TURUN YLIOPISTO

Poisson-jakauman keskeiset ominaisuudet ja riippumattomat Poisson-satunnaismuuttujat

Heidi Rintamäki

LuK-tutkielma
Huhtikuu 2026

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Tarkastajat:
Aleksi Winstén

TURUN YLIOPISTO, Matematiikan ja tilastotieteen laitos

LuK-tutkielma
Pääaine: Sovellettu matematiikka
Tekijä: Heidi Rintamäki
Otsikko: Poisson-jakauman keskeiset ominaisuudet ja riippumattomat Poisson-satunnaismuuttujat
Ohjaaja: Aleksi Winstén
Sivumäärä: 13 sivua
Aika: Huhtikuu 2026

Tässä LuK-tutkielmassa johdetaan Poisson-jakauma kahdella eri menetelmällä, binomijakauman raja-arvona ja Poisson-prosessin kautta. Huomiota kiinnitetään Poisson-jakauman keskeisiin ominaisuuksiin, tarkemmin sen odotusarvoon ja varianssiin. Lisäksi tarkastellaan kahden riippumattoman satunnaismuuttujan X ja Y summajakaumaa sekä kovarianssia ja korrelaatiota.

Asiasanat: Poisson-jakauma, binomijakauman raja-arvo, Poisson-prosessi, odotusarvo, varianssi

Sisällys

1 Johdanto

Satunnaisten tapahtumien lukumäärän mallintaminen on keskeinen osa todennäköisyyslaskentaa ja tilastollista mallinnusta. Monissa luonnontieteellisissä, teknisissä ja yhteiskunnallisissa sovelluksissa tarkastellaan tilanteita, joissa yksittäiset tapahtumat ovat harvinaisia, toisistaan riippumattomia ja syntyvät satunnaisesti ajan kuluessa.

Tällaisia ilmiöitä esiintyy esimerkiksi puhelinliikenteessä, teollisuuden prosesseissa ja tieliikenteessä. Puhelinliikenteestä voidaan tutkia esimerksi montako puhelua saapuu tiettyyn paikkaan tietyllä aikavälillä. Erilaisissa asiakaspalvelutilanteissa saapuvien puheluiden määrä tiettynä ajankohtana voi olla oleellinen tieto, jotta pystytään määrittämään resurssitarpeet. Teollisuudessa tuotantolinjalla voi esiintyä viallisia tuotteita ja on hyvä tietää kuinka todennäköistä se on, jotta voidaan esimerkiksi arvioida riskejä. Taas tieliikenteessä voidaan tarkastella liikenneonnettomuuksien määrää tietyllä alueella eri ajanjaksoina. Nämä tiedot auttavat esimerkiksi liikenneturvallisuuden suunnittelussa tai yleisesti onnettomuusennusteissa.

Tälläisten tapahtumien mallintaminen ja ratkaiseminen ei ole aina niin suoraviivaista. Vaikka voidaan olettaa, että nämä kaikki tapahtumat ovat satunnaisia ja riippumattomia niin esimerkiksi liikenneonnettomuuksiin voi vaikuttaa ruuhka-ajat. Samoin myös puhelinliikenteeseen voi vaikuttaa kellonajat ja viallisten tuotteiden määrään voi vaikuttaa esimerkiksi teollisuuslaitteen rikkoutuminen. [6]

Näihin kaikkiin tilanteisiin voidaan hyödyntää Poisson-prosesseja, koska voidaan kuitenkin olettaa, että nämä tilanteet ovat satunnaisesti riippumattomia ja harvinaisia. Poisson-jakauma on diskreetti todennäköisyysjakauma, jota käytetään mallintamaan satunnaisten tapahtumien lukumäärää aika- tai mittayksikössä. Poisson-jakauman parametrina toimii $\lambda > 0$, joka kuvaa keskimääräistä esiintymisnopeutta aika- tai mittayksikköä kohden. [2]

Poisson-jakauman on keksinyt nimen mukaisesti Siméon-Denis Poisson (1781-1840). Hän on oivaltanut monia nykypäivän merkityksellisiä saavutuksia matematiikan ja fysiikan aloille sekä hän toimi Ranskassa. Poisson-jakauma esiintyy vuonna 1837 hänen koko tuonannossaan *Recherches sur la Probablité des Jugments* vain yhden kerran sivulla 206 [10]. Jakauma esiintyy Poissonin teoksessa binommijakauman raja-arvona tilanteessa, jossa kokeiden lukumäärä on suuri ja yksittäisen tapahtuman todennäköisyys on hyvin pieni. [3, 6]

Poisson-jakauman merkitys perustuu sen matemaattisiin ominaisuuksiin sekä siihen, että se voidaan johtaa useista toisistaan riippumattomista lähtökohdista. Tässä tutkielmassa esitetään kaksi johtamistapaa, joista toinen johtaminen esitetään binomijakauman raja-arvona ja toinen tapa esitetään Poisson-prosessin kautta.

Tämän työn tavoitteena on esitellä miten Poisson-jakaumaan päädytään eli sen johtamistavat, Poisson-jakauman keskeiset ominaisuudet sekä tarkastella myös kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan summaa. Tämän tutkielman kirjoituksessa on hyödynnetty ChatGPT-tekoälytyökalua kielenhuoltoon.

2 Poisson-jakauman johtaminen

Klassinen tapa johtaa Poisson-jakauma on tarkastella binomijakauman rajajakaumaa, jonka nojalla voidaan ajatella Poisson-jakauma sopivaksi harvinaisen tapahtuman esiintymiskertojen lukumääräksi suuressa populaatiossa [3]. Poisson-jakauma voidaan myös johtaa Poisson-prosessin oletuksista, jotka pohjautuvat hyvin lyhyiden aikavälien mallintamiseen. [5, 4]

2.1 Binomijakauman raja-arvo

Tässä alaluvussa havainnollistetaan, miten Poisson-jakauma syntyy binomijakauman raja-arvona harvinaisten tapahtumien tapauksessa. Alaluku pohjautuu kolmeen eri lähteeseen [2, 3, 6].

Olkoon yksittäinen satunnaiskoe Bernoulli-koe, jossa onnistumistodennäköisyys on p ja epäonnistumisen todennäköisyys $1-p$. Kun koe toistetaan n kertaa toisistaan riippumattomasti ja onnistumisten lukumäärää merkitään satunnaismuuttujalla X, saadaan

$$X \sim \mathrm{Bin}(n, p),$$

ja pistetodennäköisyys on

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n. \tag{1}$$

Seuraava tulos kuvaa binomijakauman raja-arvoa.

Väite. Olkoon $X_n \sim \mathrm{Bin}(n, p_n)$ ja oletetaan, että

$$p_n \to 0, \qquad n \to \infty,$$

siten, että

$$\lim_{n\to\infty} np_n = \lambda > 0. \tag{2}$$

Tällöin jokaiselle $k = 0, 1, 2, \dots$ pätee, että

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}. \tag{3}$$

Toisin sanoen binomijakauman pistetodennäköisyydet suppenevat Poisson-jakauman pistetodennäköisyyksiin parametrilla λ.

Todistus. Merkitään $\lambda_n = np_n$ ja k on vakio, jolloin $\lambda_n = \lim_{n\to\infty} \lambda$. Tällöin

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}. \tag{4}$$

Binomikerroin voidaan kirjoittaa muodossa

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!}. \tag{5}$$

Näin saadaan

$$P(X_n = k) = \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k}. \tag{6}$$

Voidaan merkitä $p_n = \frac{\lambda}{n}$ ja $p_n^k = (\frac{\lambda}{n})^k = \frac{\lambda^k}{n^k}$. Tällöin

$$= \frac{n(n-1)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot (1-\frac{\lambda}{n})^n \cdot (1-\frac{\lambda}{n})^{-k}. \tag{7}$$

Voidaan tarkastella tulon tekijöitä yksitellen tulon raja-arvosäännön nojalla, koska jokaisella tekijällä on olemassa äärellinen raja-arvo. Kaksi ensimmäistä tulon tekijää voidaan ryhmitellä uudelleen, jolloin ensimmäinen tulon tekijä lähestyy arvoa 1 ja toinen on vakio

$$\lim_{n\to\infty} (\frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!}) = 1 \cdot \frac{\lambda^k}{k!}. \tag{8}$$

Seuraavana tarkastellaan kahta viimeistä tulon tekijää yksitellen, jolloin

$$\lim_{n\to\infty} (1-\frac{\lambda}{n})^n = e^{-\lambda}. \tag{9}$$

Lisäksi

$$\lim_{n\to\infty} (1-\frac{\lambda}{n})^{-k} = 1, \tag{10}$$

joten näin ollen

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{11}$$

Siis X_n suppenee melkein varmasti kohti poisson jakauman pistetodennäköisyysfunktiota parametrilla λ. $\square$

2.2 Poisson-prosessi

Tässä luvussa johdetaan Poisson-jakauma Poisson-prosessin aksioomista. Aluksi määritellään Poisson-prosessi ja sitä koskevat oletukset, minkä jälkeen näistä oletuksista johdetaan Poisson-jakauman differentiaaliyhtälöt ja ratkaistaan ne.

2.2.1 Poisson-prosessin aksioomat

Poisson-prosessi $X(t)$ on stokastinen prosessi, jossa tapahtumat esiintyvät satunnaisesti ajan kuluessa siten, että niiden väliset ajat eivät ole säännöllisiä. Prosessille ominaista on, että tapahtumien keskimääräinen esiintymisnopeus pysyy vakiona. Prosessia kuvataan aksioomien avulla, jotka heijastavat harvinaisten ja toisistaan riippumattomien tapahtumien perusominaisuuksia. Olkoon $0 \leq t \in \mathbb{R}$ jatkuva aikamuuttuja, joka kuvaa prosessia. [4, 6] Poisson-prosessin aksioomat ovat seuraavat:

1. Jos $[t_1, t_2]$ ja $[t_3, t_4]$ ovat erillisiä aikavälejä niin niillä tapahtuvien tapahtumien lukumäärät ovat toisistaan riippumattomia:

$$X(t_2) - X(t_1) \text{ ja } X(t_4) - X(t_3) \text{ ovat riippumattomia.} \tag{12}$$

2. Tapahtumien esiintymisen todennäköisyys lyhyellä aikavälillä $[t, t+h]$ on suunnilleen verrannollinen muuttujan $h > 0$ kanssa, verrannollisuussuhteen ollessa $\lambda > 0$. Tarkemmin sanottuna:

$$\lim_{h \to 0} \frac{P(X(t+h) - X(t) = 1)}{h} = \lambda. \tag{13}$$

3. Kahden tapahtuman esiintymisen todennäköisyys lyhyellä aikavälillä $[t, t + h]$ on paljon pienempi kuin yhden tapahtuman todennäköisyys. Tarkemmin sanottuna:

$$\lim_{h \to 0} \frac{P(X(t+h) - X(t) \geq 2)}{h} = 0. \tag{14}$$

Merkitään $P_0(t)$ todennäköisyyttä, että aikavälillä $[0, t]$ ei tapahdu yhtään tapahtumaa. Jos pidemmällä aikavälillä $[0, t+h]$ ei tapahdu tapahtumia, niin tapahtumia ei esiinny myöskään erillisillä aikaväleillä $[0, t]$ ja $[t, t+h]$. Nämä aikavälit ovat erillisiä ja Poisson-prosessin inkrementtien riippumattomuus tarkoittaa, että näillä väleillä tapahtumattomuus on riippumaton. Siksi

$$P_0(t+h) = P_0(t)P_0(h). \tag{15}$$

Derivaatan määritelmän perusteella $P_0'(t)$ on

$$P_0'(t) = \lim_{h \to 0} \frac{P_0(t+h) - P_0(t)}{h} = P_0(t) \lim_{h \to 0} \frac{P_0(h) - 1}{h}. \tag{16}$$

Lisäksi todennäköisyys $P_0(h) = P(X(t+h) - X(t) = 0) = 1 - P(X(t+h) - X(t) \geq 1)$, joten

$$P_0'(t) = -P_0(t) \lim_{h \to 0} \frac{P(X(t+h) - X(t) \geq 1)}{h}. \tag{17}$$

Aksiooman 2. ja 3. nojalla viimeinen raja-arvo on yhtä kuin λ. Nyt saadaan yhtälö

$$P_0'(t) = -\lambda P_0(t). \tag{18}$$

Tämä on ensimmäisen kertaluvun differentiaaliyhtälö, jonka yleinen ratkaisu on muotoa $P_0(t) = Ae^{-\lambda t}$. Alkuarvosta $P_0(0) = 1$ seuraa $A = 1$, joten

$$P_0(t) = e^{-\lambda t}. \tag{19}$$

Edellä johdettu tulos (19) osoittaa, että nollan tapahtuman todennäköisyys vähenee eksponentiaalisesti ajan funktiona. Seuraavaksi johdetaan yleisemmin kaikkien tapahtumamäärien todennäköisyydet.

2.2.2 Poisson-jakauman johtaminen differentiaaliyhtälön avulla

Tässä alaluvussa tavoitteena on osoittaa induktiolla, että todennäköisyys sille, että aikavälillä $[0, t]$ tapahtuu täsmälleen n tapahtumaa, on muotoa

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{20}$$

Tätä varten johdetaan ensin differentiaaliyhtälö, joka kuvaa todennäköisyyksien $P_n(t)$ ajallista kehitystä. Todistetaan induktiolla, että $P_n(t)$ pätee kaikilla luvuilla $n \in \mathbb{N}$

Merkitään $P_n(t)$ todennäköisyyttä, että aikavälillä $[0, t]$ tapahtuu täsmälleen n tapahtumaa. Tarkastellaan lyhyttä aikaväliä h. Tällöin todennäköisyydet yhdelle ja nollalle tapahtumalle ovat aksioomien nojalla

$$P_1(h) = \lambda h, \quad P_0(h) = 1 - \lambda h. \tag{21}$$

Tarkastellaan todennäköisyyttä $P_n(t + h)$. Aikavälillä $[t, t + h]$ voi tapahtua $0, 1, 2$ tai useampia tapahtumia. Poisson-prosessin aksioomien mukaan kuitenkin kahden tai useamman tapahtuman todennäköisyys hyvin lyhyellä aikavälillä $[t, t + h]$ on huomattavasti pienempi kuin yhden tapahtuman todennäköisyys. Aksiooma 3 sanoo, että kahden tai useamman tapahtuman todennäköisyys on niin pieni, että se voidaan jättää huomiotta, kun tarkastellaan raja-arvoa $h \to 0$. Näin saadaan rekursio

$$P_n(t + h) = P_n(t)(1 - \lambda h) + P_{n-1}(t)\lambda h. \tag{22}$$

Vähentämällä yhtälöstä $P_n(t)$ ja jakamalla puolittain parametreilla h sekä siirtymällä raja-arvoon $h \to 0$ saadaan differentiaaliyhtälö

$$\frac{dP_n(t)}{dt} + \lambda P_n(t) = \lambda P_{n-1}(t). \tag{23}$$

Yhdessä yhtälön $P_0'(t) = -\lambda P_0(t)$ kanssa tämä muodostaa differentiaaliyhtälöryhmän, joka kuvaa Poisson-prosessin ajallista kehitystä. Ratkaisu johdetaan induktiolla. Perustapauksessa $n = 0$ on jo saatu

$$P_0(t) = e^{-\lambda t}. \tag{24}$$

Tarkastellaan seuraavaksi tapausta $n = 1$. Differentiaaliyhtälö saa muodon

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda P_0(t). \tag{25}$$

Sijoittamalla $P_0(t) = e^{-\lambda t}$ ja kertomalla yhtälö (25) integroivalla tekijällä $e^{\lambda t}$ saadaan

$$\frac{d}{dt}\left(e^{\lambda t} P_1(t)\right) = \lambda. \tag{26}$$

Integroimalla seuraa

$$e^{\lambda t} P_1(t) = \lambda t + C. \tag{27}$$

Alkuarvosta $P_1(0) = 0$ saadaan $C = 0$, joten

$$P_1(t) = \lambda t e^{-\lambda t}, \tag{28}$$

mikä on väitteen mukainen muoto tapauksessa $n = 1$. Tehdään nyt induktio-oletus: oletetaan, että jollakin $n \in \mathbb{N}$ pätee

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{29}$$

Tarkastellaan differentiaaliyhtälöä indeksille $n + 1$

$$\frac{dP_{n+1}(t)}{dt} + \lambda P_{n+1}(t) = \lambda P_n(t). \tag{30}$$

Kertomalla integroivalla tekijällä $e^{\lambda t}$ saadaan

$$\frac{d}{dt}\left(e^{\lambda t} P_{n+1}(t)\right) = e^{\lambda t} \lambda P_n(t). \tag{31}$$

Induktio-oletusta käyttäen yhtälön (31) oikea puoli voidaan kirjoittaa muotoon

$$e^{\lambda t} \lambda \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda \frac{(\lambda t)^n}{n!}. \tag{32}$$

Siis

$$\frac{d}{dt}\left(e^{\lambda t} P_{n+1}(t)\right) = \lambda \frac{(\lambda t)^n}{n!}. \tag{33}$$

Integroimalla saadaan

$$e^{\lambda t} P_{n+1}(t) = \int \lambda \frac{(\lambda t)^n}{n!}\, dt = \frac{(\lambda t)^{n+1}}{(n+1)!} + C. \tag{34}$$

Alkuarvosta $P_{n+1}(0) = 0$ seuraa $C = 0$, joten

$$P_{n+1}(t) = \frac{(\lambda t)^{n+1}}{(n+1)!} e^{-\lambda t}. \tag{35}$$

Täten väite pätee luvulle $n + 1$, ja induktioperiaatteen nojalla kaava

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{36}$$

pätee kaikille $n \in \mathbb{N}$. Lopputuloksena on saatu Poisson-jakauman tunnettu muoto eli Poisson-jakauman todennäköisyysfunktio. Tämä johtaminen osoittaa, että Poisson-jakauma seuraa suoraan Poisson-prosessin aksioomista ja niitä vastaavasta differentiaaliyhtälöryhmästä. [5]

3 Poisson-jakauman keskeiset ominaisuudet

Tässä luvussa keskitytään Poisson-jakauman keskeisiin ominaisuuksiin, jotka ovat olennaisia jakauman tulkinnan ja sovellusten kannalta. Erityistä huomiota kiinnitetään odotusarvoon, varianssiin sekä jakauman käyttäytymiseen parametrin λ eri arvoilla. Aluksi voidaan johtaa Poisson-jakauman odotusarvo suoraan määritelmästä. [2]

3.1 Poisson-jakauman odotusarvo ja varianssi

Lause 3.1. *$X \sim Poisson(\lambda) \Rightarrow \mathbb{E}[X] = \lambda$.*

Todistus. Aloitetaan odotusarvon määritelmästä

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!}. \tag{37}$$

Huomataan, että termi $k = 0$ on nolla, joten summa voidaan aloittaa indeksistä $k = 1$. Tämän jälkeen voidaan sieventää kertomalla ja jakamalla λ

$$\mathbb{E}[X] = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}. \tag{38}$$

Tehdään indeksin vaihto $m = k - 1$, jolloin summa muuttuu muotoon

$$\sum_{m=0}^{\infty} \frac{\lambda^m}{m!}. \tag{39}$$

Tämä on eksponenttifunktion sarjakehitelmä, joten saadaan

$$\sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{\lambda}. \tag{40}$$

Sijoittamalla tämä takaisin saadaan

$$\mathbb{E}[X] = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda. \tag{41}$$

□

Seuraavaksi voidaan todistaa varianssin kaava, jotta saadaan laskettua Poisson-jakauman varianssi. Varianssi saadaan odotusarvon nojalla varianssin kaavasta.

Lause 3.2. *Satunnaismuuttujan X varianssi voidaan ilmaista muodossa*

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \tag{42}$$

Todistus. Voidaan laskea varianssi hyödyntämällä odotusarvon lineaarisuutta. Saadaan

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] && (43)\\
&= \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] && (44)\\
&= \mathbb{E}[X^2] - \mathbb{E}[2\mathbb{E}[X]X - \mathbb{E}[X]^2] && (45)\\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 && (46)\\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2. && (47)
\end{aligned}$$

□

Seuraavaksi voidaan laskea Poisson-jakauman varianssi käyttämällä lausetta 3.2.

Lause 3.3. $X \sim Poisson(\lambda) \Rightarrow \mathrm{Var}(X) = \lambda$.

Todistus. Jotta voidaan laskea Poisson-jakauman varianssi, kirjoitetaan termi X^2 uudelleen:

$$X^2 = X(X-1) + X. \tag{48}$$

Sijoittamalla tämä varianssin peruskaavaan (47), saadaan

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[X(X-1) + X] - \mathbb{E}[X]^2 &(49)\\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2. &(50) \end{aligned}$$

Nyt voidaan laskea Poisson-jakauman varianssi hyödyntämällä tätä muotoa. Lasketaan ensiksi

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!}. \tag{51}$$

Voidaan aloittaa summa $k = 2$, sillä termit $k = 0$ ja $k = 1$ ovat nollia. Saadaan

$$\mathbb{E}[X(X-1)] = e^{-\lambda}\sum_{k=2}^{\infty} k(k-1)\,\frac{\lambda^k}{k!}. \tag{52}$$

Koska $\frac{k(k-1)}{k!} = \frac{k(k-1)}{k(k-1)(k-2)!} = \frac{1}{(k-2)!}$, niin seuraavaksi saadaan

$$\mathbb{E}[X(X-1)] = e^{-\lambda}\sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}. \tag{53}$$

Tehdään indeksinvaihto $j = k - 2$, josta tulee

$$\mathbb{E}[X(X-1)] = e^{-\lambda}\sum_{j=0}^{\infty} \frac{\lambda^{j+2}}{j!} = e^{-\lambda}\lambda^2\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}. \tag{54}$$

Tällöin

$$\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}. \tag{55}$$

Eli saadaan

$$\mathbb{E}[X(X-1)] = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2. \tag{56}$$

Sijoitetaan saatu tulos varianssin kaavaan

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2. \tag{57}$$

Tästä saadaan Poisson-jakauman varianssi, joka on

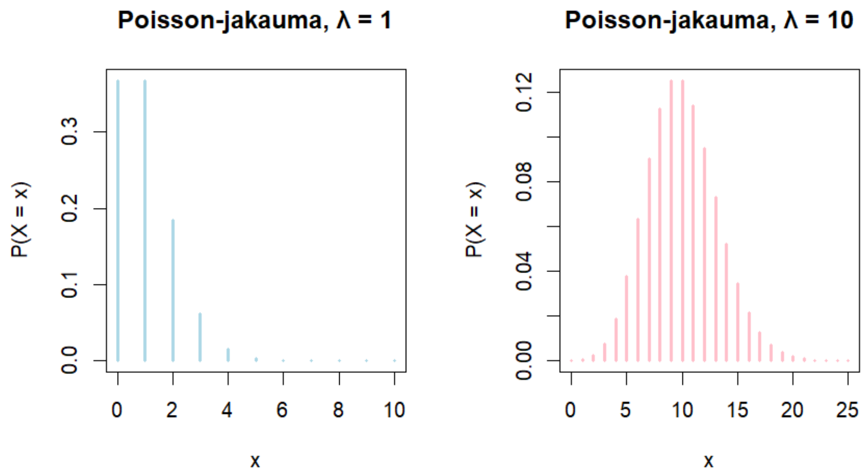
$$\mathrm{Var}(X) = \lambda. \tag{58}$$

□

Odotusarvon ja varianssin yhtäsuuruus on Poisson-jakauman tunnusomainen piirre ja erottaa sen monista muista diskreeteistä jakaumista. Keskihajonta kasvaa neliöjuurena parametrin mukana

$$\sigma = \sqrt{\lambda}, \tag{59}$$

mikä vaikuttaa suoraan jakauman leveyteen ja hajontaan. Kuvasta 1 huomataan, että pienillä λ -arvoilla jakauma on kapea ja selvästi oikealle vino, kun taas suuremmilla λ -arvoilla hajonta kasvaa ja jakauma alkaa muistuttamaan normaalijakaumaa. Poisson-jakauma voidaan tulkita binomijakauman rajatapauksena, kuten luvussa 2.1 johdettiin. Tällöin muuttuja voidaan esittää Bernoulli-muuttujien summana. Kun parametrin arvo kasvaa, summassa on paljon pieniä osatekijöitä, jolloin keskeisen raja-arvolauseen seurauksena jakauma alkaa muistuttaa normaalijakaumaa. [6]



Kuva 1: Poisson-jakauman muoto pienillä ja suurilla λ-arvoilla.

3.2 Kirjoitusvirheiden määrä tekstissä

Poisson-jakaumaa käytetään laajasti erilaisissa sovelluksissa, niin kuin luvussa 1 kerrottiin. Yksi tyypillinen esimerkki on vielä kirjoitusvirheiden lukumäärä tekstissä. Tällaisessa tilanteissa Poisson-jakauma on yksinkertainen, mutta tehokas tapa selvittää todennäköisyys.

Esimerkki 3.4. Oletetaan, että tämän tutkielman yhdellä sivulla olevien satunnaisten kirjoitusvirheiden määrä noudattaa Poisson-jakaumaa parametrilla $\lambda = \frac{1}{2}$. Parametri tarkoittaa, että keskimäärin yhdellä sivulla esiintyy puoli satunnaista kirjoitusvirhettä, eli noin yksi satunnainen virhe kahdella sivulla. Tämä vastaa Poisson-jakauman määritelmää, jossa parametri kuvaa odotusarvoa eli tapahtumien keskimääräistä määrää tarkasteltavalla alueella. Lasketaan todennäköisyys sille, että tällä sivulla on ainakin yksi satunnainen virhe.

Jos X merkitsee virheiden lukumäärää tällä sivulla, saadaan

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/2} \approx 0.395 \tag{60}$$

Satunnaisia kirjoitusvirheitä voidaan pitää Poisson-malliin sopivina tapahtumina, koska kirjoitusvirheet täyttävät Poisson-prosessin keskeiset ominaisuudet. Yhden kirjoitusvirheen todennäköisyys lyhyellä tekstiosuudella on hyvin pieni. Tämä vastaa Poisson-prosessin harvinaisten tapahtumien oletusta eli aksioomaa 2. Kahden tai useamman kirjoitusvirheen esiintyminen samalla tekstiosuudella on huomattavasti epätodennäköisempää kuin yhden kirjoitusvirheen esiintyminen, niinkuin aksiooma 3 sanoo. Kirjoitusvirheet syntyvät tyypillisesti toisistaan riippumatta, esimerkiksi näppäilyvirhe tai huolimattomuus. Virheen esiintyminen yhdessä kohdassa ei siis vaikuta virheen todennäköisyyteen muualla tekstissä, mikä vastaa Poisson-prosessin inkrementtien riippumattomuutta. [7]

Kun yhden muuttujan Poisson-jakauman matemaattinen perusta on esitetty, voidaan seuraavaksi tarkastella kahden riippumattoman Poisson-satunnaismuuttujan yhteistä käyttäytymistä. Tällöin kiinnostuksen kohteena on se, miten Poisson-jakauman ominaisuudet säilyvät tai muuttuvat, kun tarkastellaan kahta satunnaismuuttujaa samanaikaisesti.

4 Riippumattomat Poisson-satunnaismuuttujat

Tässä luvussa tarkastellaan kahta Poisson-jakautunutta riippumatonta satunnaismuuttujaa. Kahden riippumattoman Poisson-satunnaismuuttujan yhteistodennäköisyysfunktio mahdollistaa muun muassa todennäköisyyksien laskemisen eri yhdistelmille sekä summamuuttujan $X + Y$ analysoinnin. Tässä luvussa tarkastellaan esimerkkiä tästä tilanteesta. Tarkastellaan myös kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan kovarianssia ja korrelaatiota. Ensiksi kuitenkin määritellään riippumattomuus. Luku 4 pohjautuu hyvin paljon Pekka Tuomisen kirjaan *Todennäköisyyslaskenta 1* [2].

Määritelmä 4.1. X ja Y ovat riippumattomia, jos

$$P(X \cap Y) = P(X)P(Y). \tag{61}$$

Seuraavaksi voidaan määritellä yhteistodennäköisyysfunktio ja tarkastella esimerkkiä.

Määritelmä 4.2. Kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan yhteistodennäköisyysfunktio on

$$f(x, y) = P(X = i, Y = j) = \frac{\lambda_1^i e^{-\lambda_1}}{i!} \frac{\lambda_2^j e^{-\lambda_2}}{j!} = P(X = i)P(Y = j), \tag{62}$$

mikä on kahden riippumattoman Poisson pistetodennäköisyysfunktion tulo. [1]

4.1 Kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan summajakauma

Esitetään todennäköisyysmitan additiivisuus ennen kuin esitetään esimerkkiä tilanteesta, missä tarkastellaan kahden riippumattoman poisson-jakautuneen satunnaismuuttujan summajakaumaa.

Määritelmä 4.3. Jos $X_i \in \mathcal{F}(i = 1, 2, ..., n)$ ja $X_i \cap X_j = \emptyset$ $(i \neq j)$, niin

$$P(\bigcup_{i=1}^{n} X_i) = \sum_{i=1}^{n} P(X_i). \tag{63}$$

Esimerkki 4.4. Olkoon $X \perp Y$, $X \sim \text{Poisson}(\lambda_1)$ ja $Y \sim \text{Poisson}(\lambda_2)$. Tarkastellaan summan $X + Y$ jakaumaa.
Merkitään $P\{X = i\}$ ja $P\{Y = j\}$, missä $i, j \in \mathbb{N}$. Tapahtuma $\{X + Y = k\} = \{\omega \in \Omega : X(\omega) + Y(\omega) = k\}$ voidaan esittää erillisenä yhdisteenä

$$\{X + Y = k\} = \bigcup_{i=0}^{k} \{X = i\} \cap \{Y = k - i\}. \tag{64}$$

Koska joukot $\{X = i\} \cap \{Y = k - i\}$ ovat toisensa poissulkevia, todennäköisyysmitan additiivisuuden 4.3 nojalla

$$P\{X + Y = k\} = \sum_{i=0}^{k} P\{X = i, Y = k - i\}. \tag{65}$$

Riippumattomuusoletus $X \perp Y$ tarkoittaa, että $P\{X = i, Y = k - i\} = P\{X = i\} P\{Y = k - i\}$ kaikilla $i = 0, 1, ..., k$, joten

$$P\{X + Y = k\} = \sum_{i=0}^{k} P\{X = i\} P\{Y = k - i\}. \tag{66}$$

Saatu summa on diskreettien todennäkösyysjakaumien konvoluutio. Kahden riippumattoman diskreetin satunnaismuuttujan summajakauma saadaan niiden jakaumien konvoluutiona. Tämä summa esiintyy myös potenssisarjojen kertolaskussa, Cauchyn tulona.
Poisson-oletus tarkoittaa, että

$$P\{X = i\} = e^{-\lambda_1} \frac{\lambda_1^i}{i!}, \qquad P\{Y = k - i\} = e^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k - i)!}. \tag{67}$$

Kun nämä sijoitetaan yhtälöön (66), saadaan

$$P\{X + Y = k\} = \sum_{i=0}^{k} e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k - i)!} \tag{68}$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \tag{69}$$

Summa on binomilauseen mukainen, joten saadaan

$$P\{X + Y = k\} = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}, \qquad k \in \mathbb{N}. \tag{70}$$

Siis $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

4.2 Kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan kovarianssi ja korrelaatio

Seuraavana voidaan laskea näiden kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan X ja Y kovarianssi ja korrelaatio. Poisson-jakauman varianssi laskettiin aiemmin luvussa 3 lauseessa 3.3 ja se on $\mathrm{Var}(X) = \lambda_1$. Aluksi voidaan määritellä kovarianssi ja korrelaatio.

Lause 4.5. *Kahden riippumattoman satunnaismuuttujan X ja Y kovarianssi on*

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \tag{71}$$

Todistus. Koska X ja Y ovat riippumattomia, myös satunnaismuuttujat $X - \mathbb{E}[X]$ ja $Y - \mathbb{E}[Y]$ ovat riippumattomia. Riippumattomien satunnaismuuttujien tulojen odotusarvo hajoaa tuloksi:

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]]\,\mathbb{E}[Y - \mathbb{E}[Y]]. \tag{72}$$

Koska $\mathbb{E}[X - \mathbb{E}[X]] = 0$ ja $\mathbb{E}[Y - \mathbb{E}[Y]] = 0$, saadaan

$$\mathrm{Cov}(X, Y) = 0. \tag{73}$$

□

Määritelmä 4.6. Kahden satunnaismuuttujan X ja Y korrelaatio voidaan merkitä

$$\rho_{i,j} = \frac{\mathrm{Cov}(X_i, Y_j)}{\sigma_i \sigma_j}. \tag{74}$$

Esimerkki 4.7. $X \sim \mathrm{Poisson}(\lambda_1)$ ja $Y \sim \mathrm{Poisson}(\lambda_2)$ ovat riippumattomat. Lasketaan satunnaismuuttujien X ja $Z = X + Y$ kovarianssi ja korrelaatio.

$$\begin{aligned}
\mathrm{Cov}(X, Z) &= \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] && (75)\\
&= \mathbb{E}[X(X + Y)] - \mathbb{E}[X]\mathbb{E}[X + Y] && (76)\\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] && (77)\\
&= \mathrm{Var}(X) + \mathrm{Cov}(X, Y) && (78)
\end{aligned}$$

Kovarianssi $\mathrm{Cov}(X, Z)$ yhtyy satunnaismuuttujan X varianssiin λ_1. Käytetään määritelmää 4.6, jolloin saadaan korrelaatioksi

$$\rho = \frac{\lambda_1}{\sqrt{\lambda_1(\lambda_1 + \lambda_2)}} = \sqrt{\frac{\lambda_1}{\lambda_1 + \lambda_2}}. \tag{79}$$

Näin on saatu esitettyä kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan X ja Y kovarianssi ja korrelaatio. [9]

5 Yhteenveto

Poisson-jakauma on tehokas todennäköisyyslaskennan työkalu erilaisissa harvinaisissa ja satunnaisissa tapahtumissa. Tässsä LuK-tutkielmassa esitettiin, miten Poisson-jakaumaan päädytään binomijakauman raja-arvona sekä Poisson-prosessin kautta. Poisson-prosessin aksioomia voidaan hyödyntää erilaisissa tapahtumissa niin kuin esimerkissä 3.4 tutustuttiin kirjoitusvirheiden määrään tekstissä. Tutkielmassa esitettiin Poisson-jakauman keskeisiä ominaisuuksia, joita oli odotusarvo ja varianssi. Lisäksi esitettiin kahden riippumattoman Poisson-jakautuneen satunnaismuuttujan summajakaumaa ja niiden välistä kovarianssia sekä korrelaatiota. Nämä tulokset havainnolistivat esimerkiksi riippumattomuusoletuksen merkitystä. Ne ovat keskeisiä esimerkiksi monissa sovelluksissa, joissa yhdistyy kaksi Poisson-jakautunutta satunnaismuuttujaa.

Poisson-jakaumaan liittyy myös muita keskeisiä ominaisuuksia, kuten todennäköisyysgeneroiva funktio, jonka avulla voidaan johtaa jakauman momentteja ja tarkastella erilaisia rajajakaumia. Näitä ominaisuuksia ei kuitenkaan käsitelty tässä LuK-tutkielmassa, koska työn tavoitteena oli keskittyä Poisson-jakauman johtamiseen ja Poisson-prosessin perusominaisuuksiin. Todennäköisyysgeneroivien funktioiden käsittely olisi edellyttänyt laajempaa teoreettista taustaa, joka olisi vienyt tutkielman painopistettä pois sen varsinaisesta tarkoituksesta. Näihin asioihin pystyy tutustumaan esimerkiksi lähteistä [1] ja [6].

Viitteet

[1] S. Kocherlakota ja K. Kocherlakota: *Bivariate Discrete Distributions.* Marcel Dekker, New York, 1992.

[2] P. Tuominen: *Todennäköisyyslaskenta I.* Limes ry, Helsinki, 1993.

[3] Stephen M. Stigler: *Poisson on the Poisson Distribution.* Statistics & Probability Letters, 1982, 33–35.

[4] D. Joyce: *The Poisson Process.* Math 217 Probability and Statistics, Clark University, Fall 2014.

[5] Glen Cowan: *Derivation of the Poisson Distribution.* RHUL Physics, 1 December 2009.

[6] Frank A. Haight: *Handbook of the Poisson Distribution.* Wiley, 1967

[7] Sheldon Ross: *A First Course in Probability.* Macmillan, New York, 1976.

[8] H. Pesonen: *Todennäköisyyslaskennan peruskurssi.* Turun Yliopisto

[9] J. Lempa ja H. Saarinen: *Todennäköisyyslaskennan jatkokurssi.* Turun Yliopisto

[10] S.-D. Poisson: *Recherches sur la Probablité des Jugments.* Pariisi, 1837