Poisson-prosessit

Heli Taattola

Matematiikan pro gradu

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2015

Tiivistelmä: Heli Taattola *Poisson-prosessit* (engl. *Poisson processes*), matematiikan pro gradu -tutkielma, 47 s., Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, kevät 2015.

Tämän tutkielman tarkoituksena on tutustua erilaisiin Poisson-prosesseihin, joita käytetään muun muassa vakuutusmatematiikassa, jonotusteoriassa, rahoituksessa sekä kun mallinnetaan aikaa, kunnes jotakin tapahtuu. Tutkielmassa tutustutaan Poisson-prosessien tärkeimpiin jakauman tunnuslukuihin ja ominaisuuksiin todistuksineen sekä esimerkkien muodossa nähdään mitä hyötyä ominaisuuksista käytännössä on.

Ensimmäisenä tutustutaan homogeeniseen ja epähomogeeniseen Poisson-prosessiin sekä niiden ominaisuuksiin. Esimerkeissä lasketaan eri käytännön ongelmia ja lisäksi lasketaan tunnettu odottamiseen liittyvä paradoksi, joka tunnetaan muun muassa nimellä liftarin paradoksi ja josta tässä käytetään nimeä odottajan paradoksi.

Toisena tutustutaan painotettuun Poisson-prosessiin ja sen ominaisuuksiin. Huomataan myös, että sillä on osittain samoja ominaisuuksia kuin epähomogeenisella Poisson-prosessilla, mutta erojakin on. Lisäksi tutustutaan millaisissa tilanteissa kannattaisi käyttää painotettua prosessia tavallisen epähomogeenisen prosessin sijaan.

Viimeisenä tutustutaan yhdistettyyn painotettuun Poisson-prosessiin sekä sen erikoistapaukseen Cramér-Lundbergin malliin ja näiden ominaisuuksiin. Erityisesti Cramér-Lundbergin malli on paljon käytetty riskiteoriassa ja sitä soveltaen lasketaan lopuksi palovakuutusvahinkoesimerkki, jonka aineisto on simuloitu.

Avainsanoja: Poisson-prosessi, homogeeninen Poisson-prosessi, odottajan paradoksi, painotettu Poisson-prosessi, yhdistetty painotettu Poisson-prosessi, Cramér-Lundbergin malli

Sisältö

Johdanto

Tämän tutkielman tarkoituksena on tutustua Poisson-prosessiin, painotettuun Poisson-prosessiin ja yhdistettyyn painotettuun Poisson-prossessiin sekä näiden tärkeimpiin ominaisuuksiin. Poisson-prosesseilla on paljon sovelluksia muun muassa jonotusteoriassa, vakuutusmatematiikassa, rahoituksessa sekä mallinnettaessa aikaa, kunnes jotakin tapahtuu. Lukujen lopussa lasketaan esimerkkejä, joissa nähdään mitä hyötyä Poisson-prosessien eri ominaisuuksista on käytännössä.

Lukijan oletetaan osaavan todennäköisyysteorian perusteet, mutta ensimmäisessä luvussa kerrataan lyhyesti tärkeimmät tässä käytettävät määritelmät. Toisessa luvussa käsitellään homogeenista ja epähomogeenista Poisson-prosessia ja sen ominaisuuksia. Luvun lopussa on käytännön esimerkkejä eri ominaisuuksiin liittyen ja esimerkkinä ratkaistaan myös odottajan paradoksi, joka tunnetaan myös muun muassa liftarin paradoksina.

Kolmannessa luvussa tutustutaan painotettuun Poisson-prosessiin ja sen ominaisuuksiin sekä lisäksi verrataan niitä ei-painotettuun prosessiin. Tutustutaan myös millaisissa tilanteissa kannattaisi käyttää painotettua prosessia tavallisen ei-painotetun prosessin sijaan. Luvun lopussa on kaksi esimerkkiä.

Viimeisessä luvussa käsitellään yhdistettyä painotettua Poisson-prosessia ja tutustutaan sen erikoistapaukseen, riskiteoriassa paljon käytettyyn Cramér-Lundbergin malliin. Cramér-Lundbergin mallia soveltaen lasketaan lopuksi palovakuutusvahinkoesimerkki, jonka aineisto on simuloitu.

Tutkielmassa esitetyt tulokset todistetaan yhtä todistuksessa tarvittavaa aputulosta lukuunottamatta. Tutkielma pohjautuu pääosin Mikoschin kirjaan Non-life insurance mathematics [3] ja Seppälän riskiteoria-kurssin luentomuistiinpanoihin [7].

LUKU 1

Kertausta

Palautetaan aluksi mieliin muutamia todennäköisyysteorian peruskäsitteitä ennen varsinaisiin Poisson-prosesseihin siirtymistä.

Mallinnettavan tilanteen otosavaruudelle käytetään merkintää Ω. Otosavaruuden tapahtumien muodostamalle sigma-algebralle käytetään merkintää $\mathcal{F}$ ja joukkofunktiota $\mathbb{P} : \mathcal{F} \to [0, 1]$ sanotaan todennäköisyysmitaksi tai lyhyesti todennäköisyydeksi. Kolmikko $(\Omega, \mathcal{F}, \mathbb{P})$ muodostaa todennäköisyysavaruuden.

Jos funktio $M : \Omega \to \mathbb{R}$ on $\mathcal{F}$-mitallinen, eli $M^{-1}(B) = \{\omega \in \Omega : M(\omega) \in B\} \in \mathcal{F}$ kaikilla $B \in \mathcal{B}(\mathbb{R})$, missä $\mathcal{B}$ on Borellin sigma-algebra, niin funktion M sanotaan olevan satunnaismuuttuja. Erityisesti siis satunnaismuuttuja on funktio. Jos kahden satunnaismuuttujan X ja Y kertymäfunktiot ovat samat, eli $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$ kaikilla x, niin niiden sanotaan olevan jakaumaltaan samat ja sitä merkitään $X \overset{d}{=} Y$.

Jos satunnaismuuttuja M saa vain kokonaislukuarvoja ja sen pistetodennäköisyysfunktio on $\mathbb{P}(M = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$, niin sen sanotaan noudattavan Poisson-jakaumaa parametrilla $\lambda > 0$, merkitään $M \sim Pois(\lambda)$. Jos satunnaismuuttuja $M = 0$ melkein varmasti (m.v.) eli $\mathbb{P}(M = 0) = 1$, niin tällöin sen sanotaan noudattavan Poisson-jakaumaa parametrilla 0. Poisson-satunnaismuuttujalla on se harvinainen ominaisuus, että sen odotusarvo ja varianssi ovat yhtäsuuret ja tämä arvo on Poisson-parametri; jos $M \sim Pois(\lambda)$, niin $\mathbb{E}M = \lambda = Var\,M$.

Jos satunnaismuuttuja X saa arvon $a \in \mathbb{R}$ m.v. eli $\mathbb{P}(X = a) = 1$, niin satunnaismuuttujan sanotaan olevan degeneroitunut. Tällöin X ei itseasiassa ole enää satunnainen. Jos satunnaismuuttuja ei ole vakioarvoinen, niin sen sanotaan olevan ei-degeneroitunut.

Poisson-jakauman lisäksi toinen jatkossa tarvittava jakauma on eksponenttijakauma, joka on jatkuva toisin kuin Poisson-jakauma. Jos satunnaismuuttuja ε noudattaa eksponenttijakaumaa parametrilla $\lambda > 0$, merkitään $\varepsilon \sim Exp(\lambda)$, niin tällöin sen tiheysfunktio on $f_\varepsilon(x) = \lambda e^{-\lambda x}$ ja kertymäfunktio $F_\varepsilon(x) = \mathbb{P}(\varepsilon \leq x) = 1 - e^{-\lambda x}$ kaikilla $x \geq 0$.

Satunnaismuuttujien $N : \Omega \times [0, \infty) \to \mathbb{R}$ kokoelmaa sanotaan stokastiseksi prosessiksi ja sitä merkitään $N = (N(t))_{t \in [0,\infty)}$. Stokastinen prosessi $(N(t))_{t\in[0,\infty)}$ on mitallinen, jos funktio $N : \Omega \times [0, \infty) \to \mathbb{R}$, $(\omega, t) \mapsto N(\omega, t)$ on mitallinen sigma-algebrojen $\mathcal{F} \otimes \mathcal{B}([0, \infty))$ ja $\mathcal{B}(\mathbb{R})$ suhteen eli $\{(\omega, t) : N(\omega, t) \in A\} \in \mathcal{F} \otimes \mathcal{B}([0, \infty))$ kaikilla $A \in \mathcal{B}(\mathbb{R})$. Funktiota $t \mapsto N(\omega, t)$ sanotaan prosessin N poluksi kiinnitetyn muuttujan ω suhteen.

LUKU 2

Poisson-prosessi

Tässä luvussa käsitellään homogeenista ja epähomogeenista Poisson-prosessia sekä niiden ominaisuuksia. Ennen Poisson-prosessin määrittelemistä otetaan käyttöön välin kuvautumiseen liittyvä merkintä ja lisäksi keskiarvofunktion käsite, joita molempia tarvitaan jatkossa.

MERKINTÄ. Mille tahansa funktiolle f, joka on määritelty joukossa $[0,\infty)$, voidaan käyttää merkintää $f(s,t] = f(t) - f(s)$, kun $0 \leq s < t < \infty$.

Jatkossa käytetään molempia merkintöjä ja käytettävä merkintä valitaan tilanteen mukaan.

MÄÄRITELMÄ 2.1. *Funktio $\mu : [0,\infty) \to [0,\infty)$ on keskiarvofunktio, jos sillä on seuraavat ominaisuudet:*

(1) $\mu(0) = 0$
(2) $\mu(s) \leq \mu(t)$ *kaikilla* $0 \leq s \leq t$
(3) μ *on oikealta jatkuva.*

Määritellään sitten Poisson-prosessi:

MÄÄRITELMÄ 2.2. *Stokastinen prosessi $N = (N(t))_{t\in[0,\infty)}$ on Poisson-prosessi, jos se täyttää seuraavat ehdot:*

(1) $N(0) = 0$ *m.v., eli prosessi alkaa nollasta.*
(2) *Prosessilla N on riippumattomat lisäykset, eli jos $0 = t_0 < t_1 < \cdots < t_n$, niin $N(t_n) - N(t_{n-1}), N(t_{n-1}) - N(t_{n-2}), \ldots, N(t_1) - N(t_0)$ ovat riippumattomia.*
(3) *On olemassa keskiarvofunktio μ siten, että*
$$\mathbb{P}(N(t) - N(s) = m) = e^{-(\mu(t)-\mu(s))}\frac{(\mu(t)-\mu(s))^m}{m!}$$
kaikilla $m = 0, 1, 2, \ldots$ ja $t > 0$ kun $0 \leq s \leq t$, eli
$N(t) - N(s) \sim Pois(\mu(t) - \mu(s))$.
(4) *Prosessin N polut ovat oikealta jatkuvia ja niillä on vasemmanpuoleiset raja-arvot.*

Jos N on Poisson-prosessi, jolle $N(t) = m$, niin sanotaan, että hetkeen t mennessä on tullut m saapumista.

HUOMAUTUS 2.3. Poisson-prosessin erikoistapauksia:

(1) Poisson-prosessia sanotaan homogeeniseksi, jos $\mu(t) = \lambda t$ jollakin $\lambda > 0$.
(2) Homogeenisen Poisson-prosessin lisäykset ovat riippumattomuuden lisäksi stationaarisia, eli millä tahansa $0 \leq s < t$,
$$N(t) - N(s) \stackrel{d}{=} N(t-s) - N(0) \sim Pois(\lambda(t-s))$$

eli siis Poisson-parametri riippuu vain välin pituudesta, ei sen sijainnista.

Todistus.

$$\begin{aligned}\mathbb{P}(N(t)-N(s)=m) &= e^{-(\lambda t-\lambda s)}\frac{(\lambda t-\lambda s)^m}{m!}\\ &= e^{-(\lambda(t-s)-\lambda\cdot 0)}\frac{(\lambda(t-s)-\lambda\cdot 0)^m}{m!}\\ &= \mathbb{P}(N(t-s)-N(0)=m).\end{aligned}$$

□

(3) Homogeenista Poisson-prosessia sanotaan standardiksi, jos $\lambda = 1$.

Seuraavassa lauseessa tutkitaan epähomogeenisen ja standardin homogeenisen Poisson-prosessin yhteyttä. Lauseen avulla voidaan helposti vaihtaa epähomogeenisesta Poisson-prosessista standardiin homogeeniseen Poisson-prosessiin ja toisin päin. Sovelluksissa on usein hyödyllistä voida vaihtaa epähomogeeninen prosessi teoreettisesti helpommin käsiteltävään standardiin homogeeniseen prosessiin. Ennen lauseen muotoilua määritellään mitä tarkoittaa, että prosesseilla on samat äärellisulotteiset jakaumat:

Määritelmä 2.4. *Olkoot $X = (X(t))_{t\in[0,\infty)}$ ja $Y = (Y(t))_{t\in[0,\infty)}$ stokastisia prosesseja todennäköisyysavaruudessa $(\Omega, \mathcal{F}, \mathbb{P})$. Prosesseilla X ja Y sanotaan olevan samat äärellisulotteiset jakaumat, merkitään $X \stackrel{d}{=} Y$, jos*

$$\mathbb{P}\big((X(t_1),...,X(t_n)) \in A\big) = \mathbb{P}\big((Y(t_1),...,Y(t_n)) \in A\big)$$

kaikilla $n \in \mathbb{N}$, kaikilla $0 \le t_1 < t_2 < ... < t_n < \infty$ ja kaikilla $A \in \mathcal{B}(\mathbb{R}^n)$.

Lause 2.5. *Jos μ on epähomogeenisen Poisson-prosessin N keskiarvofunktio ja $\tilde{N}$ on standardi homogeeninen Poisson-prosessi, niin*

(1) $(N(t))_{t\in[0,\infty)} \stackrel{d}{=} (\tilde{N}(\mu(t)))_{t\in[0,\infty)}$.
(2) *Jos μ on jatkuva, kasvava ja $\lim_{t\to\infty}\mu(t) = \infty$, niin* $(N(\mu^{-1}(t)))_{t\in[0,\infty)} \stackrel{d}{=} (\tilde{N}(t))_{t\in[0,\infty)}$.

Todistus. (1) Osoitetaan, että prosesseilla on sama pistetodennäköisyysfunktio, jolloin prosessit ovat jakaumaltaan samat:

$$\begin{aligned}\mathbb{P}(N(t)-N(s)=m) &= e^{-(\mu(t)-\mu(s))}\frac{(\mu(t)-\mu(s))^m}{m!}\\ &= \mathbb{P}(\tilde{N}(\mu(t))-\tilde{N}(\mu(s))=m).\end{aligned}$$

Lisäksi, kun $A \subset \{0, 1, 2, ...\}$ eli $A = \{a_1, a_2, a_3, ...\}$, niin

$$\begin{aligned}
\mathbb{P}(N(t) \in A) &= \sum_{k=1}^{\infty} \mathbb{P}(N(t) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(N(t) - N(0) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(\tilde{N}(\mu(t)) - \tilde{N}(\mu(0)) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(\tilde{N}(\mu(t)) = a_k) \\
&= \mathbb{P}(\tilde{N}(\mu(t)) \in A).
\end{aligned}$$

(2) Funktion μ oletuksista seuraa, että sillä on olemassa käänteisfunktio. Nyt vastaavasti kuin kohdassa (1), saadaan

$$\begin{aligned}
\mathbb{P}[N(\mu^{-1}(t)) - N(\mu^{-1}(s)) = m] &= e^{-[\mu(\mu^{-1}(t)) - \mu(\mu^{-1}(s))]} \frac{[\mu(\mu^{-1}(t)) - \mu(\mu^{-1}(s))]^m}{m!} \\
&= e^{-(t-s)} \frac{(t-s)^m}{m!} \\
&= \mathbb{P}(\tilde{N}(t) - \tilde{N}(s) = m)
\end{aligned}$$

ja

$$\begin{aligned}
\mathbb{P}(\tilde{N}(t) \in A) &= \sum_{k=1}^{\infty} \mathbb{P}(\tilde{N}(t) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(\tilde{N}(t) - \tilde{N}(0) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(N(\mu^{-1}(t)) - N(\mu^{-1}(0)) = a_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(N(\mu^{-1}(t)) = a_k) \\
&= \mathbb{P}(N(\mu^{-1}(t)) \in A).
\end{aligned}$$

□

MÄÄRITELMÄ 2.6. *Olkoon μ epähomogeenisen Poisson-prosessin N keskiarvofunktio. Jos μ on absoluuttisesti jatkuva, eli millä tahansa $s < t$ lisäys $\mu(t) - \mu(s)$ voidaan esittää muodossa*

$$\mu(t) - \mu(s) = \int_s^t \lambda(y)\, dy,$$

jollakin ei-negatiivisella ja mitallisella funktiolla λ, niin tällöin funktion λ sanotaan olevan prosessin N intensiteettifunktio.

Erityisesti ylläolevasta määritelmästä saadaan, että jos prosessilla on intensiteettifunktio, niin μ on jatkuva funktio, ei pelkästään oikealta jatkuva. Lisäksi homogeeniselle Poisson-prosessille saadaan intensiteettifunktioksi $\lambda = \lambda(y)$, sillä tällöin $\mu(t) - \mu(s) = \lambda t - \lambda s = \int_s^t \lambda \, dy$.

Jos prosessin N intensiteettifunktio λ ei ole vakio, niin aika "nopeutuu" kun $\lambda > 1$, eli prosessin saapumiset tulevat tiheämmin. Vastaavasti, jos $\lambda < 1$ niin aika "hidastuu" ja saapumisia tulee harvemmin. Vakuutussovelluksissa ei-vakio intensiteetti λ voi viitata kausivaihteluun tai trendiin.

Osoitetaan seuraavaksi kaksi aputulosta, joita tarvitaan homogeenisen Poisson-prosessin toisenlaisen määritelmän todistuksessa.

Lemma 2.7. *Olkoon $(\varepsilon_i)_{i=1}^n$ jono riippumattomasti samoin jakautuneita satunnaismuuttujia siten, että $\varepsilon_i \sim Exp(\lambda)$ kaikilla $i = 1, ..., n$. Kun $x > 0$, niin*

$$\mathbb{P}(\varepsilon_1 + ... + \varepsilon_n \leq x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}.$$

Todistus. Todistetaan väite induktiolla:
Osoitetaan ensin, että väite pätee, kun $n = 1$:

$$\mathbb{P}(\varepsilon_1 \leq x) = \int_0^x \lambda e^{-\lambda y_1} \, dy_1 = 1 - e^{-\lambda x}.$$

Induktio-oletuksena oletetaan, että väite $\mathbb{P}(\varepsilon_1 + ... + \varepsilon_{n-1} \leq x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-2} \frac{(\lambda x)^k}{k!}$ pätee. Osoitetaan lopuksi väitteen pitävän paikkansa kun summassa on n kappaletta termejä:

$$\begin{aligned}
&\mathbb{P}(\varepsilon_1 + ... + \varepsilon_n \leq x) \\
=&\int_0^x \int_0^{x-y_n} \dots \int_0^{x-y_n-...-y_2} \lambda e^{-\lambda y_n} \lambda e^{-\lambda y_{n-1}} \dots \lambda e^{-\lambda y_1} \, dy_1 \, dy_2 ... dy_n \\
=&\int_0^x \lambda e^{-\lambda y_n} \left[\int_0^{x-y_n} \dots \int_0^{x-y_n-...-y_2} \lambda e^{-\lambda y_{n-1}} \dots \lambda e^{-\lambda y_1} \, dy_1 \, dy_2 ... dy_{n-1} \right] dy_n \\
=&\int_0^x \lambda e^{-\lambda y_n} \Big(\mathbb{P}(\varepsilon_1 + ... \varepsilon_{n-1} \leq x - y_n) \Big) dy_n \\
=&\int_0^x \lambda e^{-\lambda y_n} \left(1 - e^{-\lambda(x-y_n)} \sum_{k=0}^{n-2} \frac{(\lambda(x-y_n))^k}{k!} \right) dy_n \qquad \text{(induktio-oletus)} \\
=&\int_0^x \lambda e^{-\lambda y_n} \, dy_n - \lambda e^{-\lambda x} \sum_{k=0}^{n-2} \frac{\lambda^k}{k!} \int_0^x (x-y_n)^k \, dy_n \\
=&1 - e^{-\lambda x} - \lambda e^{-\lambda x} \sum_{k=0}^{n-2} \frac{\lambda^k}{k!} \Big/_0^x - \frac{(x-y_n)^{k+1}}{k+1} \\
=&1 - e^{-\lambda x} - \lambda e^{-\lambda x} \sum_{k=0}^{n-2} \frac{\lambda^k x^{k+1}}{(k+1)!}
\end{aligned}$$

$$=1 - e^{-\lambda x} - e^{-\lambda x}\sum_{k=0}^{n-2}\frac{(\lambda x)^{k+1}}{(k+1)!}$$
$$=1 - e^{-\lambda x}\sum_{k=0}^{n-1}\frac{(\lambda x)^{k}}{k!}.$$

□

LEMMA 2.8. *Olkoon* $(\varepsilon_i)_{i=1}^n$ *jono riippumattomasti samoin jakautuneita satunnaismuuttujia siten, että* $\varepsilon_i \sim Exp(\lambda)$ *kaikilla* $i = 1, ..., n$ *ja olkoon* $T_n = \varepsilon_1 + ... + \varepsilon_n$. *Tällöin*

$$\mathbb{P}(a < T_n \leq b) = \int_a^b \lambda^n e^{-\lambda y}\frac{y^{n-1}}{(n-1)!}dy.$$

TODISTUS. Lemman 2.7. nojalla $\mathbb{P}(T_n \leq x) = 1 - e^{-\lambda x}\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!} =: F_{T_n}(x)$. Funktion T_n kertymäfunktiota derivoimalla saadaan sen tiheysfunktio:

$$\begin{aligned}
F'_{T_n}(x) &= \lambda e^{-\lambda x}\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!} - e^{-\lambda x}\sum_{k=1}^{n-1}\frac{\lambda k(\lambda x)^{k-1}}{k!}\\
&= \lambda e^{-\lambda x}\left(\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!} - \sum_{k=1}^{n-1}\frac{(\lambda x)^{k-1}}{(k-1)!}\right)\\
&= \lambda e^{-\lambda x}\left(\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!} - \sum_{k=0}^{n-2}\frac{(\lambda x)^{k}}{k!}\right)\\
&= \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}.
\end{aligned}$$

Kertymäfunktion arvo jollakin välillä saadaan integroimalla tiheysfunktiota yli kyseisen välin eli saadaan siis

$$\mathbb{P}(a < T_n \leq b) = \int_a^b \lambda^n e^{-\lambda y}\frac{y^{n-1}}{(n-1)!}\,dy.$$

□

Muotoillaan seuraavaksi homogeeniselle Poisson-prosessille toinen esitystapa, jota voi joissakin tapauksissa olla helpompi käytää kuin Määritelmää 2.2. Lausessa 2.11. osoitetaan, että Määritelmän 2.9. prosessi $\hat{N}(t)$ todella vastaa homogeenista Poisson-prosessia.

MÄÄRITELMÄ 2.9. *Olkoon* $(\varepsilon_i)_{i=1}^n$ *jono riippumattomasti samoin jakautuneita satunnaismuuttujia siten, että* $\varepsilon_i \sim Exp(\lambda)$ *kaikilla* $i = 1, ..., n$. *Määritellään* $T_n = \varepsilon_1 + ... + \varepsilon_n$, $n \geq 1$, *ja*

$$\hat{N}(t) = \#\{i \geq 1, T_i \leq t\}, t \geq 0.$$

LEMMA 2.10. *Olkoon* $\hat{N}(t)$ *kuten Määritelmässä 2.8. Tällöin kaikille* $n = 0, 1, 2, \ldots$ *ja* $t \geq 0$ *pätee, että*

$$\mathbb{P}(\{\omega \in \Omega : \hat{N}(t, \omega) = n\}) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

eli $\hat{N}(t) \sim Pois(\lambda t)$.

TODISTUS. Prosessin $\hat{N}$ määritelmän nojalla

$$\begin{aligned}\{\omega \in \Omega : \hat{N}(t,\omega) = n\} &= \{\omega \in \Omega : \#\{i \geq 1 : T_i(\omega) \leq t\} = n\} \\ &= \{\omega \in \Omega : T_n(\omega) \leq t < T_{n+1}(\omega)\} \\ &= \{\omega \in \Omega : T_n(\omega) \leq t\} \setminus \{\omega \in \Omega : T_{n+1}(\omega) \leq t\}.\end{aligned}$$

Koska $T_n \leq T_{n+1}$, niin $\{\omega \in \Omega : T_{n+1}(\omega) \leq t\} \subseteq \{\omega \in \Omega : T_n(\omega) \leq t\}$. Nyt Lemman 2.7. nojalla saadaan, että

$$\begin{aligned}\mathbb{P}(\hat{N}(t) = n) &= \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t) \\ &= 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - 1 + e^{-\lambda t} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!}.\end{aligned}$$

□

LAUSE 2.11. (1) *Määritelmän 2.9. prosessi* $(\hat{N}(t))_{t\in[0,\infty)}$ *on homogeeninen Poisson-prosessi intensiteetillä* $\lambda > 0$.

(2) *Mikä tahansa homogeeninen Poisson-prosessi intensiteetillä* $\lambda > 0$ *voidaan esittää muodossa*

$$N(t) = \#\{i \geq 1, T_i \leq t\}, \qquad t \geq 0,$$

missä $T_n = \varepsilon_1 + ... + \varepsilon_n$, $n \geq 1$ *ja* $(\varepsilon_i)_{i=1}^n$ *on jono riippumattomasti samoin jakautuneita satunnaismuuttujia siten, että* $\varepsilon_i \sim Exp(\lambda)$ *kaikilla* $i = 1, ..., n$.

TODISTUS. **(1)** Tarkastetaan Määritelmän 2.2. ehdot (1)-(4):

(1) Lemman 2.7. nojalla $\mathbb{P}(\varepsilon_1 \leq^{-1}) = 1 - e^{-\lambda n^{-1}}$, jolloin $\mathbb{P}(\varepsilon_1 > n^{-1}) = e^{-\lambda n^{-1}}$ kaikilla $n \geq 1$. Kun $n \to \infty$, niin $\mathbb{P}(\varepsilon_1 > 0) \geq \mathbb{P}(\varepsilon_1 > n^{-1}) \to 1$, joten $\hat{N}(t,\omega) = 0$ kaikilla $t < T_1(\omega) = \varepsilon_1(\omega)$. Koska $\varepsilon_1 > 0$ melkein varmasti, niin $\hat{N}(0,\omega) = 0$ melkein varmasti eli kohta (1) pätee.

(2) Osoitetaan lisäysten riippumattomuus suoraan laskemalla vain yhdelle lisäykselle, jolla todistuksen metodi tulee selville. Erilainen todistus, jossa riippumattomuus todistetaan mielivaltaisen monelle lisäykselle, lötyy esimerkiksi Resnickin kirjasta [5]. Osoitetaan nyt, jos $0 < s < t$, niin

$$\mathbb{P}\big(\hat{N}(s) = l, \hat{N}(t) - \hat{N}(s) = m\big) = \mathbb{P}\big(\hat{N}(s) = l\big)\mathbb{P}\big(\hat{N}(t) - \hat{N}(s) = m\big),$$

kun $l, m \geq 0$. Jaetaan todistus neljään eri tapaukseen ja käytetään lyhyyden vuoksi merkintöjä $\{\omega : t < T_l(\omega)\} =: (t < T_l)$ ja $\{\omega : t < \varepsilon_l(\omega)\} =: (t < \varepsilon_l)$.

Tapaus $l = m = 0$:

$$\begin{aligned}
\mathbb{P}\Big(\hat{N}(s) = 0, \hat{N}(t) - \hat{N}(s) = 0\Big) &= \mathbb{P}\Big(\hat{N}(s) = 0, \hat{N}(t) = 0\Big) \\
&= \mathbb{P}\left(s < T_1, t < T_1\right) \\
&\overset{s \le t}{=} \mathbb{P}\left(t < \varepsilon_1\right) \\
&= \int_t^\infty \lambda e^{-\lambda x}\, dx \\
&= e^{-\lambda t} \\
&= e^{-\lambda s} e^{-\lambda(t-s)} \\
&= \mathbb{P}\Big(\hat{N}(s) = 0\Big)\mathbb{P}\Big(\hat{N}(t-s) = 0\Big).
\end{aligned}$$

Koska nyt

$$\mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) - \hat{N}(s) = m\big) = \mathbb{P}(\hat{N}(s) = l)\, \mathbb{P}(\hat{N}(t-s) = m),$$

niin summaamalla prosessin $\hat{N}$ mahdollisten arvojen yli ajanhetkellä s saadaan

$$\begin{aligned}
&\sum_{l=0}^{\infty} \mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) - \hat{N}(s) = m\big) = \sum_{l=0}^{\infty} \mathbb{P}\big(\hat{N}(s) = l\big)\, \mathbb{P}\big(\hat{N}(t) - \hat{N}(s) = m\big) \\
&\iff \qquad \mathbb{P}\big(\hat{N}(t) - \hat{N}(s) = m\big) = \mathbb{P}\big(\hat{N}(t-s) = m\big),
\end{aligned} \tag{2.1}$$

joten väite pätee kun $l = m = 0$.

Tapaus $l \geq 1$ ja $m = 0$:

$$\begin{aligned}
\mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) - \hat{N}(s) = 0\big) &= \mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) = l\big) \\
&= \mathbb{P}\big(T_l \le s < T_{l+1},\ T_l \le t < T_{l+1}\big) \\
&= \mathbb{P}\big(T_l \le s, t < T_l + \varepsilon_{l+1}\big) \\
&= \int_0^s \int_{t-y}^\infty \lambda^l e^{-\lambda y} \frac{y^{l-1}}{(l-1)!} \lambda e^{-\lambda x}\, dx\, dy,
\end{aligned}$$

missä ulompi integraali seuraa Lemmasta 2.8. ja sisempi integraali saadaan suoraan eksponenttifunktion tiheysfunktiona, koska T_l ja ε_{l+1} ovat riippumattomia. Integraalit laskemalla saadaan edelleen

$$\begin{aligned}\mathbb{P}\big(\hat{N}(s)=l,\ \hat{N}(t)-\hat{N}(s)=0\big) &= \int_0^s e^{-\lambda y}\frac{\lambda(\lambda y)^{l-1}}{(l-1)!}e^{-\lambda(t-y)}\,dy\\ &= e^{-\lambda t}\int_0^s \frac{\lambda(\lambda y)^{l-1}}{(l-1)!}\,dy\\ &= e^{-\lambda t}\frac{(\lambda s)^l}{l!}\\ &= e^{-\lambda(s+t-s)}\frac{(\lambda s)^l}{l!}\\ &= \left(e^{-\lambda s}\frac{(\lambda s)^l}{l!}\right)\left(e^{-\lambda(t-s)}\frac{(\lambda(t-s))^0}{0!}\right)\\ &= \mathbb{P}(\hat{N}(s)=l)\,\mathbb{P}(\hat{N}(t-s)=0).\end{aligned}$$

Väite seuraa yhtälön 2.1 nojalla.

Tapaus $l=0, m\geq 1$:

$$\begin{aligned}&\mathbb{P}(\hat{N}(s)=0,\ \hat{N}(t)-\hat{N}(s)=m)\\ =&\mathbb{P}(\hat{N}(s)=0,\ \hat{N}(t)=m)\\ =&\mathbb{P}(T_0=0\leq s<T_1=\varepsilon_1,\ T_m\leq t<T_m+\varepsilon_{m+1})\\ =&\mathbb{P}(s<\varepsilon_1<\infty,\ 0\leq T_m\leq t,\ t-T_m<\varepsilon_{m+1}<\infty)\\ =&\int_s^\infty\int_0^{t-x_1}\int_{t-x_1-x_2}^\infty h_3(x_3)\,h_2(x_2)\,h_1(x_1)\,dx_3\,dx_2\,dx_1,\end{aligned}$$

missä kaikilla $x_i>0, i=1,2,3,$ h_1 ja h_3 ovat eksponenttifunktion tiheysfunktioita ja tiheysfunktio h_2 seuraa Lemmasta 2.8., koska ε_1, T_m ja ε_{m+1} ovat riippumattomia. Jos $x_i\leq 0$, niin $h_i(x_i)=0$ kaikilla $i=1,2,3$. Nyt saadaan

$$\begin{aligned}&\mathbb{P}(\hat{N}(s)=0,\ \hat{N}(t)-\hat{N}(s)=m)\\ =&\int_s^\infty\int_0^{t-x_1}\int_{t-x_1-x_2}^\infty \lambda e^{-\lambda x_3}\,\lambda e^{-\lambda x_2}\,\frac{(\lambda x_2)^{m-2}}{(m-2)!}\,\lambda e^{-\lambda x_1}\,dx_3\,dx_2\,dx_1\\ =&\int_s^\infty\int_0^{t-x_1} -e^{-\lambda(t-x_1-x_2)}\,\lambda e^{-\lambda x_2}\,\frac{(\lambda x_2)^{m-2}}{(m-2)!}\lambda e^{-\lambda x_1}\,dx_2\,dx_1\\ =&\int_s^\infty -e^{-\lambda t}\,\frac{(\lambda(t-x_1))^{m-1}}{(m-1)!}\lambda\,dx_1\\ =&e^{-\lambda t}\,\frac{(\lambda(t-s))^m}{m!}\\ =&e^{-\lambda s}\left(e^{-\lambda(t-s)}\frac{(\lambda(t-s))^m}{m!}\right)\\ =&\mathbb{P}(\hat{N}(s)=0)\,\mathbb{P}(\hat{N}(t-s)=m).\end{aligned}$$

Yhtälön (2.1) nojalla väite pätee tässäkin tapauksessa.

Tapaus $l, m \geq 1$:

$$\begin{aligned}\mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) - \hat{N}(s) = m\big) &= \mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) = m + l\big)\\ &= \mathbb{P}\big(T_l \leq s < T_{l+1},\ T_{l+m} \leq t < T_{l+m+1}\big).\end{aligned}$$

Merkitsemällä

$$\begin{aligned}T_l &=: f_1\\ T_{l+1} &= T_l + \varepsilon_{l+1} =: f_1 + f_2\\ T_{l+m} &= T_l + \varepsilon_{l+1} + \varepsilon_{l+2} + ... + \varepsilon_{l+m} =: f_1 + f_2 + f_3\\ T_{l+m+1} &= T_l + \varepsilon_{l+1} + ... + \varepsilon_{l+m} + \varepsilon_{l+m+1} =: f_1 + f_2 + f_3 + f_4\end{aligned}$$

saadaan satunnaismuuttujien f_1, f_2, f_3 ja f_4 riippumattomuuden nojalla edelleen, että

$$\begin{aligned}&\mathbb{P}\big(\hat{N}(s) = l,\ \hat{N}(t) - \hat{N}(s) = m\big)\\ =&\mathbb{P}\big(\hat{N}(s) = l, \hat{N}(t) = m + l\big)\\ =&\mathbb{P}\big(f_1 \leq s < f_1 + f_2,\ f_1 + f_2 + f_3 \leq t < f_1 + f_2 + f_3 + f_4\big)\\ =&\mathbb{P}\big(0 \leq f_1 \leq s, s - f_1 < f_2 < \infty,\\ &0 \leq f_3 \leq t - f_1 - f_2, t - f_1 - f_2 - f_3 < f_4 < \infty\big)\\ =&\int_0^s \int_{s-x_1}^{\infty} \int_0^{t-x_1-x_2} \int_{t-x_1-x_2-x_3}^{\infty} h_4(x_4)h_3(x_3)h_2(x_2)h_1(x_1)dx_4dx_3dx_2dx_1,\end{aligned}$$

missä h_i on satunnaismuuttujan f_i tiheysfunktio kaikilla $i = 1, 2, 3, 4$, missä

$$h_i(x_i) = \begin{cases} \lambda e^{-\lambda x_1} \frac{(\lambda x_1)^{l-1}}{(l-1)!} & \text{kun } i = 1\\ \lambda e^{-\lambda x_2} & \text{kun } i = 2\\ \lambda e^{-\lambda x_3} \frac{(\lambda x_3)^{m-2}}{(m-2)!} & \text{kun } i = 3\\ \lambda e^{-\lambda x_4} & \text{kun } i = 4, \end{cases}$$

kun $x_i > 0$ ja $h_i(x_i) = 0$ kun $x_i \leq 0$. Tässä funktiot h_2 ja h_4 saadaan taas suoraan eksponenttifunktion tiheysfunktioina ja funktiot h_1 ja h_3 seuraavat jälleen Lemmasta 2.8. Laskemalla integraalit vaiheittain sisimmästä aloittaen ja huomioiden ehdot $0 < x_3 < t - x_1 - x_2$ ja $0 < x_2 < t - x_1$, saadaan sisimmästä integraalista

$$\int_{t-x_1-x_2-x_3}^{\infty} h_4(x_4)\ dx_4 = e^{-\lambda(t-x_1-x_2-x_3)}.$$

Funktioiden h_3 ja h_4 tulon integraaliksi saadaan

$$\begin{aligned}&\int_0^{t-x_1-x_2} \int_{t-x_1-x_2-x_3}^{\infty} h_4(x_4)\ dx_4\ h_3(x_3)\ dx_3\\ =&\int_0^{t-x_1-x_2} e^{-\lambda(t-x_1-x_2-x_3)} \lambda^{m-1} \frac{x_3^{m-2}}{(m-2)!} e^{-\lambda x_3} \mathbb{1}_{[0,t-x_1)}(x_2)\ dx_3\\ =&\ e^{-\lambda(t-x_1-x_2)} \lambda^{m-1} \frac{(t - x_1 - x_2)^{m-1}}{(m-1)!} \mathbb{1}_{[0,t-x_1)}(x_2),\end{aligned}$$

ja toisiksi uloimmasta integraalista

$$\begin{aligned}
&\int_{s-x_1}^{\infty}\int_0^{t-x_1-x_2}\int_{t-x_1-x_2-x_3}^{\infty} h_4(x_4)\,dx_4\,h_3(x_3)\,dx_3\,h_2(x_2)\,dx_2\\
&=\int_{s-x_1}^{\infty} e^{-\lambda(t-x_1-x_2)}\lambda^{m-1}\frac{(t-x_1-x_2)^{m-1}}{(m-1)!}\mathbb{1}_{[0,t-x_1)}(x_2)\lambda e^{-\lambda x_2}\mathbb{1}_{[0,s)}(x_1)\,dx_2\\
&=\lambda^m e^{-\lambda(t-x_1)}\frac{(t-s)^m}{m!}\mathbb{1}_{[0,s)}(x_1).
\end{aligned}$$

Lopulta koko alkuperäiseksi intregraaliksi saadaan

$$\begin{aligned}
&\mathbb{P}\big(\hat{N}(s)=l,\ \hat{N}(t)-\hat{N}(s)=m\big)\\
&=\int_0^s\int_{s-x_1}^{\infty}\int_0^{t-x_1-x_2}\int_{t-x_1-x_2-x_3}^{\infty} h_4(x_4)dx_4h_3(x_3)dx_3h_2(x_2)dx_2h_1(x_1)dx_1\\
&=\int_0^s \lambda^m e^{-\lambda(t-x_1)}\frac{(t-s)^m}{m!}\lambda^l\frac{x_1^{l-1}}{(l-1)!}e^{-\lambda x_1}dx_1\\
&=\lambda^m\lambda^l e^{-\lambda t}\frac{(t-s)^m}{m!}\frac{s^l}{l!}\\
&=\left(e^{-\lambda s}\frac{(\lambda s)^l}{l!}\right)\left(e^{-\lambda(t-s)}\frac{(\lambda(t-s))^m}{m!}\right)\\
&=\mathbb{P}\big(\hat{N}(s)=l\big)\mathbb{P}\big(\hat{N}(t-s)=m\big)\\
&=\mathbb{P}\big(\hat{N}(s)=l\big)\mathbb{P}\big(\hat{N}(t)-\hat{N}(s)=m\big)\qquad\text{(yhtälö (2.1)).}
\end{aligned}$$

Määritelmän 2.2. kohta (2) on nyt todistettu.

(3) Seuraa Lemmasta 2.10. ja yhtälöstä (2.1)

(4) Seuraa konstruktiosta, sillä $\{\hat{N}(t)=n\}=\{T_n\le t<T_{n+1}\}$.

(2) Väitteen mukaan $\varepsilon_1,\varepsilon_2,...,\varepsilon_n$ ovat riippumattomia ja $\varepsilon_i\sim Exp(\lambda)$ kaikilla $i=1,...,n$. Merkitään $A_n=\{\varepsilon_1>t_1,...,\varepsilon_n>t_n\}$, $t_i>0$, jolloin pitäisi olla

$$\mathbb{P}(A_n)=e^{-\lambda\sum_{i=1}^n t_i},\qquad n\ge 1. \tag{2.2}$$

Todistetaan yhtälö (2.2.) induktiolla:

Jos n=1, niin väite pätee, sillä

$$\mathbb{P}(A_1)=\mathbb{P}(\varepsilon_1>t_1)=\mathbb{P}(N(t_1)=0)\overset{(*)}{=}e^{-\lambda t},$$

missä kohta $(*)$ seuraa Lemmasta 2.10. Oletaan sitten induktio-oletuksena, että väite $\mathbb{P}(A_{n-1})=e^{-\lambda\sum_{i=1}^{n-1}t_i}$ pätee. Osoitetaan vielä, että väite pätee myös joukolle A_n. Muotoillaan ensin tarkasteltavaa todennäköisyyttä hieman toisenlaiseksi:

$$\mathbb{P}(A_n)=\mathbb{P}(A_{n-1},\varepsilon_n>t_n)=\mathbb{P}(A_{n-1})\ \mathbb{P}(\varepsilon_n>t_n|A_{n-1}).$$

Kirjoitetaan sitten joukko A_{n-1} toisella tavalla käyttäen tietoa $\varepsilon_i = T_i - T_{i-1}$:

$$\begin{aligned}A_{n-1} &= \{\varepsilon_1 > t_1, \ldots, \varepsilon_{n-1} > t_{n-1}\}\\ &= \{T_1 - T_0 > t_1, \ldots, T_{n-1} - T_{n-2} > t_{n-1}\}\\ &= \{N(T_0, T_0 + t_1] = 0, \ldots, N(T_{n-2}, T_{n-2} + t_{n-1}] = 0\},\end{aligned}$$

koska $T_i > T_{i-1} + t_i$. Kirjoitetaan myös joukko $\varepsilon_n > t_n$ muodossa $N(T_{n-1}, T_{n-1} + t_n] = 0$, jolloin saadaan

$$\begin{aligned}&\mathbb{P}(\varepsilon_n > t_n | A_{n-1})\\ =&\mathbb{P}\left(N(T_{n-1}, T_{n-1} + t_n] = 0 | N(T_0, T_0 + t_1] = 0, \ldots, N(T_{n-2}, T_{n-2} + t_{n-1}] = 0\right)\\ =&\mathbb{P}\left(N(T_{n-1}, T_{n-1} + t_n] = 0\right) \qquad \text{(lisäysten riippumattomuus)}\\ =&\mathbb{P}\left(N(t_n) = 0\right) \qquad \text{(lisäysten stationaarisuus)}\\ =&e^{-\lambda t_n}\end{aligned}$$

Yhdistämällä edelliset, saadaan

$$\mathbb{P}(A_n) = \mathbb{P}(A_{n-1})\, \mathbb{P}(\varepsilon_n > t_n | A_{n-1}) = e^{-\lambda \sum_{i=1}^{n-1} t_i}\, e^{-\lambda t_n} = e^{-\lambda \sum_{i=1}^{n} t_i}.$$

□

2.1. Tunnuslukuja

Esitetään seuraavaksi Poisson-prosessin jakauman sijaintiin ja muotoon liittyviä tunnuslukuja. Lause todistetaan luvussa kolme.

LAUSE 2.12. *Olkoon N Poisson-prosessi keskiarvofunktiolla μ. Tällöin prosessille N pätee*

1. *Odotusarvo* $\mathbb{E}N(t) = \mu(t)$.
2. *Varianssi* $Var\, N(t) = \mu(t)$.
3. *Vinous* $\gamma_1\left(N(t)\right) = \frac{1}{\mu(t)^{1/2}}$
4. *Huipukkuus* $\gamma_2\left(N(t)\right) = \frac{1}{\mu(t)}$.

TODISTUS. Katso Lauseen 3.2. todistus ja Huomautus 3.3. □

Vinousluku mittaa jakauman vinoutta suhteessa symmetriseen jakaumaan, jolle vinous on 0. Koska Poisson-prosessille N on $\gamma_1(N(t)) = \mu(t)^{-1/2} > 0$, niin sanotaan, että prosessi on oikealle vino. Suuri osa arvoista on siis pienempää kuin keskiarvo. Vastaavasti, jos vinous olisi negatiivinen, niin prosessi olisi vasemmalle vino ja suuri osa arvoista olisi suurempaa kuin keskiarvo.

Tässä käytetty huipukkuusluku $\gamma_2(X) = \frac{\mathbb{E}\left[(X-\mathbb{E}X)^4\right]}{(Var\, X)^2} - 3$ mittaa jakauman huipukkutta suhteessa normaalijakaumaan, jolle huipukkuus on 0. Huipukkuuden voisi määritellä myös kaavalla $\hat{\gamma}_2(X) = \frac{\mathbb{E}\left[(X-\mathbb{E}X)^4\right]}{(Var\, X)^2}$, jolloin normaalijakauman huipukkuus olisi 3. Juuri tästä tulee huipukkuuden kaavassa esiintyvä termi -3. Jos huipukkuusluku on positiivinen kuten Poisson-prosessilla, niin prosessi on normaalijakaumaa terävähuippuisempi ja vastaavasti jos huipukkuus olisi negatiivinen, niin jakauma olisi laakeampi kuin normaalijakauma.

2.2. Ominaisuuksia

Seuraavaksi käydään läpi joitakin Poisson-prosessin tärkeimpiä ominaisuuksia. Luvun lopussa on esimerkkejä, joissa näitä päästään hyödyntämään.

LAUSE 2.13. *(Summautuvuus)*
Olkoon prosessit N_j riippumattomia Poisson-prosesseja, joilla vastaavat keskiarvofunktiot ovat μ_j kaikilla $j = 1, 2, ..., n$ ja merkitään $M(t) = \sum_{j=1}^{n} N_j(t)$. Tällöin $M(t)$ on Poisson-prosessi keskiarvofunktiolla $\sum_{j=1}^{n} \mu_j(t)$.

TODISTUS. Katso Lauseen 3.6. todistus. □

Seuraavaksi osoitettava Markov-ominaisuus tarkoittaa, että saapumisten lukumäärä seuraavalla hetkellä riippuu vain tästä hetkestä, ei aikaisemmista saapumisista.

LAUSE 2.14. *(Markov-ominaisuus)*
Olkoon N Poisson-prosessi. Tällöin mille tahansa jonolle $(t_i)_{i=0}^{n}$, jolle $0 = t_0 < t_1 < ... < t_n$, ja ei-vähenevälle luonnollisten lukujen jonolle $(k_i)_{i=1}^{n}$, $k_i \geq 0$ kaikilla $i = 1, 2, ..., n$ ja $n \geq 2$ pätee

$$\mathbb{P}\big(N(t_n) = k_n | N(t_1) = k_1, ..., N(t_{n-1}) = k_{n-1}\big) = \mathbb{P}\big(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}\big).$$

TODISTUS. (1) Osoitetaan ensin, että

$$\mathbb{P}\big(N(t_n) = k_n | N(t_1) = k_1, ..., N(t_{n-1}) = k_{n-1}\big) = \mathbb{P}\big(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}\big).$$

Tehdään tämä suoraan laskemalla:

$$\begin{aligned}
&\mathbb{P}(N(t_n) = k_n | N(t_1) = k_1, ..., N(t_{n-1}) = k_{n-1}) \\
&\overset{(i)}{=} \frac{\mathbb{P}(N(t_1) = k_1, ..., N(t_n) = k_n)}{\mathbb{P}(N(t_1) = k_1, ..., N(t_{n-1}) = k_{n-1})} \\
&\overset{(ii)}{=} \frac{\mathbb{P}(N(t_1) - N(t_0) = k_1 - k_0, ..., N(t_n) - N(t_{n-1}) = k_n - k_{n-1})}{\mathbb{P}(N(t_1) - N(t_0) = k_1 - k_0, ..., N(t_{n-1}) - N(t_{n-2}) = k_{n-1} - k_{n-2})} \\
&\overset{(iii)}{=} \frac{\mathbb{P}(N(t_1) - N(t_0) = k_1 - k_0), ..., \mathbb{P}(N(t_n) - N(t_{n-1}) = k_n - k_{n-1})}{\mathbb{P}(N(t_1) - N(t_0) = k_1 - k_0), ..., \mathbb{P}(N(t_{n-1}) - N(t_{n-2}) = k_{n-1} - k_{n-2})} \\
&= \mathbb{P}(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}),
\end{aligned}$$

missä (i) seuraa ehdollisen odotusarvon laskukaavasta , (ii) tiedosta $0 = N(0) = N(t_0) = k_0$ ja (iii) Poisson-prosessin lisäysten riippumattomuudesta.

(2) Osoitetaan sitten, että

$$\mathbb{P}\big(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}\big) = \mathbb{P}\big(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}\big).$$

Tehdään tämä käyttämällä samoja ominaisuuksia kuin kohdassa (1):

$$\begin{aligned}
&\mathbb{P}(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}) \\
&\overset{(i)}{=} \frac{\mathbb{P}(N(t_n) = k_n, N(t_{n-1}) = k_{n-1})}{\mathbb{P}(N(t_{n-1}) = k_{n-1})} \\
&\overset{(ii)}{=} \frac{\mathbb{P}(N(t_{n-1}) - N(t_0) = k_{n-1} - k_0, N(t_n) - N(t_{n-1}) = k_n - k_{n-1})}{\mathbb{P}(N(t_{n-1}) - N(t_0) = k_{n-1} - k_0)} \\
&\overset{(iii)}{=} \frac{\mathbb{P}(N(t_{n-1}) - N(t_0) = k_{n-1} - k_0)\,\mathbb{P}(N(t_n) - N(t_{n-1}) = k_n - k_{n-1})}{\mathbb{P}(N(t_{n-1}) - N(t_0) = k_{n-1} - k_0)} \\
&= \mathbb{P}(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}).
\end{aligned}$$

Väite seuraa yhdistämällä kohdat (1) ja (2). □

Viimeinen tässä käytävä tärkeä Poisson-prosessin ominaisuus on järjestystunnuslukuominaisuus. Järjestetyllä otoksella tarkoitetaan yksinkertaisesti sitä, että havainnot on järjestetty suuruusjärjestykseen pienimmästä alkaen. Järjestettyjä otoksia käytetään paljon parametrittomissa malleissa, mutta niitä käytetään myös Poisson-prosessien yhteydessä, jos esimerkiksi halutaan laskea saapumishetkien ehdollinen jakauma, kun niiden lukumäärä tiedetään. Ennen järjestystunnuslukuominaisuuden tarkkaa muotoilua ja todistusta osoitetaan siinä tarvittava aputulos, joka koskee järjestetyn otoksen yhteistiheysfunktiota sekä kerrataan tarvittava Radon-Nikodym-lause.

Lause 2.15. *(Radon-Nikodym)*
Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja μ avaruuden $(\Omega, \mathcal{F})$ merkkimitta, joka on absoluuttisesti jatkuva todennäköisyysmitan $\mathbb{P}$ suhteen. Tällöin on olemassa mitallinen $L : \Omega \to \mathbb{R}$ siten, että $\int_\Omega |L| \, d\mathbb{P} < \infty$ ja

$$\mu(A) = \int_A L \, d\mathbb{P} \qquad \textit{kaikilla } A \in \mathcal{F}.$$

Mitallinen kuvaus L on yksikäsitteinen siinä mielessä, että jos L ja L' ovat molemmat Radon-Nikodym -lauseen toteuttavia satunnaismuuttujia, niin $\mathbb{P}(L \neq L') = 0$.

Todistus. Katso esimerkiksi [2], luku 7. □

Lemma 2.16. *(Järjestetyn otoksen yhteistiheysfunktio)*
Jos riippumattomasti samoin jakautuneilla satunnaismuuttujilla X_i on sama tiheysfunktio f, niin tällöin niiden yhteistiheysfunktio on

$$f_{X_{(1)},...,X_{(n)}}(x_1, ..., x_n) = n! \prod_{i=1}^{n} f(x_i) \mathbb{1}_{\{x_1 < ... < x_n\}}.$$

Huomautus 2.17. Järjestetyn otoksen määritelmän nojalla vektorin $(X_{(1)}, ...X_{(n)})$ kantaja on joukossa

$$C_n = \{(x_1, ..., x_n) : x_1 \leq ... \leq x_n\} \subset \mathbb{R}^n,$$

ja tämän vuoksi tiheysfunktio $f_{X_{(1)},...,X_{(n)}}$ katoaa joukon C_n ulkopuolella. Koska satunnaismuuttujien X_i tiheysfunktion olemassaolosta seuraa, että riippumattomasti samoin jakautuneen otoksen $X_1, ..., X_n$ komponentit ovat erisuuria melkein varmasti, niin joukon C_n määritelmässä olevat epäyhtälöt $\leq$ voitaisiin korvata aidoilla epäyhtälöillä $<$.

Todistetaan sitten Lemma 2.16:

Todistus. Osoitetaan ensin, että riippumattomasti samoin jakautuneet satunnaismuuttujat, joilla on olemassa yhteistiheysfunktio, ovat erisuuria. Kun merkitään

$$\tilde{\Omega} = \{X_{(1)}, ..., X_{(n)}\} = \{X_i \neq X_j \text{ kaikilla } 1 \leq i < j \leq n\},$$

niin väite saadaan muotoon $\mathbb{P}(\tilde{\Omega}) = 1$, joka on yhtäpitävä väitteen $\mathbb{P}(\tilde{\Omega}^C) = 0$ kanssa. Koska väitteet ovat yhtäpitäviä, niin riittää osoittaa vain toinen. Osoitetaan väite joukolle $\tilde{\Omega}^C$:

$$\mathbb{P}(\tilde{\Omega}^C) = \mathbb{P}\left(\bigcup_{1\leq i<j\leq n} \{X_i = X_j\}\right) \leq \sum_{1\leq i<j\leq n} \mathbb{P}(X_i = X_j) = 0.$$

Tämä seuraa suoraan siitä, että kaikilla $i \neq j$ on

$$\mathbb{P}(X_i = X_j) = \mathbb{E}[\mathbb{P}(X_i = X_j | X_j)] = \int_{\mathbb{R}} \mathbb{P}(X_i = y) f(y)\, dy = 0,$$

koska $\mathbb{P}(X_i = y) = \int_{\{y\}} f(z)\, dz = 0$.

Osoitetaan sitten varsinainen Lemman väite. Käytetään joukon $\{1, ..., n\}$ järjestysten π joukolle merkintää Π_n. Kiinnitetään sitten arvot $x_1 < ... < x_n$. Tällöin

$$\mathbb{P}(X_{(1)} \leq x_1, ..., X_{(n)} \leq x_n) = \mathbb{P}\left(\bigcup_{\pi\in\Pi_n} A_\pi\right), \tag{2.3}$$

missä

$$A_\pi = \{X_{\pi(i)} = X_{(i)}, i = 1, ..., n\} \cap \tilde{\Omega} \cap \{X_{\pi(1)} \leq x_1, ..., X_{\pi(n)} \leq x_n\}.$$

Yhtälö (2.3) tarkoittaa, että järjestetty otos $X_{(1)} < ... < X_{(n)}$ olisi voinut tulla mistä tahansa järjestetystä otoksesta $X_{\pi(1)} < ... < X_{\pi(n)}$ kun $\pi \in \Pi_n$. Tässä kuitenkin käytetään myös tietoa, että kaikki muuttujat X_i ovat erisuuria, jolloin itseasiassa vain yksi joukon A_π järjestyksistä kelpaa. Koska joukot A_π ovat erillisiä, saadaan

$$\mathbb{P}\left(\bigcup_{\pi\in\Pi_n} A_\pi\right) = \sum_{\pi\in\Pi_n} \mathbb{P}(A_\pi).$$

Edelleen, koska muuttujat X_i ovat riippumattomasti samoin jakautuneita, niin

$$\begin{aligned}
\mathbb{P}(A_\pi) &= \mathbb{P}\big((X_{\pi(1)}, ..., X_{\pi(n)}) \in C_n \cap (\infty, x_1] \times \cdots \times (-\infty, x_n]\big) \\
&= \mathbb{P}\big((X_{(1)}, ..., X_{(n)}) \in C_n \cap (\infty, x_1] \times \cdots \times (-\infty, x_n]\big) \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f(y_i) \mathbb{1}_{\{y_1 < \cdots < y_n\}}\, dy_n \ldots dy_1.
\end{aligned}$$

Nyt, koska joukon $\{1, ..., n\}$ alkiot voidaan järjestää $n!$ eri tavalla, on joukossa Π_n $n!$ alkiota, joten saadaan

$$\begin{aligned}&\mathbb{P}(X_{(1)} \leq x_1, ..., X_{(n)} \leq x_n)\\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} n! \prod_{i=1}^{n} f(y_i) \mathbb{1}_{\{y_1 < \cdots < y_n\}}\, dy_n \ldots dy_1.\end{aligned}$$

Radon-Nikodym -lauseen ja Huomautuksen 2.17 nojalla haluttu tiheysfunktio saadaan edellisen integraalin integrandina. Radon-Nikodym -lauseen nojalla integrandi on melkein varmasti yksikäsitteinen vektorin $\left(X_{(1)}, ..., X_{(n)}\right)$ tiheysfunktio. □

Olkoon N Poisson-prosessi keskiarvofunktiolla μ. Tällöin järjestystunnuslukuominaisuus tarkoittaa, että jos tiedetään aikavälillä $(0, \mu(t))$ tulevan n saapumista prosessin N mukaisesti, niin saapumisia voidaan simuloida ottamalla n riippumatonta arvoa jakaumasta, jonka tiheysfunktio on $\lambda(x)/\mu(t)$, $0 < x \leq t$, ja asettamalla nämä arvot suuruusjärjestykseen.

LAUSE 2.18. *(Järjestystunnuslukuominaisuus)*
Olkoon N Poisson-prosessi, jolla on jatkuva ja melkein kaikkialla positiivinen intensiteettifunktio λ, ja saapumishetket $0 < T_1 < T_2 < ...$ m.v.. Olkoon lisäksi satunnaismuuttujat $X_1, ..., X_n$ riippumattomasti samoin jakautuneita siten, että tiheysfunktio kaikilla X_i, $i = 1, 2, ..., n$, on $\lambda(x)/\mu(t)$, $0 < x \leq t$. Tällöin vektorin $(T_1, ..., T_n)$ ehdollinen jakauma annetulla $N(t) = n$ on järjestetyn otoksen $(X_{(1)}, ..., X_{(n)})$ yhteistiheysfunktio eli

$$(T_1, ..., T_n | N(t) = n) \stackrel{d}{=} \left(X_{(1)}, ..., X_{(n)}\right).$$

Toisin sanoen siis vasemman puolen vektorin ehdollinen tiheysfunktio on

$$f_{T_1,...,T_n}(x_1, ..., x_n | N(t) = n) = \frac{n!}{(\mu(t))^n} \prod_{i=1}^{n} \lambda(x_i), \tag{2.4}$$

missä $0 < x_1 < ... < x_n < t$.

TODISTUS. Osoitetaan, että raja-arvo

$$\lim_{\substack{h_i \downarrow 0 \\ i=1,...,n}} \frac{\mathbb{P}\left(T_1 \in (x_1, x_1 + h_1], ..., T_n \in (x_n, x_n + h_n] | N(t) = n\right)}{h_1 \ldots h_n} \tag{2.5}$$

on olemassa ja se on muuttujien x_i jatkuva funktio. Täysin vastaavasti osoitettaisiin samanlainen väite väleille $(x_i - h_i, x_i]$ ja saataisiin sama raja-funktio. Koska todistus on sama, niin tehdään se vain väleille $(x_i, x_i + h_i]$. Raja-arvo voidaan tulkita tiheysfunktiona saapumisaikojen $(T_1, ..., T_n)$ ehdolliselle todennäköisyysjakaumalle ehdolla $\{N(t) = n\}$.

Koska $0 < x_1 < ... < x_n < t$, niin muuttujat h_i voidaan valita niin pieniksi, että väleistä $(x_i, x_i + h_i] \subset [0, t]$, kun $i = 1, ..., n$, tulee erillisiä. Tällöin saadaan suoraan seuraava yhtälö:

$$\begin{aligned}&\{T_1 \in (x_1, x_1 + h_1], ..., T_n \in (x_n, x_n + h_n], N(t) = n\}\\ =&\{N(0, x_1] = 0, N(x_1, x_1 + h_1] = 1, N(x_1 + h_1, x_2] = 0,\\ &N(x_2, x_2 + h_2] = 1, ..., N(x_{n-1} + h_{n-1}, x_n] = 0,\\ &N(x_n, x_n + h_n] = 1, N(x_n + h_n, t] = 0\}.\end{aligned}$$

Otetaan seuraavaksi molemmilta puolilta todennäköisyydet ja käytetään yhtälön oikealle puolelle Poisson-prosessin N riippumattomien lisäysten ominaisuutta. Saadaan

$$\begin{aligned}
&\mathbb{P}(T_1 \in (x_1, x_1+h_1], \ldots, T_n \in (x_n, x_n+h_n], N(t)=n)\\
=&\mathbb{P}(N(0,x_1]=0)\,\mathbb{P}(N(x_1,x_1+h_1]=1)\,\mathbb{P}(N(x_1+h_1,x_2]=0)\\
&\mathbb{P}(N(x_2,x_2+h_2]=1)\ \ldots\ \mathbb{P}(N(x_{n-1}+h_{n-1},x_n]=0)\\
&\mathbb{P}(N(x_n,x_n+h_n]=1)\,\mathbb{P}(N(x_n+h_n,t]=0)\\
=&e^{-\mu(x_1)}\left[\mu(x_1,x_1+h_1]\,e^{-\mu(x_1,x_1+h_1]}\right]e^{-\mu(x_1+h_1,x_2]}\\
&\left[\mu(x_2,x_2+h_2]\,e^{-\mu(x_2,x_2+h_2]}\right]\ldots e^{-\mu(x_{n-1}+h_{n-1},x_n]}\\
&\left[\mu(x_n,x_n+h_n]\,e^{-\mu(x_n,x_n+h_n]}\right]e^{-\mu(x_n+h_n,t]}\\
=&e^{-\mu(t)}\,\mu(x_1,x_1+h_1]\ldots\mu(x_n,x_n+h_n].
\end{aligned} \tag{2.6}$$

Lasketaan sitten raja-arvo (2.5) käyttäen ehdollisen todennäköisyyden kaavaa ja tietoa $\mathbb{P}(N(t)=n) = e^{-\mu(t)}\frac{\mu(t)^n}{n!}$:

$$\begin{aligned}
&\frac{\mathbb{P}\left(T_1 \in (x_1,x_1+h_1],...,T_n \in (x_n,x_n+h_n] | N(t)=n\right)}{h_1\ldots h_n}\\
=&\frac{\mathbb{P}\left(T_1 \in (x_1,x_1+h_1],...,T_n \in (x_n,x_n+h_n], N(t)=n\right)}{h_1\ldots h_n\,\mathbb{P}(N(t)=n)}\\
=&\frac{e^{-\mu(t)}\,\mu(x_1,x_1+h_1]\ldots\mu(x_n,x_n+h_n]}{h_1...h_n\,e^{-\mu(t)}\frac{\mu(t)^n}{n!}} \qquad \text{(yhtälö (2.6))}\\
=&\frac{n!}{\mu(t)^n}\,\frac{\mu(x_1,x_1+h_1]}{h_1}\ldots\frac{\mu(x_n,x_n+h_n]}{h_n}\\
\to&\ \frac{n!}{\mu(t)^n}\,\lambda(x_1)\ldots\lambda(x_n), \qquad \text{kun } h_i \downarrow 0 \text{ ja } i=1,...,n.
\end{aligned}$$

Tämä on haluttu yhtälö (2.4). Todistuksen viimeinen vaihe seuraa Määritelmästä 2.6. ja intensiteettifunktion λ jatkuvuudesta, jolloin $\mu'(x_i) = \lambda(x_i)$. □

Seuraus 2.19. *(Järjestystunnuslukuominaisuus homogeeniselle Poisson-prosessille) Olkoon N homogeeninen Poisson-prosessi intensiteetillä $\lambda > 0$. Tällöin Lauseen 2.18. mukaan saapumisten T_i ehdollinen yhteistiheysfunktio on*

$$f_{T_1,...,T_n}(x_1,...,x_n|N(t)=n) = \frac{n!}{t^n}, \qquad \textit{kun } 0 < x_1 < ... < x_n < t.$$

Todistus. Koska N on homogeeninen Poisson-prosessi, niin sen intensiteettifunktio on $\lambda = \lambda(x_i)$ kaikilla $i = 1,2,...,n$ ja keskiarvofunktio $\mu(t) = \lambda t$. Nyt Lauseen 2.18. nojalla saadaan

$$f_{T_1,...,T_n}(x_1,...,x_n|N(t)=n) = \frac{n!}{(\lambda t)^n}\prod_{i=1}^{n}\lambda = \frac{n!}{t^n}, \qquad \text{kun } 0 < x_1 < ... < x_n < t.$$

□

Erityisesti Seurauksesta 2.19. saadaan, että homogeenisen Poisson-prosessin järjestystunnuslukuominaisuus ei riipu intensiteetistä λ.

2.3. Esimerkkejä

Esimerkki 2.20. ([4] Ongelma 5.1.7.) Järjestelmään tulee häiriöitä homogeenisen Poisson-prosessin N mukaisesti. Oletetaan että järjestelmä selviää jokaisesta häiriöstä muista häiriöistä riippumatta todennäköisyydellä α eli todennäköisyys, että järjestelmä selviää k kappaleesta häiriöitä on α^k. Mikä on todennäköisyys että järjestelmä toimii hetkellä t?

Ratkaisu:

Järjestelmä selviää, jos yhtään häiriötä ei ole tullut tai se on selviytynyt kaikista häiriöistä. Todennäköisyys, että yhtään häiriötä ei ole tullut on

$$\mathbb{P}(N(t)=0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Odotusarvo saapuneiden häiriöiden lukumäärälle ja että järjestelmä on selviytynyt kaikista on

$$\begin{aligned}\sum_{k=0}^{\infty} k\,\mathbb{P}(N(t)=k)\,\alpha^k &= \sum_{k=0}^{\infty} k\,e^{-\lambda t}\frac{(\lambda t)^k}{k!}\,\alpha^k \\ &= e^{-\lambda t}\sum_{k=0}^{\infty} k\,\frac{(\alpha\lambda t)^k}{k!} \\ &= \alpha\lambda t e^{-\lambda t}\sum_{k=1}^{\infty}\frac{(\alpha\lambda t)^{k-1}}{(k-1)!} \\ &= \alpha\lambda t e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\alpha\lambda t)^k}{k!} \\ &= \alpha\lambda t e^{-\lambda t}\,e^{\alpha\lambda t} \\ &= \alpha\lambda t e^{\alpha}.\end{aligned}$$

Todennäköisyydeksi, että järjestelmä toimii hetkellä t saadaan siis $e^{-\lambda t} + \alpha\lambda t e^{\alpha}$.

Esimerkki 2.21. ([4] Tehtävä 5.1.6) Viestejä saapuu asiakaspalveluun homogeenisen Poisson-prosessin mukaisesti, keskiarvoisesti kuusi viestiä tunnissa.

(1) Mikä on todennäköisyys, että aamupäivällä klo 8-12 ei saavu yhtään viestiä?
(2) Mikä on aamun ensimmäisen viestin saapumisen jakauma?

Ratkaisu:

(1) Viestien saapuminen N_1 noudattaa siis homogeenista Poisson-prosessia parametrilla 6 joka tunti. Koska aikavälillä 8-12 on neljä tuntia, niin tällä välillä viestien saapumisen eli prosessin N_4 parametriksi saadaan Lauseen 2.13. nojalla $4 \cdot 6 = 24$. Todennäköisyys, että aamupäivällä ei saavu yhtään viestiä on siis

$$\mathbb{P}(N_4([8,12])=0) = \mathbb{P}(N_4(4)=0) = e^{-24\cdot 4}\frac{(24\cdot 4)^0}{0!} = e^{-24\cdot 4} \approx 0.$$

(2) Lauseen 2.11. nojalla homogeeninen Poisson-prosessi voidaan esittää muodossa $N(t) = \#\{i \geq 1, T_i \leq t\}$, kun $t \geq 0$, $T_n = \varepsilon_1 + ... + \varepsilon_n$, $\varepsilon_1, \varepsilon_2, ...$ ovat riippumattomia ja $\varepsilon_1, \varepsilon_2, ... \sim Exp(6)$. Koska tiheysfunktio määrittää jakauman yksikäsitteisesti, niin lasketaan se käyttämällä prosessille yllä olevaa

muotoilua:

$$\mathbb{P}(N(t)=1)=\mathbb{P}(T_1 \le t)=\mathbb{P}(\varepsilon_1 \le t),$$

missä yhtälöketjun oikea puoli noudattaa eksponenttijakaumaa parametrilla 6 Lauseen 2.11. mukaan ja siten aamun ensimmäisen viestin saapumisen jakauma on eksponenttijakauma parametrilla 6.

Seuraavassa esimerkissä tutustutaan erääseen tunnettuun homogeeniseen Poisson-prosessiin liittyvään paradoksiin. Paradoksi tunnetaan muun muassa liftarin paradoksina ja englannin kielisissä teoksissa nimellä the inspection paradox. Koska liftarin tilanteessa ei voi tietää pääsekö hän seuraavan ohiajavan auton kyytiin, niin tässä versiossa autot on vaihdettu linja-autoiksi.

Esimerkki 2.22. (*Odottajan paradoksi*)
Linja-autoja kulkee pysäkin ohi homogeenisen Poisson-prosessin mukaisesti ja linja-autojen keskimääräinen väliaika on 10 minuuttia. Henkilö tulee pysäkille satunnaisella ajanhetkellä.

(1) Milloin edellinen linja-auto meni ohi?
(2) Mikä on henkilön keskimääräinen odotusaika seuraavan auton saapumiseen? Viisi minuuttia?

Ratkaisu:
Lauseen 2.11. nojalla linja-autojen saapumisten väliset ajat $\varepsilon_n = T_n - T_{n-1}$, $n \ge 1$, ja $T_0 = 0$, muodostavat jonon riippumattomasti samoin eksponentiaalisesti jakautuneita satunnaismuuttujia, joille $\mathbb{E}\,\varepsilon_n = 10 = \frac{1}{1/10}$, eli $\varepsilon_n \sim Exp(1/10)$. Merkitään, että henkilö tulee pysäkille hetkellä t. Tällöin edellinen linja-auto meni ohi hetkellä $T_{N(t)}$ ja seuraava auto menee pysäkin ohi hetkellä $T_{N(t)+1}$. Halutaan siis vastaukset seuraaviin kysymyksiin:

(1) Mikä on muuttujan $B(t) = t - T_{N(t)}$ jakauma, eli mikä on aikavälin $(T_{N(t)}, t]$ pituus, siis kuinka kauan sitten edellinen auto meni ohi?
(2) Mikä on muuttujan $F(t) = T_{N(t)+1} - t$ jakauma, eli mikä on aikavälin $(t, T_{N(t)+1}]$ pituus, siis kuinka kauan pysäkillä täytyy odottaa seuraavaa autoa?

Suuretta $B(t)$ käytetään usein viitattaessa edellisestä tapahtumasta kuluneeseen aikaan ja $F(t)$ kun viitataan aikaan, joka kuluu ennen seuraavaa tapahtumaa.

Intuitiivisesti, koska ajanhetki t on jossakin kahden auton saapumisajan välissä ja koska autojen saapumisten väliset ajat ovat riippumattomasti samoin eksponentiaalisesti jakautuneita, niin voisi ehkä odottaa että $\mathbb{P}(B(t) \le x_1) < 1 - e^{-\lambda x_1}$, kun $x_1 < t$ ja vastaavasti $\mathbb{P}(F(t) \le x_2) < 1 - e^{-\lambda x_2}$, kun $x_2 > 0$. Tämä ei kuitenkaan pidä paikkaansa, kuten seuraavaksi huomataan laskemalla yhteiskertymäfunktio

$$G_{B(t),F(t)}(x_1,x_2) = \mathbb{P}(B(t) \le x_1, F(t) \le x_2), \qquad \text{kun } x_1, x_2 \ge 0.$$

Koska muuttujan $B(t)$ määritelmän nojalla $B(t) \le t$ m.v., niin käsitellään tapaukset $x_1 < t$ ja $x_1 \ge t$ erikseen. Aloitetaan tapauksesta $x_1 < t$ ja $x_2 > 0$. Kirjoitetaan ensin joukot $\{B(t) \le x_1\}$ ja $\{F(t) \le x_2\}$ toisella tavalla, kun huomioidaan ehto $T_{N(t)} \le t < T_{N(t)+1}$:

$$\{B(t) \le x_1\} = \{t - T_{N(t)} \le x_1\} = \{t - x_1 \le T_{N(t)} \le t\} = \{N(t - x_1, t] \ge 1\},$$
$$\{F(t) \le x_2\} = \{T_{N(t)+1} - t \le x_2\} = \{t < T_{N(t)+1} \le t + x_2\} = \{N(t, t + x_2] \ge 1\}.$$

Koska homogeenisen Poisson-prosessin lisäykset ovat riippumattomia ja stationaarisia, niin saadaan

$$
\begin{aligned}
G_{B(t),F(t)}(x_1, x_2) &= \mathbb{P}(N(t - x_1, t] \geq 1,\ N(t, t + x_2] \geq 1) \\
&= \mathbb{P}(N(t - x_1, t] \geq 1)\ \mathbb{P}(N(t, t + x_2] \geq 1) \\
&= \mathbb{P}(N(x_1) \geq 1)\ \mathbb{P}(N(x_2) \geq 1) \\
&= [1 - \mathbb{P}(N(x_1) = 0)]\ [1 - \mathbb{P}N(x_2) = 0)] \\
&= \left(1 - e^{-\lambda x_1}\right)\ \left(1 - e^{-\lambda x_2}\right).
\end{aligned}
$$

Tapaukselle $x_1 \geq t$ ja $x_2 > 0$ joukko $\{F(t) \leq x_2\}$ on sama kuin edellä, koska sen ehto ei ole muuttunut. Koska tiedetään, että $B(t) \leq t$ m.v., niin nyt $\mathbb{P}(B(t) \leq x_1) = 1$. Yhdistämällä tapaukset $x_1 < t$ ja $x_1 \geq t$ saadaan

$$
G_{B(t),F(t)}(x_1, x_2) = \left[\,\left(1 - e^{-\lambda x_1}\right) \mathbb{1}_{[0,t)}(x_1) + \mathbb{1}_{[t,\infty)}(x_1)\right] \left(1 - e^{-\lambda x_2}\right).
$$

Koska $B(t)$ ja $F(t)$ ovat riippumattomia, niin yhteiskertymäfunktiosta nähdään, että satunnaismuuttuja $B(t)$ noudattaa katkaistua eksponenttijakaumaa, jossa katkaisu on hetkellä t, eli

$$
\mathbb{P}(B(t) \leq x_1) = 1 - e^{-\lambda x_1},\ x_1 < t, \qquad \text{ja} \qquad \mathbb{P}(B(t) = t) = e^{-\lambda t}.
$$

Satunnaismuuttuja $F(t)$ noudattaa samaa jakaumaa kuin Poisson-prosessin N saapumisten väliset ajat ε_i eli $F(t) \sim Exp(\lambda)$. Tämä ominaisuus liittyy läheisesti eksponenttijakauman muistittomuuteen ja lisäksi Poisson-prosessin lisäysten riippumattomuuteen.

Nyt voidaan vastata esitettyihin kysymyksiin:

(1) Edellisestä linja-autosta kulunut aika on odotusarvoisesti

$$
\begin{aligned}
\mathbb{E}\, B(t) &= \int_0^t x\lambda e^{-\lambda x}\, dx + te^{-\lambda t} \\
&= -\Big/_{0}^{t} xe^{-\lambda x} + \int_0^t e^{-\lambda x}\, dx + te^{-\lambda t} \\
&= \frac{1 - e^{-\lambda t}}{\lambda} \\
&\overset{\lambda=1/10}{=} 10 - 10\, e^{-t/10}
\end{aligned}
$$

eli edellisestä autosta kulunut aika riippuu siitä, milloin odottaja on saapunut pysäkille.

(2) Seuraavan linja-auton saapumiseen on odotusarvoisesti

$$
\mathbb{E}\, F(t) = \frac{1}{1/10} = 10 \text{ minuuttia},
$$

koska F(t) noudattaa eksponenttijakaumaa parametrilla 1/10. Eksponenttijakauman muistittomuudesta seuraa, että aika seuraavan linja-auton saapumiseen on samalla tavalla eksponenttijakautunut riippumatta siitä, paljonko aikaa edellisestä autosta on kulunut ja erityisesti se ei riipu henkilön saapumisajasta. Siten pysäkillä olevan henkilön odotusajan odotusarvo on 10 minuuttia, ei suinkaan viisi minuuttia.

Esimerkki 2.23. Pankkiin tulee aamupäivällä asiakkaita homogeenisen Poisson-prosessin mukaisesti keskimäärin 4 asiakasta 15 minuutissa. Välillä 9:00-9:15 asiakkaita tuli 7, seuraavan vartin aikana 3 ja seuraavan 4. Mikä on todennäköisyys, että välillä 9:45-10:00 asiakkaita tulee enemmän kuin 5, kun asiakkaiden saapumiset välillä 9:00-9:45 tunnetaan?

Ratkaisu:

Merkitään kellonaikoja siten, että 9:00=: t_0, 9:15=: t_1, ... , 10:00=: t_4. Saadaan

$$\begin{aligned}
&\mathbb{P}\left(N(t_3,t_4] > 5 | N(t_0,t_1] = 7, N(t_1,t_2] = 3, N(t_2,t_3] = 4\right)\\
=&\mathbb{P}\left(N(t_3,t_4] > 5 | N(t_2,t_3] = 4\right) \qquad \text{(Markov-ominaisuus)}\\
=&\frac{\mathbb{P}(N(t_3,t_4] > 5)\,\mathbb{P}(N(t_2,t_3] = 4)}{\mathbb{P}(N(t_2,t_3] = 4)}\\
=&1 - \mathbb{P}(N(t_3,t_4] \leq 5),
\end{aligned}$$

missä $\mathbb{P}(N(t_3,t_4] = k) \sim Pois(4)$. Kysytyksi todennäköisyydeksi saadaan siten

$$1 - \sum_{k=0}^{5} e^{-4}\frac{4^k}{k!} \approx 0,215.$$

Järjestystunnuslukuominaisuuden soveltamiseen palataan painotettujen Poisson-prosessien esimerkkien yhteydessä luvussa kolme.

LUKU 3

Painotettu Poisson-prosessi

Tässä luvussa määritellään painotettu Poisson-prosessi ja verrataan sen ominaisuuksia edellisessä luvussa esitetyn ei-painotetun Poisson-prosessin ominaisuuksiin.

MÄÄRITELMÄ 3.1. *Olkoon $\tilde{N}$ standardi homogeeninen Poisson-prosessi ja $\mu : [0,\infty) \to [0,\infty)$ Poisson-prosessin N keskiarvofunktio. Olkoon $\theta > 0$ m.v. ei-degeneroitunut satunnaismuuttuja, joka on riippumaton prosessista $\tilde{N}$. Tällöin prosessia*

$$N(t) = \tilde{N}(\theta\mu(t)), \qquad t \geq 0,$$

kutsutaan painotetuksi Poisson-prosessiksi painotusmuuttujalla θ.

Jos painotetun Poisson-prosessin $\hat{N}$ painotusmuuttuja θ on kiinnitetty, eli $\theta = \xi$ jollakin vakiolla ξ, niin prosessi $\hat{N}$ on tavallinen ei-painotettu Poisson-prosessi parametrilla $\xi\mu(t)$. Lisäksi, jos havaitaan vain yksi painotetun Poisson-prosessin polku $\tilde{N}(\theta(\omega)\mu(t),\omega)$, niin sitä ei pystyisi erottamaan ei-painotetun Poisson-prosessin, jonka keskiarvofunktio on $\theta(\omega)\mu$, polusta. Kun polkuja on useita, esimerkiksi 10, niin tällöin huomataan ero polkujen vaihtelun suuruudessa.

Vakuutusten yhteydessä painotettua Poisson-prosessia käytetään korvausvaatimusten lukumäärän prosessina, jos ei uskota yhden tietyn Poisson-prosessin mallintavan saapumisten lukumäärää hyvin. Ajatellaan esimerkiksi autovakuutusten korvausvaatimusten lukumäärän prosessia kokoelmana itsenäisiä polkuja, jotka vastaavat eri henkilöitä, jotka ovat ottaneet autovakuutuksen. Tällöin satunnaismuuttujan θ voisi ajatella vastaavan eri henkilöiden ominaisuuksia, kuten esimerkiksi ajotaitoa, ajokokemusta ja ikää, jotka vaikuttavat korvausvaatimusten todennäköisyyteen.

Miten painotusmuuttuja sitten valitaan? Painotusmuuttujan jakauman tulee pohjautua aiempiin havaintoihin (tai odotuksiin) tapahtumasta, kun huomioidaan myös ympäristötekijät. Seuraavat kolme tapaa ovat yleisesti painotusmuuttujan valinnassa käytettyjä tekniikoita ([1]):

(1) Painotusmuuttuja θ on analyyttisessä muodossa. Käytännössä painotusmuuttujasta on usein vain vähän tietoa, jolloin tiheysfunktion voi joutua määrittämään vain vaihtelun perusteella. Jos funktion oikeellisuudelle saa tukea vastaavanlaisesta aiemmasta tapauksesta, niin analyyttistä muotoa kannattaa käyttää. Painotusmuuttujan jakaumana käytetään usein Gamma-jakaumaa tai negatiivista binomijakaumaa.
(2) Painotusmuuttuja on taulukkomuodossa. Arvojoukko jaetaan sopivan mittaisiin väleihin ja jokaiselle välille annetaan oma painotusmuuttuja. Ajatellaan esimerkiksi kesän metsäpaloja, joihin vaikuttaa sää; jos kesä on hyvin kuiva, niin metsäpaloja on enemmän kuin normaalisti ja sadekesänä paloja on

vähemmän. Painotusmuuttujan taulukko voisi olla esimerkiksi ([1]):

Säätyyppi	Painotusmuuttuja θ
Erittäin kuiva	$3,00$
Kuiva	$1,75$
Normaali	$0,80$
Sateinen	$0,60$
Erittäin sateinen	$0,30$

Menetelmää voidaan käyttää, jos tilastollista tietoa on runsaasti.

(3) Painotusmuuttuja saadaan momenttimenetelmällä. Tarkkaa painotusmuuttujan funktiota ei tiedetä, mutta tärkeimmät tunnusluvut, erityisesti keskiarvo, keskihajonta ja vinous, tunnetaan. Tunnusluvut saadaan nimittäin laskettua suoraan empiirisestä aineistosta, vaikka aineistoa olisi vain niukasti. Näin ollen tarkkaa tiheysfunktiota ei tarvita.

Palataan vielä edellä olleeseen autovakuutustilanteeseen. Vakuutusyhtiöllä on olemassa korvaustilastot usean kymmenen vuoden ajalta, joista voi selvittää keille yleisimmin sattuu korvausvaatimukseen johtavia tilanteita. Voisi siis ajatella menetelmän 2 sopivan painotusmuuttujan jakauman valintaan. Tässä tapauksessa voisi myös ajatella nuorilla, vasta ajokortin saaneilla olevan paljon suurempi aineistosta laskettu painotusmuuttuja kuin yli 15 vuotta ajaneilla henkilöillä.

3.1. Tunnuslukuja

Lasketaan seuraavaksi painotetulle Poisson-prosessille vastaavat tunnusluvut kuin ei-painotetulle. Todistetaan ne momentit generoivan funktion avulla, joka määritellään satunnaismuuttujalle X olevan

$$m_X(z) = \mathbb{E}\left[e^{zX}\right].$$

Samalla saadaan todistettua ei-painotetun Poisson-prosessin tunnuslukujen kaavat.

Lause 3.2. *Olkoon N painotettu Poisson-prosessi keskiarvofunktiolla μ ja painotusmuuttujalla θ. Tällöin prosessille N pätee*

(1) *odotusarvo* $\mathbb{E}N(t) = \mu(t)\mathbb{E}\theta$

(2) *varianssi* $Var\, N(t) = \mathbb{E}N(t) + \mu(t)^2 Var\,\theta$

(3) *vinous*

$$\gamma_1(N(t)) = \frac{Var\, N(t) - \mathbb{E}N(t)^3 + 2\mu(t)^2 Var\,\theta + \mu(t)^3\left(\mathbb{E}(\theta^3) - 3\mathbb{E}\theta\, Var\,\theta\right)}{Var\, N(t)^{3/2}}$$

(4) *huipukkuus*

$$\begin{aligned}\gamma_2(N(t)) =& \Big(\mathbb{E}N(t) + \mu(t)^2\left[7\mathbb{E}(\theta^2) - 4(\mathbb{E}\theta)^2\right] + \mu(t)^3\left[6\mathbb{E}(\theta^3) - 12\mathbb{E}\theta\mathbb{E}(\theta^2) + 6(\mathbb{E}\theta)^3\right] \\ &+ \mu(t)^4\left[\mathbb{E}(\theta^4) - 4\mathbb{E}\theta\mathbb{E}(\theta^3) + 6(\mathbb{E}\theta)^2\mathbb{E}(\theta^2) - 3(\mathbb{E}\theta)^4\right]\Big) \\ &\cdot \Big(Var\, N(t)\Big)^{-2} - 3.\end{aligned}$$

TODISTUS. Merkitään painotusmuuttujan θ kertymäfunktiota $\mathbb{P}(\theta \leq \xi) = F_\theta(\xi)$. Prosessin N ehdollinen todennäköisyys ehdolla $\theta = \xi$ saadaan suoraan Poisson-jakauman avulla ja se on

$$\mathbb{P}(N(t) = k | \theta = \xi) = e^{-\xi\mu(t)} \frac{(\xi\mu(t))^k}{k!}.$$

Prosessin N pistetodennäköisyys ilman ehtoa $\theta = \xi$ saadaan integroimalla kaikkien muuttujan θ mahdollisten arvojen yli eli

$$\mathbb{P}(N(t) = k) = \int_0^\infty e^{-\xi\mu(t)} \frac{(\xi\mu(t))^k}{k!} \, dF_\theta(\xi).$$

Näiden tietojen avulla lasketaan nyt prosessin N momentit generoiva funktio $m_{N(t)}$. Sen avulla saadaan osoitettua kohdat (1)-(4), sillä momentit generoivalle funktiolle pätee

$$m_X^{(k)}(0) = \mathbb{E}X^k.$$

Momentit generoivaksi funktioksi saadaan

$$\begin{aligned}
m_{N(t)}(z) &= \mathbb{E}\left[e^{zN(t)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{zN(t)} \middle| \theta\right]\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{\infty} e^{zk} e^{-\theta\mu(t)} \frac{(\theta\mu(t))^k}{k!}\right] \\
&= \mathbb{E}\left[e^{-\theta\mu(t)} \sum_{k=0}^{\infty} \frac{(e^z\theta\mu(t))^k}{k!}\right] \\
&= \mathbb{E}\left[e^{-\theta\mu(t)} e^{\theta\mu(t)e^z}\right] \\
&= \int_0^\infty e^{\xi\mu(t)(e^z-1)} \, dF_\theta(\xi) \qquad \text{(odotusarvo muuttujan } \xi \text{ suhteen)} \\
&= \mathbb{E}\left[e^{h(z)\theta}\right] \\
&= m_\theta(h(z)),
\end{aligned}$$

missä $h(z) = \mu(t)\,(e^z - 1)$. Prosessin N momentit generoiva funktio pisteessä z on siis sama kuin muuttujan θ momentit generoiva funktio pisteessä $h(z)$. Huomataan vielä, että funktiolle h pätee

$$h'(z) = \mu(t)e^z = h^{(k)}(z) \qquad \text{kaikilla } k = 1, 2, ...$$

ja lisäksi $h(0) = 0$ ja $h^{(k)}(0) = \mu(t)$. Osoitetaan sitten lauseen kohdat (1)-(4):

(1) Muuttujan odotusarvo saadaan siis laskemalla sen momentit generoivan funktion derivaatta pisteessä 0. Prosessin N momentit generoivan funktion derivaataksi saadaan

$$m'_{N(t)}(z) = h'(z)m'_\theta(h(z)) = \mu(t)e^z m'_\theta(h(z))$$

ja sijoittamalla $z = 0$ saadaan odotusarvoksi

$$\mathbb{E}N(t) = m'_{N(t)}(0) = \mu(t)m'_\theta(0) = \mu(t)\mathbb{E}\theta.$$

(2) Varianssin osoittamiseksi lasketaan ensin $\mathbb{E}N(t)^2$. Momentit generoivan funktion toiseksi derivaataksi saadaan

$$m''_{N(t)}(z) = h''(z)m'_\theta(h(z)) + h'(z)^2 m''_\theta(h(z)).$$

Sijoittamalla taas $z = 0$ saadaan

$$\mathbb{E}N(t)^2 = m''_{N(t)}(0) = \mu(t)m'_\theta(0) + \mu(t)^2 m''_\theta(0) = \mu(t)\mathbb{E}\theta + \mu(t)^2\mathbb{E}\left(\theta^2\right),$$

jolloin varianssiksi tulee

$$\begin{aligned} Var\, N(t) &= \mathbb{E}N(t)^2 - (\mathbb{E}N(t))^2 \\ &= \mu(t)\mathbb{E}\theta + \mu(t)^2\mathbb{E}\left(\theta^2\right) - \mu(t)^2\left(\mathbb{E}\theta\right)^2 \\ &= \mathbb{E}N(t) + \mu(t)^2\left(\mathbb{E}(\theta^2) - (\mathbb{E}\theta)^2\right) \\ &= \mathbb{E}N(t) + \mu(t)^2\, Var\,\theta. \end{aligned}$$

(3) Prosessin N vinous saadaan laskettua kaavalla

$$\begin{aligned} \gamma_1\left(N(t)\right) &= \mathbb{E}\left[\left(\frac{N(t) - \mathbb{E}N(t)}{(Var\, N(t))^{1/2}}\right)^3\right] \\ &= \frac{\mathbb{E}\left[N(t)^3 - 3N(t)^2\mathbb{E}N(t) + 3N(t)(\mathbb{E}N(t))^2 - (\mathbb{E}N(t))^3\right]}{(Var\, N(t))^{3/2}} \\ &= \frac{\mathbb{E}(N(t)^3) - 3\mathbb{E}N(t)\, Var\, N(t) - (\mathbb{E}N(t))^3}{(Var\, N(t))^{3/2}}. \end{aligned}$$

Tämän laskemiseksi käytetään jälleen ensiksi kaavaa $\mathbb{E}N(t)^3 = m^{(3)}_{N(t)}(0)$. Kolmanneksi derivaataksi saadaan

$$\begin{aligned} m^{(3)}_{N(t)}(z) =& h^{(3)}(z)m'_\theta(h(z)) + h'(z)h''(z)m''_\theta(h(z)) \\ &+ 2h'(z)h''(z)m''_\theta(h(z)) + h'(z)^3 m^{(3)}_\theta(h(z)) \\ =& h^{(3)}(z)m'_\theta(h(z)) + 3h'(z)h''(z)m''_\theta(h(z)) + h'(z)^3 m^{(3)}_\theta(h(z)) \end{aligned}$$

ja odotusarvoksi

$$\begin{aligned} \mathbb{E}N(t)^3 &= m^{(3)}_{N(t)}(0) \\ &= \mu(t)m'_\theta(0) + 3\mu(t)^2 m''_\theta(0) + \mu(t)^3 m^{(3)}_\theta(0) \\ &= \mu(t)\mathbb{E}\theta + 3\mu(t)^2\mathbb{E}(\theta^2) + \mu(t)^3\mathbb{E}(\theta^3). \end{aligned}$$

Vinouden laskukaavan osoittaja saadaan nyt muotoon

$$\begin{aligned} &\mathbb{E}[N(t)^3] - 3\mathbb{E}N(t)\, Var\, N(t) - (\mathbb{E}N(t))^3 \\ =& \mu(t)\mathbb{E}\theta + 3\mu(t)^2\mathbb{E}(\theta^2) + \mu(t)^3\mathbb{E}(\theta^3) \\ &- 3\mu(t)\mathbb{E}\theta(\mu(t)\mathbb{E} + \mu(t)^2\, Var\,\theta) - (\mu(t)\mathbb{E}\theta)^3 \\ =& \mu(t)\mathbb{E}\theta + 3\mu(t)^2(\mathbb{E}(\theta^2) - (\mathbb{E}\theta)^2) + \mu(t)^3(\mathbb{E}(\theta^3) - 3\mathbb{E}\theta\, Var\,\theta - (\mathbb{E}\theta)^3) \\ =& \mathbb{E}N(t) + \mu(t)^2 Var\,\theta + 2\mu(t)^2 Var\,\theta - (\mathbb{E}N(t))^3 + \mu(t)^3(\mathbb{E}(\theta^3) - 3\mathbb{E}\theta Var\,\theta), \end{aligned}$$

jolloin vinoudeksi saadaan

$$\gamma_1(N(t)) = \frac{Var\,N(t) + 2\mu(t)^2 Var\,\theta - (\mathbb{E}N(t))^3 + \mu(t)^3(\mathbb{E}(\theta^3) - 3\mathbb{E}\theta Var\,\theta)}{(Var\,N(t))^{3/2}}.$$

(4) Prosessin N huipukkuus saadaan kaavalla

$$\begin{aligned}\gamma_2(N(t)) =& \frac{\mathbb{E}\left[\left(N(t) - \mathbb{E}N(t)\right)^4\right]}{(Var\,N(t))^2} - 3 \\ =& \Big(\mathbb{E}[N(t)^4] - 4\mathbb{E}[N(t)^3]\mathbb{E}N(t) + 6\mathbb{E}[N(t)^2](\mathbb{E}N(t))^2 \\ &- 4\mathbb{E}N(t)(\mathbb{E}N(t))^3 + (\mathbb{E}N(t))^4\Big)\Big(Var\,N(t)\Big)^{-2} - 3,\end{aligned}$$

Lasketaan prosessin N momentit generoivan funktion neljäs derivaatta,

$$\begin{aligned}m_{N(t)}^{(4)}(z) =& h^{(4)}(z)m_\theta'(h(z)) + h'(z)h^{(3)}(z)m_\theta''(h(z)) + 3h''(z)^2 m_\theta''(h(z)) \\ &+ 3h'(z)\left[h^{(3)}(z)m_\theta''(h(z)) + h'(z)h''(z)m_\theta^{(3)}(h(z))\right] \\ &+ 3h'(z)^2h''(z)m_\theta^{(3)}(h(z)) + h'(z)^4 m_\theta^{(4)}(h(z))\end{aligned}$$

ja sijoitetaan $z = 0$

$$\begin{aligned}\mathbb{E}[N(t)^4] =& m_{N(t)}^{(4)}(0) \\ =& \mu(t)\mathbb{E}\theta + \mu(t)^2\mathbb{E}(\theta^2) + 3\mu(t)^2\mathbb{E}(\theta^2) \\ &+ 3\mu(t)\left[\mu(t)\mathbb{E}(\theta^2) + \mu(t)^2\mathbb{E}(\theta^3)\right] + 3\mu(t)^3\mathbb{E}(\theta^3) + \mu(t)^4\mathbb{E}(\theta^4) \\ =& \mu(t)\mathbb{E}\theta + 7\mu(t)^2\mathbb{E}(\theta^2) + 6\mu(t)^3\mathbb{E}(\theta^3) + \mu(t)^4\mathbb{E}(\theta^4).\end{aligned}$$

Huipukkuuden kaavan osoittajaksi saadaan nyt

$$\begin{aligned}&\mathbb{E}[N(t)^4] - 4\mathbb{E}[N(t)^3]\mathbb{E}N(t) + 6\mathbb{E}[N(t)^2](\mathbb{E}N(t))^2 - 4\mathbb{E}N(t)(\mathbb{E}N(t))^3 + (\mathbb{E}N(t))^4 \\ =& \mu(t)\mathbb{E}\theta + 7\mu(t)^2\mathbb{E}(\theta^2) + 6\mu(t)^3\mathbb{E}(\theta^3) + \mu(t)^4\mathbb{E}(\theta^4) \\ &- 4\mu(t)\mathbb{E}\theta\left[\mu(t)\mathbb{E}\theta + 3\mu(t)^2\mathbb{E}(\theta^2) + \mu(t)^3\mathbb{E}(\theta^3)\right] \\ &+ 6\mu(t)^2(\mathbb{E}\theta)^2\left[\mu(t)\mathbb{E}\theta + \mu(t)^2\mathbb{E}(\theta^2)\right] - 4\mu(t)^4(\mathbb{E}\theta)^4 + \mu(t)^4(\mathbb{E}\theta)^4 \\ =& \mu(t)\mathbb{E}\theta + \mu(t)^2\left[7\mathbb{E}(\theta^2) - 4(\mathbb{E}\theta)^2\right] + \mu(t)^3\left[6\mathbb{E}(\theta^3) - 12\mathbb{E}\theta\mathbb{E}(\theta^2) + 6(\mathbb{E}\theta)^3\right] \\ &+ \mu(t)^4\left[\mathbb{E}(\theta^4) - 4\mathbb{E}\theta\mathbb{E}(\theta^3) + 6(\mathbb{E}\theta)^2\mathbb{E}(\theta^2) - 3(\mathbb{E}\theta)^4\right],\end{aligned}$$

jolloin huipukkuus on

$$\begin{aligned}\gamma_2(N(t)) =& \Big(\mathbb{E}N(t) + \mu(t)^2\left[7\mathbb{E}(\theta^2) - 4(\mathbb{E}\theta)^2\right] + \mu(t)^3\left[6\mathbb{E}(\theta^3) - 12\mathbb{E}\theta\mathbb{E}(\theta^2) + 6(\mathbb{E}\theta)^3\right] \\ &+ \mu(t)^4\left[\mathbb{E}(\theta^4) - 4\mathbb{E}\theta\mathbb{E}(\theta^3) + 6(\mathbb{E}\theta)^2\mathbb{E}(\theta^2) - 3(\mathbb{E}\theta)^4\right]\Big) \\ &\cdot \Big(Var\,N(t)\Big)^{-2} - 3.\end{aligned}$$

□

HUOMAUTUS 3.3. Sijoittamalla edellisessä lauseessa $\theta \equiv 1$ saadaan tunnusluvut ei-painotetulle Poisson-prosessille $\hat{N}$, jonka keskiarvofunktio on $\mu(t)$:

(1) Odotusarvo $\mathbb{E}\hat{N}(t) = \mu(t)$.
(2) Varianssi $Var\, \hat{N}(t) = \mathbb{E}\hat{N}(t)$.
(3) Vinous

$$\gamma_1\left(\hat{N}(t)\right) = \frac{Var\, \hat{N}(t) - \mathbb{E}\hat{N}(t)^3 + \mu(t)^3}{Var\, \hat{N}(t)^{3/2}} = \frac{1}{\mu(t)^{1/2}}.$$

(4) Huipukkuus

$$\gamma_2\left(\hat{N}(t)\right) = \frac{\mathbb{E}\hat{N}(t) + 3\mu(t)^2}{Var\, \hat{N}(t)^2} - 3 = \frac{1}{\mu(t)}.$$

3.2. Ominaisuuksia

Painotettu Poisson-prosessi N perii joitakin ominaisuuksia ei-painotetulta Poisson-prosessilta. Painotetut prosessit summautuvat samoin kuin ei-painotetut, Markov-ominaisuus pätee ja myös järjestystunnuslukuominaisuus pätee edelleen. Kaikki ei-painotetun Poisson-prosessin ominaisuudet eivät kuitenkaan periydy, vaan joitakin hyviä ominaisuuksia menetetään. Painotuksen myötä prosessin lisäykset eivät ole enää riippumattomia, prosessin jakauma ei noudata Poisson-jakaumaa ja kuten edellä laskettiin, odotusarvo ja varianssi eivät enää ole samat. Painotetun Poisson-prosessin ominaisuutta $Var\, N(t) > \mathbb{E}N(t)$ kaikilla $t > 0$, sanotaan ylihajonnaksi ja se on yksi tärkeimmistä painotetun ja ei-painotetun Poisson prosessin eroista yhdessä jakauman kanssa.

Osoitetaan ensin, että painotetun Poisson-prosessin jakauma ei ole Poisson-jakauma ja että sen lisäykset ovat riippuvia.

LAUSE 3.4. *Olkoon N painotettu Poisson-prosessi keskiarvofunktiolla μ ja painotusmuuttujalla $\theta > 0$ m.v.. Tällöin prosessin N jakauma ei yleisesti noudata Poisson-jakaumaa.*

TODISTUS. Merkitään satunnaismuuttujan θ kertymäfunktiota $F_\theta(\xi)$. Lauseen 3.2. todistuksen alussa saatiin

$$\mathbb{P}(N(t) = k) = \int_0^\infty e^{-\xi\mu(t)} \frac{(\xi\mu(t))^k}{k!}\, dF_\theta(\xi),$$

josta nähdään, ettei kyseessä ole Poisson-jakauman pistetodennäköisyysfunktio. □

LAUSE 3.5. *(Riippuvat lisäykset)*
Olkoon N painotettu Poisson-prosessi keskiarvofunktiolla μ ja painotusmuuttujalla θ, ja $0 = t_0 < t_1 < \cdots < t_n$. Tällöin lisäykset $N(t_n) - N(t_{n-1}), N(t_{n-1}) - N(t_{n-2}), \ldots,$ $N(t_1) - N(t_0)$ ovat riippuvia.

TODISTUS. Olkoon $0 \leq s < t$ ja $0 \leq x < y$ siten, että välit $(s, t]$ ja $(x, y]$ ovat erillisiä ja olkoon θ ei-vakio. Jos lisäykset $N(s, t)$ ja $N(x, y)$ olisivat riippumattomia, niin tällöin olisi

$$\mathbb{E}\left[N(s, t)\, N(x, y)\right] - \mathbb{E}N(s, t)\, \mathbb{E}N(x, y) = 0.$$

Lauseen 3.2. nojalla tiedetään, että $\mathbb{E}N(s,t) = \mu(s,t)\mathbb{E}\theta$ ja $\mathbb{E}N(x,y) = \mu(x,y)\mathbb{E}\theta$. Lasketaan $\mathbb{E}\,[N(s,t)\,N(x,y)]$:

$$\begin{aligned}
\mathbb{E}[N(s,t), N(x,y)] &= \mathbb{E}\big[\mathbb{E}[N(s,t), N(x,y)\,|\,\theta\,]\big] \\
&= \mathbb{E}\big[\mathbb{E}[N(s,t)|\theta]\,\mathbb{E}[N(x,y)|\theta]\big] \qquad \text{(ehdollinen riippumattomuus)} \\
&= \mathbb{E}[\mu(s,t)\theta\,\mu(x,y)\theta] \\
&= \mu(s,t)\mu(x,y)\mathbb{E}[\theta^2].
\end{aligned}$$

Nyt saadaan siis

$$\mathbb{E}\,[N(s,t),\,N(x,y)] - \mathbb{E}N(s,t)\,\mathbb{E}N(x,y) = \mu(s,t)\mu(x,y)\big(\mathbb{E}[\theta^2] - [\mathbb{E}\theta]^2\big) > 0,$$

koska oletuksen mukaan θ ei ole vakio, joten $Var\,\theta = \mathbb{E}[\theta^2] - [\mathbb{E}\theta]^2 > 0$.

□

Osoitetaan sitten painotettujen Poisson-prosessien summautuvuus. Sama todistus osoittaa myös ei-painotettujen Poisson-prosessien summautuvuuden, kun sijoitetaan $\theta \equiv 1$.

LAUSE 3.6. *(Summautuvuus)*
Olkoon μ_j painotetun Poisson prosessin N_j keskiarvofunktio ja θ_j painotusmuuttuja. Lisäksi prosessit N_j ovat riippumattomia kaikilla $j = 1, 2, ..., n$ ja merkitään $M(t) = \sum_{j=1}^{n} N_j(t)$. Tällöin $M(t)$ on painotettu Poisson-prosessi parametrilla $\sum_{j=1}^{n} \theta_j \mu_j(t)$.

TODISTUS. Todistetaan väite käyttämällä momentit generoivaa funktiota. Lauseen 3.2. todistuksessa laskettiin momentit generoiva funktio yhdelle painotetulle Poisson-prosessille ja funktioksi saatiin

$$m_{N_j(t)}(z) = \mathbb{E}[e^{zN_j(t)}] = \mathbb{E}\left[e^{(e^z-1)\theta_j\mu_j(t)}\right].$$

Lasketaan sitten momentit generoiva funktio prosessille $M(t)$:

$$\begin{aligned}
m_{M(t)}(z) &= \mathbb{E}\left[e^{zM(t)}\right] = \mathbb{E}\left[e^{z\sum_{j=1}^{n} N_j(t)}\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{zN_j(t)}\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{(e^z-1)\theta_j\mu_j(t)}\right] \\
&= \mathbb{E}\left[e^{(e^z-1)\sum_{j=1}^{n}\theta_j\mu_j(t)}\right],
\end{aligned}$$

josta nähdään, että $M(t)$ on painotettu Poisson-prosessi parametrilla $\sum_{j=1}^{n} \theta_j \mu_j(t)$.

□

HUOMAUTUS 3.7. Jos Lauseen 3.6. tilanteessa kaikilla painotetuilla Poisson-prosesseilla N_j olisi sama painotusmuuttuja θ ja prosessit olisivat riippumattomia ehdolla θ, niin tällöin $M(t)$ olisi painotettu Poisson-prosessi keskiarvofunktiolla $\sum_{j=1}^{n} \mu_j(t)$ ja painotusmuuttujalla θ.

LAUSE 3.8. *(Markov-ominaisuus)*
Olkoon N painotettu Poisson-prosessi. Tällöin mille tahansa jonolle $(t_i)_{i=0}^{n}$, jolle $0 = t_0 < t_1 < ... < t_n$, ja ei-vähenevälle luonnollisten lukujen jonolle $(k_i)_{i=1}^{n}$, $k_i \geq 0$ kaikilla $i = 1, 2, ..., n$ ja $n \geq 2$ pätee

$$\mathbb{P}(N(t_n) = k_n | N(t_1) = k_1, ..., N(t_{n-1}) = k_{n-1}) = \mathbb{P}(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}).$$

Todistus. Merkitään taas painotusmuuttujan θ kertymäfunktiota $F_\theta(\xi)$. Tällöin

$$\begin{aligned}
&\mathbb{P}\big(N(t_n)=k_n \,\big|\, N(t_1)=k_1,...,N(t_{n-1})=k_{n-1}\big)\\
=&\int_0^\infty \mathbb{P}\big(N(t_n)=k_n \,\big|\, N(t_1)=k_1,...,N(t_{n-1})=k_{n-1},\theta=\xi\big)\,dF_\theta(\xi)\\
\overset{(*)}{=}&\int_0^\infty \mathbb{P}\big(N(t_n)=k_n \,\big|\, N(t_{n-1})=k_{n-1},\theta=\xi\big)\,dF_\theta(\xi)\\
=&\mathbb{P}\big(N(t_n)=k_n \,\big|\, N(t_{n-1})=k_{n-1}\big),
\end{aligned}$$

missä $(*)$ seuraa ei-painotetun Poisson-prosessin Markov-ominaisuudesta, koska ehdollistettuna painotusmuuttujalla θ prosessi N vastaa ei-painotettua Poisson-prosessia. □

Lause 3.9. *(Järjestystunnuslukuominaisuus)*
Olkoon N painotettu Poisson-prosessi, jolla on jatkuva ja melkein kaikkialla positiivinen intensiteettifunktio λ, ja saapumishetket $0<T_1<T_2<...$ m.v.. Olkoon lisäksi satunnaismuuttujat $X_1,...,X_n$ riippumattomasti samoin jakautuneita siten, että tiheysfunktio kaikilla X_i, $i=1,2,...,n$, on $\lambda(x)/\mu(t)$, $0<x\le t$. Tällöin vektorin $(T_1,...,T_n)$ ehdollinen jakauma annetulla $N(t)=n$ on järjestetyn otoksen $(X_{(1)},...,X_{(n)})$ yhteistiheysfunktio eli

$$(T_1,...,T_n|N(t)=n)\overset{d}{=}\big(X_{(1)},...,X_{(n)}\big)\,.$$

Toisin sanoen siis vasemman puolen vektorin ehdollinen tiheysfunktio on

$$f_{T_1,...,T_n}(x_1,...,x_n|N(t)=n)=\frac{n!}{(\mu(t))^n}\prod_{i=1}^n\lambda(x_i), \tag{3.1}$$

missä $0<x_1<...<x_n<t$.

Todistus. Olkoon $0<x_1<...<x_n<t$ ja $h_i>0$ kaikilla $i=1,...,n$ siten, että välit $(x_i,x_i+h_i]$ ja $(x_{i+1},x_{i+1}+h_{i+1}]$ ovat erillisiä kaikilla $i=1,...,n-1$. Todistus olisi täysin vastaavanlainen väleille $(x_i-h_i,x_i]$, joten riittää osoittaa väite väleille $(x_i,x_i+h_i]$. Lasketaan väitteen vasemman puolen ehdollinen todennäköisyys:

$$\begin{aligned}
&\mathbb{P}\big(T_1\in(x_1,x_1+h_1],...,T_n\in(x_n,x_n+h_n]|N(t)=n\big)\\
=&\mathbb{E}[\mathbb{1}_{\{T_1\in(x_1,x_1+h_1],...,T_n\in(x_n,x_n+h_n]\}}|N(t)=n]\\
=&\mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{T_1\in(x_1,x_1+h_1],...,T_n\in(x_n,x_n+h_n]\}}|N(t)=n,\theta=\xi]\big]\\
=&\int_0^\infty \mathbb{P}\big(T_1\in(x_1,x_1+h_1],...,T_n\in(x_n,x_n+h_n]|N(t)=n,\theta=\xi\big)\,d\mathbb{P}_\theta(\xi)\\
\overset{(*)}{=}&\int_0^\infty \frac{n!}{(\xi\mu(t))^n}\frac{\xi\mu(x_1,x_1+h_1]}{h_1}\cdots\frac{\xi\mu(x_n,x_n+h_n]}{h_n}\,d\mathbb{P}_\theta(\xi)\\
=&\int_0^\infty \frac{n!}{(\mu(t))^n}\frac{\mu(x_1,x_1+h_1]}{h_1}\cdots\frac{\mu(x_n,x_n+h_n]}{h_n}\,d\mathbb{P}_\theta(\xi)\\
=&\frac{n!}{(\mu(t))^n}\frac{\mu(x_1,x_1+h_1]}{h_1}\cdots\frac{\mu(x_n,x_n+h_n]}{h_n}\\
\to&\ \frac{n!}{(\mu(t))^n}\lambda(x_1)\dots\lambda(x_n), \qquad \text{kun } h_i\downarrow 0,\ i=1,...,n,
\end{aligned}$$

missä raja-arvo seuraa taas Määritelmästä 2.5. ja intensiteettifunktion λ jatkuvuudesta, jolloin $\mu'(x_i) = \lambda(x_i)$. Kohta $(*)$ seuraa Lauseen 2.17. todistuksen lopusta, koska ehdollistettuna muuttujalla θ prosessi vastaa ei-painotettua Poisson-prosessia. □

Järjestystunnuslukuominaisuus on siis täysin sama painotetulle ja ei-painotetulle Poisson-prosessille.

3.3. Esimerkkejä

Esimerkki 3.10. Luvun alussa oli esimerkkinä kesän metsäpalojen lukumäärä. Oletetaan, että kesän metsäpalojen lukumäärä noudattaa painotettua Poisson-prosessia N keskiarvofunktiolla μ ja painotusmuuttujalla θ. Oletetaan kullekkin säätyypille satunnaismuuttujan θ arvo ja todennäköisyys alla olevan taulukon mukaisesti:

Säätyyppi	Painotusmuuttuja θ	Todennäköisyys p
Erittäin kuiva	$3,00$	$0,05$
Kuiva	$1,75$	$0,22$
Normaali	$0,80$	$0,40$
Sateinen	$0,60$	$0,25$
Erittäin sateinen	$0,30$	$0,08$

Montako metsäpaloa on kesän odotusarvo?

Ratkaisu:

Lauseen 3.2. nojalla

$$\begin{aligned}\mathbb{E}N(t) &= \mu(t)\mathbb{E}\theta \\ &= \mu(t)\,(3\cdot 0,05 + 1,75\cdot 0,22 + 0,80\cdot 0,40 + 0,60\cdot 0,25 + 0,30\cdot 0,08) \\ &= 1,029\,\mu(t)\end{aligned}$$

Tässä painotetun Poisson-prosessin odotusarvo on siis n. 3% suurempi kuin vastaavan painottamattoman Poisson-prosessin.

Toinen esimerkki mukailee Resnickin kirjan [5] esimerkkiä 4.49 ja Rossin kirjan [6] viidennen luvun esimerkkiä 64.

Esimerkki 3.11. Asiakkaiden saapuminen linja-autoasemalle noudattaa painotettua homogeenista Poisson-prosessia N keskiarvofunktiolla λ ja painotusmuuttujalla θ. Edellinen linja-auto lähti hetkellä 0 ja seuraava lähtee hetkellä t. Mikä on kaikkien niiden asiakkaiden odotettu kokonaisodotusaika, jotka saapuvat aikavälillä $(0, t)$?

Ratkaisu:

Järjestystunnuslukuominaisuuden nojalla (Lause 3.9.) annetulla $N(t) = n$ jokainen saapuminen T_i, $i = 1, ..., n$, noudattaa tasajakaumaa välillä $(0, t)$: koska N on painotettu homogeeninen Poisson-prosessi, niin Lauseen 3.9. muuttujien X_i tiheysfunktio saadaan muotoon

$$\frac{\lambda(x)}{\mu(t)} = \frac{\lambda}{\lambda t} = \frac{1}{t},$$

joka on tasajakauman tiheysfunktio välillä $(0, t)$.

Merkitään muuttujalla Y kaikkien niiden asiakkaiden yhteenlaskettua odotusaikaa, jotka saapuvat välillä $(0, t)$. Odotetuksi kokonaisodotusajaksi saadaan nyt

$$\mathbb{E}Y = \mathbb{E}\big[\mathbb{E}[Y|N(t)]\big] \overset{(*)}{=} \mathbb{E}\left[N(t)\int_0^t \frac{t-s}{t}\,ds\right] = \mathbb{E}\left[N(t)\frac{t}{2}\right] = \frac{t}{2}\mathbb{E}[N(t)] = \frac{t}{2}\lambda\mathbb{E}\theta,$$

missä $(*)$ seuraa siitä, että saapumisia on $N(t)$ kappaletta ja kun saapumisaika on s, niin odotusaika on $t-s$.

LUKU 4

Yhdistetty painotettu Poisson-prosessi

Edellä esitellyissä Poisson-prosesseissa oltiin kiinnostuneita vain saapumisten lukumäärän jakaumasta, esimerkiksi kuinka monta korvausvaatimusta tulee vakuutusyhtiölle hetkeen t mennessä. Vakuutusyhtiön tapauksessa kiinnostuksen kohteena olisi korvausvaatimusten lukumäärän lisäksi vaatimusten suuruus eli kuinka paljon vakuutusyhtiö kaikkiaan joutuu maksamaan korvauksia. Tätä kokonaismäärää mallinnetaan satunnaissummalla, missä sekä summattavat että summattavien lukumäärä on satunnainen.

MÄÄRITELMÄ 4.1. *Olkoon N painotettu Poisson-prosessi keskiarvofunktiolla μ ja painotusmuuttujalla θ ja olkoon satunnaismuuttujat X_i riippumattomasti samoin jakautuneita kaikkilla i siten, että jono (X_i) on riippumaton prosessista N. Määritellään prosessi $S(t)$ siten, että*

$$S(t) = \sum_{i=1}^{N(t)} X_i, \qquad \text{kun } t \geq 0.$$

Tällöin prosessi $S(t)$ on yhdistetty painotettu Poisson-prosessi.

Yllä olevassa määritelmässä prosessi $N(t)$ siis kuvaa kuinka monta saapumista hetkeen t mennessä on tullut ja (X_i) kuvaa niiden suuruutta. Muuttujien (X_i) riippumattomuus, myös prosessista N, tarkoittaa, että edellisten saapumisten suuruus ei vaikuta seuraavan saapumisen suuruuteen ja toisaalta myöskään saapumisten lukumäärä ei vaikuta saapumisten suuruuteen.

Osoitetaan seuraavaksi kolme ominaisuutta yhdistetylle painotetulle Poisson-prosessille S. Huomataan, että nämä prosessin S ominaisuudet ovat samat kuin Määritelmän 2.2. ehdot (1), (2) ja (4) eli ehdot, jotka määrittelevät ei-painotetun Poisson-prosessin.

LAUSE 4.2. *Yhdistetylle painotetulle Poisson-prosessille S pätee*

(1) *S(0)=0 m.v..*
(2) *Prosessilla S on riippumattomat lisäykset.*
(3) *Prosessi S on oikealta jatkuva ja sillä on vasemman puoleiset raja-arvot.*

TODISTUS. (1) Koska N on Poisson prosessi, niin $N(0) = 0$ melkein varmasti, jolloin $S(0) = \sum_{i=1}^{0} X_i = 0$ m.v..

(2) Olkoon $0 \leq r < t$.

$$\begin{aligned}\mathbb{P}\big(S(r) = l, S(t) - S(r) = m\big) &= \mathbb{P}\left(\sum_{i=1}^{N(r)} X_i = l, \sum_{N(r)+1}^{N(t)} X_i = m\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{N(r)} X_i = l\right)\mathbb{P}\left(\sum_{N(r)+1}^{N(t)} X_i = m\right)\\ &= \mathbb{P}\big(S(r) = l\big)\mathbb{P}\big(S(t) - S(r) = m\big),\end{aligned}$$

koska muuttujat X_i ovat riippumattomasti samoin jakautuneita ja lisäksi N ja (X_i) ovat riippumattomia.

(3) Osoitetaan ensin, että prosessi S on oikealta jatkuva:

$$\mathbb{P}\left(\lim_{t\downarrow r} S(t) = S(r)\right) = \mathbb{P}\left(\lim_{t\downarrow r}\sum_{i=1}^{N(t)} X_i = \sum_{i=1}^{N(r)} X_i\right) = 1,$$

koska prosessi N on oikealta jatkuva.

Osoitetaan vielä, että prosessilla S on vasemman puoleiset raja-arvot:

$$\mathbb{P}\left(\exists \lim_{t\uparrow r} S(t)\right) = \mathbb{P}\left(\exists \lim_{t\uparrow r}\sum_{i=1}^{N(t)} X_i\right) = \mathbb{P}\left(\exists \lim_{t\uparrow r} N(t)\right) = 1,$$

koska prosessilla N on vasemman puoleiset raja-arvot.

□

Lasketaan prosessille S jakauman tunnusluvuista odotusarvo ja varianssi käyttämällä momentit generoivaa funktiota.

Lause 4.3. *Yhdistetyn painotetun Poisson-prosessin $S(t) = \sum_{i=1}^{N(t)} X_i$ odotusarvo on $\mathbb{E}S(t) = \mathbb{E}N(t)\,\mathbb{E}X_1$ ja varianssi $var\, S(t) = (\mathbb{E}X_1)^2\, var\, N(t) + \mathbb{E}N(t)\, var\, X_1$.*

Todistus. Prosessin S momentit generoiva funktio ehdolla $N(t) = k$ on

$$m_{S(t)}(r|N(t) = k) = \mathbb{E}\left[e^{rS(t)}\,|N(t) = k\right] = \mathbb{E}\left[e^{r\sum_{i=1}^{k} X_i}\right] \overset{X_i\, i.i.d.}{=} \prod_{i=1}^{k}\mathbb{E}\left[e^{rX_1}\right] = m_{X_1}(r)^k,$$

jolloin

$$\begin{aligned}m_{S(t)}(r) &= \sum_{k=0}^{\infty}\mathbb{P}(N(t) = k)\, m_{S(t)}(r|N(t) = k) = \mathbb{E}\left[m_{S(t)}(r|N(t))\right] = \mathbb{E}\left[m_{X_1}(r)^{N(t)}\right]\\ &= \mathbb{E}\left[e^{N(t)\log(m_{X_1}(r))}\right] = m_{N(t)}\big(\log(m_{X_1}(r))\big) \overset{(*)}{=} m_\theta\big(\mu(t)(e^{\log[m_{X_1}(r)]} - 1)\big)\\ &= m_\theta\big(\mu(t)(m_{X_1}(r) - 1)\big),\end{aligned}$$

missä kohta $(*)$ saadaan Lauseen 3.2. todistuksesta, jossa laskettiin painotetun Poisson-prosessin N momentit generoiva funktio.

Derivoimalla momentit generoivaa funktiota saadaan

$$m'_{S(t)}(r) = m'_\theta\big(\mu(t)[m_{X_1}(r) - 1]\big)\mu(t)m'_{X_1}(r),$$

ja siten odotusarvo on

$$
\begin{aligned}
\mathbb{E}S(t) = m'_{S(t)}(0) &= m'_\theta\big(\mu(t)[m_{X_1}(0)-1]\big)\,\mu(t)\,m'_{X_1}(0)\\
&= m'_\theta(0)\,\mu(t)\,m'_{X_1}(0)\\
&= \mathbb{E}\theta\,\mu(t)\,\mathbb{E}X_1 = \mathbb{E}N(t)\,\mathbb{E}X_1.
\end{aligned}
$$

Varianssin osoittamiseksi lasketaan ensin momentit generoivan funktion toinen derivaatta

$$
m''_{S(t)}(r) = m''_\theta\big(\mu(t)[m_{X_1}(r)-1]\big)\,\big(\mu(t)m'_{X_1}(r)\big)^2 + m'_\theta\big(\mu(t)[m_{X_1}(r)-1]\big)\mu(t)m''_{X_1}(r),
$$

jolloin

$$
\begin{aligned}
\mathbb{E}S(t)^2 = m''_{S(t)}(0) &= m''_\theta(0)\mu(t)^2 m'_{X_1}(0)^2 + m'_\theta(0)\mu(t)m''_{X_1}(0)\\
&= \mathbb{E}\theta^2\mu(t)^2(\mathbb{E}X_1)^2 + \mathbb{E}\theta\mu(t)\mathbb{E}X_1^2.
\end{aligned}
$$

Varianssiksi saadaan siis

$$
\begin{aligned}
\mathrm{var}(S(t)) &= \mathbb{E}S(t)^2 - (\mathbb{E}S(t))^2\\
&= \mathbb{E}\theta^2\mu(t)^2(\mathbb{E}X_1)^2 + \mathbb{E}\theta\mu(t)\mathbb{E}X_1^2 - (\mathbb{E}X_1)^2(\mathbb{E}N(t))^2\\
&= (\mu(t)\mathbb{E}X_1)^2\mathrm{var}\,\theta + \mathbb{E}\theta\mu(t)\mathbb{E}X_1^2\\
&= (\mu(t)\mathbb{E}X_1)^2\mathrm{var}\,\theta + \mathbb{E}N(t)\mathbb{E}X_1^2 + (\mathbb{E}X_1)^2\mathbb{E}N(t) - (\mathbb{E}X_1)^2\mathbb{E}N(t)\\
&= (\mathbb{E}X_1)^2\big(\mu(t)^2\mathrm{var}\,\theta + \mathbb{E}N(t)\big) + \mathbb{E}N(t)\big(\mathbb{E}X_1^2 - (\mathbb{E}X_1)^2\big)\\
&= (\mathbb{E}X_1)^2\,var\,N(t) + \mathbb{E}N(t)\,var\,X_1.
\end{aligned}
$$

□

4.1. Cramér-Lundbergin malli

Cramér-Lundbergin malli on yhdistetyn painotetun Poisson-prosessin erikoistapaus, jossa saapumiset tulevat homogeenisen Poisson-prosessin mukaisesti. Malli on tärkeässä osassa vakuutusmatematiikassa; se on riskiteoriassa yksi suosituimmista ja hyödyllisimmistä malleista. Yksinkertaisuudestaan huolimatta malli kuvaa käytännössä havaittavien vakuutuskorvausvaatimusten kokonaismäärän prosessin välttämättömiä ominaisuuksia hyvin.

Cramér-Lundbergin malli. Olkoon $S(t) = \sum_{i=1}^{N(t)} X_i$ yhdistetty Poisson-prosessi. Prosessi S on Cramér-Lundbergin mallin mukainen, jos

(1) korvausvaatimukset tulevat homogeenisen Poisson-prosessin $N(t) = \#\{i \geq 1 : T_i \leq t\}, t \geq 0$ saapumishetkillä $0 \leq T_1 \leq T_2 \leq ....$
(2) hetkellä T_i saapuva korvausvaatimus on suuruudeltaan X_i. Korvausvaatimusten suuruudet $X_1, X_2, ...$ ovat riippumattomasti samoin jakautuneita ja $X_i \geq 0$ kaikilla $i = 1, 2, ...$
(3) saapumishetket $(T_i)_{i=1}^\infty$ ja vaatimusten suuruudet $(X_i)_{i=1}^\infty$ ovat keskenään riippumattomia eli N ja (X_i) ovat riippumattomia.

Seuraavassa Lauseessa osoitetaan, että Cramér-Lundbergin mallissa prosessin S lisäykset ovat stationaarisia. Aivan kuten Lauseen 4.2. yhteydessä huomattiin prosessilla S olevan samoja ominaisuuksia kuin ei-painotetulla Poisson-prosessilla, niin seuraava Lause 4.4. osoittaa, että Cramér-Lundbergin mallin mukaisella prosessilla on sama ominaisuus kuin homogeenisella Poisson-prosessilla (Huomautus 2.3.(2)).

LAUSE 4.4. *Cramér-Lundbergin mallin prosessille S pätee Lauseen 4.2. lisäksi, että prosessin S lisäykset ovat stationaariset.*

TODISTUS. Pitää siis osoitaa, että

$$
\begin{aligned}
& \mathbb{P}(S(t)-S(r)\leq z)=\mathbb{P}(S(t-r)-S(0)\leq z)\\
\Leftrightarrow\quad & \mathbb{P}\left(\sum_{i=1}^{N(t)}X_i-\sum_{i=1}^{N(r)}X_i\leq z\right)=\mathbb{P}\left(\sum_{i=1}^{N(t-r)}X_i\leq z\right)\\
\overset{\text{yhtälö (2.1)}}{\Leftrightarrow}\quad & \mathbb{P}\left(\sum_{i=N(r)+1}^{N(t)}X_i\leq z\right)=\mathbb{P}\left(\sum_{i=1}^{N(t)-N(r)}X_i\leq z\right).
\end{aligned}
$$

Osoitetaan tämä laskemalla vasemman puoleinen todennäköisyys:

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{i=N(r)+1}^{N(t)}X_i\leq z\right)\\
=&\sum_{k=1}^{\infty}\sum_{q=k}^{\infty}\mathbb{P}\left(N(r)=k,N(t)=q,\sum_{i=k+1}^{q}X_i\leq z\right)\\
\overset{(*)}{=}&\sum_{k=1}^{\infty}\sum_{q=k}^{\infty}\mathbb{P}(N(r)=k)\,\mathbb{P}(N(t)=q|N(r)=k)\,\mathbb{P}\left(\sum_{i=k+1}^{q}X_i\leq z|N(r)=k,N(t)=q\right)\\
=&\sum_{k=1}^{\infty}\sum_{q=k}^{\infty}\mathbb{P}(N(r)=k)\,\mathbb{P}(N(t)=q|N(r)=k)\,\mathbb{P}\left(\sum_{i=1}^{q-k}X_i\leq z\right)\\
=&\sum_{k=1}^{\infty}\sum_{q=k}^{\infty}\mathbb{P}(N(r)=k)\,\mathbb{P}(N(t)=q|N(r)=k)\,\mathbb{P}\left(\sum_{i=1}^{q-k}X_i\leq z|N(r)=k,N(t)=q\right)\\
=&\mathbb{P}\left(\sum_{i=1}^{N(t)-N(r)}X_i\leq z\right),
\end{aligned}
$$

missä kohta $(*)$ seuraa ehdollisen todennäköisyyden määritelmästä. □

Cramér-Lundbergin mallissa odotusarvo ja varianssi saadaan yksinkertaisempaan muotoon kuin yleisessä yhdistetyssä painotetussa Poisson-prosessissa, koska homogeeniselle Poisson-prosessille odotusarvo ja varianssi ovat samat. Seuraavassa Lauseessa 4.5. nämä on laskettu.

LAUSE 4.5. *Cramér-Lundbergin mallissa odotusarvo on $\mathbb{E}S(t)=\lambda t\mathbb{E}X_1$ ja varianssi var $S(t)=\lambda t\mathbb{E}X_1^2$.*

TODISTUS. Koska N on homogeeninen Poisson-prosessi parametrilla λ, niin $\mathbb{E}N(t)=\lambda t=\operatorname{var}N(t)$, jolloin odotusarvo seuraa Lauseesta 4.3. Saman lauseen nojalla

$$\begin{aligned}\operatorname{var} S(t) &= (\mathbb{E}X_1)^2\, var\, N(t) + \mathbb{E}N(t)\, var\, X_1 \\ &= (\mathbb{E}X_1)^2\lambda t + \mathbb{E}X_1^2\lambda t - (\mathbb{E}X_1)^2\lambda t \\ &= \lambda t\mathbb{E}X_1^2.\end{aligned}$$

□

Luvuissa 2 ja 3 ei-painotetulle ja painotetulle Poisson-prosessille osoitettiin summautuvuusominaisuudet (Lauseet 2.13. ja 3.6). Kun yhdistetyssä Poisson-prosessissa S saapumiset noudattavat homogeenista Poisson-prosessia, niin silloin myös prosessien S summalle saadaan laskettua jakauma.

Lause 4.6. *Olkoot $S_k(t) = \sum_{i=1}^{N_k(t)} X_i^{(k)}$, $k = 1, ..., n$, Cramér-Lundbergin mallin mukaisia riippumattomia yhdistettyjä Poisson-prosesseja, eli $N_k(t)$ on homogeeninen Poisson-prosessi parametrilla $\lambda_k t$ ja $(X_i^{(k)})_{i\geq 1}$ on jono riippumattomasti samoin jakautuneita satunnaismuuttujia (riippumattomia myös prosessista N_k) kaikilla $k = 1, ..., n$. Tällöin $S(t) := \sum_{k=1}^{n} S_k(t)$ on yhdistetty Poisson-prosessi ja se voidaan kirjoittaa muodossa*

$$S(t) \stackrel{d}{=} \sum_{l=1}^{N(t)} Y_l(t),$$

missä $N(t)$ on homogeeninen Poisson-prosessi parametrilla $\lambda nt := \sum_{k=1}^{n} \lambda_k t$ ja $(Y_l)_{l\geq 1}$ on jono riippumattomasti samoin jakautuneita satunnaismuuttujia (riippumattomia myös prosessista N) ja

$$Y_l(t) \stackrel{d}{=} \sum_{k=1}^{n} \mathbb{1}_{\{J(t)=k\}} X_1^{(k)}, \qquad \mathbb{P}(J(t) = k) = \frac{\lambda_k t}{\lambda n t}$$

ja lisäksi J on riippumaton jonosta $(X_1^{(k)})_{k=1}^{n}$.

Todistus. Osoitetaan, että prosessin S ja summan $\sum_{l=1}^{N} Y_l$ momentit generoivat funktiot ovat samat.

Lauseen 4.3. todistuksessa laskettiin prosessin S_k momentit generoivaksi funktioksi

$$m_{S_k(t)}(r) = m_\theta\left(\mu_k(t)\big(m_{X_1^{(k)}}(r) - 1\big)\right).$$

Koska nyt $N_k(t)$ on homogeeninen Poisson-prosessi, niin $\mu_k(t) = \lambda_k t$ ja koska $\theta \equiv 1$, niin $m_\theta(r) = e^r$, joten saadaan

$$m_{S_k(t)}(r) = \exp\left(\lambda_k t\big(m_{X_1^{(k)}}(r) - 1\big)\right),$$

ja näin ollen prosessin S momentit generoivaksi funktioksi saadaan

$$\begin{aligned} m_{S(t)}(r) &= \mathbb{E}\left[e^{rS(t)}\right] = \mathbb{E}\left[e^{r\sum_{k=1}^{n} S_k(t)}\right] \overset{\perp\!\!\!\perp}{=} \prod_{k=1}^{n} \mathbb{E}\left[e^{rS_k(t)}\right] = \prod_{k=1}^{n} \mathbb{E}\left[m_{S_k(t)}(r)\right] \\ &= \prod_{k=1}^{n} \exp\Big(\lambda_k t\big(m_{X_1^{(k)}}(r) - 1\big)\Big) \\ &= \exp\left(\sum_{k=1}^{n} \lambda_k t\big(m_{X_1^{(k)}}(r) - 1\big)\right). \end{aligned}$$

Lasketaan sitten prosessin $T(t) := \sum_{l=1}^{N(t)} Y_l(t)$ momentit generoiva funktio. Vastaavasti kuin prosessille S_k, saadaan

$$m_{T(t)}(r) = \exp\left(\lambda n t(m_{Y_1}(r) - 1)\right).$$

Koska satunnaismuuttujalle Y_1 saadaan momentit generoivaksi funktioksi

$$\begin{aligned} m_{Y_1(t)}(r) = \mathbb{E}\left[e^{rY_1(t)}\right] &= \mathbb{E}\left[\exp\left(r\sum_{k=1}^{n} \mathbb{1}_{\{J(t)=k\}} X_1^{(k)}\right)\right] \\ &= \mathbb{E}\left[\sum_{l=1}^{n} \exp\left(r\sum_{k=1}^{n} \mathbb{1}_{\{J(t)=k\}} X_1^{(k)}\right) \mathbb{1}_{\{J(t)=l\}}\right] \\ &= \sum_{l=1}^{n} \mathbb{E}\left[e^{rX_1^{(l)}} \mathbb{1}_{\{J(t)=l\}}\right] \\ &\overset{\perp\!\!\!\perp}{=} \sum_{l=1}^{n} m_{X_1^{(l)}}(r)\frac{\lambda_l t}{\lambda n t}, \end{aligned}$$

niin

$$\begin{aligned} m_{T(t)}(r) &= \exp\left(\lambda n t\left(\sum_{l=1}^{n} m_{X_1^{(l)}}(r)\frac{\lambda_l t}{\lambda n t} - 1\right)\right) \\ &= \exp\left(\sum_{l=1}^{n} m_{X_1^{(l)}}(r)\lambda_l t - \lambda n t\right) \\ &= \exp\left(\sum_{l=1}^{n} \lambda_l t\big(m_{X_1^{(l)}}(r) - 1\big)\right). \end{aligned}$$

Koska prosessien S ja T momentit generoivat funktiot ovat samat, niin prosessit ovat jakaumaltaan samat. □

4.2. Esimerkkejä

Esimerkki 4.7. ([5] tehtävä 4.48) Moottoriajoneuvoja saapuu tietullille homogeenisen Poisson-prosessin mukaisesti keskiarvoisesti 2 ajoneuvoa minuutissa. Kuskit maksavat tullia 1, 2 tai 5 euroa riippuen siitä, mihin kolmesta painoluokasta ajoneuvo kuuluu. Oletetaan painoluokkien 1, 2 ja 3 todennäköisyyksiksi 1/2, 1/3 ja 1/6 tässä järjestyksessä.

(1) Mikä on annetun tunnin tullimaksujen odotusarvo ja varianssi?
(2) Mikä on todennäköisyys, että täsmälleen 3 euroa kerätään joka minuutti kolmen peräkkäisen minuutin ajan?
(3) Mikä on todennäköisyys, että kahden 5 euroa maksavan ajoneuvon väli on yli 10 minuuttia?

Ratkaisu:
Merkitään kerättyjen tullimaksujen kokonaissummaa $S(t) = \sum_{i=1}^{N(t)} X_i$, missä N on saapumisten lukumäärä ja X_i ajoneuvon i maksaman tullimaksun suuruus.

(1) Koska ajoneuvoja saapuu tietullille joka minuutti homogeenisen Poisson-prosessin, jonka parametri on 2, mukaisesti, niin tunnissa ajoneuvoja saapuu homogeenisen Poisson-prosessin mukaisesti odotusarvoisesti $60 \cdot 2 = 120$. Koska saapumisprosessi on homogeeninen Poisson-prosessi, voidaan käyttää Cramér-Lundbergin mallia ja saadaan

$$\mathbb{E}S(t) = 120 \cdot \mathbb{E}X = 120 \cdot \left(1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6}\right) = 240\text{e/tunti}$$

ja

$$\operatorname{Var} S(t) = 120 \cdot \mathbb{E}X_1^2 = 120 \cdot \left(1 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{3} + 5^2 \cdot \frac{1}{6}\right) = 720,$$

eli tunnissa kerätyn summan suuruus on keskimäärin $\sqrt{720} = 27$ euron päässä odotusarvosta.

(2) Halutaan siis laskea todennäköisyys

$$\mathbb{P}\big(S(t_k) - S(t_{k-1}) = 3,\ S(t_{k+1}) - S(t_k) = 3,\ S(t_{k+2}) - S(t_{k+1}) = 3\big).$$

Koska prosessin N lisäykset ovat riippumattomia ja stationaarisia, satunnaismuuttujat X_i ovat riippumattomasti samoin jakautuneita ja lisäksi $S(0) = 0$ m.v., niin yllä oleva todennäköisyys voidaan kirjoittaa muotoon

$$\begin{aligned}&\mathbb{P}(S(t_k) - S(t_{k-1}) = 3)\,\mathbb{P}(S(t_{k+1}) - S(t_k) = 3)\,\mathbb{P}(S(t_{k+2}) - S(t_{k+1}) = 3)\\ &= \mathbb{P}(S(1) = 3)^3.\end{aligned}$$

Kolme euroa voidaan kerätä siten, että tulee joko kolme painoluokan yksi autoa tai tulee kaksi autoa, joista toinen kuuluu painoluokkaan yksi ja toinen painoluokkaan kaksi. Tämän tapahtuman todennäköisyydeksi minuutin aikana saadaan

$$\begin{aligned}&\mathbb{P}(N = 3) \cdot \left(\frac{1}{2}\right)^3 + \mathbb{P}(N = 2) \cdot 2 \cdot \frac{1}{2} \cdot \frac{1}{3}\\ &= e^{-2}\,\frac{2^3}{3!} \cdot \frac{1}{8} + e^{-2}\,\frac{2^2}{2!} \cdot \frac{1}{3}\\ &= e^{-2}\,\frac{5}{6},\end{aligned}$$

jolloin kysytyksi todennäköisyydeksi saadaan $\left(e^{-2} \cdot \frac{5}{6}\right)^3 \approx 0,0014$.

(3) Koska ajoneuvojen saapumiset noudattavat homogeenista Poisson-prosessia, niin vain välin pituudella on väliä, ei sen sijainnilla. Voidaan siis olettaa, että hetkellä t_0 tullista meni ajoneuvo, joka maksoi tullia 5 euroa. Kysytään todennäköisyyttä, että seuraavan 10 minuutin aikana ei tule toista painoluokan 3 ajoneuvoa. Todennäköisyys, että saapuva auto ei kuulu painoluokkaan 3 on 5/6 ja 10 minuutin aikavälillä Poisson-prosessin parametri on $10 \cdot 2 = 20$, joten kysytyksi todennäköisyydeksi saadaan

$$\begin{aligned}\sum_{k=0}^{\infty} \mathbb{P}(N = k) \cdot \left(\frac{5}{6}\right)^k &= \sum_{k=0}^{\infty} e^{-20} \frac{20^k}{k!} \left(\frac{5}{6}\right)^k = e^{-20} \sum_{k=0}^{\infty} \frac{\left(20 \cdot \frac{5}{6}\right)^k}{k!} \\ &= e^{-20} \cdot e^{50/3} = e^{-10/3} \approx 0,036.\end{aligned}$$

Viimeisessä esimerkissä sovitetaan simuloituun palovahinkoaineistoon painotettu Poisson-prosessi ja verrataan aineistosta laskettuja tunnuslukuja teoreettisiin jakaumasta tuleviin tunnuslukuihin. Lisäksi lasketaan kuinka todennäköisesti vakuutusyhtiö joutuu yhden kesän aikana maksamaan korvauksia enemmän kuin 1,2 kertaa odotetun verran. Esimerkissä korvausvaatimusten suuruuksien oletetaan noudattavan log-normaalia jakaumaa, joten ennen esimerkkiin siirtymistä esitetään log-normaalin jakauman tiheysfunktio sekä lasketaan sille odotusarvo ja varianssi.

PROPOSITIO 4.8. *Olkoon satunnaismuuttuja X log-normaalisti jakautunut parametreilla μ ja σ^2, merkitään $X \sim logN(\mu, \sigma^2)$, ja sen tiheysfunktio on*

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\sigma} exp\left(-\frac{(log\, x - \mu)^2}{2\sigma^2}\right).$$

Tällöin odotusarvo $\mathbb{E}X = e^{\mu + \frac{1}{2}\sigma^2}$ ja varianssi $var\, X = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

TODISTUS. Lasketaan ensin odotusarvo:

$$\begin{aligned}\mathbb{E}X &= \int_0^{\infty} \frac{x}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) dx \\ &\overset{(*)}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} + y\right) dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2 - 2y\mu + \mu^2 - y2\sigma^2 - (\mu+\sigma^2)^2 + (\mu+\sigma^2)^2}{2\sigma^2}\right) dy \\ &= \exp\left(\frac{(\mu+\sigma^2)^2 - \mu^2}{2\sigma^2}\right) \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y - (\mu+\sigma^2)\right)^2}{2\sigma^2}\right) dy \\ &\overset{(**)}{=} \exp\left(\frac{\sigma^4 + 2\sigma^2\mu + \mu^2 - \mu^2}{2\sigma^2}\right) \cdot 1 \\ &= e^{\mu+\frac{1}{2}\sigma^2},\end{aligned}$$

missä kohta $(*)$ seuraa muuttujanvaihdosta $y = \log x$ ja kohta $(**)$ seuraa siitä, että integroitavana on normaalijakauman tiheysfunktio odotusarvolla $\mu+\sigma^2$ ja varianssilla σ^2 ja tiheysfunktion integraali yli määrittelyjoukkonsa on aina 1.

Varianssin laskemiseksi lasketaan ensin $\mathbb{E}(X^2)$:

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_0^\infty \frac{x^2}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) dx \\
&\overset{(\star)}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(y - \frac{(y-\mu)^2}{2\sigma^2} + y\right) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2 - 2y\mu + \mu^2 - 4y\sigma^2 - (\mu + 2\sigma^2)^2 + (\mu + 2\sigma^2)^2}{2\sigma^2}\right) dy \\
&= \exp\left(\frac{(\mu + 2\sigma^2)^2 - \mu^2}{2\sigma^2}\right) \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(y - (\mu + 2\sigma^2))^2}{2\sigma^2}\right) dy \\
&\overset{(\star\star)}{=} \exp\left(\frac{\mu^2 + 4\mu\sigma^2 + 4\sigma^4 - \mu^2}{2\sigma^2}\right) \cdot 1 \\
&= e^{2\mu + 2\sigma^2},
\end{aligned}
$$

missä kohta $(\star)$ seuraa taas muuttujanvaihdosta $y = \log x$ ja kohta $(\star\star)$ tiheysfunktion integraalista. Varianssiksi saadaan siis

$$\operatorname{var} X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

□

Esimerkki 4.9. Oletetaan, että yhden kesän aikana tulipaloista aiheutuvien vahinkojen väliset ajat jakaantuvat seuraavasti:

Aika päivissä	Osuus sateisena kesänä (%)	Osuus aurinkoisena kesänä (%)
1	0,20	9,73
2	0,60	17,21
3	2,51	22,75
4	5,52	18,34
5	9,04	16,19
6	12,25	7,89
7	13,55	4,92
8	14,26	1,64
9	13,05	1,22
10	9,74	0,10
11	7,73	0
12	5,12	0
13	3,51	0
14	2,00	0
15	0,90	0

Oletetaan lisäksi, että korvausvaatimusten (X_i) suuruudet noudattavat log-normaalia jakaumaa odotusarvolla $\mu = 0,71$ ja keskihajonnalla $\sigma = 1,72$, kun kaikki vaatimukset on jaettu luvulla 100 000 laskujen selkeyttämiseksi. Yksittäisen korvausvaatimuksen odotusarvo on siis 71 000 euroa. Keskihajontaan aineiston lineaarinen muunnos ei vaikuta. Oletetaan vielä, että aurinkoinen ja sateinen kesä ovat yhtä todennäköisiä.

(1) Miten aineistosta lasketut saapumisten keskiarvo ja varianssi eroavat teoreettisesta odotusarvosta ja varianssista?
(2) Mikä on todennäköisyys, että yhden kesän korvausvaatimusten kokonaissumma on suurempaa kuin sen 1,2 kertainen odotusarvo?

Ratkaisu:

(1) Aineistosta laskemalla aurinkoisen kesän keskiarvoksi saadaan 3,74 päivää ja varianssiksi 3,25. Sateiselle kesälle saadaan keskiarvoksi 8,07 päivää ja varianssiksi 7,34. Vaikka aineistossa on alihajontaa, sitä ei kuitenkaan ole niin paljoa, että alihajonta haittaisi Poisson-jakaumaoletusta.

Käytetään aineistosta laskettuja keskiarvoja Poisson-prosessin parametreina eri kesille ja merkitään $\lambda_A = 3,74$ ja $\lambda_S = 8,07$. Kun aineistot yhdistetään, niin yhdistetty aineisto noudattaa painotettua Poisson-prosessia N painotusmuuttujalla θ, kun valitaan ensin prosessin N keskiarvofunktioksi $\mu(t) = \lambda_A$, ja asetetaan sitten

$$\theta = \begin{cases} \theta_A = 1 & \text{jos kesä on aurinkoinen} \\ \theta_S = \frac{\lambda_S}{\lambda_A} & \text{jos kesä on sateinen.} \end{cases}$$

Lauseen 3.2 nojalla prosessin N teoreettinen odotusarvo yhdelle kesälle on

$$\begin{aligned} \mathbb{E}N(t) &= \mu(t)\mathbb{E}\theta = \lambda_A(0,5 \cdot \theta_A + 0,5 \cdot \theta_S) \\ &= 0,5 \cdot \lambda_A + 0,5 \cdot \lambda_S \approx 5,905 \end{aligned}$$

ja varianssi

$$\begin{aligned} \mathrm{var}N(t) &= \mathbb{E}N(t) + \mu(t)^2 \mathrm{var}\theta = \lambda_A \mathbb{E}\theta + \lambda_A^2 \mathrm{var}\theta \\ &= \lambda_A\big(0,5\,\theta_A + 0,5\,\theta_S\big) + \lambda_A^2\big(0,5\,\theta_A^2 + 0,5\,\theta_S^2 - (0,5\,\theta_A + 0,5\,\theta_S)^2\big) \\ &= 0,5\,(\lambda_A + \lambda_S + \lambda_A^2 + \lambda_S^2 - 0,5\,\lambda_A^2 - 0,5\,\lambda_S^2 - \lambda_A\lambda_S) \\ &= 0,5\,(3,74 + 8,07 + 3,74^2 + 8,07^2 - 0,5 \cdot 3,74^2 - 0,5 \cdot 8,07^2 - 3,74 \cdot 8,07) \\ &\approx 10,592. \end{aligned}$$

Aineistosta laskettu keskiarvo on pyöristettynä $5,905$, joka on sama kuin teoreettinen odotusarvo. Aineistosta laskettu varianssi taas on $9,996$, joka on hieman pienempi kuin teoreettinen varianssi, mutta lähellä sekin.

(2) Korvausvaatimusten kokonaissumman $S(t) = \sum_{i=1}^{N(t)} X_i$ odotusarvoksi yhdelle kesälle saadaan Lauseen 4.3. nojalla

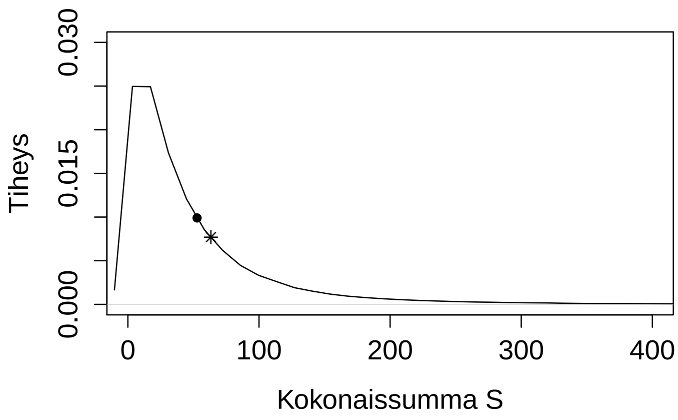
$$\mathbb{E}S(t) = \mathbb{E}N(t)\mathbb{E}X_1 = 5,91\, e^{0,71+{}^1\!/_2 \cdot 1,72^2} = 52,76,$$

eli alkuperäiseen yksikköön palautettuna noin $5,3$ miljoonaa euroa.

Mallinnetaan todennäköisyyttä $\mathbb{P}\left(S(T) > 1,2 \cdot \mathbb{E}[S(T)]\right)$ simuloimalla. Simuloimalla R-ohjelmalla 1 000 000 arvoa muuttujalle $S(T)$, saadaan kysytyksi todennäköisyydeksi noin 24%.

Vakuutusyhtiön korvausvaateet yhdeltä kesältä ovat siis 24% todennäköisyydellä enemmän kuin 1,2 kertainen odotusarvo. Tästä seuraa, että vakuutusyhtiön on toimintakykyisenä pysyäkseen kerättävä vakuutusmaksuja enemmän kuin mitä se keskimäärin maksaa korvauksia, vaikka tarkoituksena ei edes olisi tehdä voittoa.

Alla olevassa kuvassa on vielä piirrettynä kokonaissumman S tiheysfunktio. Käyrälle piirretty pallo merkitsee odotusarvoa ja tähti on 1,2 kertainen odotusarvo.



KUVA 4.1. Kokonaissumman S tiheysfunktio.

Lähdeluettelo

[1] Daykin, C.D., Pentikäinen, T. and Pesonen, M.: *Practical risk theory for actuaries.* Toinen laitos, Chapman & Hall, 1995.

[2] Geiss, Christel and Geiss, Stefan: *An introduction to probability theory, luentomuistiinpanot* Jyväskylän yliopisto 2015

[3] Mikosch, Thomas : *Non-life insurance mathematics.* Springer, 2009.

[4] Pinsky, Mark A. and Karlin, Samuel: *An introduction to stochastic modelling.* 4. laitos, Elsevier inc. 2011.

[5] Resnick, Sidney I.: *Adventures in stochastic processes.* Birkhäuser Boston, 1992.

[6] Ross, Sheldon M.: *Introduction to probability models.* 9. laitos, Elsevier 2007.

[7] Seppälä, Heikki : *Riskiteoria, luentomuistiinpanot.* Jyväskylän yliopisto, 2013.