

<b>5</b>	<b>Diskreettejä yksiulotteisia jakaumia</b>	<b>111</b>
5.1	Diskreetti satunnaismuuttuja . . . . .	111
5.2	Bernoullin kokeet ja binomijakauma . . . . .	113
5.2.1	Jakauman symmetria . . . . .	118
5.3	Odotusaikojen jakaumat . . . . .	119
5.3.1	Odotusajat Bernoullin kokeissa . . . . .	119
5.3.2	Geometrinen jakauma ja negatiivinen binomijakauma . . . . .	122
5.3.3	Odotusajat peräkkäisotannassa . . . . .	125
5.3.4	Hypergeometrinen jakauma ja negatiivinen hypergeometrinen jakauma . . . . .	127
5.3.5	Tasajakauma . . . . .	129
5.4	Poissonin jakauma . . . . .	129
5.5	Poissonin prosessi . . . . .	135
5.5.1	Laskuri prosessi . . . . .	135
5.5.2	Poissonin prosessin määrittely . . . . .	136
5.5.3	Satunnaistapahtumat tila-avaruudessa . . . . .	138
	Yhteenveto . . . . .	139
	Harjoituksia . . . . .	141

# Luku 5

## Diskreettejä yksiulotteisia jakaumia

Diskreetti satunnaismuuttuja määriteltiin alaluvussa 2.6.1. Olemme jo edellisissä luvuissa käsitelleet hypergeometrista jakaumaa (alaluku 2.7.1), binomijakaumaa (alaluvut 2.9 ja sen erikoistapauksena Bernoullin jakaumaa sekä diskreettiä tasajakaumaa (alaluku 2.17), jotka kaikki ovat esimerkkejä *diskreetteistä jakaumista*.

### 5.1 Diskreetti satunnaismuuttuja

Otosavaruudessa  $\Omega$  määritellyn diskreetin satunnaismuuttujan  $X$  arvojoukko  $S \subset \mathbb{R}$  on numeroituva ja  $P(X \in S) = 1$ . Joukon  $S$  pisteillä on positiivinen todennäköisyys ja ne ovat  $X$ :n kertymäfunktion  $F$  *hyppypisteitä* ja näiden pisteiden todennäköisyydet ovat  $F$ :n hyppyjä.

Määritellään nyt yksinkertainen *hyppyfunktio*  $\varepsilon(x)$  seuraavasti:

$$\varepsilon(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Olkoon  $X$ :n arvoalue  $S = \{1, 2, 3, \dots\}$  ja  $P(x = i) = p_i$ ,  $i \geq 1$ . Silloin  $X$ :n kertymäfunktio  $F(X)$  voidaan kirjoittaa muodossa

$$(5.1.1) \quad F(x) = \sum_{i=1}^{\infty} p_i \varepsilon(x - i).$$

Vaikka usein tarkastelemme vain kokonaislukuarvoisia satunnaismuuttujia, se ei ole oleellinen rajoitus. Olkoon  $S^* = \{x_1, x_2, x_3, \dots\}$  diskreetin satunnaismuuttujan arvojoukko. Silloin joukkojen  $S$  ja  $S^*$  välillä on bijektiivinen vastaavuus  $g(x_i) = i$  ja  $P(X = x_i) = P(g(X) = i)$ , joten voimme aina tarvittaessa siirtyä tarkastelemaan vastaavaa kokonaislukuarvoista satunnaismuuttujaa.

**Esimerkki 5.1** Yksinkertaisin satunnaismuuttuja  $X$  on sellainen, jonka arvoalue  $S = \{c\}$  on yksi piste, jolloin  $P(X = c) = 1$ . Silloin  $X$ :n kertymäfunktio on

$$F(x) = \varepsilon(x - c) = \begin{cases} 1, & x \geq c; \\ 0, & x < c. \end{cases}$$

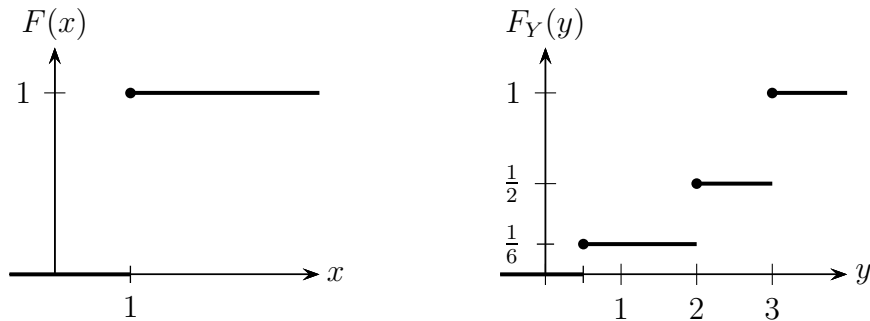
Olkoon  $Y$ :n todennäköisyysfunktio

$$P(Y = \frac{1}{2}) = \frac{1}{6}, \quad P(Y = 2) = \frac{1}{3} \quad \text{ja} \quad P(Y = 3) = \frac{1}{2}.$$

Silloin  $Y$ :n kertymäfunktio on

$$F_Y(y) = \frac{1}{6} \varepsilon(y - \frac{1}{2}) + \frac{1}{3} \varepsilon(y - 2) + \frac{1}{2} \varepsilon(y - 3).$$

□



**Kuvio 5.1.** Funktioiden  $F(x) = \varepsilon(x - 1)$  ja  $F_Y(y)$  kuvaajat.

**Esimerkki 5.2** Hatussa on  $N$  arpalippua, jotka on numeroitu juoksevasti ykkösestä lähtien. Valitaan hatusta arpa satunnaisesti palauttaen  $n$  kertaa ja merkitään valittujen arpojen numerot muistiin. Olkoon  $X$  suurin valittujen arpojen numeroista. Silloin  $P(X \leq r) = (r/N)^n$  ja

$$\begin{aligned} P(X = r) &= P(X \leq r) - P(X \leq r - 1) \\ &= \left(\frac{r}{N}\right)^n - \left(\frac{r-1}{N}\right)^n. \end{aligned}$$

Määritelmän mukaan  $X$ :n odotusarvo on

$$\begin{aligned} E(X) &= N^{-n} \sum_{r=1}^N [r^n - (r-1)^n] r \\ &= N^{-n} \sum_{r=1}^N [r^{n+1} - (r-1)^n r] \\ &= N^{-n} \sum_{r=1}^N [r^{n+1} - (r-1)^n ((r-1) + 1)] \end{aligned}$$

$$\begin{aligned}
&= N^{-n} \sum_{r=1}^N [r^{n+1} - (r-1)^{n+1} - (r-1)^n] \\
&= N^{-n} \left[ N^{n+1} - \sum_{r=1}^N (r-1)^n \right].
\end{aligned}$$

□

## 5.2 Bernoullin kokeet ja binomijakauma

Alaluvussa 2.9 binomijakauma esiteltiin tarkastelemalla otantaa palauttaen ja alaluvussa 4.7 binomijakauma liitettiin Bernoullin kokeisiin. Bernoullin koe on satunnaiskoe, jolla on täsmälleen kaksi toisensa poissulkevaa tulosvaihtoehtoa (onnistuminen ja epäonnistuminen — lyhyesti O ja E). Esimerkiksi mielipidetiedustelussa henkilö kannattaa tai ei kannata ehdokasta, laatukontrollissa tuote on virheetön tai viallinen, hoidon tuloksena potilas paranee tai ei parane.

Satunnaismuuttuja  $X$  noudattaa *Bernoullin jakaumaa*, kun

$$(5.2.1) \quad X = \begin{cases} 1 & \text{todennäköisyydellä } p, \\ 0 & \text{todennäköisyydellä } 1 - p, \end{cases}$$

missä  $0 \leq p \leq 1$ . Nyt siis  $X$  on 'onnistumisen' indikaattorifunktio. Onnistumistodennäköisyys on  $P(X = 1) = p$  ja vastaavasti epäonnistumisen todennäköisyys on  $P(X = 0) = 1 - p$ , jota merkitään usein  $q = 1 - p$ . Bernoullin jakaumaa noudattavan satunnaismuuttujan  $X$  odotusarvo ja varianssi ovat

$$E(X) = p \quad \text{ja} \quad \text{Var}(X) = pq,$$

sillä

$$E(X) = p \cdot 1 + q \cdot 0 = p, \quad E(X^2) = p \cdot 1^2 + q \cdot 0^2 = p$$

ja

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p) = pq.$$

Merkitsemme  $X \sim \text{Ber}(p)$ , kun  $X$  noudattaa Bernoullin jakaumaa, jonka odotusarvo on  $p$ .

Jos  $X \sim \text{Ber}(p)$ , niin  $X$ :n kertymäfunktio on

$$F(x) = (1 - p) \varepsilon(x) + p \varepsilon(x - 1).$$

Yleisesti  $X$ :n  $r$ . momentti

$$E(X^r) = (1 - p) \cdot 0^r + p \cdot 1^r = p$$

on tässä tapauksessa hyvin helppo laskea. Bernoullin jakauman  $\text{Ber}(p)$  momenttifunktio on

$$\begin{aligned}
M(t) &= E(e^{tX}) = P(X = 0)e^{t \cdot 0} + P(X = 1)e^{t \cdot 1} \\
&= (1 - p) + pe^t = 1 + p(e^t - 1),
\end{aligned}$$

joka on määritelty kaikilla  $t \in \mathbb{R}$ .

**Esimerkki 5.3 (Sabharwal 1969).** Olkoon  $n:n$  Bernoullin kokeen jonossa  $X_1, X_2, \dots, X_n$  onnistumistodennäköisyys  $P(O) = p$  ja vastaavasti  $P(E) = 1 - p$  ( $E =$  epäonnistuminen). Olkoon  $Y_n$  tapahtuman OE (osajono) esiintymisten lukumäärä koejonossa. Mikä on tällaisten osajonojen lukumäärän odotusarvo  $E(Y_n)$ ? Määritellään ensin uusi satunnaismuuttuja

$$Z_i = h(X_i, X_{i+1}) = \begin{cases} 1, & \text{jos } X_i = O \text{ ja } X_{i+1} = E; \\ 0 & \text{muulloin,} \end{cases}$$

kun  $i = 1, 2, \dots, n - 1$ . Silloin

$$Y_n = \sum_{i=1}^{n-1} Z_i$$

ja

$$\begin{aligned} E Y_n &= \sum_{i=1}^{n-1} E(Z_i) \\ &= \sum_{i=1}^{n-1} p(1-p) = (n-1)p(1-p). \end{aligned}$$

Jos esimerkiksi  $p = \frac{1}{2}$  ja  $n = 101$ , niin

$$E(Y_n) = \frac{n-1}{4} = 25.$$

□

Tehdään  $n$  riippumatonta Bernoullin koetta, joissa jokaisessa onnistumistodennäköisyys on  $p$ . Olkoon  $i$ . Bernoullin kokeen tulos satunnaismuuttuja  $X_i$ , joka saa arvon 1 tai 0. Silloin koesarjan tulos on riippumattomien samaa Bernoullin jakaumaa noudattavien satunnaismuuttujien jono  $X_1, X_2, \dots, X_n$ , missä  $P(X_i = 1) = p$  ja  $P(X_i = 0) = q$ ,  $i = 1, 2, \dots, n$ . Kun koe on tehty, tulos voisi olla esimerkiksi 111011000...110. Tällaisen tuloksen todennäköisyys (ennen koetta) olisi

$$ppp(1-p)p(1-p)(1-p)ppp \cdots pp(1-p) = p^k(1-p)^{n-k},$$

missä  $k$  on onnistumisten lukumäärä ja  $n - k$  epäonnistumisten lukumäärä. Olkoon  $X$  onnistumisten lukumäärä  $n:ssä$  riippumattomassa Bernoullin kokeessa. Alaluvussa 4.7 totesimme, että  $X$  noudattaa binomijakaumaa parametrein  $n$  ja  $p$ . Silloin merkitään  $X \sim \text{Bin}(n, p)$ . Binomijakauman todennäköisyysfunktio on

$$(5.2.2) \quad f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Esitetään nyt edellä mainittu binomijakauman luonnehdinta Bernoullin kokeiden avulla lauseen muodossa. Jatkossa oletetaan, että Bernoullin kokeet ovat toisistaan riippumattomat, vaikkei oletusta erikseen mainittaisikaan.

**Lause 5.1** *Tehdään  $n$  riippumatonta Bernoullin koetta, joissa jokaisessa onnistumistodennäköisyys on  $p$ . Olkoon  $X$  onnistumisten lukumäärä. Silloin*

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

**Todistus.** Koska  $X$  on onnistumisten lukumäärä  $n$ :ssä riippumatossa Bernoullin kokeessa, niin  $X = X_1 + X_2 + \dots + X_n$ , missä  $X_i \sim \text{Ber}(p) = \text{Bin}(1, p)$ ,  $i = 1, 2, \dots, n$  ovat riippumattomat ja noudattavat samaa Bernoullin jakaumaa. Merkitään nyt  $X = S_n$  ja

$$S_n = X_1 + X_2 + \dots + X_n = S_{n-1} + X_n.$$

Todistamme väitteen induktiolla.

Kun  $n = 1$ , niin oletuksen mukaan  $X = X_1 \sim \text{Ber}(p) = \text{Bin}(1, p)$ , joten väite pitää paikkansa tapauksessa  $n = 1$ . Teemme nyt induktiooletuksen  $S_{n-1} \sim \text{Bin}(n-1, p)$  ja näytämme, että  $S_n \sim \text{Bin}(n, p)$ .

Tapahtuma  $\{S_{n-1} + X_n = k\}$  voidaan lausua yhdisteenä

$$\{S_{n-1} + X_n = k\} = \{S_{n-1} = k, X_n = 0\} \cup \{S_{n-1} = k-1, X_n = 1\},$$

missä  $\{S_{n-1} = k, X_n = 0\}$  ja  $\{S_{n-1} = k-1, X_n = 1\}$  ovat erillisiä tapahtumia. Silloin yhteenlaskusäännön nojalla

$$P(S_{n-1} + X_n = k) = P(S_{n-1} = k, X_n = 0) + P(S_{n-1} = k-1, X_n = 1).$$

Satunnaismuuttujat  $S_{n-1}$  ja  $X_n$  ovat oletuksen mukaan riippumattomat, joten

$$\begin{aligned} P(S_{n-1} + X_n = k) &= P(S_{n-1} = k) P(X_n = 0) + P(S_{n-1} = k-1) P(X_n = 1) \\ &= \binom{n-1}{k} p^k (1-p)^{n-1-k} (1-p) + \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} p \\ &= \binom{n-1}{k} p^k (1-p)^{n-k} + \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= \left[ \binom{n-1}{k} + \binom{n-1}{k-1} \right] p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

missä viimeinen yhtäsuuruus seuraa siitä, että  $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$  [Pascalin kolmio]. Näin on lause todistettu.  $\square$

**Esimerkki 5.4** Erään kasvin siementen itämistodennäköisyydeksi on ilmoitettu 0.8. Siemenen itäminen on tässä ”onnistuminen” ja itämistodennäköisyys on onnistumistodennäköisyys. Jos kylvetään 10 siementä ja siementen itämistapahtumat ovat toisistaan riippumattomat, niin kylvöä voidaan pitää

kymmenenä riippumattomana Bernoullin kokeena, joissa onnistumistodennäköisyys on 0.8. Silloin itävien siementen lukumäärä  $X \sim \text{Bin}(10, 0.8)$ , eli

$$f(x) = \binom{10}{x} 0.8^x \cdot 0.2^{10-x}, \quad x = 0, 1, \dots, 10.$$

Mikä on todennäköisyys, että vähemmän kuin 9 jyvää itää? Todennäköisyys

$$\begin{aligned} P(X < 9) &= P(X \leq 8) = 1 - \sum_{k=9}^{10} P(X = k) \\ &= 1 - 10 \cdot 0.8^9 \cdot 0.2 - 0.8^{10} = 0.6242. \end{aligned}$$

□

Laskemme usein muotoa  $P(X \leq x)$  olevia todennäköisyyksiä, kuten edellisessä esimerkissä. Todennäköisyydet  $P(X \leq x)$  määrittelevät jakauman kertymäfunktion

$$F(x) = P(X \leq x).$$

Kertymäfunktio määriteltiin alaluvussa ???. Binomijakauman kertymäfunktion arvot pisteissä  $x = 0, 1, \dots, n$  ovat

$$F(x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}.$$

**Lause 5.2** Jos  $X \sim \text{Bin}(n, p)$ , niin

1.  $X$ :n todennäköisyysfunktio  $f(x)$  on

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

kaikilla  $n \in \mathbb{N}$  ja kaikilla  $p \in [0, 1]$ ;

2.  $X$ :n kertymäfunktio  $F(y)$  on

$$F(y) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \varepsilon(y-x)$$

kaikilla  $y \in \mathbb{R}$ , missä  $\varepsilon(y)$  on hyppyfunktio;

3.  $X$ :n odotusarvo, varianssi ja momenttifunktio ovat

$$\begin{aligned} \mu &= E(X) = np, & \text{Var}(X) &= np(1-p), \\ M(t) &= E(e^{tX}) = (1-p + pe^t)^n, & -\infty < t < \infty. \end{aligned}$$

**Todistus.** 1. Binomijakauman todennäköisyysfunktio johdettiin Lauseen 5.1 todistuksessa.

2. Odotusarvo ja varianssi. Koska  $X = X_1 + X_2 + \dots + X_n$  on riippumattomien Bernoullin muuttujien  $X_i \sim \text{Ber}(p)$  summa, niin

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \dots + E(X_n) \\ &= p + p + \dots + p = np \end{aligned}$$

ja

$$\begin{aligned} \text{Var}(X) &= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) \\ &= p(1-p) + p(1-p) + \dots + p(1-p) = np(1-p). \end{aligned}$$

3. Momenttifunktio on

$$\begin{aligned} M(t) &= E(e^{tX}) \\ &= E(e^{t(X_1+X_2+\dots+X_n)}) = E(e^{tX_1+tX_2+\dots+tX_n}) \\ &= E(e^{tX_1}e^{tX_2}\dots e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2})\dots E(e^{tX_n}), \end{aligned}$$

missä viimeinen yhtäsuuruus seuraa lauseista 4.5 ja 4.8. Koska  $X_i$  ja  $X_j$  ( $i \neq j$ ) ovat riippumattomat, niin  $e^{tX_i}$  ja  $e^{tX_j}$  ovat riippumattomat (Lause 4.5) ja riippumattomien satunnaismuuttujien  $e^{tX_1}, e^{tX_2}, \dots, e^{tX_n}$  tulon odotusarvo on yksittäisten tulon tekijöiden odotusarvojen tulo (Lause 4.8). Koska

$$M_{X_i}(t) = E(e^{tX_i}) = 1 - p + pe^t, \quad i = 1, 2, \dots, n,$$

niin

$$M(t) = (1 - p + pe^t)^n \quad \text{kaikilla } t \in \mathbb{R}.$$

Momenttifunktio itse asiassa määrittelee yksikäsitteisesti todennäköisyysfunktion (Lause 4.10). Näytämme kuitenkin vielä eksplisiittisesti, että binomitodennäköisyydet määrittelevät todennäköisyysfunktion. Koska Binomilauseen 2.6 perusteella

$$[p + (1-p)]^n = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = 1$$

kaikilla  $p \in [0, 1]$ , niin todennäköisyydet  $f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$  määrittelevät todennäköisyysfunktion kaikilla  $p \in [0, 1]$  ja  $n \geq 1$ . Huomaa myös, että

$$M(0) = (1 - p + pe^0)^n = [p + (1-p)]^n.$$

□

**Seuraus 5.1** Jos  $X_1 \sim \text{Bin}(n_1, p)$  ja  $X_2 \sim \text{Bin}(n_2, p)$  ovat riippumattomat, niin  $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$ .



**Todistus.** Koska Lauseen 5.2 mukaan  $X_1$ :n momenttifunktio on  $(1 - p + pe^t)^{n_1}$  ja  $X_2$ :n momenttifunktio on  $(1 - p + pe^t)^{n_2}$ , niin satunnaismuuttujan  $X_1 + X_2$  momenttifunktio on Lauseen 4.11 mukaan  $(1 - p + pe^t)^{n_1+n_2}$ . Mutta Lauseen 5.2 perusteella  $(1 - p + pe^t)^{n_1+n_2}$  on binomijakuman  $\text{Bin}(n_1 + n_2, p)$  momenttifunktio. Tästä seuraa momenttifunktion yksikäsitteisyyden (Lause 4.10) nojalla, että  $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$ .  $\square$

Seurauslauseen 5.1 todistuksessa on käytetty esimerkin vuoksi yleistä momenttifunktiootekniikkaa. Tässä tapauksessa tulos saadaan kuitenkin helposti turvautumatta noin voimakkaisiin menetelmiin. Koska  $X_1$  esittää onnistumisten lukumäärää  $n_1$ :ssä Bernoullin kokeessa ja  $X_2$  onnistumisten lukumäärää  $n_2$ :ssa kokeessa, missä  $p$  on jokaisen kokeen onnistumistodennäköisyys, niin riippumattomien satunnaismuuttujien  $X_1$  ja  $X_2$  summa  $X_1 + X_2$  esittää onnistumisen lukumäärää  $(n_1 + n_2)$ :ssa kokeessa. Tämän perusteella saadaan tulos  $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$ . Analyttisesti tulos voidaan tarkistaa laskemalla lauseke

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{i=0}^{n_1} P(X_1 = i, X_2 = k - i) \\ &= \sum_{i=0}^{n_1} P(X_1 = i) P(X_2 = k - i) \\ &= \sum_{i=0}^{n_1} \binom{n_1}{i} p^i (1 - p)^{n_1 - i} \binom{n_2}{k - i} p^{k - i} (1 - p)^{n_2 - k + i}, \end{aligned}$$

missä  $\binom{n_2}{k - i} = 0$  kaikilla  $k - i > n_2$ . Tästä seuraa

$$P(X_1 + X_2 = k) = p^k (1 - p)^{n_1 + n_2 - k} \sum_{i=0}^{n_1} \binom{n_1}{i} \binom{n_2}{k - i}.$$

Soveltamalla hypergeometrista identiteettiä (ks. Lause 2.7)

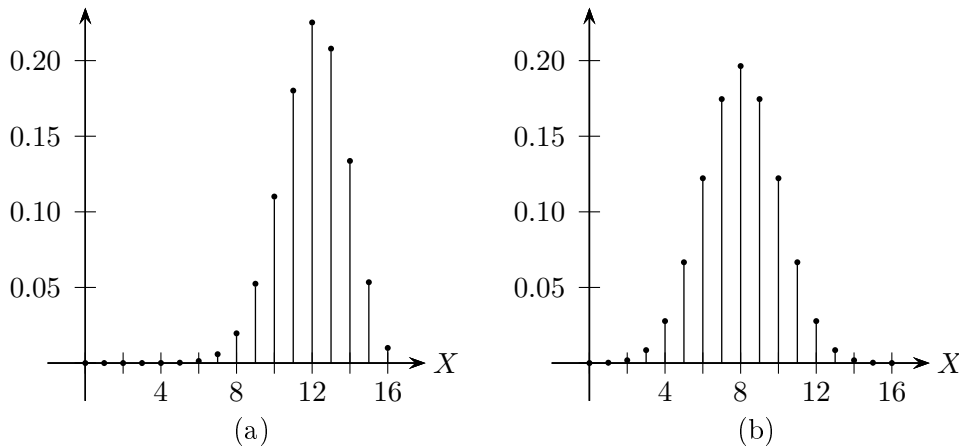
$$\binom{n_1 + n_2}{k} = \sum_{i=0}^{n_1} \binom{n_1}{i} \binom{n_2}{k - i}$$

saadaan kaivattu tulos.

### 5.2.1 Jakauman symmetria

Symmetriaan perustuvaa argumentointia voidaan usein hyödyntää todennäköisyyksien laskemisessa.

**Symmetria pisteen suhteen.** Jos  $P(X = b + x) = P(X = b - x)$  kaikilla  $x$ , niin  $X$ :n jakauma on *symmetrinen pisteen  $b$  suhteen*. Satunnaismuuttuja  $X$



**Kuvio 5.2.** Binomijakauman kuvaajat, kun (a)  $X \sim \text{Bin}(16, 0.75)$   
(b)  $X \sim \text{Bin}(16, 0.50)$ .

on symmetrinen  $b$ :n suhteen jos ja vain jos  $X - b$  on symmetrinen origon suhteen. Silloin

$$P(X \leq b - x) = P(X \geq b + x).$$

Esimerkiksi binomijakauma  $\text{Bin}(16, 0.50)$  on symmetrinen pisteen 8 suhteen, mutta binomijakauma  $\text{Bin}(16, 0.75)$  ei ole symmetrinen (Kuvio 5.2). Binomijakaumassa  $\text{Bin}(16, 0.50)$  on

$$P(X = 8 + x) = P(X = 8 - x)$$

kaikilla  $x$ . Silloin jokaista  $a \in \mathbb{R}$  kohti

$$P(X \leq 8 - a) = P(X \geq 8 + a).$$

## 5.3 Odotusaikojen jakaumat

Monissa sovelluksissa on kiinnostuksen kohteena odotusaika siihen hetkeen, että jokin tietty tapahtuma sattuu. Tässä alaluvussa käsitellään Bernoullin kokeisiin ja yksinkertaiseen satunnaisotantaan liittyviä odotusaikatehtäviä.

### 5.3.1 Odotusajat Bernoullin kokeissa

Tarkastellaan riippumattomien samaa Bernoullin jakaumaa noudattavien satunnaismuuttujien jonoa  $X_1, X_2, \dots, X_n$ , missä  $X_i \sim \text{Ber}(p)$ . Määritellään satunnaismuuttujat  $S_n$  ja  $W_r$  seuraavasti:

$$S_n = X_1 + X_2 + \dots + X_n,$$

$$W_r = r\text{:ään onnistumiseen tarvittavien yrittysten määrä.}$$

Jos ajattemme, että yhteen Bernoullin kokeeseen kuluu yhden yksikön pituinen aika, niin  $S_n$  vie  $n$  aikayksikköä. Nyt siis  $W_r$  on  $r$ :n onnistumisen

saavuttamiseen tarvittava aika eli *odotusaika* ja sen mahdolliset arvot ovat  $r, r + 1, r + 2, \dots$ . Tiedämme, että  $S_n \sim \text{Bin}(n, p)$ , mutta mikä on  $W_r$ :n jakauma?

**Esimerkki 5.5** Heitetään harhatonta lanttia, kunnes saadaan kruuna (R). Olkoon  $W_1$  tarvittavien heittojen lukumäärä. Tapahtuma  $\{W_1 = x\}$  sattuu vain silloin, kun  $(x - 1)$ :llä ensimmäisellä heitolla on saatu pelkkiä klaavoja (L) ja  $x$ . heitolla saadaan kruuna:

$$\underbrace{\text{LLL} \dots \text{L}}_{x-1 \text{ kertaa}} \text{R.}$$

Tästä seuraa, että

$$P(W_1 = x) = \frac{1}{2^x}, \quad x = 1, 2, \dots$$

Satunnaismuuttujan  $W_1$  odotusarvo on määritelmän mukaan

$$(5.3.1) \quad E(W_1) = \sum_{x=1}^{\infty} \frac{x}{2^x}.$$

Tiedämme, että

$$(5.3.2) \quad \sum_{x=0}^{\infty} p^x = 1 + p + p^2 + p^3 + \dots = \frac{1}{1-p}, \quad \text{kun } |p| < 1.$$

Kun derivoimme sarjan (5.3.2) termeittäin, saamme

$$(5.3.3) \quad 0 + 1 + 2p + 3p^2 + \dots = \sum_{x=0}^{\infty} (x+1)p^x = \frac{1}{(1-p)^2}, \quad \text{kun } |p| < 1.$$

Koska sarjan (5.3.2) suppenemissäde on 1, suppenee derivointioperaation tuloksena saatu sarja (5.3.3) arvoilla  $|p| < 1$ . Sijoittamalla  $p = \frac{1}{2}$  sarjaan (5.3.3) saadaan

$$\sum_{x=0}^{\infty} (x+1) \left(\frac{1}{2}\right)^x = 4,$$

joka voidaan esittää muodossa

$$\sum_{x=0}^{\infty} x \left(\frac{1}{2}\right)^x + \sum_{x=0}^{\infty} \left(\frac{1}{2}\right)^x = \sum_{x=0}^{\infty} x \left(\frac{1}{2}\right)^x + 2 = 4,$$

missä summa  $\sum_{x=0}^{\infty} \left(\frac{1}{2}\right)^x = 2$  saadaan kaavasta (5.3.2). Nyt siis odotusarvo (5.3.1) on 2.

Jos kruunan todennäköisyys on  $p$ , niin silloin

$$P(W_1 = x) = \underbrace{(1-p)(1-p) \dots (1-p)}_{x-1 \text{ kertaa}} p = (1-p)^{x-1} p$$

ja

$$\begin{aligned} E(W_1) &= \sum_{x=1}^{\infty} x(1-p)^{x-1}p = p \sum_{x=0}^{\infty} (x+1)(1-p)^x \\ &= p \cdot \frac{1}{[1-(1-p)]^2} = \frac{1}{p}, \end{aligned}$$

missä sarjan summa saadaan (5.3.3):n avulla. Satunnaismuuttuja  $W_1$  on siis kruunan tai yleisemmin 'onnistumisen' odotusaika. Jakaumaa

$$(5.3.4) \quad P(W_1 = x) = (1-p)^{x-1}p, \quad x = 1, 2, \dots$$

kutsutaan *geometriseksi jakaumaksi*. Todennäköisyydet (5.3.4) todellakin määrittelevät jakauman, koska

$$\sum_{x=1}^{\infty} P(W_1 = x) = \sum_{x=1}^{\infty} (1-p)^{x-1}p = p \cdot \sum_{x=0}^{\infty} (1-p)^x = p \cdot \frac{1}{p} = 1.$$

□

Tapahtuma  $\{W_r = x\}$  sattuu, kun  $(x-1)$ :ssä ensimmäisessä kokeessa on saatu  $r-1$  onnistumista ja  $x$ . kokeessa saadaan onnistuminen:

$$\begin{array}{l} \underbrace{\text{OOEOE} \dots \text{EO}} \\ \left. \begin{array}{l} x-1 \text{ koetta,} \\ r-1 \text{ onnistumista,} \\ \text{kokeiden järjestys} \\ \text{mielivaltainen} \end{array} \right\} \begin{array}{l} x. \text{ koe,} \\ r. \text{ onnistuminen} \end{array} \end{array}$$

Nyt siis  $\{W_r = x\} = \{S_{x-1} = r-1, X_x = 1\}$ . Koska  $X_i$ :t ( $i = 1, 2, \dots, x$ ) ovat riippumattomat, niin myös  $S_{x-1}$  ja  $X_x$  ovat riippumattomat. Silloin

$$\begin{aligned} (5.3.5) \quad P(W_r = x) &= P(S_{x-1} = r-1) P(X_x = 1) \\ &= \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r} p = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \end{aligned}$$

koska  $S_{x-1} \sim \text{Bin}(x-1, p)$ . Todennäköisyydet (5.3.5) määrittelevät ns. *negatiivisen binomijakauman*. Soveltamalla identiteettiä [ks. (2.4.6)]

$$\frac{r}{x} \binom{x}{r} = \binom{x-1}{r-1}$$

saadaan

$$P(W_r = x) = \frac{r}{x} P(S_x = r).$$

Toinen usein käyttökelpoinen identiteetti on

$$P(W_r > x) = P(S_x < r).$$

### 5.3.2 Geometrinen jakauma ja negatiivinen binomijakauma

Sanomme, että satunnaismuuttuja  $X$  noudattaa *negatiivista binomijakaumaa* parametrein  $r$  ja  $p$ , jos

$$(5.3.6) \quad P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, r+2, \dots$$

Merkitsemme silloin

$$X \sim \text{NBin}(r, p).$$

Edellisessä pykälässä huomasimme, että odotusaika  $W_r \sim \text{NBin}(r, p)$ . Kun  $r = 1$ , sanomme negatiivista binomijakaumaa *geometriseksi jakaumaksi*. Geometrisen jakauman todennäköisyysfunktio on siis

$$(5.3.7) \quad f(x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \dots$$

Kun siis  $X \sim \text{NBin}(1, p)$ , niin  $X$ :n noudattaa geometrista jakaumaa parametrilla  $p$ . Merkitsemme silloin  $X \sim \text{Geo}(p)$ .

**Lause 5.3** *Oletetaan, että  $X \sim \text{NBin}(r, p)$ .*

1. *Funktio (5.3.6) on negatiivisen binomijakauman todennäköisyysfunktio kaikilla positiivisilla kokonaisluvuilla  $r$  ja kaikilla  $0 < p < 1$  ja*

2.

$$E(X) = \frac{r}{p}, \quad \text{Var}(X) = \frac{r(1-p)}{p^2},$$

$$M(t) = E(e^{tX}) = \frac{(pe^t)^r}{[1 - (1-p)e^t]^r}, \quad t < -\log(1-p).$$

**Todistus.** Johdamme ensin negatiivisen binomijakauman momenttifunktion suoraan määritelmän nojalla. Koska  $M(t) = E(e^{tX})$ , niin momenttifunktio

on

$$\begin{aligned}
E(e^{tX}) &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\
&= p^r \sum_{y=0}^{\infty} e^{t(y+r)} \binom{r+y-1}{r-1} p^r (1-p)^y \\
&= p^r e^{tr} \sum_{y=0}^{\infty} e^{ty} \binom{r+y-1}{y} (1-p)^y \\
&= p^r e^{tr} \sum_{y=0}^{\infty} e^{ty} (-1)^y \binom{-r}{y} (1-p)^y \\
&= p^r e^{tr} \sum_{y=0}^{\infty} \binom{-r}{y} [-(1-p)e^t]^y \\
&= p^r e^{tr} [1 - (1-p)e^t]^{-r} = \left[ \frac{pe^t}{1 - (1-p)e^t} \right]^r.
\end{aligned}$$

Binomisarja  $\sum_{y=0}^{\infty} \binom{-r}{y} [-(1-p)e^t]^y$  suppenee, kun  $(1-p)e^t < 1$ , joka on yhtäpitävä epäyhtälön  $t < -\log(1-p)$  kanssa.

Koska  $M(0) = 1$  kaikilla positiivisilla kokonaisluvuilla  $r$  ( $r \in \mathbb{N}$ ) ja kaikilla  $0 < p < 1$ , niin (5.3.6) on todennäköisyysfunktio kaikilla  $r \in \mathbb{N}$  ja kaikilla  $0 < p < 1$ . Odotusarvo ja varianssi saadaan laskemalla ensin  $M(t)$ :n 1. ja 2. derivaatta ja niiden avulla

$$E(X) = M'(0) \quad \text{ja} \quad \text{Var}(X) = M''(0) - [M'(0)]^2.$$

□

**Seuraus 5.2** Jos  $X \sim \text{NBin}(1, p)$  eli  $X$  noudattaa geometrista jakaumaa  $\sim \text{Geo}(p)$  ja

1. geometrisen jakauman todennäköisyysfunktio on (5.3.7) kaikilla  $0 < p < 1$  ja
- 2.

$$\begin{aligned}
E(X) &= \frac{1}{p}, & \text{Var}(X) &= \frac{1-p}{p^2}, \\
M(t) = E(e^{tX}) &= \frac{pe^t}{[1 - (1-p)e^t]}, & t &< -\log(1-p).
\end{aligned}$$

Olkoon  $Y$  epäonnistumisten lukumäärä Bernoullin toistokokeessa, ennen kuin saadaan  $r$ . onnistuminen. Koska  $r$ . onnistumiseen tarvittavien yritysten määrä  $W_r \sim \text{NBin}(r, p)$ , niin

$$Y = W_r - r \quad \text{ja} \quad E(Y) = E(W_r) - r = \frac{r}{p} - r = \frac{r(1-p)}{p}.$$

$Y$ :n varianssi on tietysti sama kuin  $W_r$ :n varianssi. Nyt siis  $P(Y = y) = P(W_r = r + y)$  kaikilla  $y = 0, 1, 2, \dots$

Nimitys ”negatiivinen binomijakauma” on peräisin esitystavasta

$$1 = p^r \cdot p^{-r} = p^r [1 - (1 - p)]^{-r} = p^r \sum_{y=0}^{\infty} \binom{-r}{y} [-(1 - p)]^y,$$

mistä saadaan todennäköisyydet  $P(W_r = y + r)$ ,  $y = 0, 1, 2, \dots$ . Merkintä  $\binom{-r}{y}$  on määritelmänsä mukaan

$$\binom{-r}{y} = \frac{(-r)^{(y)}}{y!} = (-1)^y \binom{r + y - 1}{y},$$

missä  $r > 0$  ja  $y \geq 0$  ovat kokonaislukuja.

**Esimerkki 5.6** Geometrisella jakaumalla ja negatiivisella binomijakaumalla on tärkeä merkitys esimerkiksi jonoteoriassa. Oletetaan, että joukko asiakkaita jonottaa pääsyä palvelutiskille. Olkoon todennäköisyys  $p$ , että jokaisella pienellä aikavälillä tulee 1 uusi asiakas (0 uutta asiakasta todennäköisyydellä  $1 - p = q$ ). Silloin seuraavan asiakkaan odotusaika  $W \sim \text{Geo}(p)$ . Todennäköisyys  $P(W > k)$ , että seuraavan  $k$ :n aikayksikön aikana ei tule asiakasta, on

$$\begin{aligned} P(W > k) &= \sum_{j=k+1}^{\infty} q^{j-1} p = q^k (p + qp + q^2 p + \dots) \\ &= q^k = 1 - P(W \leq k). \end{aligned}$$

□

Geometrisen jakauman kertymäfunktio on siis

$$\begin{aligned} F(k) &= P(W \leq k) = \sum_{i=1}^k (1 - p)^{i-1} p \\ &= 1 - P(W > k) = 1 - q^k, \end{aligned}$$

missä  $q = 1 - p$  ja  $k = 1, 2, \dots$ . Geometrisen jakauman kertymäfunktion arvot saadaan geometrisesta sarjasta, josta jakauman nimi tulee.

Usein oletetaan, että myös asiakkaan palvelemiseen käytetty aika (palveluaika) noudattaa geometrista jakaumaa. Palveluajan jakaumalla on tietysti yleensä eri parametrin  $p$  arvo kuin palvelun odotusajan jakaumalla. Geometrisella jakaumalla on ”unohtamisominaisuus”, joka havaitaan laskemalla seuraava ehdollinen todennäköisyys:

$$(5.3.8) \quad P(W > k + s \mid W > k) = \frac{P(W > k + s)}{P(W > k)} = \frac{q^{k+s}}{q^k} = q^s.$$

Nyt siis todennäköisyys, että asiakkaan palveleminen kestää vielä  $s$  aikayksikköä, ei riipu siitä, kuinka kauan häntä on jo palveltu. Onneksi kuitenkin käytännössä palveluaika ei aina täysin noudata geometrista jakaumaa.

**Esimerkki 5.7 Banachin tulitikkuongelma.** Piippua polttelevalle matemaatikolla oli tapana pitää yksi tulitikkulaatikko oikeassa ja yksi vasemmassa taskussa. Joka kerta tikkua tarvitessaan hän valitsi taskun täysin satunnaisesti, joten kummankin taskun valintatodennäköisyys on  $\frac{1}{2}$ . Tarkastellaan tapahtumaa, että matemaatikko huomaa laatikon olevan tyhjä. Oletetaan, että kummassakin laatikossa oli alunperin  $N$  tikkua. Mikä on todennäköisyys, että toisessa laatikossa on täsmälleen  $k$  tikkua ( $k = 0, 1, \dots, N$ ) silloin, kun matemaatikko havaitsee toisen laatikon olevan tyhjä?

Olkoon  $A$  tapahtuma, että matemaatikko huomaa oikeanpuoleisen laatikon olevan tyhjä ja samalla vasemman taskun laatikossa on  $k$  tikkua. Tapahtuma voi sattua täsmälleen silloin, kun oikeanpuoleisen taskun laatikosta valitaan tikku ( $N+1$ ). kerran ja yhteensä valintoja on tehty  $N+1+N-k$  kappaletta. Teemme siis valintoja palauttamatta. Molemmista laatikoissa on  $N$  tikkua, joten tapahtuma  $A$  on ekvivalentti tapahtuman  $\{W_{N+1} = N+1+N-k\}$  kanssa. Saamme kaavalla (5.3.6) todennäköisyydeksi

$$P(W_{N+1} = N+1+N-k) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k+1}.$$

Koska myös todennäköisyys, että vasemmanpuoleinen laatikko huomataan tyhjäksi ja oikeanpuoleisessa on  $k$  tikkua, on  $P(W_{N+1} = N+1+N-k)$ , niin vastaus kysymykseen on

$$2P(W_{N+1} = N+1+N-k) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k}.$$

□

### 5.3.3 Odotusajat peräkkäisotannassa

Oletetaan, että populaatiossa on kahdenlaisia alkioita. Valitaan populaatiosta peräkkäisotos. Käytetään nyt apuna urnamallia. Olkoon urnassa  $a$  valkoista palloa ja  $b$  mustaa palloa eli yhteensä  $a+b=N$  palloa. Poimitaan satunnaisvalinnalla palloja urnasta yksitellen. Määritellään satunnaismuuttujat

$S_n$  = valkoisten pallojen (onnistumisten) lukumäärä  
 $n$ :ssä ensimmäisessä nostossa;

$W_r$  =  $r$ :n valkoisen pallon saamiseksi tarvittavien nostojen määrä.

Jos ajatellaan, että nostoon menee yksi aikayksikkö, niin  $W_r$  on  $r$ :n valkoisen pallon saamiseksi tarvittava odotusaika.

Jos otanta tehdään *palauttaen*, niin peräkkäiset nostot ovat riippumattomia Bernoullin kokeita, joissa onnistumistodennäköisyys on  $p = a/N$ . Tässä tapauksessa voidaan suoraan soveltaa edellä esitettyjä Bernoullin kokeita koskevia tuloksia.

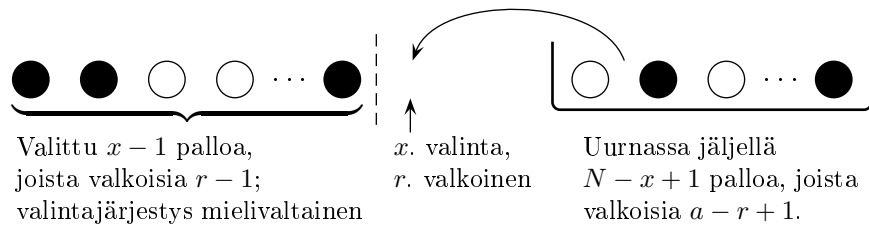


Kun otanta tehdään *palauttamatta*, peräkkäiset nostot eivät ole riippumattomia, koska valkoisten pallojen suhteellinen osuus urnassa riippuu siitä, mitä sieltä on jo valittu. Alaluvussa 2.7.1 osoitimme, että  $S_n$  noudattaa hypergeometrista jakaumaa, kun otanta tehdään palauttamatta (ks. myös alaluku 3.3.3). Silloin

$$(5.3.9) \quad P(S_n = x) = \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}},$$

kun  $x = 0, 1, \dots, n$ . Mikä on todennäköisyys, että saamme  $x$ . nostossa  $r$ . valkoisen pallon?

Tapahtuma  $\{W_r = x\}$  sattuu täsmälleen silloin, kun  $x - 1$  ensimmäisessä nostossa on saatu  $r - 1$  valkoista ja  $x$ . nostossa saadaan valkoinen:



Voimme siis kirjoittaa  $\{W_r = x\} = \{S_{x-1} = r - 1, X_x = 1\}$ , missä  $S_{x-1} \sim \text{HGeo}(x - 1, N, a/x)$  [ks. Esimerkki 4.5 ja (4.1.5)] ja  $X_x = 1$ , kun valitaan valkoinen pallo  $x$ . nostossa. Tästä seuraa, että

$$(5.3.10) \quad \begin{aligned} P(W_r = x) &= P(S_{x-1} = r - 1, X_x = 1) \\ &= P(S_{x-1} = r - 1) P(X_x = 1 \mid S_x = r - 1) \\ &= \frac{\binom{a}{r-1} \binom{N-a}{x-r}}{\binom{N}{x-1}} \cdot \frac{a - r + 1}{N - x + 1}, \end{aligned}$$

kun  $x = r, r + 1, \dots, N$ .

Todennäköisyys (5.3.10) voidaan kirjoittaa lausekkeena

$$(5.3.11) \quad P(W_r = x) = \binom{x-1}{r-1} \frac{\binom{N-x}{a-r}}{\binom{N}{a}},$$

joka on *negatiivisen hypergeometrisen jakauman* todennäköisyysfunktio. Koska  $\binom{x-1}{r-1} = \frac{r}{x} \binom{x}{r}$ , niin

$$P(W_r = x) = \frac{r}{x} \cdot \frac{\binom{x}{r} \binom{N-x}{a-r}}{\binom{N}{a}} = \frac{r}{x} P(S_x = r),$$

missä  $S_x \sim \text{HGeo}(x, N, a/N)$ . Vastaavanlainen tulos saatiin otannassa palauttaen. Samoin on jälleen helppo nähdä, että

$$P(W_r > x) = P(S_x < r).$$

Merkitään  $W_r \sim \text{NHGeo}(r, N, p)$ , missä  $p = a/N$ .

### 5.3.4 Hypergeometrinen jakauma ja negatiivinen hypergeometrinen jakauma

Olemme esitelleet hypergeometrisen jakauman tarkastelemalla otantaa palauttamatta (alaluku 2.7.1). Jakauman avulla voidaan siis ratkaista otantaan liittyviä todennäköisyystehtäviä. Hypergeometrisen jakauman momenttifunktiolla  $M(t)$  ei ole olemassa siistiä lauseketta, vaikka se tietysti voidaan lausua määritelmänsä mukaan äärellisenä summana, koska satunnaismuuttujan arvojoukko on äärellinen. Hypergeometrisen jakauman odotusarvon ja varianssin laskeminen ei myöskään ole aivan helppo tehtävä.

Olemme merkinneet populaation alkioden lukumäärää  $N = a + b$ , joista  $a$  kappaletta on tyyppiä A ja  $b$  kappaletta tyyppiä B. Esimerkiksi tuotepopulaatiossa on  $a$  viallista. Tyyppiä A olevien alkioden suhteellinen osuus on  $p = a/N$ . Tyyppiä A olevan alkion valinta on ”onnistuminen” ja tyyppiä B valinta ”epäonnistuminen”. Valitaan populaatiosta  $n$ :n alkion otos palauttamatta. Olkoon  $X$  onnistuneiden valintojen lukumäärä otoksessa. On selvää, että  $0 \leq X \leq n$ . Koska populaatiossa on  $pN$  kappaletta tyyppiä A olevia alkioita ja  $(1-p)N$  kappaletta tyyppiä B, niin  $X \leq pN$  ja  $n - X \leq (1-p)N$ . Siksi  $X$ :n arvoalue  $S$  on ehdon

$$\max\{0, n - (1-p)N\} \leq x \leq \min\{n, pN\}$$

toteuttavien kokonaislukujen  $x$  joukko.

Kun  $X$  noudattaa hypergeometrista jakaumaa  $\text{HGeo}(n, N, p)$ , niin  $X$ :n todennäköisyysfunktio on

$$(5.3.12) \quad f(x) = P(X = x) = \frac{\binom{Np}{x} \binom{N-Np}{n-x}}{\binom{N}{n}}, \quad x \in S.$$

Huomattakoon, että todennäköisyys (5.3.12) on määritelty myös arvoilla  $x \notin S$ , mutta silloin  $f(x) = 0$ .

**Lause 5.4** *Oletetaan, että  $X \sim \text{HGeo}(n, N, p)$ . Silloin*

$$E(X) = np \quad \text{ja} \quad \text{Var}(X) = \frac{N-n}{N-1} np(1-p).$$

**Todistus.** Hypergeometrisen jakauman odotusarvo laskettiin esimerkissä 4.5 ja alaluvussa 3.3.3. Varianssi voidaan laskea vastaavalla tavalla.  $\square$

**Lause 5.5** *Oletetaan, että  $Y \sim \text{NHGeo}(r, N, p)$ . Silloin*

$$E(Y) = r \cdot \frac{N+1}{Np+1} \quad \text{ja} \quad \text{Var}(Y) = \frac{rN(N+1)(1-p)(Np+1-r)}{(Np+1)^2(Np+2)}.$$

Mainitsimme jo alaluvussa 2.9.1, että binomijakaumaa voidaan käyttää hypergeometrisen jakauman likiarvona, kun  $N$  on suuri. Erityisesti, kun  $N$  on ääretön tai hyvin suuri (verrattuna otoskokoon), on yhdentekevää, käytetäänkö otantaa palauttaen vai palauttamatta. Oletetaan nyt, että

$$X_N \sim \text{HGeo}(n, N, p) \quad \text{ja} \quad X \sim \text{Bin}(n, p).$$

Kun parametrit  $n$  ja  $p$  ovat annettuja vakioita ja  $N$  kasvaa rajatta, voimme osoittaa, että  $X_N$ :n jakauma lähestyy  $X$ :n jakaumaa. Silloin siis

$$X_N \xrightarrow{d} X, \quad \text{kun} \quad N \rightarrow \infty.$$

Koska  $X \sim \text{Bin}(n, p)$ , niin

$$X_N \xrightarrow{d} \text{Bin}(n, p),$$

eli  $X_N$ :n jakauma lähestyy binomijakaumaa, jonka parametrit ovat  $n$  ja  $p$ . Sanomme myös, että  $X_N$ :n jakauma suppenee kohti  $X$ :n jakaumaa  $N$ :n kasvaessa. Kutsumme  $X$ :n jakaumaa  $X_N$ :n *asymptoottiseksi jakaumaksi*.

Lauseen 4.4 mukaan satunnaismuuttujilla on sama jakauma, jos niillä on sama kertymäfunktio. Voimme nyt tarkastella satunnaismuuttujien jonoa

$$\{X_N; N = 1, 2, \dots\} = X_1, X_2, \dots$$

ja vastaavaa kertymäfunktioiden jonoa

$$\{F_N; N = 1, 2, \dots\} = F_1, F_2, \dots,$$

missä  $F_N(x)$  on  $X_N$ :n kertymäfunktio.

**Määritelmä 5.1** Jono  $\{X_N; N = 1, 2, \dots\}$  suppenee jakaumaltaan kohti satunnaismuuttujaa  $X$ , jos

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

kaikissa pisteissä  $x \in \mathbb{R}$ , joissa  $X$ :n kertymäfunktio  $F(x)$  on jatkuva.

Diskreettien satunnaismuuttujien tapauksessa voidaan helposti todistaa tulos, joka osoittaa, että suppenemista jakaumamieleessä voidaan tarkastella yhtä hyvin myös todennäköisyysfunktioiden avulla.

**Lause 5.6** *Olkoon  $\{X_N; N = 1, 2, \dots\}$  sellainen epänegatiivisten kokonaislukuarvoisten satunnaismuuttujien jono, että  $X_N$ :n todennäköisyysfunktio on  $f_N(k)$ ,  $N = 1, 2, \dots$ . Olkoon  $X$  epänegatiivinen kokonaislukuarvoinen satunnaismuuttuja, jonka todennäköisyysfunktio on  $f(k)$ . Silloin*

$$X_N \xrightarrow{d} X \Leftrightarrow \lim_{N \rightarrow \infty} f_N(k) = f(k)$$

*kaikilla epänegatiivisilla kokonaisluvuilla  $k$ .*

**Todistus.** Jätetään harjoitustehtäväksi. □

**Lause 5.7** Jos  $X_N \sim \text{HGeo}(n, N, p)$ , niin

$$X_N \xrightarrow{d} \text{Bin}(n, p), \quad \text{kun } N \rightarrow \infty.$$

**Todistus.** Käytetään lausetta 5.6 ja osoitetaan, että  $P(X_N = k) = f_N(k) \rightarrow f(k)$  kaikilla epänegatiivisilla kokonaisluvuilla  $k$ , kun  $N \rightarrow \infty$ . Yksityiskohdat jätetään lukijan pohdittavaksi. □

### 5.3.5 Tasajakauma

Diskreetti tasajakauma esiteltiin ensimmäisen kerran alaluvussa 2.17. Satunnaismuuttuja  $X$ , jonka arvoavaruus on  $S = \{1, 2, \dots, N\}$ , noudattaa diskreettiä tasajakaumaa, jos

$$P(X = x) = \frac{1}{N}, \quad x = 1, 2, \dots, N.$$

Silloin merkitään  $X \sim \text{Tasd}(1, 2, \dots, N)$ , missä  $N \geq 1$  on annettu positiivinen kokonaisluku. Jos  $X \sim \text{Tasd}(1, 2, \dots, N)$ , niin

$$E(X) = \frac{N+1}{2} \quad \text{ja} \quad \text{Var}(X) = \frac{(N+1)(N-1)}{12}.$$

## 5.4 Poissonin jakauma

Satunnaismuuttuja  $X$ , jonka todennäköisyysfunktio on

$$(5.4.1) \quad f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots$$

noudattaa *Poissonin jakaumaa* parametrilla  $\lambda > 0$ , joka on Poissonin jakauman odotusarvo. Silloin merkitään

$$X \sim \text{Poi}(\lambda).$$

Poissonin jakaumalla on runsaasti sovelluksia eri aloilla. Sitä voidaan käyttää myös binomijakauman  $\text{Bin}(n, p)$  likiarvona, kun  $n$  on suuri ja  $p$  pieni. Silloin siis pätee

$$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{e^{-np} (np)^x}{x!}.$$

**Lause 5.8** Olkoon  $X \sim \text{Poi}(\lambda)$ . Silloin

1. funktio (5.4.1) on Poissonin jakauman todennäköisyysfunktio kaikilla  $\lambda > 0$  ja

2.

$$\begin{aligned}\mu &= E(X) = \lambda, & \text{Var}(X) &= \lambda, \\ M(t) &= E(e^{tX}) = \exp(\lambda e^t - \lambda).\end{aligned}$$

**Todistus.** Sovelletaan eksponenttifunktion sarjakehitelmää

$$(5.4.2) \quad \exp(\lambda) = e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}.$$

1. Ensinnäkin  $f(x) \geq 0$  kaikilla  $x = 0, 1, 2, \dots$ , ja eksponenttifunktion sarjakehitelmän (5.4.2) perusteella

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^\lambda = 1.$$

2. Johdetaan ensin momenttifunktion  $M(t)$  lauseke:

$$\begin{aligned}M(t) &= E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} \cdot \exp(\lambda e^t) = \exp(\lambda e^t - \lambda).\end{aligned}$$

Odotusarvo ja varianssi saadaan sitten laskemalla  $M(t)$ :n 1. ja 2. derivaatta ja soveltamalla identiteettejä

$$E(X) = M'(0) \quad \text{ja} \quad \text{Var}(X) = M''(0) - [M'(0)]^2.$$

□

Riippumattomien Poissonin jakaumaa noudattavien satunnaismuuttujien summa noudattaa myös Poissonin jakaumaa.

**Lause 5.9** *Olkoot  $X_1, X_2, \dots, X_n$  riippumattomat ja  $X_i \sim \text{Poi}(\lambda_i)$ ,  $i = 1, 2, \dots, n$ . Olkoon  $Y = X_1 + X_2 + \dots + X_n$ . Silloin*

$$Y \sim \text{Poi}(\lambda),$$

missä  $\lambda = \sum_{i=1}^n \lambda_i$ .

**Todistus.** Seurauslauseen 4.1 mukaan

$$\begin{aligned}M_Y(t) &= \prod_{i=1}^n M_{X_i}(t) \\ &= \prod_{i=1}^n \exp(\lambda_i e^t - \lambda_i) = \exp[(e^t - 1)\lambda],\end{aligned}$$

missä  $\lambda = \sum_{i=1}^n \lambda_i$ . Lauseesta 4.10 seuraa sitten väite  $Y \sim \text{Poi}(\lambda)$ .

□

Jos riippumattomat  $X_1, X_2, \dots, X_n$  noudattavat samaa Poissonin jakaumaa  $\text{Poi}(\lambda)$ , niin Lauseen 5.9 mukaan niiden summa  $Y = X_1 + X_2 + \dots + X_n$  noudattaa Poissonin jakaumaa  $\text{Poi}(n\lambda)$ . Poissonin jakauma on hyvä binomijakauman  $\text{Bin}(n, p)$  likiarvo silloin, kun  $n$  on suuri ja  $p$  pieni.

Kun  $X \sim \text{Bin}(n, p)$ , niin binomitodennäköisyys on

$$(5.4.3) \quad f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n.$$

Annetaan nyt  $p$ :n riippua  $n$ :stä ja merkitään lausekkeessa (5.4.3)  $p = p_n$ . Valitaan erityisesti

$$p_n = \frac{\lambda}{n}, \quad n \geq 1.$$

Tarkastellaan nyt binomijakaumien jonoa

$$\text{Bin}(1, p_1), \text{Bin}(2, p_2), \text{Bin}(3, p_3), \dots$$

ja vastaavaa satunnaismuuttujien  $X_1, X_2, X_3, \dots$  jonoa, missä  $X_n \sim \text{Bin}(n, p_n)$ ,  $n \geq 1$ . Nyt siis

$$(5.4.4) \quad P(X_n = x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, \quad 0 \leq x \leq n.$$

Merkitään todennäköisyyttä (5.4.4) lyhyesti  $b_x(n)$

Kiinnitetään nyt  $x$  ja annetaan  $n$ :n kasvaa rajatta. Osoittautuu, että  $b_x(n)$  suppenee kaikilla  $x$ . Valitaan ensin  $x = 0$ . Silloin saamme

$$(5.4.5) \quad \lim_{n \rightarrow \infty} b_0(n) = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Se on eräs keskeinen eksponenttifunktioon liittyvä kaava, joka pitäisi analysoida kurssien perusteella muistaa. Tulos (5.4.5) saadaan esimerkiksi Taylorin sarjan

$$\log(1-p) = -\sum_{n=1}^{\infty} \frac{p^n}{n}$$

avulla, kun sijoitetaan  $p = \frac{\lambda}{n}$ :

$$(5.4.6) \quad \begin{aligned} \log\left(1 - \frac{\lambda}{n}\right)^n &= n \log\left(1 - \frac{\lambda}{n}\right) = n\left(-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2} - \frac{\lambda^3}{3n^3} - \dots\right) \\ &= -\lambda - \frac{\lambda^2}{2n} - \frac{\lambda^3}{3n^2} - \dots \\ &= -\lambda - \frac{1}{n}\left(\frac{\lambda^2}{2} + \frac{\lambda^3}{3n} + \dots\right). \end{aligned}$$

Kun  $n \rightarrow \infty$ , niin  $\frac{1}{n}\left(\frac{\lambda^2}{2} + \frac{\lambda^3}{3n} + \dots\right) \rightarrow 0$  ja siksi  $\log\left(1 - \frac{\lambda}{n}\right)^n \rightarrow -\lambda$ .

Lasketaan seuraavaksi  $b_x(n)$ :n raja-arvo, kun  $x > 0$ . Tarkastellaan peräkkäisten binomitodennäköisyyksien suhdetta

$$\frac{b_{x+1}(n)}{b_x(n)} = \frac{n-x}{x+1} \left(\frac{\lambda}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-1} = \frac{\lambda}{x+1} \left(\frac{n-x}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-1},$$

missä  $\frac{n-x}{n} \rightarrow 1$  ja  $1 - \frac{\lambda}{n} \rightarrow 1$ , kun  $n \rightarrow \infty$ . Tästä seuraa, että

$$(5.4.7) \quad \lim_{n \rightarrow \infty} \frac{b_{x+1}(n)}{b_x(n)} = \frac{\lambda}{x+1}.$$

Kun lähdetään tuloksesta (5.4.5) ja käytetään hyväksi raja-arvoa (5.4.7), saadaan

$$\begin{aligned} \lim_{n \rightarrow \infty} b_1(n) &= \frac{\lambda}{1} \lim_{n \rightarrow \infty} b_0(n) = \lambda e^{-\lambda}, \\ \lim_{n \rightarrow \infty} b_2(n) &= \frac{\lambda}{2} \lim_{n \rightarrow \infty} b_1(n) = \frac{\lambda^2}{1 \cdot 2} e^{-\lambda}, \\ &\vdots \\ \lim_{n \rightarrow \infty} b_x(n) &= \frac{\lambda}{x} \lim_{n \rightarrow \infty} b_{x-1}(n) = \frac{\lambda^x}{1 \cdot 2 \cdots x} e^{-\lambda}. \end{aligned}$$

Olemme siis näyttäneet, että

$$(5.4.8) \quad \lim_{n \rightarrow \infty} b_x(n) = \frac{\lambda^x}{x!} e^{-\lambda},$$

missä raja-arvo on  $P(X = x)$ , kun  $X \sim \text{Poi}(\lambda)$ . Tulos (5.4.8) tunnetaan *Poissonin raja-arvolakina*.

Satunnaismuuttujat noudattavat samaa jakaumaa, kun niillä on sama kertymäfunktio (Lause 4.4). Jos diskreetit satunnaismuuttujat noudattavat samaa jakaumaa, niin niillä on sama todennäköisyysfunktio. Jos satunnaismuuttujan  $X_n$  jakauma lähenee  $X$ :n jakaumaa  $n$ :n kasvaessa rajatta, niin  $X_n$ :n todennäköisyysfunktio lähenee  $X$ :n todennäköisyysfunktioita, mikäli jakaumat ovat diskreettejä (Lause 5.6). Vaikka edellä olemmekin johtaneet Poissonin raja-arvolain (5.4.8), esitetään tulos vielä *Poissonin lauseena*.

**Lause 5.10 (Poissonin lause)** *Olkoon  $X_n \sim \text{Bin}(n, p)$ . Silloin*

$$X_n \xrightarrow{d} \text{Poi}(\lambda),$$

*kun  $n \rightarrow \infty$  siten, että  $np = \lambda$ .*

**Todistus.** Koska  $np = \lambda$ , voimme merkitä  $p = \lambda/n$ . Todistus perustuu

Lauseeseen 5.6. Jos  $X_n \sim \text{Bin}(n, p)$ , niin

(5.4.9)

$$\begin{aligned} f_{X_n}(x) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{(n-x)! n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left[ \binom{n}{n} \binom{n-1}{n} \cdots \binom{n-x+1}{n} \right] \left(1 - \frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Kiinteällä  $x$ :n arvolla

$$\lim_{n \rightarrow \infty} \left[ \binom{n}{n} \binom{n-1}{n} \cdots \binom{n-x+1}{n} \right] = 1$$

ja

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1.$$

Näistä tuloksista yhdessä raja-arvon (5.4.5) kanssa seuraa

$$\lim_{n \rightarrow \infty} f_{X_n}(x) = \frac{e^{-\lambda} \lambda^x}{x!}.$$

Satunnaismuuttujan  $X_n$  jakauma lähestyy siis Poissonin jakaumaa  $\text{Poi}(\lambda)$ , kun  $n \rightarrow \infty$ . □

Poissonin jakaumaa sanotaan usein harvinaisten tapahtumien laiksi. Tämä luonnehdinta perustuu edellisessä lauseessa esitettyyn ominaisuuteen. Jos tehdään suuri määrä riippumattomia Bernoullin kokeita, joissa onnistumistodennäköisyys on hyvin pieni, niin silloin Lauseen 5.10 mukaan onnistumisten lukumäärä noudattaa likimain Poissonin jakaumaa. Esimerkiksi suuri määrä ihmisiä on päivittäin alttiina liikenneonnettomuuksille. Yksittäisen henkilön todennäköisyys (onnistumistodennäköisyys!) joutua onnettomuuteen on pieni, mutta onnettomuuksille alttiina olevien henkilöiden lukumäärä  $n$  on suuri. Silloin onnettomuuksien lukumäärä noudattaa likimain Poissonin jakaumaa.

**Lause 5.11** *Olkoot  $X$  ja  $Y$  sellaiset riippumattomat satunnaismuuttujat, että  $X \sim \text{Poi}(\lambda_1)$  ja  $Y \sim \text{Poi}(\lambda_2)$ . Silloin  $X$ :n ehdollinen jakauma ehdolla  $X + Y$  on binomijakauma.*

**Todistus.** Olkoot  $m$  ja  $n$  sellaiset epänegatiiviset kokonaisluvut, että  $m < n$ .



Silloin

$$\begin{aligned}
 P(X = m \mid X + Y = n) &= \frac{P(X = m, X + Y = n)}{P(X + Y = n)} \\
 &= \frac{P(X = m, Y = n - m)}{P(X + Y = n)} \\
 &= \frac{P(x = m) P(Y = n - m)}{P(X + Y = n)} \\
 &= \frac{e^{-\lambda_1} (\lambda_1^m / m!) e^{-\lambda_2} [\lambda_2^{n-m} / (n - m)!]}{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n / n!} \\
 &= \binom{n}{m} \frac{\lambda_1^m \lambda_2^{n-m}}{(\lambda_1 + \lambda_2)^n} \\
 &= \binom{n}{m} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^m \left( 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{n-m}
 \end{aligned}$$

on binomitodennäköisyys kaikilla  $m = 0, 1, \dots, n$ . Näin on lause todistettu.  $\square$

Lauseella 5.11 on tärkeä merkitys esimerkiksi frekvenssiaineistojen analyysissä.

**Esimerkki 5.8** Tiedetään, että auto-onnettomuuksien lukumäärä aikayksikössä (esimerkiksi kuukaudessa) noudattaa Poissonin jakaumaa. Tarkastellaan eräällä tieosuudella lokakuussa sattuvien onnettomuuksien lukumäärää. Aikaisempien tilastojen perusteella voidaan olettaa, että auto-onnettomuuksien lukumäärä  $Z$  kyseisellä tieosuudella (kuukaudessa) noudattaa Poissonin jakaumaa  $\text{Poi}(\lambda)$ . Onnettomuudet luokitellaan mahdollisten henkilövahinkojen mukaan vakaviin ja lieviin (jokainen onnettomuus kuuluu toiseen näistä luokista). Vakavien onnettomuuksien lukumäärä  $X \sim \text{Poi}(\lambda_1)$  ja lievien lukumäärä  $Y \sim \text{Poi}(\lambda_2)$ . Lisäksi  $X$  ja  $Y$  ovat toisistaan riippumattomat. Koska  $Z = X + Y$ , niin  $E(Z) = E(X) + E(Y)$  eli  $\lambda = \lambda_1 + \lambda_2$ .

Tutkijat valitsivat poliisin tiedostoista satunnaisesti valitun kuukauden (vuonna 2003) onnettomuudet. He havaitsivat onnettomuuksien lukumääräksi 120 ( $n = 120$ ), mutta he eivät olleet vielä luokitelleet onnettomuuksia. Mitä jakaumaa noudattaa vakavien onnettomuuksien lukumäärä? Lauseen 5.11 perusteella

$$P(X = m \mid Z = 120) = \binom{120}{m} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^m \left( 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{120-m},$$

$m = 0, 1, \dots, 120$ . Vakavien onnettomuuksien lukumäärä noudattaa siis binomijakaumaa  $\text{Bin}(120, \frac{\lambda_1}{\lambda_1 + \lambda_2})$ . Aikaisempien onnettomuustilastojen perusteella voidaan arvioida parametrit  $\lambda_1$  ja  $\lambda_2$ . Kun tutkijat olivat luokitelleet nuo 120 onnettomuutta, aineistossa havaittiin 15 vakavaa onnettomuutta. Koska  $E(X \mid Z = 120) = 120 \frac{\lambda_1}{\lambda_1 + \lambda_2}$ , niin binomijakaumasta voidaan arvioida parametrin  $\frac{\lambda_1}{\lambda_1 + \lambda_2}$  arvo.  $\square$

## 5.5 Poissonin prosessi

### 5.5.1 Laskuriprosessi

Stokastinen prosessi  $\{N(t), t \geq 0\}$  on *laskuriprosessi*, jos  $N(t)$  on ajankohtaan  $t$  mennessä sattuneiden ”tapahtumien” lukumäärä.

**Esimerkki 5.9** Seuraavassa luetellaan esimerkkejä laskuriprosesseista.

1. Jos  $N(t)$  on annetulla tieosuudella hetkeen  $t$  mennessä sattuneiden onnettomuuksien lukumäärä, niin  $\{N(t), t \geq 0\}$  on tapahtumaan ”onnettomuus” liittyvä laskuriprosessi.
2. Olkoon  $N(t)$  palvelutiskille tulleiden asiakkaiden lukumäärä hetkeen  $t$  mennessä. Tapahtuma on ”asiakkaan tulo palvelutiskille” ja  $\{N(t), t \geq 0\}$  on tapahtumaan liittyvä laskuriprosessi.
3.  $N(t)$  on vuoden alusta hetkeen  $t$  mennessä syntyneiden lasten lukumäärä kaupungissa  $A$ .
4.  $N(t)$  on jalkapallojoukkueen  $A$  tekemien maalien lukumäärä kauden alusta ajankohtaan  $t$  mennessä.

□

Laskuriprosessin tulee toteuttaa seuraavat ominaisuudet:

1.  $N(t) \geq 0$ .
2.  $N(t) \in \mathbb{N}$ , eli  $N(t)$  on kokonaislukuarvoinen.
3. Jos  $s < t$ , niin  $N(s) \leq N(t)$ .
4. Kun  $s < t$ , niin  $N(t) - N(s)$  on välillä  $(s, t]$  sattuneiden tapahtumien lukumäärä.

Laskuriprosessi on *riippumattomien lisäysten* prosessi, jos erillisillä aikaväleillä sattuvien tapahtumien lukumäärät ovat riippumattomat. Esimerkiksi satunnaismuuttujat  $N(2)$  ja  $N(10) - N(2)$  ovat riippumattomat, jos  $N(t)$  on riippumattomien lisäysten laskuriprosessi. Laskuriprosessin *lisäykset ovat stationaariset*, jos millä tahansa välillä sattuvien tapahtumien lukumäärän jakauma riippuu vain välin pituudesta. Jos  $N(t)$  on stationaarinen laskuriprosessi, niin satunnaismuuttujilla  $N(t_2) - N(t_1)$  ja  $N(t_2 + s) - N(t_1 + s)$  on sama jakauma kaikilla väleillä  $(t_1, t_2]$  ja  $(t_1 + s, t_2 + s]$ , missä  $t_2 > t_1$  ja  $s > 0$ .

### 5.5.2 Poissonin prosessin määrittely

Poissonin prosessi on yksi tärkeimpiä laskuriprosesseja. Se määritellään seuraavasti:

**Määritelmä 5.2** Laskuriprosessi  $\{N(t), t \geq 0\}$  on Poissonin prosessi, jonka intensiteetti on  $\lambda$  ( $\lambda > 0$ ), jos

1.  $N(0) = 0$ .
2. Prosessin lisäykset ovat riippumattomat.
3. Tapahtumien lukumäärä jokaisella  $h$ :n pituisella välillä noudattaa Poissonin jakaumaa, jonka odotusarvo on  $\lambda h$ :

$$P[N(h+t) - N(t) = x] = e^{-\lambda h} \frac{(\lambda h)^x}{x!}, \quad x = 0, 1, \dots$$

kaikilla  $h, t \geq 0$ .

Laskuriprosessin osoittaminen Poissonin prosessiksi Määritelmän 5.2 avulla saattaa olla hankalaa. Ei ole mitään yksinkertaista keinoa tarkistaa esimerkiksi ehdon 3 pätevyyttä. Siksi esitetään vielä toinen määritelmä, jonka avulla voi olla helpompaa tunnistaa prosessi. Voidaan osoittaa, että määritelmät 5.2 ja 5.3 ovat yhtäpitävät.

**Määritelmä 5.3** Laskuriprosessi  $\{N(t), t \geq 0\}$  on Poissonin prosessi, jonka intensiteetti on  $\lambda$  ( $\lambda > 0$ ), jos

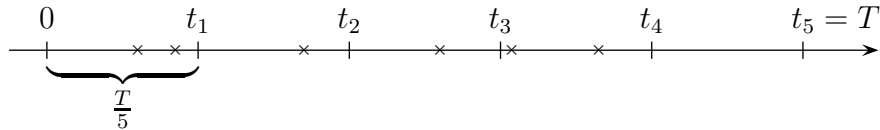
1.  $N(0) = 0$ .
2. Prosessin lisäykset ovat stationaariset ja riippumattomat.
3.  $P(N(t+h) - N(t) = 1) = \lambda h + o(h)$ .
4.  $P(N(t+h) - N(t) \geq 2) = o(h)$ .

Määritelmässä 5.3 käytetään merkintää  $o(h)$ . Sanomme, että funktio  $f(\cdot) = o(h)$ , jos

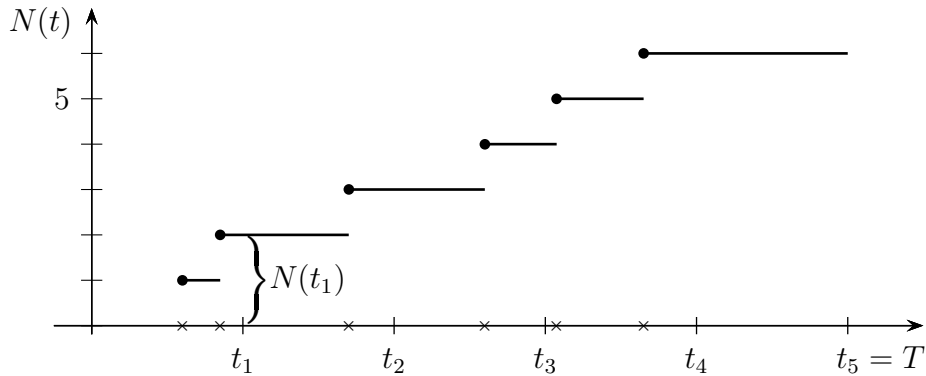
$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0.$$

**Esimerkki 5.10 Tieliikenneonnettomuudet.** Havainnoidaan esimerkiksi jollain tieosuudella sattuvien auto-onnettomuuksien lukumäärää. Onnettomuuksien määrä noudattaa tavallisesti varsin hyvin Poissonin prosessia.  $\square$

Tarkastellaan nyt hieman lähemmin Poissonin prosessin oletuksia. Oletetaan, että onnettomuuksien lukumäärä eräällä tieosuudella noudattaa aikavälillä  $(0, T)$  Poissonin prosessia, jonka intensiteetti on  $\lambda$ . Aikaväli voi olla esimerkiksi ruuhka-aika tietyinä perjantai-iltapäivinä klo 15–19 ja tieosuus jokin ulosmenotie. Oheisessa kuviossa on havaitut onnettomuudet merkitty aika-akselille.

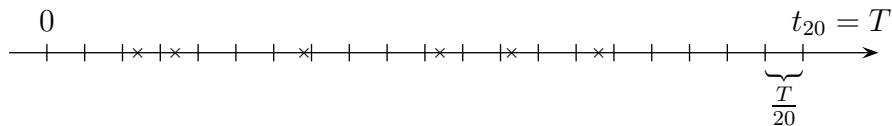


Tarkasteluväli  $(0, T]$  on jaettu viiteen yhtä pitkään osaväliin, joiden pituudet ovat  $T/5$ . Nyt esimerkiksi 1. osavälillä sattuneiden onnettomuuksien lukumäärä on  $N(t_1) - N(0) = N(t_1)$ , joka on siis hetkeen  $t$  mennessä sattuneiden onnettomuuksien lukumäärä. Kuvioon 5.3 on piirretty prosessin  $\{N(t), t \in (0, T]\}$  realisaatio, missä havaintoina ovat kyseiset onnettomuudet.



**Kuvio 5.3.** Poissonin prosessin  $\{N(t), t \in (0, T]\}$  erään realisaation kuvaaja.

Määritelmän 5.3 oletuksen 2 mukaan lisäykset  $N(t_1) - N(0)$ ,  $N(t_2) - N(t_1)$ ,  $N(t_3) - N(t_2)$ ,  $N(t_4) - N(t_3)$  ja  $N(t_5) - N(t_4)$  ovat riippumattomat ja noudattavat samaa jakaumaa. Määritelmän 5.3 oletukset 3 ja 4 tarkoittavat, että tapahtumat (onnettomuudet) sattuvat yksittäin ja samalla intensiteetillä koko tarkastelujakson ajan. Koska tapahtumat ovat erillisiä pisteitä, niin aina voidaan valita niin hienojakoinen välin ositus, että kullakin osavälillä on korkeintaan 1 tapahtuma. Jos tarkastelemassamme esimerkkitapauksessa valitaan osavälin pituudeksi  $T/20$ , sattuu tässä osituksessa kullekin osavälille korkeintaan 1 tapahtuma. Riippuen tietysti kulloisestakin havaintojaksosta, kuinka hienojakoinen ositus tarvitaan.



Todennäköisyys, että  $T/n$ :n pituiselle osavälille sattuu havainto, on Määritelmän 5.3 oletuksen 3 mukaan

$$P\left[N\left(t + \frac{T}{n}\right) - N(t) = 1\right] = \lambda \cdot \frac{T}{n} + o\left(\frac{T}{n}\right).$$

Vastaavasti todennäköisyys, että osavälillä sattuu enemmän kuin yksi havainto, on häviävän pieni, sillä Määritelmän 5.3 oletuksen 4 mukaan

$$P\left[N\left(t + \frac{T}{n}\right) - N(t) \geq 2\right] = o\left(\frac{T}{n}\right).$$

Voimme siis olettaa, että kullakin osavälillä sattuu vain 0 tai 1 tapahtumaa, kun  $n$  on riittävän suuri.

Määritellään nyt satunnaismuuttujat

$$X_i = N\left(\frac{iT}{n}\right) - N\left(\frac{(i-1)T}{n}\right), \quad i = 1, 2, \dots, n.$$

Muuttujia  $X_i$  voidaan käsitellä toisistaan riippumattomina Bernoullin jakaumaa noudattavina satunnaismuuttujina:

$$X_i \sim \text{Ber}\left(\frac{\lambda T}{n}\right), \quad i = 1, 2, \dots, n.$$

Koko välillä  $(0, T]$  havaittujen tapahtumien lukumäärä on

$$S_n = X_1 + X_2 + \dots + X_n,$$

joka noudattaa binomijakaumaa  $\text{Bin}(n, \frac{\lambda T}{n})$ . Koska  $E(S_n) = n \cdot \frac{\lambda T}{n} = \lambda T$  kaikilla  $n \in \mathbb{N}$ , niin  $E(S_n) = \lambda T$ , kun  $n \rightarrow \infty$ . Voimme siis soveltaa Poissonin lausetta (Lause 5.10), jonka mukaan  $S_n$  noudattaa Poissonin jakaumaa  $\text{Poi}(\lambda T)$ , kun  $n$  kasvaa rajatta. Näin esimerkiksi todennäköisyys, että välillä  $(0, T]$  sattuu  $x$  onnettomuutta, on

$$P(N(T) = x) = \frac{e^{-\lambda T} (\lambda T)^x}{x!}.$$

Todennäköisyys riippuu vain välin pituudesta  $T$  ja intensiteetistä  $\lambda > 0$ .

### 5.5.3 Satunnaistapahtumat tila-avaruudessa

Poissonin prosessilla mallinnetaan myös ilmiöitä, jotka tapahtuvat satunnaisesti tila-avaruudessa. Silloin Määritelmän 5.3 ehdot voidaan luonnehtia seuraavasti:

1. *Riippumattomuus.* Erillisillä alueilla sattuvien tapahtumien lukumäärät ovat riippumattomat.
2. *Yksittäisyys.* Todennäköisyys, että alueella sattuu enemmän kuin yksi tahtuma, on häviävän pieni.
3. *Homogeenisuus.* Tapahtumat sattuvat samalla intensiteetillä koko tarkasteltavalla alueella.

Tarkastellaan esimerkiksi Poissonin prosessia tasossa. Silloin todennäköisyys, että pinta-alaltaan  $A$ :n kokoisella alueella sattuu  $x$  tapahtumaa, on

$$f_A(x) = \frac{e^{-\lambda A}(\lambda A)^x}{x!}, \quad x = 0, 1, \dots,$$

missä  $\lambda$  on tapahtumien lukumäärän odotusarvo yhtä pinta-alayksikköä kohti. Jos Poissonin prosessia noudattavat tapahtumat sattuvat kolmiulotteisessa avaruudessa, niin silloin  $V$ :n kokoiseen tilaan osuu  $x$  tapahtumaa todennäköisyydellä

$$f_V(x) = \frac{e^{-\lambda V}(\lambda V)^x}{x!}, \quad x = 0, 1, \dots,$$

missä  $\lambda$  on tapahtumien lukumäärän odotusarvo yhtä tilavuus-yksikköä kohti.

**Esimerkki 5.11** Leipomo valmistaa suuren erän pullataikinaa, josta tehdään rusinapullia. Leipuri haluaa, että ainakin 95 % pullista sisältää vähintään 2 rusinaa. Kuinka monta rusinaa pullaa kohti pitäisi sekoittaa taikinaan?

Olkoon pullan tilavuus  $V = 1$ . Kun rusinat sekoitetaan hyvin taikinaan, on kaikilla pullilla sama todennäköisyys sisältää rusinoita (homogeenisuus). Koska taikina on suuri, ovat eri pulliin sattuvien rusinoiden lukumäärät toisistaan riippumattomat. Todennäköisyys, että pieneen pullaan sattuu enemmän kuin yksi rusina, on hyvin pieni.

Tässä tilanteessa on kyse Poissonin prosessista 3-ulotteisessa tila-avaruudessa. Pullassa on  $x$  rusinaa todennäköisyydellä

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

ja ainakin 2 rusinaa todennäköisyydellä

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-\lambda} - e^{-\lambda}\lambda. \end{aligned}$$

Leipuri vaatii, että

$$1 - e^{-\lambda} - e^{-\lambda}\lambda \geq 0.95.$$

Epäyhtälö toteutuu, kun  $\lambda \geq 4.74$ , joten rusinoita on sekoitettava taikinaan 5 rusinaa pullaa kohti.  $\square$

## Diskreetit jakaumat: Yhteenveto

<b>Bernoulli</b>	$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$
Ber( $p$ )	$E(X) = p, \quad \text{Var}(X) = p(1-p)$
	$M(t) = 1 - p + pe^t$

<b>Binomi</b> Bin( $n, p$ )	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$ $E(X) = np, \quad \text{Var}(X) = np(1-p)$ $M(t) = (1-p + pe^t)^n$
<b>Negatiivinen binomi</b> NBin( $r, p$ )	$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots$ $E(X) = \frac{r}{p}, \quad \text{Var}(X) = \frac{r(1-p)}{p^2}$ $M(t) = \left[ \frac{pe^t}{1-(1-p)e^t} \right]^r, \quad t < -\log(1-p)$
<b>Geometrisen</b> Geo( $p$ )	$f(x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \dots$ $E(X) = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}$ $M(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -\log(1-p)$
<b>Hypergeometrisen</b> HGeo( $n, N, p$ )	$f(x) = \frac{\binom{Np}{x} \binom{N-Np}{n-x}}{\binom{N}{n}}, \quad \begin{array}{l} X \leq pN \text{ ja} \\ n-X \leq N-Np \end{array}$ $E(X) = np, \quad \text{Var}(X) = \frac{N-n}{N-1} np(1-p),$
<b>Negatiivinen hypergeometrisen</b> NHGeo( $r, N, p$ )	$f(x) = \binom{x-1}{r-1} \frac{\binom{N-x}{Np-r}}{\binom{N}{Np}}, \quad x = r, r+1, \dots, N$ $E(X) = r \cdot \frac{N+1}{Np+1}$ $\text{Var}(X) = \frac{r(1-p)N(N+1)(Np+1-r)}{(Np+1)^2(Np+2)}$
<b>Poisson</b> Poi( $\lambda$ )	$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots$ $E(X) = \lambda, \quad \text{Var}(X) = \lambda$ $M(t) = \exp(\lambda e^t - \lambda), \quad -\infty < t < \infty$

**Poissonin prosessi.** Laskuri prosessi  $\{N(t), t \geq 0\}$ , jonka intensiteetti  $\lambda$ .

1.  $N(0) = 0$ .
2. Prosessin lisäykset ovat riippumattomat.
3. Tapahtumien lukumäärä jokaisella  $t$ :n pituisella välillä noudattaa Poissonin jakaumaa, jonka odotusarvo on  $\lambda t$ :

$$P[N(t+s) - N(s) = x] = e^{-\lambda t} \frac{(\lambda t)^x}{x!}, \quad x = 0, 1, \dots$$

kaikilla  $s, t \geq 0$ .

## Harjoituksia

1. Olkoon  $X$  satunnaismuuttuja, jonka arvojoukko on  $S_X = \{x_1, x_2\}$  ja todennäköisyysfunktio  $P(X = x_1) = p$ ,  $P(X = x_2) = 1 - p$ .

  - (a) Laske  $E(X^r)$ ,  $r = 1, 2$  ja
  - (b)  $\text{Var}(X)$ .
  - (c) Määritä  $X$ :n momenttifunktio.
2. Olkoon  $X \sim \text{Ber}(p)$  ja  $Y$  sellainen satunnaismuuttuja, jonka arvojoukko on  $S_Y = \{y_1, y_2\}$  ja todennäköisyysfunktio  $P(Y = y_1) = p$ ,  $P(Y = y_2) = 1 - p$ . Lausu  $Y$  satunnaismuuttujan  $X$  avulla.
3. Heitetään lanttia  $n$  kertaa ( $n$  riippumattonta Bernoullin koetta). Olkoon kruunan (R) todennäköisyys  $p$  ja  $X$  toistosten RR lukumäärä heittosarjassa.

  - (a) Mitä on  $E(X)$ ? Mikä on  $E(X)$ :n arvo, kun  $n = 200$ ?
  - (b) Laske  $\text{Var}(X)$ .
  - (c) Mitä on toistosten RRRR lukumäärän odotusarvo?

(Vihje: Katso Esimerkki 5.2.)
4. Jos  $X$  noudattaa binomijakaumaa, jonka odotusarvo on 6 ja varianssi 2.4, niin mitä on  $P(X = 5)$ ?
5. Hatussa on  $N$  yhdestä lähtien juoksevasti numeroitua arpalippua. Valitaan hatusta  $n$ :n arvan satunnaisotos palauttamatta (ks. Esimerkki 5.2). Olkoon  $X$  suurin valittujen arpalippujen järjestysnumeroista.

  - (a) Piirrä  $X$ :n todennäköisyys- ja kertymäfunktion kuvaajat, kun  $N = 100$  ja  $n = 10$ .
  - (b) Piirrä  $X$ :n odotusarvon kuvaaja  $n$ :n funktiona, kun  $N = 100$ .
6. Valitaan satunnaisesti ja toisistaan riippumatta 2000 pistettä yksikköneliöstä  $\{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$ . Olkoon  $Z$  yksikköympyrään  $\{(x, y) \mid x^2 + y^2 \leq 1\}$  osuvien pisteiden lukumäärä.

  - (a) Mitä jakaumaa  $Z$  noudattaa?
  - (b) Laske  $Z$ :n odotusarvo ja hajonta.
  - (c) Satunnaismuuttujan  $\frac{Z}{500}$  odotusarvo?
  - (d) Generoi 2000 satunnaislukuparia. Määritä  $Z$ :n arvo ja laske sen avulla  $\pi$ :n likiarvo.
7. Erääseen 90:n virheettömän kännykän tuote-erään oli sekaantunut 10 viallista. Valitaan tästä 100:n kännykän joukosta 30 kännykän otos palauttamatta. Olkoon  $X$  viallisten lukumäärä otoksessa.



- (a) Määritä  $X$ :n todennäköisyysfunktio.
- (b) Laske  $P(X = 10)$ .
- (c) Valitaan kännyköitä testaukseen satunnaisotannalla yksitellen palauttamatta, kunnes kaikki vialliset on löydetty. Olkoon  $Y$  tarvittavien testien lukumäärä. Laske  $P(Y \geq 20)$ , eli todennäköisyys, että tarvitaan ainakin 20 testiä.

8. Heitetään harhatonta lanttia, kunnes havaitaan toistos RR (kaksi kruunaa peräkkäin). Olkoon  $X$  tarvittavien heittojen lukumäärä. Olkoon  $f_n$   $n$ . Fibonaccin luku, joka määritellään siten, että  $f_1 = f_2 = 1$  ja  $f_n = f_{n-1} + f_{n-2}$ ,  $n = 3, 4, \dots$

- (a) Osoita, että  $X$ :n todennäköisyysfunktio on

$$f(x) = \frac{f_{x-1}}{2^x}, \quad x = 2, 3, 4, \dots$$

- (b) Osoita tuloksen

$$f_x = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^x - \left( \frac{1 - \sqrt{5}}{2} \right)^x \right]$$

avulla, että  $\sum_{x=2}^{\infty} = 1$ .

- (c) Osoita, että  $E(X) = 6$ .
- (d) Osoita, että  $E[X(X - 1)] = 52$  ja  $\text{Var}(X) = 22$ .
- (e) Simuloi  $X$ :n arvoja ja tarkastele, vastaavatko simuloinnin tulokset teoreettisia tuloksia.

9. Eräessä vaalissa 4000:sta äänestäjästä 100 kannatti ehdokasta  $A$ . Jos valitaan 50 alkion otos äänestäjistä esitutkimukseen, niin millä todennäköisyydellä haastatelluista korkeintaan 5 kannattaa  $A$ :ta?

10. Yritykseen tulee lähetys, joka sisältää 1000 varaosaa. Tarkistus suunnitelman mukaan  $n = 100$  satunnaisesti valittua (palauttamatta) varaosaa on tarkistettava. Tuote-erä hyväksytään, jos tarkistuksessa ei löydy kahta viallista enempää. Mikä on todennäköisyys, että tuote-erä hyväksytään? Laske todennäköisyys

- (a) hypergeometrisen jakauman avulla.
- (b) Laske sitten sama todennäköisyys käyttäen hypergeometrisen jakauman likiarvona binomijakaumaa
- (c) ja Poissonin jakaumaa.

11. Oletetaan, että  $X \sim \text{Poi}(\lambda)$ . Osoita, että  $E(X) = \lambda$  ja  $\text{Var}(X) = \lambda$ .

12. Leipomossa valmistetaan suuri taikina, josta tehdään rusinaleivoksia. Leipoyrittäjä haluaa, että 95 % leivoksista sisältää ainakin 2 rusinaa. Kuinka monta rusinaa leivosta kohti hänen pitää sekoittaa taikinaan?
13. Laboratoriohiiriin ruiskutetaan kahta eri liuosta. Ensimmäisessä liuoksessa on keskimäärin  $c$  kappaletta  $C$ -tyypin organismeja millitrassa ja toisessa liuoksessa keskimäärin  $d$  kappaletta  $D$ -tyypin organismeja millitrassa. Organismit ovat jakautuneet nesteeseen täysin satunnaisesti. Jokaiseen hiireen ruiskutetaan kumpaakin liuosta yksi millilitra. Hiiri säilyy hengissä jos ja vain jos kummassakaan ruiskeessa ei ole yhtään organismeja.
- (a) Millä todennäköisyydellä hiiri jää eloon?
- (b) Millä todennäköisyydellä kuolleista hiiristä löytyy molempia organismeja?
- (Vihje: Käytä Poissonin jakaumaa.)
14. Tehtaalla sattuu keskimäärin 1.5 onnettomuutta kuukaudessa. Määritä seuraavien tapahtumien todennäköisyydet:
- (a) Ei onnettomuuksia tammikuussa,
- (b) yhteensä neljä onnettomuutta helmikuussa ja maaliskuussa,
- (c) ainakin yksi onnettomuus vuoden jokaisena kuukautena.
- (Vihje: Käytä Poissonin jakaumaa.)
15. Olkoot  $X$  ja  $Y$  toisistaan riippumattomat Poissonin jakaumaa noudattavat satunnaismuuttujat. Olkoon  $E(X) = 1$  ja  $E(Y) = 2$ .
- (a) Laske todennäköisyys  $P(X + Y) = 5$ .
- (b) Millä kokonaislukuarvolla  $n$  todennäköisyys  $P(X + Y) = n$  saavuttaa maksiminsa?
- (c) Lausu todennäköisyys  $P(X + Y) = 5$  satunnaismuuttujien  $X$  ja  $Y$  todennäköisyysfunktioiden avulla.
16. Kirjassa on 200 sivua. Painovirheiden lukumäärä jokaisella sivulla noudattaa Poissonin jakaumaa, jonka keskiarvo on 0.01. Painovirheiden lukumäärät eri sivuilla ovat toisistaan riippumattomat.
- (a) Mikä on virheettömien sivujen lukumäärän odotusarvo ja hajonta?
- (b) Kirjan oikolukija havaitsee minkä tahansa annetun virheen todennäköisyydellä 0.9. Mikä on oikolukijan havaitsemien virheellisten sivujen lukumäärän odotusarvo?