulka. Üldkogumiks võib olla näiteks Eesti rahvastik, kõik Eesti üliõpilased, ühe tehase poolt valmistatud teatud mõõtmetega autorehvid, mingi taimeliik, loomaliik (Eestis, Euroopas, terves maailmas) jne.

Üldkogumi uurimine on ühel juhul kulukas (mille tõttu kogu rahvastiku uurimine viiakse läbi mitte sagedamini kui üks kord kümne aasta jooksul) või sootuks võimatu (kui katsetada kõiki tehase poolt toodetud autorehve, pole midagi müüa).

Selle tõttu võetakse üldkogumist mingi alamhulk, nn *valim*. Statistiliste otsuste tegemiseks peab valim olema piisavalt suur ja see ei või olla tendentslik, st peab esindama üldkogumit. Näiteks kui uuritakse Eesti üliõpilasi, siis peab valimis olema erinevatest rahvustest üliõpilasi, mõlemast soost õpilasi, linnast ja maalt pärit üliõpilasi, kõikide maakondade üliõpilasi, üliõpilasi erinevate sissetulekutega perekondadest jne.

Matemaatilise statistika põhiülesandeks on valimi põhjal järelduste tegemine üldkogumi kohta.

Üldkogumit uuritakse valimi abil teatud *tunnuse* seisukohalt. Kui uuritakse Eesti üliõpilasi, võib tunnuseks olla üliõpilase pikkus, kaal, kinga number, kõrgema matemaatika hinne, elukoha kaugus koolist, silmade värv, vanemate päritolu, õpilase rahulolu kooliga jne.

Tunnuseid jaotatakse *kvantitatiivseteks* ja *kvalitatiivseteks*. Kvantitatiivsed tunnused on mõõdetavad, st iseloomustatavad arvudega. Nendeks on näiteks üliõpilase pikkus, kaal, kinga number, matemaatika hinne jne.

Kvantitatiivsed tunnused jaotuvad *diskreetseteks* ja *pidevateks* tunnusteks. Õpilase kaal ja pikkus on pidevad tunnused, kinga number ja matemaatika hinna aga diskreetsed tunnused.

Kvalitatiivsed tunnused on omakorda jaotatavad *ordinaalseteks* ja *nominaalseteks*. Ordinaalsed tunnused on mingil moel loogiliselt järjestatavad tunnused. Näiteks üliõpilase rahulolu oma õppekavas pakutuga võib anda hinnanguks *olen väga rahul*, *olen rahul*, *neutraalne*, *ei ole rahul* või *pole üldse rahul*. Nominaalseid tunnuse tulemusi omavahel ei saa võrrelda. No-

minaalseteks tunnusteks on näiteks üliõpilase sugu, silmade värv, vanemate päritolu.

Tunnuseid hakkame tähistama sarnaselt tõenäosusteoorias vaadeldus juhusliku suurusega tähestiku lõpu suurte tähtedega X, Y, Z, U, V, W .

Järgnevaid mõisteid vaatleme näite varal. Oletame, et õpetaja Kobra seisis ühe kuu jooksul igal koolipäeva hommikul kooli sissepääsu juures ja registreeris esimesse tundi hilinejate arvu. Õpetaja Kobral oli kuu aja pärast järgmine tulemuste rida

17, 6, 11, 14, 10, 18, 7, 16, 17, 22, 14, 4, 12, 14, 20, 12, 6, 19, 14, 12, 15.

Olgu tunnus X - hilinejate arv. Rida, mis saadakse tunnuse väärtuste registreerimisel nende esinemise järjekorras, nimetatakse *statistiliseks reaks*. Objektide (antud näites koolipäevade) arvu statistilises reas nimetatakse *valimi mahuks*. Näites on valimi maht $n = 21$, st vaadeldavas kuus on 21 koolipäeva.

Variatsioonreaks nimetatakse korrastatud statistilist rida, kus tulemused on toodud kasvavas järjekorras.

Tunnuse X - hilinejate arv - variatsioonrida on

4, 6, 6, 7, 10, 11, 12, 12, 12, 14, 14, 14, 14, 15, 16, 17, 17, 18, 19, 20, 22.

Variatsioonrea põhjal moodustame *sagedustabeli*, milles ülemisse ritta kirjutame tunnuse väärtused ja alumisse nende väärtuste sagedused f_i . Esimene väärtus 4 esineb variatsioonreas ühe korra, st selle sagedus on $f_1 = 1$, teine väärtus 6 esineb variatsioonreas 2 korda, st selle sagedus on $f_2 = 2$, ... , seitsmes väärtus 14 esineb variatsioonreas 4 korda, st selle sagedus $f_7 = 4$ jne.

X	4	6	7	10	11	12	14	15	16	17	18	19	20	22
f_i	1	2	1	1	1	3	4	1	1	2	1	1	1	1

Kui tunnuse väärtusi on palju, siis väärtused jaotatakse *väärtuste klassidesse*. Klasside arvu m võib määrata klassifitseerija, kuid, kui väärtused ei anna head ideed klasside arvu valikuks, võib kasutada reeglit $m = 1 + \log_2 n$ ja ümardada tulemus lähima täisarvuni. Meie näites oleks klasside arv on $m = 1 + \log_2 21 = 5,39 \approx 5$.

Klassi ulatuseks on loomulik valida variatsioonreas esineva suurima väärtuse ja vähima väärtuse vahe jagatud klasside arvuga $\frac{x_{\max} - x_{\min}}{m}$. See ümardatakse alati ülespoole, selleks, et kõik tunnuse väärtused jääksid moodustatud klassidesse. Meie näites on klassi ulatus $\frac{22 - 4}{5} = 3,6$, st klassi ulatuseks valime 4 ühikut. Et aga kogu variatsioonrea ulatus on 18 ühikut, viide klassi klassi ulatusega 4 mahub 20 ühikut, võtame esimeseks väärtuste klassiks poollõigu $[3; 7)$, vaatamata sellele, et väärtust 3 variatsioonreas ei ole. Moodustame viie klassiga sagedustabeli, kus ülemisse ritta kirjutame väärtuste klassid ja alumisse ritta klassi kuuluvate tunnuse väärtuste arvu, *sageduse*.

X	[3; 7)	[7; 11)	[11; 15)	[15; 19)	[19; 23]
f_i	3	2	8	5	3

Väärtuste klassideks on poollõigud, sest iga tunnuse väärtus võib kuuluda ainult ühte klassi. Tunnuse väärtus 7 kuulub teise väärtuste klassi, tunnuse väärtus 11 kuulub kolmandasse väärtuste klassi jne. Erandiks on viimane väärtuste klass, mis on lõik, st sisaldab mõlemat oma otspunkti. Meie näite puhul see ei ole oluline, sest väärtust 23 nagunii variatsioonreas ei ole.

Kui sagedustabelis klassi kuulumise sagedused asendada suhteliste sagedustega, siis saame tunnuse *jaotustabeli*.

X	[3; 7)	[7; 11)	[11; 15)	[15; 19)	[19; 23]
$\frac{f_i}{n}$	$\frac{1}{7}$	$\frac{2}{21}$	$\frac{8}{21}$	$\frac{5}{21}$	$\frac{1}{7}$

Jaotustabelil on samasugune omadus, nagu juhusliku suuruse jaotustabelil: kui kõik alumises reas olevad suhtelised sagedused kokku liita, saame tulemuseks ühe.

Edasi vaatleme statistilise andmestiku graafilist esitamist. Jaotustabeli põhjal koostame *histogrammi*, so tulpdiagrammi, kus abstsissel on väärtuste klassid ja ordinaatteljel suhtelised sagedused.

Kui soovitakse näha, kui suure osa moodustavad erinevad väärtuste klassid kõikide väärtuste hulgas, kasutatakse graafilise esitusena sektordiagrammi. Iga klass kujutatakse sektorina, mille kesknurk moodustab 360-st kraadist suhtelise sageduse suuruse osa. Klasse [3; 7) ja [19; 23) kujutavate sektorite kesknurk on $\frac{1}{7} \cdot 360^\circ = 51,4^\circ$, klassi [7; 11) kujutava sektori kesknurk on $\frac{2}{21} \cdot 360^\circ = 34,3^\circ$, klassi [11; 15) kujutava sektori kesknurk on $\frac{8}{21} \cdot 360^\circ = 137,1^\circ$ ning klassi [15; 19) kujutava sektori kesknurk on $\frac{5}{21} \cdot 360^\circ = 85,7^\circ$.

2.1.2 Empiiriline jaotusfunktsioon

Oletame, et uurime teatud objektide üldkogumit valimi abil ja valimi maht on n . Üldkogumi *empiiriliseks jaotusfunktsiooniks* ninetatakse fuktsiooni

$$F^*(x) = \sum_{x_i < x} \frac{f_i}{n}$$

Mida suurem on valimi maht, seda täpsemalt iseloomustab valimi põhjal koostatud empiiriline jaotusfunktsioon üldkogumi jaotusfunktsiooni $F(x)$. Kehtib teoreem.

Teoreem. Valimi mahu tõkestamatul kasvamisel koondub empiiriline jaotusfunktsioon tõenäosuse jargi üldkogumi jaotusfunktsiooniks, st

$$\lim_{n \rightarrow \infty} P(|F^*(x) - F(x)| < \varepsilon) = 1$$

iga $\varepsilon > 0$ korral.

Empiirilisel jaotusfunktsioonil on sarnased omadused juhusliku suuruse jaotusfunktsiooniga.

- $0 \leq F^*(x) \leq 1$;
- $F^*(x)$ on monotoonselt kasvav, st kui $x_2 > x_1$, siis $F^*(x_2) \geq F^*(x_1)$;
- $F^*(x) = 0$, kui $x \leq x_{min}$ ja $F^*(x) = 1$, kui $x > x_{max}$.

Empiirilise jaotusfunktsiooni graafik on treppjoon. Kooostame eelmise punkti näites toodud tunnuse X - hilinejate arv klassifitseeritud jaotustabeli põhjal empiirilise jaotusfunktsiooni.

$$F^*(x) = \begin{cases} 0, & \text{kui } x \leq 3 \\ \frac{1}{7}, & \text{kui } 3 < x \leq 7 \\ \frac{5}{21}, & \text{kui } 7 < x \leq 11 \\ \frac{13}{21}, & \text{kui } 11 < x \leq 15 \\ \frac{18}{21}, & \text{kui } 15 < x \leq 19 \\ 1, & \text{kui } x > 19 \end{cases}$$

2.1.3 Keskväärtuse, dispersiooni ja standardhälbe punkthinnangud

Punkthinnanguid saame vaadelda kvantitatiivsete tunnuste puhul. Punkt-hinnangud jagunevad paiknemist ja hajuvust iseloomustavateks. Paiknemist iseloomustavad *aritmeetiline keskmine* (keskväärtus), *mood* ja *mediaan*. Hajuvust iseloomustavad dispersioon ja standardhälve.

Üldkogumi keskvaartuse hinnanguks on valimi aritmeetiline keskmine. Kui valimi variatsioonrida on

$$x_1, x_2, x_3, \dots, x_n$$

st valimi maht on n , siis aritmeetiline keskmine on

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Keskvaartuse punkthinnang on valimist sõltuv juhuslik suurus!

Keskvaartuse punkthinnang sagedustabeli järgi, kus on m erinevat väärtust,

X	x_1	x_2	x_3	$\dots$	x_m
f_i	f_1	f_2	f_3	$\dots$	f_m

leitakse valemiga

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m x_i f_i \quad (2.1)$$

Siin kõikide sageduste summa võrdub valimi mahuga, st

$$f_1 + f_2 + \dots + f_m = n$$

Eelmises punktis vaadeldud näite tunnuse X - hilinejate arv - keskväärtus on

$$\bar{x} = \frac{4 + 6 \cdot 2 + 7 + 10 + 11 + 12 \cdot 3 + 14 \cdot 4 + 15 + 16 + 17 \cdot 2 + 18 + 19 + 20 + 22}{21} = 13\frac{1}{3}$$

Kui arvutatakse aritmeetiline keskmine klassifitseeritud sagedustabelist, siis valitakse arvutamisel klassi esindajaks klassi keskmine, st, kui klass on pool-
lõik $[a; b)$, valitakse arvutamiseks väärtus $x = \frac{a+b}{2}$. Arvutades sama kesk-
väärtuse klassifitseeritud sagedustabelist, saame tulemuseks

$$\bar{x} = \frac{5 \cdot 3 + 9 \cdot 2 + 13 \cdot 8 + 17 \cdot 5 + 21 \cdot 3}{21} = 13,571$$

mis mõnevõrra küll erineb eelmisest tulemusest, kuid suhteliselt vähe.

Mis tahes parameetri hinnang peab olema *nihutamata*, st hinnangu kesk-
väärtus peab võrduma hinnatava parameetriga. Keskväärtuse puhul

$$E\bar{x} = EX$$

Seda tuleb mõista nii: keskväärtuse punkthinnang $\bar{x}$ on valimist sõltuv ju-
huslik suurus ja selle juhusliku suurus keskväärtus võrdub üldkogumi kesk-
väärtusega.

Tunnuse X mood M_oX on tunnuse suurima sagedusega väärtus. Eelmise
punkti näite tunnuse mood on $M_oX = 14$, sest selle sagedus on 4 ja kõikide
teiste väärtuste sagedus on väiksem. Tunnusel võib mood ka puududa ning
tunnusel võib olla ka mitu moodi, st mitu väärtust, millel on suurim sagedus.

Tunnuse X mediaan M_eX on tunnuse väärtus, millest suuremaid ja väik-
semaid on võrdselt. Eelmise punkti näite tunnuse mediaan $M_eX = 14$, sest
variatsioonreas väärtusest 14 on 10 eespool ja 10 tagapool.

Juhul, kui variatsioonreas on paarisarv väärtusi, siis on mediaan võrdne
kahe keskmise väärtuse aritmeetilise keskmisega. Näiteks, kui tunnuse Y -
üliõpilase pikkus - variatsioonrida on

$$159, 162, 164, 165, 167, 168, 170, 171$$

$$\text{siis tunnuse } Y \text{ mediaan } M_eY = \frac{165 + 167}{2} = 166.$$

Olgu valimi variatsioonrida on

$$x_1, x_2, x_3, \dots, x_n$$

Tunnuse hälveteks nimetatakse vahesid $x_i - \bar{x}$ ja hälvete ruutudeks nende ruutused $(x_i - \bar{x})^2$. Tunnuse dispersiooni hinnanguks D^*X nimetatakse hälvete ruutude keskväärtust

$$D^*X = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Teisendades saame

$$\begin{aligned} D^*X &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \bar{x}^2 = \\ &= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 = \overline{x^2} - \bar{x}^2 \end{aligned}$$

Kui tunnuse X valimi sagedustabel on

X	x_1	x_2	x_3	$\dots$	x_m
f_i	f_1	f_2	f_3	$\dots$	f_m

leitakse dispersiooni hinnang valemiga

$$D^*X = \overline{x^2} - \bar{x}^2 \quad (2.2)$$

kus

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^m x_i^2 f_i$$

ja $\bar{x}$ arvutatakse valemi (2.1) abil.

Osutub, et dispersiooni hinnang D^*X ei ole nihutamata hinnang. Nimelt, on võimalik tõestada, et

$$E(D^*X) = \frac{n-1}{n} DX$$

so D^*X on nihutatud hinnang. Nihutamata hinnanguks on

$$s^2 = \frac{n}{n-1} D^*X \quad (2.3)$$

Standardhälbeks nimetatakse ruutjuurt dispersioonist

$$s = \sqrt{s^2}$$

Oletame, et tunnuse väärtused (näiteks õpilase pikkus) on sentimeetrites. Sellisel juhul ka aritmeetiline keskmine on sentimeetrites. Tunnuse väärtuste ruudud on ruutsentimeetrites, ruutsentimeetrites on ka väärtuste ruutude keskvärtus ja keskvärtuse ruut. Järelikult on hajuvust iseloomustav suurus - dispersioon - ruutsentimeetrites. Standardhälve, kui ruutjuur dispersioonist on aga sel juhul sentimeetrites nagu juhusliku suuruse väärtused, st sama mõõtmega nagu juhusliku suuruse väärtused.

Leiame eelmise punkti näites vaadeldud tunnuse X - hilinemiste arv - dispersiooni ja standardhälbe. Arvutuste mahu kokkuhoiu mõttes kasutame selleks klassifitseeritud väärtuste sagedustabelit. Keskvärtuseks saime $\bar{x} = 13,571$. Leiame ruutude aritmeetilise keskmise

$$\overline{x^2} = \frac{5^2 \cdot 3 + 9^2 \cdot 2 + 13^2 \cdot 8 + 17^2 \cdot 5 + 21^2 \cdot 3}{21} = 207,476$$

dispersiooni nihutatud hinnangu

$$D^*X = 207,476 - 13,571^2 = 23,304$$

dispersiooni nihutamata hinnangu

$$s^2 = \frac{21}{20} \cdot 23,304 = 24,469$$

ja standardhälbe nihutamata hinnangu

$$s = \sqrt{24,469} = 4,95$$

Variatsioonikordaja kasutamine hajuvuse iseloomustamiseks on vajalik selleks, et erinevastes mõõtühikutes saadud standardhälbeid omavahel võrrelda (näiteks kui üks tunnus on õpilase pikkus, teine kaal, kolmas kinga number).

Variatsioonikordajaks v nimetatakse tunnuse standardhälbe ja aritmeetilise keskmise suhet, st

$$v = \frac{\sigma}{\bar{x}}$$

Variatsioonikordaja iseloomustab kui suur on hajuvus tunnuse mõõtühikutes aritmeetilise keskmise ühe ühiku suhtes. Variatsioonikordaja on ühikuta suurus ja see esitatakse tavaliselt protsentides. Tunnuse X - hilinemiste arv - variatsioonikordaja on

$$v = \frac{4,83}{13,571} = 0,36 = 36\%$$

2.1.4 Keskvärtuse vahemikhinnang

Valimi põhjal leitud keskvärtuse punkthinnang on juhuslik ja ei ole alust arvata, et see on võrdne üldkogumi keskvärtusega. Seetõttu kasutatakse lisaks keskvärtuse punkthinnangule keskvärtuse *vahemikhinnangut*.

Vahemikhinnang on valimi põhjal leitud vahemik, millesse hinnatav parameeter, so üldkogumi keskvärtus kuulub ette antud tõenäosusega. Seda ette antud tõenäosust nimetatakse *usaldusnivooks*, ja tähistatakse β . Reeglina antakse ette usaldusnivoo 0,9 või 0,95 või 0,99.

Olgu $\bar{x}$ valimi põhjal arvutatud keskvärtuse punkthinnang ja a üldkogumi, so hinnatava parameetri, keskvärtus (mis ei ole teada). Keskvärtuse sümmeetriliseks usalduspiirkonnaks nimetatakse vahemikku

$$(\bar{x} - \varepsilon; \bar{x} + \varepsilon)$$

mis katab hinnatava keskvärtuse tõenäosusega β , st

$$P(|\bar{x} - a| < \varepsilon) = \beta$$

Keskvärtuse punkthinnang on valimist sõltuv juhuslik suurus, seepärast kasutame selle jaoks tähistust $\bar{X}$.

2.1.5 Normaaljaotuse keskvärtuse usaldusvahemik, kui standardhälve on teada

On võimalik tõestada, et kui juhuslik suurus on normaaljaotusega, siis ka selle keskvärtuse punkthinnang kui juhuslik suurus $\bar{X}$ on normaaljaotusega ja üldkogumi keskvärtuse a usaldusvahemik leitakse seosest

$$P(|\bar{X} - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma(\bar{X})}\right)$$

Siin juhusliku suuruse $\bar{X}$ dispersioon on

$$D\bar{X} = D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n Dx_i = \frac{DX}{n}$$

Seega standardhälve

$$\sigma(\bar{X}) = \sqrt{D\bar{X}} = \frac{\sqrt{DX}}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$$

Järelikult

$$P(|\bar{X} - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = \beta \quad (2.4)$$

Näide. Oletame, et üliõpilane mõõdab teatud vedeliku keemistemperatuuri (kraadides Celsiuse järgi) ja saab 6 erinevat tulemust 102,5; 101,7; 103,1; 100,9; 100,5 ja 102,2. Üliõpilane teab, et selle vedeliku keemistemperatuuride standardhälve on 1,2 kraadi. Milline on keskvärtuse usaldusvahemik usaldusnivool 0,95.

Arvutame valimi keskväärtuse

$$\bar{x} = \frac{102,5 + 101,7 + 103,1 + 100,9 + 100,5 + 102,2}{6} = 101,82$$

Suuruse ε leiame valemi (2.4) abil

$$2\Phi\left(\frac{\varepsilon\sqrt{6}}{1,2}\right) = 0,95$$

millest

$$\Phi\left(\frac{\varepsilon\sqrt{6}}{1,2}\right) = 0,475$$

Laplace'i funktsiooni väärtuste tabelist

$$\frac{\varepsilon\sqrt{6}}{1,2} = 1,96$$

st $\varepsilon = 0,96$. Seega keskväärtuse usalduspiirkond usaldusnivool 0,95 on

$$(101,82 - 0,96; 101,82 + 0,96) = (100,86; 102,78)$$

st keemistemperatuuride keskväärtus jääb tõenäosusega 0,95 just nimelt sellesse vahemikku. Kui tõenäosust suurendada, siis peab loomulikult suurendama ka usaldusvahemikku.

2.1.6 Normaaljaotuse keskväärtuse usaldusvahemik, kui standardhälve ei ole teada. Studenti jaotus

Vaatleme normaaljaotuse keskväärtuse usalduspiirkonna leidmist, kui üldkogumi standardhälve ei ole teada.

Elmises punkti saime, et $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$. Nüüd asendame tundmatu σ tema hinnanguga s ja moodustame juhusliku suuruse

$$T = \frac{\bar{X} - a}{\sigma(\bar{X})} = \frac{(\bar{X} - a)\sqrt{n}}{s} \quad (2.5)$$

Juhuslikku suurust T nimetatakse *Studenti jaotusega* juhuslikuks suuruseks *vabadusastmete arvuga* $k = n - 1$, kus n on valimi maht. Tähistame Studenti jaotusega juhusliku suuruse T võimalikke väärtusi t . On võimalik tõestada, et valimi mahu kasvades läheneb Studenti jaotuse jaotustihedus $s_k(t)$ normeeritud normaaljaotuse jaotustihedusele

$$\lim_{k \rightarrow \infty} s_k(t) = \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

ja, nagu ka normeeritud normaaljaotuse jaotustihedus, on ka Studenti jaotuse jaotustihedus paarisfunktsioon.

Keskväertuse usalduspiirkonna leidmiseks läheme valemis

$$P(|\bar{X} - a| < \varepsilon) = \beta$$

üle Studenti jaotusele

$$P(|\bar{X} - a| \frac{\sqrt{n}}{s} < \varepsilon \frac{\sqrt{n}}{s}) = P(|T| < \varepsilon \frac{\sqrt{n}}{s}) = \beta$$

2.1.7 Kahe tunnuse võrdlemine. Korrelatsioon ja regressioon

Olgu mingi valim, mida uurime mitme tunnuse seisukohalt. *Korrelatsioon* on tunnuste vastastikune mõju, mille tõttu ühe tunnuse väärtuste muutused kutsuvad esile ka teise tunnuse väärtuste muutused.

Korrelatsioonanalüüsi kasutatakse tunnuste vahelise seose olemasolu, tugevuse ja iseloomu mõõtmiseks. Suurusi vaadeldakse sümmeetriliselt, s.t. ei eeldata, et üks tunnustest on põhjus, teine tagajärg.

Korrelatsioon kahe tunnuse vahel tähendab seda, et ühe tunnuse suurematele väärtustele vastavad sagedamini teise tunnuse suuremad (positiivne korrelatsioon) või väiksemad (negatiivne korrelatsioon) väärtused. Kui seost ei ole, on korrelatsioon null.

Oletame, et meil on ühe klassi õpilased. Esimeseks tunnuseks olgu õpilase pikkus, teiseks õpilase kaal ja kolmandaks matemaatika kursusehinne. Võib oletada, et õpilase pikkus kaalu mõjutab, st üldjuhul pikem õpilane kaalub rohkem. Samas õpilase kaal matemaatika kursusehinnet vaevalt et eriti mõjutab. Seega võib eeldada, et õpilase pikkus ja kaal on omavahel tugevas korrelatsioonis, aga kaal ja matemaatika kursusehinne on nõrgas korrelatsioonis.

Oletame, et meil on ühe valimi kaks tunnust X ja Y . Kandes tunnuse X väärtused abstsisssteljele ja tunnuse Y väärtused ordinaatteljele, saame punktihulga koordinaattasandil, mida nimetatakse *korrelatsiooniväljaks*.

Näitena vaatleme ülesandes 108 toodud kahte tunnust: X - liblika tiibade siruulatus (mm) - ja Y - liblika kaal (g). Kanname koordinaatteljestikku punktid, mille abstsiss on tiibade siruulatus ja ordinaat kaal.

Saadud korrelatsiooniväljas on 15, mitte 18 punkti, sest kolm väärtuste paari (23; 16), (24; 20) ja (25; 18) esinevad korrelatsiooniväljas kahekordselt.

Kahe tunnuse vahelise korrelatsiooni tugevust iseloomustab *lineaarne korrelatsioonikordaja*. Tähistame tunnuse X aritmeetilise keskmise ja standardhälbe $\bar{x}$ ja σ_x ning tunnuse Y aritmeetilise keskmise ja standardhälbe vastavalt $\bar{y}$ ja σ_y . Lineaarne korrelatsioonikordaja arvutatakse valemist

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

kus korrutise aritmeetiline keskmine leitakse valemist

$$\overline{xy} = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n}$$

Lineaarsel korrelatsioonikordajal r on järgmised omadused

- $-1 \leq r \leq 1$;
- Kui $r > 0$, on tegemist positiivse korrelatsiooniga, st ühe tunnuse väärtuste suurenemine tingib teise tunnuse väärtuste suurenemise;
- Kui $r < 0$, on tegemist negatiivse korrelatsiooniga, st ühe tunnuse väärtuste suurenemine tingib teise tunnuse väärtuste vähenemise;
- kui tunnused on sõltumatud, siis $r = 0$;
- kui $|r| < 0,3$, loetakse korrelatsiooni nõrgaks;
- kui $0,3 < |r| < 0,7$, loetakse korrelatsiooni keskmiseks;
- kui $|r| > 0,7$, loetakse korrelatsiooni tugevaks.

Arvutame ülesande 108 kahe tunnuse, liblika tiibade siruulatuse ja kaalu, vahelise korrelatsioonikordaja. Teame, et $\bar{x} = 23,78$, $\sigma_x = 1,78$, $\bar{y} = 18$ ja $\sigma_y = 2,08$. Lisaks arvutame ülesandes 108 toodud tabeli alusel korrutise aritmeetilise keskmise

$$\overline{xy} = \frac{25 \cdot 19 + 23 \cdot 16 + 26 \cdot 19 + \dots + 25 \cdot 18}{18} = 431,2$$

ning seejärel lineaarse korrelatsioonikordaja

$$r = \frac{431,2 - 23,78 \cdot 18}{1,78 \cdot 2,08} = 0,8535$$

mis tähendab seda, et liblika tiibade siruulatuse ja kaalu vahel on tugev positiivne korrelatsioon.

Regressioonisirgeks nimetatakse sirget, mis korrelatsiooniväljas olevat punktihulka tervikuna teatud mõttes kõige paremini iseloomustaks. Olgu regressioonisirge võrrand $\hat{y} = kx + b$, kus k on sirge tõus ja b algordinaat. Tõus ja algordinaat leitakse sellise nõude järgi, et korrelatsiooniväljas olevate y

väärtuste y_i ja regressioonsirgel x vastavate väärtuste x_i korral arvutatud $\hat{y}$ väärtuste $\hat{y}_i$ vahede ruutude summa

$$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2$$

oleks vähim. Kui lähtuda sellest nõudest, on võimalik näidata, et regressioonsirge tõus k leitakse valemist

$$k = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

ja algordinaat b valemist

$$b = \bar{y} - k\bar{x}$$

Kui asendada algordinaat regressioonsirge võrrandisse, saame

$$\hat{y} = kx + \bar{y} - k\bar{x}$$

Kui selles võtta x väärtuseks aritmeetiline keskmine $\bar{x}$, saame

$$\hat{y} = k\bar{x} + \bar{y} - k\bar{x} = \bar{y}$$

st regressioonsirgel väärtusel $\bar{x}$ vastab väärtus $\bar{y}$ ehk regressioonsirge üheks punktiks on punkt koordinaatidega $(\bar{x}; \bar{y})$.

Leiame kahte tunnust - liblika tiibade siruulatust ja kaalu - iseloomustava regressioonsirge võrrandi. Kõigepealt leiame tõusu

$$k = \frac{431,2 - 23,78 \cdot 18}{1,78^2} = 0,997$$

ja seejärel algordinaadi

$$b = 18 - 0,997 \cdot 23,78 = -5,72$$

Seega on otsitav regressioonsirge võrrand

$$\hat{y} = 0,997x - 5,72$$

Joonestame uuesti korrelatsioonivälja ja sellesse leitud regressioonsirge. Sirge joonestamisel üheks punktiks on $P_1(\bar{x}; \bar{y}) = P_1(23,78; 18)$. Teise punkti abstsissiks võtame 27 ja arvutame vastava ordinaadi $\hat{y} = 0,997 \cdot 27 - 5,72 = 21,2$, st teine punkt on $P_2(27; 21,2)$. Kumbki punkt P_1 ja P_2 korrelatsioonivälja ei kuulu.

Laplace'i funktsiooni $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ väärtuste tabel

x	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998