



TODENNÄKÖISYYSJAKAUMISTA

Anabella Kivine

LuK-tutkielma
Huhtikuu 2026

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Tarkastajat:
Aleksi Winstén

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä

TURUN YLIOPISTO, Matematiikan ja tilastotieteen laitos

LuK-tutkielma

Pääaine: Sovellettu Matematiikka

Tekijä: Anabella Kivine

Otsikko: Todennäköisyysjakaumista

Ohjaaja: Aleksi Winstén

Sivumäärä: 19 sivua

Aika: Huhtikuu 2026

Tässä tutkielmassa käsitellään todennäköisyysjakaumia yleisellä tasolla. Tutkielman alkuun määritellään satunnaismuuttuja sekä siihen liittyviä käsitteitä. Tämän jälkeen todennäköisyysjakaumat jaetaan kahteen päätyyppiin: diskreetit ja jatkuvat jakaumat. Diskreetin jakauman tapauksessa syvennytään Bernoullin jakaumaan ja sen avulla johdettaviin jakaumiin. Jatkuvan jakauman tapauksessa esitellään jatkuva tasajakauma, jolle määritellään odotusarvo ja varianssi. Lopuksi esitetään esimerkit molemmista jakaumatyypeistä, diskreetille tapaukselle binomijakauma ja jatkuvalle normaalijakauma, ja esimerkkien jälkeen sovelletaan jakaumia.

Asiasanat: satunnaismuuttuja, todennäköisyysjakauma, diskreetti jakauma, jatkuva jakauma, odotusarvo.

Sisällys

1	Johdanto	1
2	Satunnaismuuttuja ja todennäköisyysjakauma	1
2.1	Diskreetti satunnaismuuttuja	2
2.2	Jatkuva satunnaismuuttuja	3
2.3	Odotusarvo ja varianssi	3
3	Todennäköisyysjakaumien luokittelu	7
3.1	Diskreetit jakaumat	8
3.2	Jatkuvat jakaumat	10
4	Binomijakauma diskreetin muuttujan mallina	13
5	Normaalijakauma jatkuvan muuttujan mallina	15
6	Yhteenveto	18

1 Johdanto

Todennäköisyysjakaumat ovat matemaattisia malleja, jotka kuvaavat satunnaismuuttujien arvojen todennäköisyyksiä. Mallien avulla voidaan havainnollistaa monimutkaista vaihtelua.

Tutkielmassa todennäköisyysjakaumat jaetaan kahteen päätyyppiin: diskreetteihin ja jatkuviin. Jako perustuu satunnaismuuttujan arvojoukkoon eli siihen, miten todennäköisyys on jakautunut arvojen kesken. Diskreetti jakauma kuvaa satunnaismuuttujaa, jonka arvojen joukko on numeroituva, eli joko äärellinen tai numeroituvasti ääretön. Tällöin jokaisella yksittäisellä arvolla on nollasta eroava pistetodennäköisyys. Jatkuva todennäköisyysjakauma liittyy muuttujaan, jonka arvojoukko on ylinumeroituva, kuten reaalilukuväli. Keskeistä on, että todennäköisyys on jakautunut jatkuvasti tietylle välille, ja yksittäisen pisteen todennäköisyys on nolla. [1][2]

Luvussa 2 tarkastellaan käsitteitä satunnaismuuttuja, diskreetti satunnaismuuttuja, jatkuva satunnaismuuttuja, pistetodennäköisyys, tiheysfunktio, kertymäfunktio, sekä odotusarvo ja varianssi. Odotusarvon ja varianssin perusominaisuudet todistetaan.

Tutkielmassa esitellään molemmista jakaumaluokista esimerkit luvuissa 4 ja 5. Diskreetin todennäköisyysjakauman esimerkkinä toimii binomijakauma. Jatkuvan todennäköisyysjakauman esimerkki on normaalijakauma, joka on tunnetuin jatkuva jakauma. Molemmille jakaumille määritellään odotusarvot ja varianssit. Binomijakaumasta esitetään sovellus, jossa käydään läpi yliopiston sisäänkirjautuneiden valmistumismäärien todennäköisyyksien laskeminen. Normaalijakauman sovelluksena esitetään suomalaisten naisten pituuksien jakautuminen ja sen tutkiminen.

Tämän tutkielman kirjoituksessa on hyödynnetty Gemini-tekoälytyökalua kielihuoltoon.

2 Satunnaismuuttuja ja todennäköisyysjakauma

Satunnaismuuttuja on tunnettu käsite todennäköisyyslaskennassa. Satunnaismuuttuja X on kuvaus perusjoukosta eli otosavaruudesta Ω reaalilukujen joukkoon \mathbb{R} , jota merkitään $X : \Omega \rightarrow \mathbb{R}$. Otosavaruus Ω koostuu alkeistapauksista ω , jotka ovat satunnaiskokeen yksittäisiä lopputuloksia. Tämä voidaan tulkita niin, että satunnaiskokeen alkeistapausten mittauksen tulos on satunnaismuuttujan arvo. Satunnaismuuttuja antaa alkeistapaukselle ω tietyn lukuarvon.

Määritelmä 1. Satunnaismuuttuja X on mitallinen kuvaus otosavaruudesta Ω reaaliluvuille \mathbb{R} . [1]

Huomautus 1. Satunnaismuuttuja voidaan tulkita mittausmenettelynä, jossa satunnaiskokeen tuloksena saatavalle lopputulomalle suoritetaan mittaus, jonka tulos satunnaismuuttujan arvo on.

Kun tarkastella satunnaismuuttujien välistä suhdetta, yksi keskeisin käsite on riippumattomuus. Jos satunnaismuuttujalla X ei ole vaikutusta satunnaismuuttujaan Y , niin sanotaan, että ne ovat riippumattomia.

Määritelmä 2. Olkoon X ja Y satunnaismuuttujia. Ne ovat riippumattomia jos ja vain jos

$$P(X = x, Y = y) = P(X = x)P(Y = y). \quad (1)$$

Riippumattomuus siis tarkoittaa, että tieto yhden satunnaismuuttujan arvosta ei muuta toisen satunnaismuuttujan arvojen todennäköisyyksiä. Käytännössä riippumattomuus tarkoittaa, että satunnaismuuttujien yhteispistetodennäköisyysfunktio on niiden marginaalijakaumien tulo eli yhtälö 1. Riippumattomuudesta seuraa, että muuttujien tulon odotusarvo on odotusarvojen tulo

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \quad (2)$$

ja niiden välinen kovarianssi on nolla. [2]

Satunnaismuuttujan avulla satunnaiskokeen alkeistapaukset, kuten kolikonheitot, voidaan esittää numeerisessa muodossa. Numeerisella muodolla todennäköisyyttä voidaan laskea summien ja integraalien avulla. Satunnaismuuttujan yksinkertaista muotoa käytetään odotusarvon ja varianssin määrittelyissä, jotta voidaan johtaa yleisiä tapauksia.

Määritelmä 3. Satunnaismuuttuja X on yksinkertainen, jos sen arvojoukko $\{X(\omega) : \omega \in \Omega\}$ on äärellinen. [1]

Huomautus 2. Yksinkertainen satunnaismuuttuja X voidaan kirjoittaa muodossa

$$X = \sum_{i=1}^n x_i 1_{A_i} = X^{-1}(\{x_i\}). \quad (3)$$

2.1 Diskreetti satunnaismuuttuja

Satunnaismuuttuja voi olla diskreetti tai jatkuva, riippuen sen arvojoukosta. Diskreetti satunnaismuuttuja saa yksittäisiä arvoja, joille voidaan määrittää pistetodennäköisyydet. Diskreettejä satunnaismuuttujia tarkastellaan pistetodennäköisyysfunktion ja summan avulla.

Määritelmä 4. Olkoon $X : \Omega \rightarrow \mathbb{R}$, missä \mathbb{R} on maalijoukko. Satunnaismuuttujan arvojoukko $\Omega_X = \{x_1, x_2, \dots\}$ on maalijoukon numeroituva osajoukko $\Omega_X \subseteq \mathbb{R}$. Tällöin diskreetin satunnaismuuttujan pistetodennäköisyysfunktio $p_X : \mathbb{R} \rightarrow [0, 1]$ on muotoa

$$p_X(x) = \begin{cases} P(X = x), & \text{kun } x \in \Omega_X \\ 0, & \text{muulloin,} \end{cases} \quad (4)$$

missä funktio $p_X(x)$ kuvaa yksittäisen arvon esiintymistodennäköisyyttä. [1]

Tyypillinen esimerkki diskreetistä satunnaismuuttujasta on kolikonheitto, jossa tarkastellaan klaavojen määrää. Satunnaiskokeita, joilla on kaksi mahdollista lopputulosta, yleensä "onnistuminen" tai "epäonnistuminen", kuvataan Bernoullin jakaumalla. Kolikonheitossa mahdolliset lopputulokset ovat "kruuna" tai "klaava". Bernoullin jakauma on monien diskreettien jakaumien perusta.

Määritelmä 5. Merkitään satunnaismuuttujaa X niin, että $X = 0$ on kruuna ja $X = 1$ klaava. Tällöin satunnaismuuttujan X pistetodennäköisyysfunktio on

$$p(x) = P(X = x) = \begin{cases} p, & \text{kun } x = 1 \\ 1 - p & \text{kun } x = 0, \end{cases} \quad (5)$$

missä $p \in [0, 1]$ on klaavan todennäköisyys. Tällöin X on Bernoullijakautunut. [1]

Koetta, kuten kolikonheittoa, jolla on onnistumistodennäköisyys p , kutsutaan Bernoullin kokeeksi. Kun suoritetaan tällaisia toisistaan riippumattomia Bernoullin kokeita n kertaa, huomataan, että kokeet noudattavat binomijakaumaa. Binomijakauma kertoo onnistumisten, kuten klaavojen, kokonaismäärän n toistoissa. Bernoullin jakauma on binomijakauman erikoistapaus, jossa $n = 1$. Binomijakaumaa, sen ominaisuuksia ja Bernoullin jakaumaa käsitellään lisää luvussa 4.

2.2 Jatkuva satunnaismuuttuja

Toisin kuin diskreetissä tapauksessa, jatkuvan satunnaismuuttujan todennäköisyys on jakautunut jatkuvasti, jolloin yksittäisen pisteen todennäköisyys on nolla. Tällöin todennäköisyyttä tarkastellaan tiheysfunktion ja integraalin avulla. Esimerkkinä jatkuvasta satunnaismuuttujasta voidaan pitää lämpötilaa.

Määritelmä 6. Satunnaismuuttuja X on jatkuva, jos sen kertymäfunktio F_X on jatkuva ja derivoituva. [2]

Huomautus 3. Määritelmästä 6 seuraa, että kertymäfunktio F_X ja tiheysfunktio f_X voidaan esittää toistensa avulla

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad (6)$$

$$f_X(x) = \frac{d}{dx} F_X(x). \quad (7)$$

Tällöin kertymäfunktio $F_X(x)$ antaa todennäköisyyden tapahtumalle $X \leq x$. Toisin sanoen kertymäfunktio kuvaa todennäköisyyden kertymistä arvoon x asti.

2.3 Odotusarvo ja varianssi

Määritelmä 7. Olkoon satunnaismuuttuja X yksinkertainen ja muodossa

$$X = \sum_{i=1}^n x_i 1_{A_i}, \quad A_i = X^{-1}(\{x_i\}). \quad (8)$$

Tällöin

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n x_i 1_{A_i}\right] = \sum_{i=1}^n x_i P(A_i) \quad (9)$$

on satunnaismuuttujan X odotusarvo. [1]

Odotusarvon perusominaisuuksia ovat lineaarisuus, monotonisuus, kolmioepäyhtälö sekä tulosääntö.

Lause 1. *Oletetaan, että satunnaismuuttujat X ja Y ovat yksinkertaisia ja että $a, b \in \mathbb{R}$. Tällöin odotusarvolle pätee*

1. *Lineaarisuus:* $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
2. *Monotonisuus:* Jos $X \leq Y$ melkein varmasti, niin $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
3. *Kolmioepäyhtälö:* $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
4. *Tulosääntö:* Jos X ja Y ovat riippumattomia, niin $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. [1]

Todistus. Oletetaan, että $X = \sum_{i=1}^n x_i 1_{A_i}$ ja $Y = \sum_{j=1}^n y_j 1_{B_j}$, missä A_i ja B_j ovat otosavaruuden Ω partitioita.

1. Todistetaan ensin skalaarilla kertominen aX käyttäen odotusarvon määritelmää (9).

$$\mathbb{E}[aX] = \mathbb{E}\left[\sum a x_i 1_{A_i}\right] = \sum a x_i P(A_i) = a \sum x_i P(A_i) = a\mathbb{E}[X]. \quad (10)$$

Todistetaan nyt odotusarvon lineaarisuus summan $X + Y$ tapauksessa. Olkoon $C_{ij} = A_i \cap B_j$ partition leikkaus. Tällöin summa voidaan kirjoittaa muodossa $X + Y = \sum_{i,j} (x_i + y_j) 1_{C_{ij}}$ (ks. huomautus 4). Odotusarvon määritelmän (9) nojalla saadaan

$$\mathbb{E}[X + Y] = \sum_{i,j} (x_i + y_j) P(C_{ij}) = \sum_{i,j} x_i P(C_{ij}) + \sum_{i,j} y_j P(C_{ij}). \quad (11)$$

Koska $\sum_j P(A_i \cap B_j) = P(A_i)$, saadaan

$$\sum_i x_i P(A_i) + \sum_j y_j P(B_j) = \mathbb{E}[X] + \mathbb{E}[Y], \quad (12)$$

jolloin lineaarisuus on todistettu.

Huomautus 4. Edellä esitettyssä todistuksessa summa $X + Y = \sum_{i,j} (x_i + y_j) 1_{C_{ij}}$ perustuu siihen, että $\{A_i\}$ ja $\{B_j\}$ ovat otosavaruuden partitioita. Koska joukko A_i voidaan kirjoittaa muodossa $A_i = \bigcup_j (A_i \cap B_j)$, voidaan vastaava indikaattorifunktio kirjoittaa muodossa $1_{A_i} = \sum_j 1_{C_{ij}}$. Tällöin satunnaismuuttuja X voidaan kirjoittaa muodossa $\sum_{i,j} x_i 1_{C_{ij}}$ ja samalla tavalla satunnaismuuttuja Y voidaan kirjoittaa muodossa $\sum_{i,j} y_j 1_{C_{ij}}$. Yhdistämällä nämä saadaan haluttu muoto ja additiivisuuden nojalla $\sum_{i,j} x_i P(C_{ij}) = \sum_i x_i P(A_i) = \mathbb{E}[X]$.

2. Oletetaan, että $X \leq Y$. Tällöin kaikilla $\omega \in \Omega$ pätee $Y(\omega) - X(\omega) \geq 0$ eli $Y - X$ saa ei-negatiivisia arvoja $y_j - x_i$. Odotusarvon määritelmän (9) nojalla

$$\mathbb{E}[Y - X] = \sum_{i,j} (y_j - x_i)P(A_i \cap B_j) \geq 0. \quad (13)$$

Odotusarvon lineaarisuuden perusteella $\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X] \geq 0$. Tästä seuraa $\mathbb{E}[Y] \geq \mathbb{E}[X]$, näin monotonisuus on todistettu.

3. Käytetään reaalilukujen kolmioepäyhtälön kaavaa $|a + b| \leq |a| + |b|$ ja monotonisuutta. Kaikille x_i pätee $x_i \leq |x_i|$ ja $-x_i \leq |x_i|$. Tällöin

$$\mathbb{E}[X] = \sum x_i P(A_i) \leq \sum |x_i| P(A_i) = \mathbb{E}[|X|]. \quad (14)$$

Samalla tavalla $-\mathbb{E}[X] = \mathbb{E}[-X] \leq \mathbb{E}[|X|]$. Koska termit $\mathbb{E}[X]$ ja $-\mathbb{E}[X]$ ovat pienempiä tai yhtä suuria kuin $\mathbb{E}[|X|]$, täytyy päteä $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$. Täten kolmioepäyhtälö on todistettu.

4. Oletetaan, että X ja Y ovat riippumattomia. Tällöin $A_i = \{X = x_i\}$ ja $B_j = \{Y = y_j\}$ ovat riippumattomia yhtälön 1 mukaan eli $P(A_i \cap B_j) = P(A_i)P(B_j)$. Tulon XY odotusarvo on

$$\mathbb{E}[XY] = \mathbb{E} \left[\left(\sum_i x_i 1_{A_i} \right) \left(\sum_j y_j 1_{B_j} \right) \right] = \mathbb{E} \left[\sum_{i,j} x_i y_j 1_{A_i \cap B_j} \right]. \quad (15)$$

Määritelmän (9) mukaan

$$\mathbb{E}[XY] = \sum_{i,j} x_i y_j P(A_i \cap B_j). \quad (16)$$

Riippumattomuuden avulla saadaan

$$\mathbb{E}[XY] = \sum_i \sum_j x_i y_j P(A_i)P(B_j) = \left(\sum_i x_i P(A_i) \right) \left(\sum_j y_j P(B_j) \right) = \mathbb{E}[X]\mathbb{E}[Y]. \quad (17)$$

Näin ollen odotusarvon perusominaisuudet on todistettu.

□

Odotusarvo on satunnaismuuttujan arvojen painotettu aritmeettinen keskiarvo. Varianssi on kuvaileva tunnusluku, joka kuvaa keskimääräistä satunnaismuuttujan arvojen vaihtelua.

Määritelmä 8. Yksinkertaisen satunnaismuuttujan X varianssi on

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2], \quad (18)$$

jossa $\mathbb{E}[X]$ on satunnaismuuttujan X odotusarvo. [1]

Varianssin perusominaisuuksia ovat ei-negatiivisuus, laskentakaava (19), siirto ja skaalaus sekä summan varianssi.

Lause 2. Oletetaan, että satunnaismuuttujat X ja $(X_i)_{i=1}^n$ ovat yksinkertaisia ja että $\alpha, \beta \in \mathbb{R}$. Tällöin

1. $Var(X) \geq 0$.
2. $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
3. $Var(\alpha X + \beta) = \alpha^2 Var(X)$.
4. $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)$,

jossa $Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]$. [1]

Todistus. Olkoon X ja X_i yksinkertaisia satunnaismuuttujia.

1. Määritelmän mukaan $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Koska $(X(\omega) - \mathbb{E}[X])^2 \geq 0$ kaikilla $\omega \in \Omega$, on kyseessä ei-negatiivisen yksinkertaisen satunnaismuuttujan odotusarvo. Tällöin odotusarvon monotonisuuden nojalla $Var(X) \geq 0$.
2. Koska odotusarvo $\mathbb{E}[X]$ on vakio, saadaan odotusarvon lineaarisuuden nojalla:

$$\begin{aligned}
 Var(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
 &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\
 &= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2] \\
 &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
 &= \mathbb{E}[X^2] - \mathbb{E}[X]^2.
 \end{aligned} \tag{19}$$

3. Olkoon $Y = \alpha X + \beta$. Tällöin $\mathbb{E}[Y] = \alpha \mathbb{E}[X] + \beta$. Sijoitetaan varianssin määritelmään (19)

$$\begin{aligned}
 Var(\alpha X + \beta) &= \mathbb{E}[((\alpha X + \beta) - (\alpha \mathbb{E}[X] + \beta))^2] \\
 &= \mathbb{E}[(\alpha X - \alpha \mathbb{E}[X])^2] \\
 &= \mathbb{E}[\alpha^2 (X - \mathbb{E}[X])^2] \\
 &= \alpha^2 \mathbb{E}[(X - \mathbb{E}[X])^2] = \alpha^2 Var(X).
 \end{aligned}$$

4. Käytetään summan neliöönkorotuskaavaa, jolloin saadaan

$$\begin{aligned}
 Var\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] \\
 &= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] + 2 \sum_{i < j} \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\
 &= \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j).
 \end{aligned}$$

□

3 Todennäköisyysjakaumien luokittelu

Satunnaismuuttujan todennäköisyysjakauma kuvaa todennäköisyyksien jakautumista sen mahdollisille arvoille. Todennäköisyysjakauma on todennäköisyysmitan ja satunnaismuuttujan käänteiskuvan yhdiste

$$\mu_X(A) := P(X^{-1}(A)). \quad (20)$$

Todennäköisyysjakaumille on yhteistä odotusarvo, varianssi, karakteristinen funktio ja todennäköisyydet generoiva funktio (PGF). Karakteristinen funktio, jota merkitään $\varphi_X(t) = \mathbb{E}[e^{itX}]$, on apufunktio, joka on olemassa kaikille satunnaismuuttujille. Todennäköisyydet generoiva funktio (PGF), jota merkitään $G_X(s) = \mathbb{E}[s^X]$, on erityisesti diskreeteille satunnaismuuttujille tarkoitettu funktio. Sitä hyödynnetään erityisesti ei-negatiivisia kokonaislukuarvoja saavien satunnaismuuttujien tarkastelussa. [7]

Satunnaismuuttujien tunnusluvut jaetaan tyypillisesti kahteen luokkaan, keskilukuihin ja hajontalukuihin. Keskilukuja ovat mediaani, moodi ja aiemmin mainittu odotusarvo. Mediaani on keskimäinen arvo, kun jakauman arvot laitetaan suuruusjärjestykseen. Diskreetissä tapauksessa moodi on todennäköisin arvo ja jatkuvassa tapauksessa moodin välittömässä läheisyydessä on suurin todennäköisyysmassa. Moodi on tällöin tiheysfunktion maksimi. Hajontalukuja ovat keskihajonta ja aiemmin mainittu varianssi. Keskihajonta σ mittaa arvojen etäisyyttä keskiarvosta.

Hajontaluku ja keskiluku liittyvät toisiinsa varianssin avulla. Varianssi mittaa keskimääräistä etäisyyttä odotusarvosta, jolloin odotusarvo on jakauman keskipiste, jonka ympärille varianssi määrittää havaintojen hajonnan. Pieni varianssi tarkoittaa arvojen pakkautumista lähelle odotusarvoa ja suuri varianssi tarkoittaa havaintojen levittäytymistä laajalle alueelle odotusarvon ympärille. [2]

On myös tunnuslukuja, jotka kuvaavat kahden satunnaismuuttujan välistä yhteyttä. Tällainen tunnusluku on kovarianssi.

Määritelmä 9. $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$.

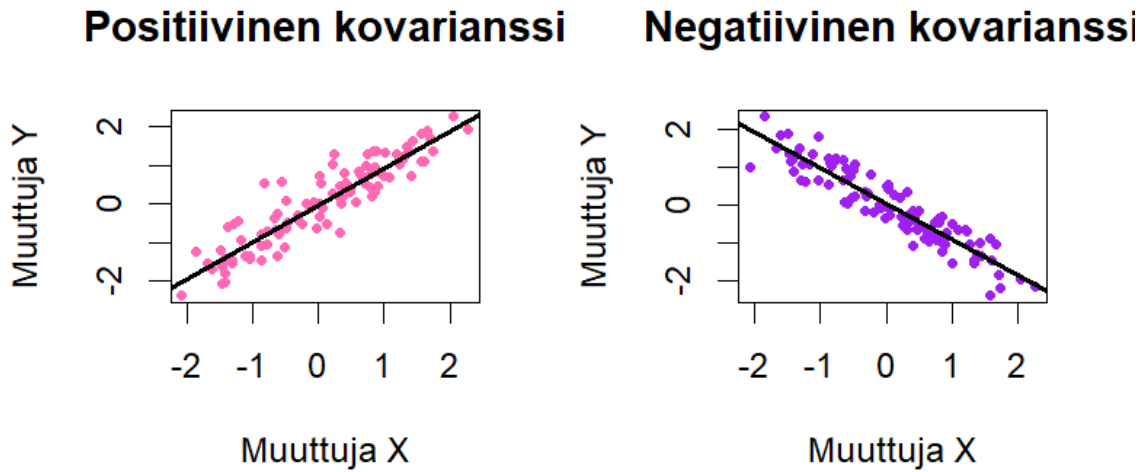
Varianssi on kovarianssin erikoistapaus, jossa muuttujaa verrataan itseensä. Olkoon $Y = X$, jolloin kovarianssin lausekkeesta saadaan suoraan varianssin määritelmä (8)

$$\text{Cov}(X, X) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))) \quad (21)$$

$$= \mathbb{E}((X - \mathbb{E}(X))^2) \quad (22)$$

$$= \text{Var}(X). \quad (23)$$

Kovarianssi voi saada positiivisia tai negatiivisia arvoja, toisin kuin varianssi, joka voi saada vain ei-negatiivisia arvoja. Kovarianssi siis mittaa, kuinka paljon X :n ja Y :n arvot muuttuvat yhdessä. Jos on todennäköistä, että molemmat muuttuvat samanaikaisesti, on kyseessä positiivinen kovarianssi. Jos taas näyttää, että toinen muuttuu suuremmaksi ja toinen pienemmäksi, niin kyseessä on negatiivinen kovarianssi, kuten kuvasta 1 nähdään. [2]



Kuva 1: Positiivinen ja negatiivinen kovarianssi

Satunnaismuuttujat jaetaan diskreetteihin ja jatkuviin niiden todennäköisyysjakaumien diskreettisuuden ja jatkuvuuden perusteella.

3.1 Diskreetit jakaumat

Diskreetti jakauma kuvaa tilannetta, jossa satunnaismuuttujan arvojoukko on numeroituva. Esimerkiksi nopanheiton silmäluku tai kolikonheitto ovat tyypillisiä diskreettejä tapauksia. Toisin kuin jatkuvassa jakaumassa, diskreetin satunnaismuuttujan pistetodennäköisyys voi olla nolasta eroava. Diskreetissä tapauksessa tapahtumien todennäköisyydet ja tunnusluvut, kuten odotusarvo, lasketaan summien avulla.

Määritelmä 10. Olkoon Ω otosavaruus ja X satunnaismuuttuja, jolle $X : \Omega \rightarrow \mathbb{R}$. Satunnaismuuttujan X todennäköisyysjakauma on diskreetti, jos sen arvojoukko on numeroituva. Tällöin jakauma voidaan määritellä pistetodennäköisyysfunktion avulla, joka liittyy jokaiseen satunnaismuuttujan saamaan arvoon x todennäköisyyden $P(X = x)$. [4]

Pistetodennäköisyysfunktion arvot ovat välillä $[0, 1]$. Koska todennäköisyysmitan P tulee toteuttaa ehto $P(\Omega) = 1$, niin diskreetissä tapauksessa $0 \leq P(x) \leq 1$. Tällöin diskreetin todennäköisyysjakauman ehtona on, että kaikkien alkeistapahtumien $P(x)$ todennäköisyyksien arvojen summa tulee olla täsmälleen 1.

$$\sum_x P(X = x) = 1. \quad (24)$$

Todennäköisyys tapahtumalle E , joka on otosavaruuden osajoukko $E \subseteq \Omega$ määritellään summana

$$P(E) = \sum_{\Omega \in E} P(x). \quad (25)$$

Diskreetit jakaumat, kuten Bernoullin jakauma, binomijakauma ja geometrinen jakauma määritellään Bernoullin kokeen avulla. Kokeessa on kaksi mahdollista lopputulosta satunnaismuuttujalle X , jotka ovat $X = 1$ onnistuu ja $X = 0$ epäonnistuu.

Määritelmä 11. Olkoon $p \in (0, 1)$ Bernoullin kokeen onnistumisen todennäköisyys. Satunnaismuuttujan X jakauma on Bernoullin jakauma, jota merkitään $X \sim \text{Bernoulli}(p)$, kun

$$P(X = k) = p^k(1 - p)^{1-k}, \quad k \in \{0, 1\}. \quad (26)$$

Lause 3. *Olkoon $X \sim \text{Bernoulli}(p)$. Tällöin odotusarvo ja varianssi ovat*

$$\begin{aligned} E(X) &= p \\ \text{Var}(X) &= p(1 - p). \end{aligned}$$

Satunnaismuuttujan ja jakauman välistä yhteyttä kuvataan yleisesti \sim symbolilla. Merkintä on muotoa

$$X \sim \{Jakauma\}. \quad (27)$$

Vasemmalla puolella on tarkasteltava satunnaismuuttuja, tässä esimerkissä X , ja oikealla puolella on jakauman nimi sekä siihen liittyvät parametrit, tässä esimerkissä $\text{Bernoulli}(p)$. Oikea puoli määrittää jakauman tarkan muodon.

Todistus. Odotusarvo lasketaan summana, jossa mahdolliset arvot k kerrotaan kaavan (26) mukaisilla todennäköisyyksillä $P(X = k)$.

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1). \quad (28)$$

Koska onnistumisen todennäköisyys on p ja epäonnistumisen $1 - p$, saadaan

$$\begin{aligned} E(X) &= 0 \cdot (1 - p) + 1 \cdot p \\ E(X) &= p \end{aligned}$$

Näin ollen odotusarvo on todistettu.

Varianssi lasketaan kaavalla $\text{Var}(X) = E(X^2) - [E(X)]^2$. Lasketaan ensin oikean puolen ensimmäinen termi.

$$\begin{aligned} E(X^2) &= 0^2 \cdot P(X = 0) + 1^2 \cdot P(X = 1) \\ E(X^2) &= 0 \cdot (1 - p) + 1 \cdot p \\ E(X^2) &= p \end{aligned} \quad (29)$$

Sijoitetaan (29) varianssin kaavaan (8), jolloin saadaan

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p), \quad (30)$$

joka on varianssi. □

Kuten aiemmin mainittiin, binomijakauma ja geometrinen jakauma määritellään riippumattomien Bernoullin kokeiden avulla. Binomijakauma kuvaa onnistumisten kokonaismäärää, kun suoritetaan n määrä riippumattomia kokeita, joiden välillä on yhtälön (1) mukainen riippumattomuus. Satunnaismuuttujien jonoa sanotaan riippumattomaksi ja samoin jakautuneeksi, jos satunnaismuuttujat ovat riippumattomia ja niillä on sama todennäköisyysjakauma.

Määritelmä 12. Olkoon X_1, \dots, X_n riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että $X_i \sim \text{Bernoulli}(p)$ kaikilla $i = 1, \dots, n$. Tällöin satunnaismuuttujien summa on

$$X = \sum_{i=1}^n X_i, \quad (31)$$

joka noudattaa binomijakaumaa ja merkitään $X \sim \text{Bin}(n, p)$. [6]

Huomautus 5. Määritelmän (31) nojalla binomijakauman odotusarvo voidaan johtaa lineaarisuuden avulla

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p = np. \quad (32)$$

Binomijakaumaan syvennyttään enemmän luvussa 4.

Geometrinen jakauma määrittää epäonnistuneiden kokeiden lukumäärän ennen ensimmäistä onnistumista. Geometrisessa jakaumassa suoritetaan siis riippumattomia Bernoullin kokeita, kunnes saavutetaan ensimmäinen onnistuminen.

Määritelmä 13. Olkoon $p \in (0, 1)$ onnistumisten todennäköisyys Bernoullin koeksessa. Satunnaismuuttuja X noudattaa geometrista jakaumaa, jota merkitään $X \sim \text{Geom}(p)$, jos se kuvaa epäonnistumisten määrää ennen ensimmäistä onnistumista. Tällöin sen pistetodennäköisyysfunktio on

$$P(X = k) = (1 - p)^k p, \quad k \in \{0, 1, 2, \dots\}. \quad (33)$$

Jokainen termi $(1 - p)$ vastaa yhtä epäonnistunutta Bernoullin koetta ja termi p vastaa ensimmäistä onnistumista, joka päättää kokeen. [6]

3.2 Jatkuvat jakaumat

Toisin kuin diskreetin jakauman tapauksessa, jatkuva satunnaismuuttuja voi saada kaikki mahdolliset arvot tietyltä reaalilukuväliltä. Jatkuvan jakauman keskeisin piirre on, että yksittäisen pisteen todennäköisyys on nolla eli $P(X = x) = 0$ kaikilla $x \in \mathbb{R}$. Tämän vuoksi todennäköisyyksiä lasketaan tiheysfunktion avulla käyttämällä integraaleja.

Määritelmä 14. Satunnaismuuttujaa $X : \Omega \rightarrow \mathbb{R}$ kutsutaan jatkuvaksi, jos on olemassa ei-negatiivinen funktio $f : \mathbb{R} \rightarrow [0, \infty)$, siten, että kaikilla $a, b \in \mathbb{R} (a \leq b)$ pätee

$$P(a \leq X \leq b) = \int_a^b f(x) dx. \quad (34)$$

Tällöin funktiota $f(x)$ kutsutaan satunnaismuuttujan X tiheysfunktiksi ja sen on toteutettava ehto

$$\int_{-\infty}^{\infty} f(x)dx = 1. \quad (35)$$

Yleisesti tapahtuman $E \subseteq \mathbb{R}$ todennäköisyys on määritelty integraalina $P(X \in E) = \int_E f(x)dx$. [4]

Tiheysfunktio määritellään usein kertymäfunktion $F(x) = P(X \leq x)$ kautta. Tiheysfunktio saadaan kertymäfunktion derivaattana

$$f(x) = \frac{d}{dx}F(x). \quad (36)$$

Kuten aiemmin luvussa 2 määriteltiin, jatkuva jakauma perustuu kertymäfunktion ja tiheysfunktion yhteyteen. Todennäköisyysjakauma on yleiskäsite, joka kuvaa todennäköisyyden jakautumista eri arvoille. Jatkuvassa tapauksessa jakaumaa voidaan kuvata tiheysfunktioilla. Kertymäfunktio $F(x)$ kertoo todennäköisyyden kertymisen pisteeseen x asti ja se määritellään tiheysfunktion avulla integraalina

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (37)$$

Jatkuvan jakauman yksinkertaisin esimerkki on jatkuva tasajakauma. Todennäköisyys jakautuu tasaisesti tietyllä välillä ja siksi tiheysfunktio on vakio tällä välillä.

Määritelmä 15. Olkoon a ja b vakioita, joille $a < b$. Satunnaismuuttujan X jakauma on jatkuva tasajakauma, jota merkitään $X \sim \text{Uni}[a, b]$, kun sen tiheysfunktio on

$$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b. \quad (38)$$

Lause 4. *Tasajakautuneella satunnaismuuttujalla $X \sim \text{Uni}[a, b]$ on odotusarvo ja varianssi*

$$\begin{aligned} E(X) &= \frac{1}{2}(a+b) \\ \text{Var}(X) &= \frac{1}{12}(b-a)^2. \end{aligned}$$

Todistus. Tasajakauman tiheysfunktio on

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b, \quad (39)$$

ja $f(x) = 0$ muualla. Jatkuvan satunnaismuuttujan odotusarvo määritellään integraalina

$$\int_{-\infty}^{\infty} x f(x)dx \quad (40)$$

Sijoitetaan tiheysfunktio:

$$E(X) = \int_a^b x \cdot \frac{1}{b-a} dx. \quad (41)$$

Koska $\frac{1}{b-a}$ on vakio, se voidaan siirtää integraalin ulkopuolelle

$$E(X) = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left(\frac{1}{2} b^2 - \frac{1}{2} a^2 \right)$$

$$E(X) = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right)$$

Käyteytään kaavaa $b^2 - a^2 = (b-a)(b+a)$:

$$E(X) = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2} = \frac{1}{2}(a+b). \quad (42)$$

Seuraavaksi varianssi. Käytetään varianssin kaavaa 19. Lasketaan ensin oikean puolen ensimmäinen termi

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{1}{3} b^3 - \frac{1}{3} a^3 \right)$$

$$E(X^2) = \frac{b^3 - a^3}{3(b-a)}.$$

Jaetaan tekijöihin kaavalla $b^3 - a^3 = (b-a)(b^2 + ab + a^2)$:

$$E(X^2) = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3} \quad (43)$$

Sijoitetaan tulokset varianssin kaavaan 19

$$Var(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2} \right)^2$$

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

Lavennetaan nimittäjään 12

$$Var(X) = \frac{4b^2 + 4ab + 4a^2 - (3a^2 + 6ab + 3b^2)}{12}$$

$$Var(X) = \frac{a^2 - 2ab + b^2}{12}.$$

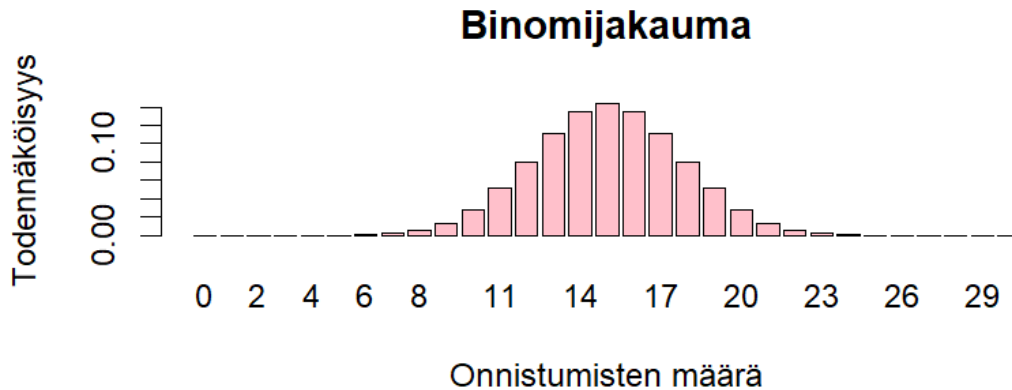
Koska $a^2 - 2ab + b^2 = (b-a)^2$ saadaan

$$Var(X) = \frac{(b-a)^2}{12} = \frac{1}{12}(b-a)^2. \quad (44)$$

□

4 Binomijakauma diskreetin muuttujan mallina

Binomijakauma on diskreetti todennäköisyysjakauma, joka saa numeroituvan määrän arvoja, kuten kokonaislukuja. Binomijakauman yksittäisellä tapahtumalla on tasan kaksi mahdollista tulosta. Esimerkiksi tehtaiden laadunvalvonnassa voidaan käyttää binomijakaumaa viallisten tuotteiden todennäköisyyksien laskemisessa, jolloin on mahdollista saada tulokseksi vain joko "ehjä" tai "viallinen".



Kuva 2: Binomijakauman kuvaaja parametrilla $p=0.5$

Määritelmä 16. Satunnaismuuttujan sanotaan olevan binomijakautunut, kun $n \in \{1, 2, \dots\}$ ja $p \in [0, 1]$, jolloin sitä merkitään $X \sim \text{Bin}(n, p)$. Tällöin pistetodennäköisyysfunktio on

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n, \quad (45)$$

missä termiä $\frac{n!}{x!(n-x)!}$ kutsutaan binomikertoimeksi. [5]

Lause 5. Binomijakauman odotusarvo on tällöin

$$\mu = E(X) = np, \quad (46)$$

jolloin saadaan

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = np(1-p), \quad (47)$$

joka on binomijakauman varianssi. [5]

Todistus. Olkoon $X \sim \text{Bin}(n, p)$. Kuten aiemmin arvattiin, X voidaan esittää riippumattomien ja samoin jakautuneiden Bernoulli-muuttujien summana $X = \sum_{i=1}^n X_i$, missä $X_i \sim \text{Bernoulli}(p)$. Tällöin yksittäisen muuttujan odotusarvo on

$$\mathbb{E}[X_i] = 1 \cdot p + 0 \cdot (1-p) = p. \quad (48)$$

Odotusarvon lineaarisuuden (4) nojalla saadaan

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p = np. \quad (49)$$

Täten odotusarvo on todistettu. Tarkastellaan sitten varianssia. Yksittäisen muuttujan varianssi on

$$\text{Var}(X_i) = \mathbb{E}(X_i^2) - [\mathbb{E}(X_i)]^2 = p - p^2 = p(1 - p). \quad (50)$$

Koska muuttujat X_i ovat riippumattomia, summan varianssi on varianssien summa

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n p(1 - p) = np(1 - p), \quad (51)$$

joka on varianssi. □

Esimerkki 1. Yliopistoon sisäänkirjautuvan valmistumistodennäköisyys on 70%. Valitaan satunnaisesti 20 opiskelijaa syksyllä 2013 aloittavista opiskelijoista. Olkoon X valitusta kahdestakymmenestä opiskelijasta valmistuvien määrä. Oletetaan että satunnaismuuttuja X on binomijakautunut, jolloin todennäköisyys, että joukossa valmistuvien määrä on k , on

$$P(X = k) = \frac{20!}{k!(20 - k)!} 0.70^k (1 - 0.70)^{20-k}, \quad k = 0, \dots, 20. \quad (52)$$

Todennäköisyysjakauman avulla voidaan laskea erilaisten tapahtumien todennäköisyyksiä. Tarkastellaan tässä muutamaa erilaista tapausta. Tämä esimerkki ja ensimmäinen tapaus ovat lähteestä [2] s. 51-52.

- Tapaus 1: "Kaikki valmistuvat" ($k = 20$)

$$P(X = 20) = 0.70^{20} = 8.0 \cdot 10^{-4}. \quad (53)$$

Todennäköisyys P on 0,0008.

- Tapaus 2: "Kukaan ei valmistu" ($k = 0$)

$$P(X = 0) = 0.30^{20} = 3.4869 \cdot 10^{-11}. \quad (54)$$

Todennäköisyys P on 0,000000000034868.

- Tapaus 3: "Alle puolet valmistuvat" ($k < 10$)

$$P(X < 10) = P(X \leq 9) = \sum_{k=0}^9 \binom{20}{k} 0.70^k 0.30^{20-k} = 0,0171\dots \quad (55)$$

Todennäköisyys P on $\approx 0,017$.

Tapauksen 3 yhtälö (55) on esimerkki binomijakauman kertymäfunktioista, jonka yleinen muoto on

$$F(k; n, p) = P(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}, \quad (56)$$

missä $\lfloor k \rfloor$ on luvun k lattia, eli suurin kokonaisluku, joka on pienempi tai yhtäsuuri kuin k . [8]

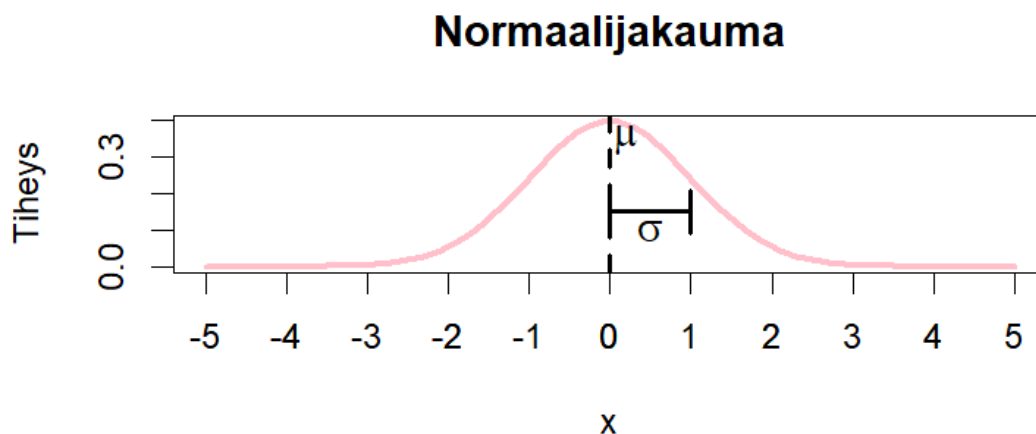
Kuten yleisestä kaavasta (56) näkee, kertymäfunktion laskeminen on erittäin työläs prosessi. Tällaisia tapauksia varten tuloksia voidaan approksimoida numeerisesti käyttämällä normaalijakaumaa $N(\mu, \sigma^2)$, missä $\mu = np = 14$ ja $\sigma = \sqrt{np(1-p)} \approx 2,05$. Jatkuvuuskorjausta käyttäen saadaan

$$P(X \leq 9) \approx P\left(Z \leq \frac{9.5 - 14}{2.05}\right) \approx 0,14 \quad , \quad (57)$$

joka on lähellä saatua arvoa tapauksessa 3 (55).

5 Normaalijakauma jatkuvan muuttujan mallina

Normaalijakauma (myös tunnettu nimellä Gaussin jakauma) on jatkuva todennäköisyysjakauma, joka on yksi tunnetuimmista ja käytetyimmistä jakaumista. Sitä kutsutaan usein kellokäyräksi sen symmetrisen muodon takia. Normaalijakauma siis nousee huippukohtaansa, joka on odotusarvo μ ja laskee samalla tavalla.



Kuva 3: Normaalijakauman kuvaaja, jossa $\mu = 0$ on odotusarvo ja $\sigma = 1$ on keskihajonta.

Lause 6. Normaalijakauman tiheysfunktio $f_X(x)$ on jatkuva kaikilla $x \in \mathbb{R}$.

Todistus. Tarkastellaan normaalijakauman tiheysfunktioita

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (58)$$

ja osoitetaan, että funktio on jatkuva kaikilla $x \in \mathbb{R}$. Tarkastellaan funktiota osissa.

Olkoon eksponentti $g(x) = -\frac{1}{2\sigma^2}(x-\mu)^2$. Funktio $g(x)$ on toisen asteen polynomi, joka on jatkuva kaikilla $x \in \mathbb{R}$. Myös eksponenttifunktio $h(u) = e^u$ on jatkuva kaikkialla määrittelyjoukossaan \mathbb{R} .

Analyysin perussääntöjen nojalla yhdistetty funktio $h(u)$ on jatkuva, sillä kahden funktion yhdistetty funktio on jatkuva. Kerroin $\frac{1}{\sqrt{2\pi\sigma^2}}$ on vakio ja vakiokertoimen ja jatkuvan funktion tulo on aina jatkuva. Koska tiheysfunktio $f_X(x)$ on jatkuvan vakion ja jatkuvan yhdistetyn funktion tulo, se on jatkuva kaikilla reaaliarvoilla. \square

Huomautus 6. Analyysin peruslauseen $F_X(x) = \int_{-\infty}^x f_X(t)dt$ nojalla tiheysfunktion $f_X(x)$ jatkuvuus takaa kertymäfunktion $F_X(x)$ jatkuvuuden.

Lause 7. *Olkoon $f_X(x)$ normaalijakauman tiheysfunktio. Tällöin*

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \quad (59)$$

toteuttaa tiheysfunktion ehdon. [3]

Todistus. Merkitään

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx. \quad (60)$$

Olkoon $z = \frac{x-\mu}{\sigma}$, jolloin $dx = \sigma dz$. Saadaan

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz. \quad (61)$$

Lasketaan integraalin neliö I^2

$$I^2 = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) \quad (62)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{z^2+y^2}{2}} dzdy. \quad (63)$$

Käytetään napakoordinaatistoa ja sijoitetaan $z = r \cos \theta$, $y = r \sin \theta$ ja $dzdy = r drd\theta$:

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r drd\theta. \quad (64)$$

Sijoitetaan $u = \frac{r^2}{2}$, jolloin $du = r dr$:

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} 1 d\theta = \frac{2\pi}{2\pi} = 1. \quad (65)$$

Koska $I^2 = 1$ ja tiheysfunktio on positiivinen ($I > 0$), täytyy olla $I = 1$. \square

Lause 8. Olkoon $X \sim \text{Normal}(\mu, \sigma^2)$. Tällöin jakauman odotusarvo ja varianssi ovat

$$\begin{aligned}\mathbb{E}(X) &= \mu \\ \text{Var}(X) &= \sigma^2.\end{aligned}$$

Todistus. Osoitetaan ensimmäiseksi odotusarvo. Lauseen 7 nojalla, tehdään sijoitus $z = \frac{x-\mu}{\sigma}$, jolloin $x = \sigma z + \mu$ ja $dx = \sigma dz$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} (\sigma z + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (66)$$

$$= \sigma \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz. \quad (67)$$

Ensimmäinen integraali on 0, sillä integroitava funktio $ze^{-\frac{z^2}{2}}$ on symmetrinen origon suhteen. Jälkimmäinen integraali todistettiin olevan 1 (6).

Seuraavaksi varianssi $\text{Var}(X) = \sigma^2$. Määritelmän (8) mukaan.

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (68)$$

Käytetään jälleen sijoitusta $z = \frac{x-\mu}{\sigma}$, jolloin $(x - \mu)^2 = \sigma^2 z^2$.

$$\text{Var}(X) = \int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz. \quad (69)$$

Käytetään osittaisintegroitua ja merkitään $u = z$ ja $dv = ze^{-z^2/2} dz$ (jolloin $v = -e^{-z^2/2}$).

$$\int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz = \left[-ze^{-\frac{z^2}{2}} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-e^{-\frac{z^2}{2}}) dz \quad (70)$$

$$= 0 + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \quad (71)$$

$$= \sqrt{2\pi}. \quad (72)$$

Sijoittamalla (72) varianssin määritelmään (8) saadaan $\text{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2$. \square

Esimerkki 2. Suomalaisten naisten pituus noudattaa likimain normaalijakaumaa, jossa odotusarvo on $\mu = 165\text{cm}$ ja keskihajonta on $\sigma = 6.5\text{cm}$, jolloin varianssi on $\sigma^2 = 42.25$. Olkoon satunnaismuuttuja X satunnaisesti valitun naisen pituus, jolloin $X \sim \text{Normal}(165, 42.25)$. Tällöin pituuden tiheysfunktio on

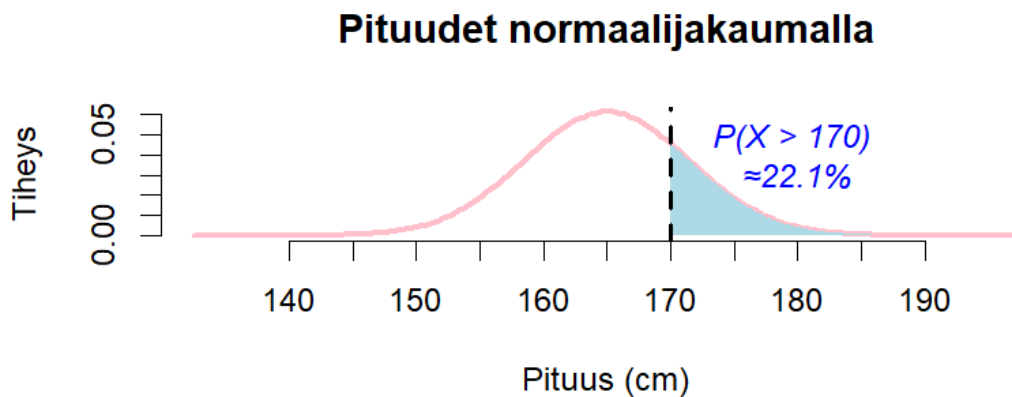
$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot 42.25}} e^{-\frac{(x-165)^2}{2 \cdot 42.25}}. \quad (73)$$

Jakauman avulla voidaan laskea erilaisia todennäköisyyksiä, kuten satunnainen nainen on yli 170cm pitkä

$$P(X > 170) = P(Z > \frac{170 - 165}{6.5}) \approx P(Z > 0.77) \approx 0,2206 \quad (74)$$

Eli noin 22,1% suomalaisista naisista on yli 170 cm pitkiä.

Huomautus 7. Kuten aiemmin luvussa 4 (1) todettiin, kertymäfunktion arvojen laskeminen on työlästä. Tämän vuoksi yhtälössä (74) hyödynnetään taulukkoarvoja.



Kuva 4: Suomalaisen naisten pituudet normaalijakaumalla, kun $\mu = 165$ ja $\sigma = 6.5$. Sininen osuus havainnollistaa, kuinka suuri osa naisista on yli 170cm pitkiä.

6 Yhteenveto

Tässä tutkielmassa tarkasteltiin todennäköisyysjakaumia keskittyen diskreetteihin ja jatkuviin todennäköisyysjakaumiin. Tutkielmassa määriteltiin odotusarvon ja varianssin käsitteet ja todistettiin niiden keskeisimmät perusominaisuudet. Jakaumiin syvennyttiin käymällä läpi diskreetin jakauman pistetodennäköisyysfunktio ja jatkuvan jakauman tiheysfunktio.

Molemmista jakaumatyypeistä esitettiin esimerkit. Diskreetin todennäköisyysjakauman esimerkkinä oli binomijakauma, joka johdettiin Bernoullin jakaumasta. Bernoullin jakauman vaikutuksiin syvennyttiin binomijakauman ohella. Binomijakaumaa sovellettiin esimerkkiin, jossa tutkittiin yliopiston sisäänkirjautuvien valmistumismääriä. Jatkuvan todennäköisyysjakauman esimerkkinä oli normaalijakauma, jonka sovelluksessa tutkittiin suomalaisten naisten pituuksia.

Tutkielman kirjoitusprosessin aikana keskeisin oivallus oli se, kuinka läheisiä diskreetit ja jatkuvat jakaumat ovat. Vaikka ne määritellään eri tavoin ja niiden todennäköisyyksille on aivan omat kaavansa, ne eivät kuitenkaan ole irrallaan toisistaan. Esimerkiksi luvussa 4 huomattiin, miten binomijakaumaa käsittelevä esimerkki käytti loppujen lopuksi normaalijakauman approksimointia laskemisen helpottamiseen.

Todennäköisyyslaskenta on valtavan suuri aihealue ja siitä löytyy paljon tietoa. Jos jakaumista haluaisi oppia lisää, luonnollisin suunta olisi kysyä miksi monet maailman ilmiöt noudattavat normaalijakaumaa ja miten se on edes mahdollista. Myös ohjelmointikielillä, kuten R:llä, voi tuottaa visuaalisesti havainnollistavia kuvia, joiden avulla on helpompi hahmottaa teoriaa.

Viitteet

- [1] J. Lempa: *Todennäköisyysteoria*.
- [2] H. Pesonen: *Todennäköisyyslaskennan peruskurssi*.
- [3] Nick T. Thomopolous: *Probability Distributions With Truncated, Log and Bivariate Extensions*. Springer, 2018.
- [4] D. Joyce: *Axioms Of Probability Math 217 Probability and Statistics*. 2014.
- [5] R. V. Hogg, S. A. Klugman: *Loss Distributions*. John Wiley & Sons, Inc, 1984.
- [6] A. S. Wagaman, R. P. Dobrow: *Probability with applications and R*. John Wiley & Sons, Inc 2021.
- [7] T. A. Severini: *Elements of Distribution Theory*. Cambridge University Press, 2005.
- [8] Wikipedia: *Binomial Distribution*. https://en.wikipedia.org/wiki/Binomial_distribution Viitattu 19.03.2026.