TILM3553 Todennäköisyyslaskennan peruskurssi

Henri Pesonen

Sisältö

Luku 1

Todennäköisyys

1.1 Johdanto

Tilastotiede on menetelmätiede, jossa tutkitaan menetelmiä joita käytetään tilastollisissa tutkimuksissa. Tilastotieteen avulla kuvataan ja selitetään erilaisia reaalimaailman ilmiöihin liittyviä mekanismeja hyödyntämällä saatavilla olevaa numeerista tai kvantitatiivista informaatiota ilmiöistä. Usein tähän informaatioon liittyy epävarmuutta tai satunnaisuutta, ja tämän vuoksi tilastotieteen tärkein työkalu on todennäköisyyslaskenta. Todennäköisyyslaskennan avulla kehitetään matemaattisia malleja erilaisille **satunnaisilmiöille**. Jos ilmiön alkutilan perusteella tiedämme tarkasti mikä tulee olemaan lopputulos, niin ilmiö on **deterministinen** (kausaalinen) eikä ilmiöön liity minkäänlaista satunnaisuutta. Jos ilmiön alkutilasta huolimatta tapahtuma A voi esiintyä tai olla esiintymättä, niin kyseessä on satunnaisilmiö (ilmiö on **stokastinen**). Kyseessä on oikeastaan hyvin filosofinen tulkinta tilanteesta. Tarkastellaan esimerkiksi kolikon heittoa. Usein tilannetta mallinnetaan siten, että lopputulos on yhtäsuurella todennäköisyydellä kruuna tai klaava. Kuitenkin jos tunnetaan ilmiön alkutila, eli heiton voimakkuus, heiton suunta, kolikon pyörimisnopeus, ilmavirtaukset,..., voisimme laskea tarkasti lopputuloksen etukäteen. Käytännössä joutuisimme kuitenkin ylivoimaisiin vaikeuksiin mallintaessamme lukemattomia kolikonheittoon vaikuttavia ilmiöitä. Tämän vuoksi todennäköisyyslaskentaa voidaan soveltaa yksinkertaistavana työkaluna reaalimaailman (deterministisiin) tapahtumiin.

1.2 Peruskäsitteitä

Tilastotieteilijän tavoitteena on usein päätelmien tekeminen jonkinlaisten objektien joukosta erilaisten satunnaiskokeiden avulla. Tilastollinen mallintaminen aloitetaan muodostamalla satunnaiskokeen kaikista mahdollisista lopputuloksista (**alkeistapahtumista**) satunnaiskokeen **otosavaruus**.

Määritelmä 1 (Alkeistapahtuma)**.** Satunnaisilmiön mahdolliset lopputulokset ovat alkeistapahtumia. Merkitsemme alkeistapahtumia $\omega_1, \omega_2, \omega_2 \dots$

Esimerkki 1. Kuusisivuisen nopan heitossa alkeistapahtumia ovat $\omega_1, \dots, \omega_6$, jossa

$$\omega_i = \text{Saadaan silmäluku } i$$

.

Määritelmä 2 (Otosavaruus)**.** Otosavaruus eli perusjoukko on kaikkien mahdollisten alkeistapahtumien joukko Ω.

Todennäköisyyslaskennassa tarvitaan joukko-opin tuloksia. Otosavaruus voi olla joko numeroituva tai ylinumeroituva. Jos jokaiselle otosavaruuden alkiolle voidaan asettaa yksittäinen luonnollinen luku, niin otosavaruus on numeroituva. Muuten se on ylinumeroituva. Jatkossa määrittelemme otosavaruuksia sanallisesti (esim. kokonaisluvut yhdestä kymmeneen) tai symbolein (esim. $\Omega = \{1, 2, 3, \dots, 10\} = \{x \in \mathbb{N} :$

$1 \le x \le 10\}$). Käytämme jatkossa usein vakiintuneita merkintöjä

Luonnollisten lukujen joukko	$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
Kokonaislukujen joukko	$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
Positiivisten kokonaislukujen joukko	$\mathbb{Z}_+ = \{1, 2, \ldots\}$
Negatiivisten kokonaislukujen joukko	$\mathbb{Z}_- = \{\ldots, -2, -1\}$
Rationaalilukujen joukko	$\mathbb{Q} = \{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\}$
Reaalilukujen joukko	$\mathbb{R}$.

Merkitsemme $x \in A$, kun alkio x kuuluu joukkoon A, ja $x \notin A$ kun ei kuulu joukkoon A. Joukko A on joukon B osajoukko (merkitään $A \subset B$), jos ja vain jos

$$x \in A \Rightarrow x \in B.$$

Yllä oleva merkintä $\Rightarrow$ tarkoittaa, että jos $x \in A$ on tosi, niin tällöin $x \in B$ on tosi. Toisin sanoen A on B:n osajoukko jos ja vain jos B sisältää kaikki A:n alkiot. Kun tässä luentomonisteessa tarkastelemme ylinumeroituvia joukkoja, tarkastelemme aina reaalilukujen joukon osajoukkoja ja käytämme merkintöjä

Avoin väli	$(a, b) = \{x \in \mathbb{R} : a < x < b\}$
Puoliavoin väli	$(a, b] = \{x \in \mathbb{R} : a < x \le b\}$
Puoliavoin väli	$[a, b) = \{x \in \mathbb{R} : a \le x < b\}$
Suljettu väli	$[a, b] = \{x \in \mathbb{R} : a \le x \le b\}$
Alhaalta rajoitettu väli	$(a, \infty) = \{x \in \mathbb{R} : a < x\}$
Ylhäältä rajoitettu väli	$(-\infty, b) = \{x \in \mathbb{R} : x < b\}$

Esimerkki 2. Kysytään kadulla satunnaisesti valitulta henkilöltä missä kuussa hänen syntymäpäivänsä on. Mahdollisten alkeistapahtumien joukko on

$$\Omega = \{\text{Tam, Hel, Maa, Huh, Tou, Kes, Hei, Elo, Syy, Lok, Mar, Jou}\},$$

joka on numeroituva.

Esimerkki 3. Tarkastellaan koehenkilöiden reaktioaikaa äkkijarrutuksen aloittamiseen liikenneturvallisuustutkimuksessa. Reaktioaika voi olla mikä tahansa positiivinen luku. Otosavaruus on

$$\Omega = \{\omega \in \mathbb{R} : \omega \ge 0\},$$

joka on ylinumeroituva.

Esimerkki 4. Tarkastellaan edellistä reaktioaikaesimerkkiä. Jos reaktioajat mitataan sekunnin kymmenesosan tarkkuudella, otosavaruus on

$$\Omega = \{0.0, 0.1, 0.2, 0.3, 0.4, \ldots\},$$

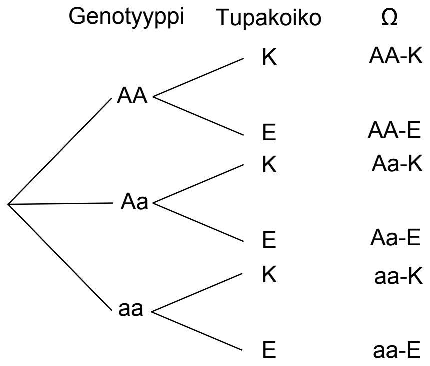
joka on numeroituva.

Satunnaisilmiöt voivat myös muodostua useamman satunnaisilmiön yhdisteenä. Tällöin saatetaan puhua yhdistetystä satunnaisilmiöstä. Kun tarkastellaan useasta satunnaisilmiöstä koostuvaa otosavaruutta, niin voidaan joskus käyttää hyödyksi puudiagrammeja, kuten Esimerkissä 5.

Esimerkki 5. Tarkastellaan satunnaisilmiötä, joka muodostuu kahdesta satunnaisilmiöstä. Ensimmäisenä satunnaisilmiönä tarkastellaan ihmisen erästä geeniä, joka muodostuu satunnaisesti kahdesta alleelista, jotka voivat olla joko A tai a. Tällöin geenin otosavaruus on {AA, aa, Aa}. Toisena satunnaisilmiönä tarkastellaan ihmisen tupakoimista, jolle otosavaruus on {K, E} (Kyllä, Ei). Kun tarkastellaan satunnaisilmiönä satunnaisesti valitun henkilön tätä geeniä sekä tupakoivuutta, niin otosavaruus on

$$\Omega = \{\text{AA-K,AA-E,Aa-K,Aa-E,aa-K,aa-E}\},$$

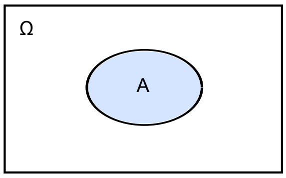
joka voidaan löytää käymällä kaikki kombinaatiot läpi esimerkiksi Kuvan 1.1 puudiagrammin avulla.



Kuva 1.1: Esimerkin 5 otosavaruus puudiagrammin avulla.

Kun tarkastelemme jotain satunnaisilmiötä, olemme kiinnostuneita esiintyykö, tai *realisoituuko*, jokin **tapahtuma**. Tapahtumat ovat otosavaruuden osajoukkoja, jotka koostuvat alkeistapahtumista, ja kuuluvat **havaittavien tapahtumien joukkoon**.

Määritelmä 3 (Tapahtuma)**.** Satunnaiskokeessa tapahtuma on perusjoukon Ω osajoukko A, joka kuuluu havaittavien tapahtumien joukkoon $\mathcal{F}$.



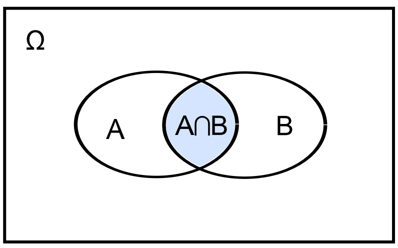
Kuva 1.2: Joukko-operaatioita voidaan havainnollistaa Venn-diagrammien avulla.

Sanomme että tapahtuma $A \subset \Omega$ esiintyy, jos satunnaisilmiön lopputulos ω kuuluu joukkoon A, eli $\omega \in A$. Tapahtumista $A \subset \Omega$ ja $B \subset \Omega$ voidaan muodoistaa uusia tapahtumia joukko-opillisten operaatioiden avulla. Joukko-operaatioita on kätevää havainnollistaa **Venn-diagrammien** avulla. Piirretään koko otosavaruus laatikoksi, jonka sisälle piirretään suljettuja kuvioita, jotka esittävät tapahtumia, kuten Kuvassa 1.2.

Joukkojen A ja B **leikkaus** $A \cap B$ määritellään joukkona, joka sisältää pisteet, jotka kuuluvat molempiin joukkoihin

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ ja } \omega \in B\}.$$

Todennäköisyyslaskennan kielellä $A \cap B$ esiintyy, jos molemmat tapahtumat A ja B esiintyvät kuten Kuvassa 1.3.

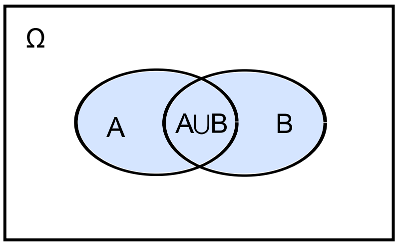


Kuva 1.3: Joukkojen A ja B leikkauksen Venn-diagrammi.

Joukkojen A ja B **unioni** $A \cup B$ määritellään joukkona, joka sisältää pisteet, jotka kuuluvat joko A:han tai B:hen, tai molempiin

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ tai } \omega \in B\}.$$

Todennäköisyyslaskennan kielellä $A \cup B$ esiintyy, jos tapahtuma A esiintyy, tapahtuma B esiintyy tai molemmat tapahtumat esiintyvät kuten Kuvassa 1.4.

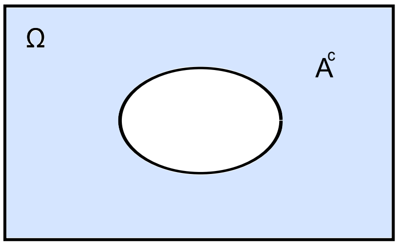


Kuva 1.4: Joukkojen A ja B unionin Venn-diagrammi.

Joukon A **komplementti** A^c määritellään joukkona, joka ei sisällä joukon A pisteitä

$$A^c := \{\omega \in \Omega : \omega \notin A\}.$$

Todennäköisyyslaskennan kielellä A^c esiintyy, jos tapahtuma A ei esiinny kuten Kuvassa 1.5.

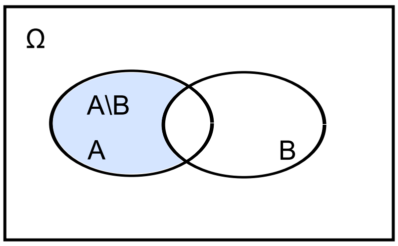


Kuva 1.5: Joukon A komplementin Venn-diagrammi.

Kahden joukon A ja B **erotus** $A \backslash B$ määritellään, joukkona, joka sisältää pisteet jotka kuuluvat joukkoon A, mutta eivät kuulu joukkoon B

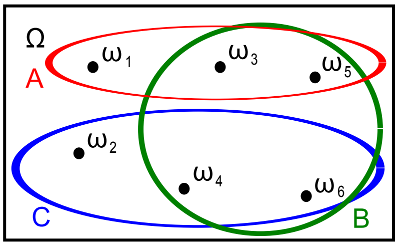
$$A \backslash B := A \cap B^c.$$

Todennäköisyyslaskennan kielellä $A \backslash B$ tapahtuu, jos A esiintyy ja B ei esiinny kuten Kuvassa 1.6.



Kuva 1.6: Joukkojen A ja B erotuksen Venn-diagrammi.

Esimerkki 6. Heitetään kuusisivuista noppaa. Alkeistapahtumat ovat määritelty Esimerkissä 1. Määritellään tapahtumat



Kuva 1.7: Esimerkin 6 tapahtumat.

$A =$ Silmäluku on parillinen $= \{\omega_2, \omega_4, \omega_6\}$
$B =$ Silmäluku on vähintään kolme $= \{\omega_3, \omega_4, \omega_5, \omega_6\}$
$C =$ Silmäluku ei ole parillinen $= \{\omega_1, \omega_3, \omega_5\}$.
Tapahtumat ovat piirretty Venn-diagrammien avulla Kuvaan 1.7

Tapahtumien A, B ja C avulla voidaan määritellä esimerkiksi tapahtumat
$A \cup B$= Silmäluku on parillinen tai vähintään kolme $= \{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$
$B \cap C =$ Silmäluku on vähintään kolme ja ei ole parillinen $= \{\omega_3, \omega_5\}$
$B^c =$ Silmäluku ei ole vähintään kolme $= \{\omega_1, \omega_2\}$.

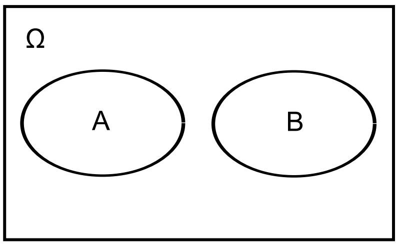
Lause 1. Joukko Ω voidaan ilmaista sen osajoukon A ja tämän komplementin A^c avulla

$$\Omega = A \cup A^c.$$

Todistus. Ei todisteta. □

Joukot A ja B ovat **toisensa poissulkevat** jos niiden leikkaus ei sisällä yhtään alkiota kuten Kuvassa 1.8. Joukkoa joka ei sisällä yhtään alkiota, eli **mahdotonta tapahtumaa**, kutsutaan tyhjäksi joukoksi $\emptyset$,

$$A \cap B = \emptyset.$$



Kuva 1.8: Toisensa poissulkevat tapahtumat.

Lause 2. Olkoon A otosavaruuden Ω tapahtuma. Otosavaruudelle ja tyhjälle joukolle pätee

$$\Omega^c = \emptyset, \quad A \cap \Omega = A, \quad A \cup \Omega = \Omega, \quad A \cap \emptyset = \emptyset, \quad A \cup \emptyset = A$$

Todistus. Ei todisteta. □

Esimerkki 7. Tarkastellaan satunnaisilmiönä kuusisivuisen nopan heittoa, jossa mielenkiinnon kohteena on heiton jälkeen suurin näkyviin jäävä silmäluku. Mikä on tässä satunnaisilmiössä otosavaruus?

Tapahtumat A ja B, joiden leikkaus on tyhjä joukko, eivät voi esiintyä samanaikaisesti. Erityisesti A ja A^c ovat toisensa poissulkevat. Tapahtumien poissulkevuus voidaan määritellä yleisemmin useammalle tapahtumalle $A_1, A_2, \ldots, A_n$. Nämä tapahtumat ovat toisensa poissulkevat, jos kaikille pareille

$$A_i \cap A_j = \emptyset, \quad i \neq j.$$

Lause 3. Jokaiselle otosavaruuden Ω tapahtumalle A, B, C pätee

Vaihdantalait	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Liitäntälait	$A \cup (B \cup C) = (A \cup B) \cup C$
	$A \cap (B \cap C) = (A \cap B) \cap C$
Osittelulait	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
De Morganin lait	$(A \cap B)^c = A^c \cup B^c$
	$(A \cup B)^c = A^c \cap B^c$

Todistus. Harjoitustehtävä. □

Kaikki Lauseen 3 pätevät myös useammalle kuin kolmelle tapahtumalle. Olkoon $A, B_1, \ldots, B_n$ tapahtumia otosavaruudessa Ω. Osittelulait ovat muotoa

$$A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} (A \cap B_i)$$
$$A \cup \left(\bigcap_{i=1}^{n} B_i \right) = \bigcap_{i=1}^{n} (A \cup B_i).$$

Käytämme merkintöjä

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$$
$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

Samoin de Morganin lait useammalle otosavaruuden Ω joukolle $A_1, \ldots, A_n$ ovat muotoa

$$\left(\bigcup_{i=1}^{n} A_i \right)^c = \bigcap_{i=1}^{n} A_i^c$$
$$\left(\bigcap_{i=1}^{n} A_i \right)^c = \bigcup_{i=1}^{n} A_i^c$$

Kun todennäköisyyksiä määritellään formaalisti otosavaruuden tapahtumille, niin näiden tapahtumien joukon täytyy olla tiettyä muotoa, jotta jokaiselle havaittavalle tapahtumalle pystytään määrään todennäköisyys.

Määritelmä 4 (Havaittavien tapahtumien joukko)**.** Havaittavien tapahtumien joukko $\mathcal{F}$ on σ-algebra, eli $\mathcal{F}$ on otosavaruuden Ω kaikkien mahdollisten osajoukkojen osajoukko, joka toteuttaa ominaisuudet

1. $\mathcal{F}$ sisältää ainakin yhden alkion
2. jos $A \in \mathcal{F}$, niin myös $A^c \in \mathcal{F}$
3. jos $A_1, A_2, A_3, \ldots \in \mathcal{F}$, niin $A_1 \cup A_2 \cup A_3 \cup \cdots \in \mathcal{F}$.

Havaittavien tapahtumien joukko sisältää siis ainakin yhden tapahtuman ominaisuuden 1. mukaisesti. Ominaisuus 2. tarkoittaa, että jos tapahtuman A esiintyminen on havaittavissa, niin havaitsemme myös, jos A ei esiinny. Lisäksi, jos useamman tapahtuman esiintyminen on havaittavissa, niin voimme myös havaita jos yksi näistä esiintyy.

Määritelmästä seuraa, että σ-algebran täytyy sisältää otosavaruus ja tyhjä joukko. Tämän kurssin puitteissa σ-algebrojen ominaisuuksiin ei tutusta tämän tarkemmin.

Esimerkki 8. Joukon $\Omega = \{\omega_1, \omega_2, \omega_3\}$ kaikkien osajoukkojen joukko on σ-algebra. Kaikkien osajoukkojen joukko Ω:lle on

$$\mathcal{F} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}\}.$$

Tarkastellaan Määritelmän 4 ehtojen toteutumista. Ehto 1) toteutuu, sillä $\mathcal{F}$ sisältää ainakin yhden alkion. Ehto 2) toteutuu, sillä

$$\begin{aligned}
\emptyset \in \mathcal{F} \quad &\text{ja} \quad \emptyset^c = \{\omega_1, \omega_2, \omega_3\} \in \mathcal{F} \\
\{\omega_1\} \in \mathcal{F} \quad &\text{ja} \quad \{\omega_1\}^c = \{\omega_2, \omega_3\} \in \mathcal{F} \\
\{\omega_2\} \in \mathcal{F} \quad &\text{ja} \quad \{\omega_2\}^c = \{\omega_1, \omega_3\} \in \mathcal{F} \\
\{\omega_3\} \in \mathcal{F} \quad &\text{ja} \quad \{\omega_3\}^c = \{\omega_1, \omega_2\} \in \mathcal{F}.
\end{aligned}$$

Ehto 3) toteutuu, sillä voimme käydä kaikkien $\mathcal{F}$ alkioiden yhdistelmien kombinaatiot läpi, ja jokainen näistä sisältyy joukkoon $\mathcal{F}$. Esimerkiksi

$$\begin{aligned}
\{\omega_2\}, \{\omega_1, \omega_3\} \in \mathcal{F} \quad &\text{ja} \quad \{\omega_2\} \cup \{\omega_1, \omega_3\} = \{\omega_1, \omega_2, \omega_3\} \in \mathcal{F} \\
\emptyset, \{\omega_2\}, \{\omega_3\} \in \mathcal{F} \quad &\text{ja} \quad \emptyset \cup \{\omega_2\} \cup \{\omega_3\} = \{\omega_2, \omega_3\} \in \mathcal{F}.
\end{aligned}$$

1.3 Todennäköisyys

Määritettyämme otosavaruuden, alkeistapahtumat sekä tapahtumat voimme määrittää matemaattisesti todennäköisyyden. Todennäköisyysmitta on funktio joka määrittää jokaiselle määritetylle tapahtumalle luvun väliltä $[0, 1]$. Todennäköisyysmitan arvoa jollekin tapahtumalle sanotaan tapahtuman todennäköisyydeksi.

Määritelmä 5 (Todennäköisyysmitta)**.** Funktio $\mathsf{P} : \mathcal{F} \to \mathbb{R}$ on todennäköisyysmitta, jos se toteuttaa seuraavat aksioomat

1. $\mathsf{P}(\Omega) = 1$
2. $\mathsf{P}(A) \geq 0, \quad \forall A \in \mathcal{F}$
3. $A_1, A_2 \ldots, \in \mathcal{F}$ ja $A_i \cap A_j = \emptyset,\ i \neq j \Rightarrow \mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$

Aksiooman 1 mukaisesti satunnaisilmiössä tapahtuu jotain todennäköisyydellä 1. Aksiooma 2 tarkoittaa, että todennäköisyys on mittana jostain suurempaa kuin 0. Aksiooman 3 perusteella toisensa poissulkevien tapausten yhdisteen todennäköisyys saadaan näiden yksittäisten tapahtumien todennäköisyyksien summana.

Lause 4. Aksioomasta 3 seuraa

$$\begin{aligned}
&A_1, \ldots, A_n \in \mathcal{F} \text{ ja } A_i \cap A_j = \emptyset,\ i \neq j \\
&\Rightarrow \mathsf{P}(A_1 \cup \cdots \cup A_n) = \mathsf{P}(A_1) + \cdots + \mathsf{P}(A_n).
\end{aligned}$$

Todistus. Valitaan $A_{n+1}, A_{n+2}, \ldots = \emptyset$, tällöin

$$\begin{aligned}
\mathsf{P}(A_1 \cup \cdots \cup A_n) &= \mathsf{P}(A_1 \cup \cdots \cup A_n \cup A_{n+1} \cup \cdots) \\
&= \mathsf{P}(A_1) + \cdots + \mathsf{P}(A_n) + \mathsf{P}(A_{n+1}) + \cdots \\
&= \mathsf{P}(A_1) + \cdots + \mathsf{P}(A_n),
\end{aligned}$$

koska $A_i \cap \emptyset = \emptyset$ ja $\mathsf{P}(\emptyset) = 0$. □

Määritelmä 6 (Todennäköisyysavaruus)**.** Otosavaruus, tapahtumien joukko ja todennäköisyysmitta muodostavat yhdessä todennäköisyysavaruuden $(\Omega, \mathcal{F}, \mathsf{P})$.

Määritelmä 7 (Partitio)**.** Jos tapahtumat $A_1, A_2, \ldots, A_n$ ovat toisensa poissulkevia, ja kattavat koko otosavaruuden, eli

$$A_i \cap A_j = \emptyset, \quad \forall i \neq j$$
$$\bigcup_{i=1}^{n} A_i = \Omega,$$

niin tapahtumien joukkoa kutsutaan partitioksi.

Lause 5. Kolmogorovin aksioomien avulla voidaan todistaa mm. seuraavat todennäköisyyden laskusäännöt

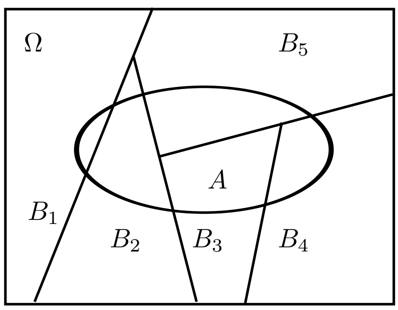
1. $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$
2. $\mathsf{P}(\emptyset) = 0$
3. Jos $A \subset B$, niin $\mathsf{P}(A) \leq \mathsf{P}(B)$
4. $0 \leq \mathsf{P}(A) \leq 1$
5. $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$
6. Jos tapahtumat $B_1, \ldots, B_k$ muodostavat partition, niin

   $$\mathsf{P}(A) = \mathsf{P}(A \cap B_1) + \cdots + \mathsf{P}(A \cap B_k).$$

   Tilannetta ollaan havainnollistettu Kuvassa 1.9.
7. Jos tapahtumat $B_1, \ldots, B_k$ muodostavat partition, niin

   $$\mathsf{P}\left(\bigcup_{i=1}^{n} B_i\right) = 1.$$

Todistus. Harjoitustehtävä. □



Kuva 1.9: Lauseen 5 laskusääntö 6.

Lause 6. Olkoon A_1, A_2 ja A_3 tapahtumia. Tapahtumien unionin todennäköisyys saadaan yhteenlaskusäännöllä

$$\begin{aligned}\mathsf{P}(A_1 \cup A_2 \cup A_3) = {} & \mathsf{P}(A_1) + \mathsf{P}(A_2) + \mathsf{P}(A_3) \\ & - \mathsf{P}(A_1 \cap A_2) - \mathsf{P}(A_1 \cap A_3) - \mathsf{P}(A_2 \cap A_3) \\ & + \mathsf{P}(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Todistus. Harjoitustehtävä. □

Yleisemmin n tapahtuman unionin todennäköisyyden laskemisessa voidaan käyttää Lauseen 7 inkluusio-ekskluusio-periaatetta.

Lause 7. Olkoon $A_1, A_2, \dots, A_3$ tapahtumia. Tapahtumien unionin todennäköisyys saadaan yhteenlaskusäännöllä

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum\sum_{i<j} \mathsf{P}(A_i \cap A_j) + \sum\sum\sum_{i<j<k} \mathsf{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

Todistus. Ei todisteta. □

Ohjeita todennäköisyyksien laskemiseen.

1. Määrittele tapahtumat symbolien avulla
2. Kirjoita ylös tunnetut tapahtumien todennäköisyydet symbolien avulla
3. Kirjoita ja laske kysytty todennäköisyys symbolien avulla
4. Käytä laskuissa hyväksi vain todennäköisyyslaskennan tunnettuja tuloksia
5. Sijoita saatuun symboliseen lausekkeeseen annetut todennäköisyydet ja sievennä
6. Anna tuloksena sekä tarkka arvo (jos mahdollista), että desimaaliarvo
7. Tarkista onko tulos järkevä!

Esimerkki 9. Tarkastellaan tavallisesta hyvin sekoitetusta korttipakasta nostettua korttia satunnaisilmiönä. Alkeistapahtumat erotetaan toisistaan kortin maalla ja arvolla $\omega = (\text{maa}, \text{arvo})$. Otosavaruus on

$$\Omega = \{(\text{maa}, \text{arvo}) : \text{maa} \in \{\text{pata}, \text{risti}, \text{ruutu}, \text{hertta}\}, \text{arvo} \in \{1, 2, \dots, 13\}\}.$$

Oletetaan jokainen 52 alkeistapahtumasta yhtä todennäköiseksi. Määritellään tapahtumat (ässä lasketaan kuvakortiksi, joten kuvakorttien arvot ovat $1, 11, 12$ ja 13)

$$\begin{aligned} A &= \text{kortin maa on hertta} \\ B &= \text{kortti on kuvakortti} \\ C &= \text{kortti on ässä} \\ D &= \text{kortin arvo on 2,3,4,5 tai 6} \end{aligned}$$

Lasketaan

$$\begin{aligned} \mathsf{P}(A) &= \frac{13}{52}, \quad \mathsf{P}(A^c) = \mathsf{P}(\text{kortin maa ei ole hertta}) = \frac{39}{52} \\ \mathsf{P}(B) &= \frac{16}{52}, \quad \mathsf{P}(C) = \frac{4}{52}, \quad \mathsf{P}(D) = \frac{20}{52} \\ \mathsf{P}(A \cup B) &= \mathsf{P}(\text{kortin maa on hertta tai kortti on kuvakortti}) = \frac{25}{52} \\ \mathsf{P}(A \cap B) &= \mathsf{P}(\text{kortin maa on hertta ja kortti on kuvakortti}) = \frac{4}{52} \end{aligned}$$

Lauseen 5 voidaan tarkistaa pätevän

$$\begin{aligned} \mathsf{P}(A^c) &= \frac{39}{52} = 1 - \frac{13}{52} = 1 - \mathsf{P}(A) \\ \mathsf{P}(A \cup B) &= \frac{25}{52} = \frac{13}{52} + \frac{16}{52} - \frac{4}{52} = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B) \\ \mathsf{P}(C) &= \frac{4}{52} \leq \frac{16}{52} = \mathsf{P}(B) \end{aligned}$$

Esimerkki 10. Olkoon A tapahtuma, että opiskelija käy opintojensa ohessa töissä, ja olkoon B tapahtuma, että opiskelijan suorittamien kurssien keskimääräinen arvosana on 4 tai suurempi. Miten ilmaistaan symbolisesti todennäköisyys tapahtumalle, että

1. opiskelija käy opintojensa ohessa töissä?
2. opiskelijan suorittamien kurssien keskimääräinen arvosana on 4 tai suurempi?
3. opiskelijan suorittamien kurssien keskimääräinen arvosana on alle 4?
4. opiskelija ei käy opintojen ohessa töissä ja hänen suorittamien kurssien keskimääräinen arvosana on 4 tai suurempi?

Esimerkki 11. A ja B ovat toisensa poissulkevat tapahtumat, joille $\mathsf{P}(A) = 0.3$ ja $\mathsf{P}(B) = 0.4$. Lasketaan todennäköisyydet tapahtumille

1. $\mathsf{P}(A \cup B)$
2. $\mathsf{P}(A \cap B)$
3. $\mathsf{P}(A^c \cap B)$

1.4 Todennäköisyyden määrääminen

Todennäköisyysmitan matemaattisen määritelmän mukaan se on funktio, joka liittää jonkin arvon väliltä $[0, 1]$ kaikille mahdollisille tapahtumille. Miten tämän luvun määräämme tapahtumille vaihtelee todennäköisyyden tulkinnasta toiseen.

1.4.1 Klassinen todennäköisyys

Todennäköisyyden klassisen tulkinnan mukaan tapahtuman todennäköisyys on suotuisten tapahtumien lukumäärän suhde kaikkien mahdollisten tapahtumien lukumäärään. Alkeistapahtumia oletetaan olevan äärellinen määrä ja kaikki alkeistapahtumat $\omega_1, \omega_2, \dots, \omega_n$ yhtä mahdollisiksi. Tällöin jokaisen alkeistapahtuman todennäköisyys on

$$\mathsf{P}(\omega_i) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Olkoon A tapahtuma, joka esiintyy yhden alkeistapahtumista $\omega_1, \dots, \omega_k$ esiintyessä. Yhteensä otosavaruudessa Ω on tapahtumat $\omega_1, \dots, \omega_n$. Tällöin klassisen todennäköisyyden mukaisesti todennäköisyys tapahtumalle A on

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|} = \frac{k}{n},$$

jossa merkinnät $|A|$ ja $|\Omega|$ tarkoittavat tapahtumalle A suotuisten alkeistapahtumien sekä kaikkien alkeistapahtumien lukumääriä, vastaavasti. Näiden tapahtumien lukumäärien laskemisessa voidaan käyttää hyödyksi kombinatoriikan tuloksia.

Klassisessa todennäköisyydessä vaaditaan että vaihtoehtoiset tapaukset ovat **yhtä mahdollisia**. Usein tätä perustellaan geometrisen symmetrian avulla. Esimerkkinä klassisesta todennäköisyydestä voidaan ajatella nopanheittoa. Todennäköisyys että saadaan nopanheitossa *kuusi* on yhden sivun suhde kuuteen sivuun, kun oletamme että jokainen sivu on symmetriaan pohjautuen yhtä mahdollinen. Mitä kuitenkaan tarkoittaa että tapahtumat ovat 'yhtä mahdollisia', jos tapahtumien todennäköisyyttä ollaan vasta määrittelemässä? Klassinen todennäköisyys on käyttökelpoinen, kun mahdollisia tapauksia on äärellinen määrä, ja tapauksille voidaan perustella yhtäsuuret todennäköisyydet. Kun nämä ehdot eivät toteudu, todennäköisyyksien tulkinta on mahdotonta. Esimerkiksi miten määrättäisiin klassisessa mielessä mielekkäästi todennäköisyys, että huomenna sataa vettä?

Kombinatoriikkaa

Kun alkeistapauksia on äärelline määrä voidaan todennäköisyyksien määräämiseksi hyödyntää kombinatoriikan tuloksia.

Esimerkki 12. Tarkastellaan tilannetta jossa valitsemme satunnaisesti yksitellen kolme kertaa kuulan laatikosta, joka sisältää valkoisen (V), sinisen (S) ja punaisen (P) kuulan. Tarkastellaan ensin tapausta, jossa kuulan poimimisen jälkeen sen väri kirjataan ylös ja kuula palautetaan laatikkoon ennen seuraavan poimimista. Mahdolliset kolmen poiminnan tapaukset ovat

VVV, VVS, VSV, SVV, VVP, VPV, PVV,
SSS, SSV, SVS, VSS, SSP, SPS, PSS,
PPP, PPV, PVP, VPP, PPS, PSP, SPP,
VSP, VPS, PVS, PSV, SVP, SPV

joita on 27 erilaista vaihtoehtoa. Jos poimittua kuulaa ei palauteta laatikkoon, mahdollisten otosten määrä pienenee. Mahdolliset otokset ovat

VSP, VPS, PVS, PSV, SVP, SPV,

joita on ainoastaan 6 erilaista. Tarkastellaan seuraavaksi tilannetta, jossa poimimme jälleen kolme kuulaa palauttaen poimimisten välillä kuulan takaisin laatikkoon, mutta emme ole kiinnostuneita siitä missä järjestyksessä kuulat poimittiin. Nyt erilaisia mahdollisia otoksia ovat

VVV, VVS, VVP,
SSS, SSV, SSP,
PPP, PPV, PPS,
VSP,

joita on 10 vaihtoehtoa. Neljäntenä vaihtoehtona tarkastelemme tapausta, jossa otos muodostetaan palauttamatta poimittuja kuulia laatikkoon, ja tekemättä eroa samat kuulat sisältävien otosten välillä. Vaihtoehtoja on vain yksi (VSP).

Esimerkistä 12 huomaamme, että samankaltaisesta tilanteesta voi muodostua hyvin erilaisia otosavaruuksia. Kombinatoriikka on matematiikan osa-alue, jonka avulla voidaan tutkia monellako eri tavalla joukko alkioita voidaan järjestää erilaisiin ryhmiin. Ennen tapahtumien muodostamista, täytyy meidän tietää minkälainen satunnaisilmiö on kyseessä. Lisäksi tutkitaan minkälainen **otanta** alkeistapahtumien joukosta suoritetaan. Ensiksi meidän täytyy selvittää, palautetaanko valittu alkio takaisin valintansa jälkeen. Otanta voidaan siis tehdä **palauttaen** tai **ilman palautusta**. Toiseksi selvitämme onko alkioiden valinnan järjestyksellä merkitystä tapahtumissa. Otanta voi olla siis **järjestetty** tai **järjestämätön**.

Eri tapahtumien määrän laskemisessa voidaan usein hyödyntää tulosääntöä, jonka mukaan kokonaismäärä voidaan laskea osissa.

Määritelmä 8 (Tulosääntö)**.** Jos otos voidaan valita yhteensä k. vaiheessa, ja i. vaihe voidaan valita n_i tavalla, niin erilaisia koetuloksia on

$$n_1 n_2 \cdots n_k.$$

Esimerkki 13. Heitetään kolikkoa k kertaa. Tarkastellaan erilaisia heittosarjoja. Koska jokaisella heitolla on kaksi mahdollista tulosta, niin tuloperiaatteen nojalla erilaisia heittosarjoja on

$$\underbrace{2 \cdot 2 \cdots 2}_{k \text{ kpl}} = 2^k.$$

Kolikon symmetriaan vedoten voidaan jokaista heittosarjaa pitää yhtä todennäköisenä. Jos kolikkoa heitetään 10 kertaa peräkkäin niin esimerkiksi tapahtuman $A =$ saadaan heittosarja HTHHHTHTTT, todennäköisyys on

$$\mathsf{P}(A) = \frac{1}{2^{10}} = \frac{1}{1024}.$$

Esimerkki 14. Tällä kurssilla useimmin käytetty σ-algebra on kaikkien järjestämättömien osajoukkojen joukko (power set). Tämä voidaan muodostaa hakemalla kaikki vaihtoehdot, joissa alkeistapahtuma joko kuuluu tai ei kuulu yksittäiseen tapahtumaan. Oletetaan, että otosavaruus koostuu n alkeistapahtumasta. Jokaisella alkeistapahtumalla on siis kaksi vaihtoehtoa ja eri tapausten kokonaismäärä saadaan tuloperiaatteen mukaisesti

$$\underbrace{2 \cdot 2 \cdots 2}_{n \text{ kpl}} = 2^n.$$

Esimerkki 15. Eräässä wok-ravintolassa voi valita annokseen yhden joukosta {riisi, nuudeli}, yhden joukosta { possu, kana, tofu } ja yhden joukosta {satay, teriyaki, chow mein, hoisin}. Mies on pyytänyt vaimoaan tuomaan kyseisestä ravintolasta minkä tahansa annoksen, mutta oikeasti mies haluasi nuudeleita, possua ja satay-kastiketta. Tuloperiaatteen nojalla erilaisia annoksia on $2 \cdot 3 \cdot 4 = 24$ kpl. Olettaen, että vaimo tuo täysin satunnaisesti valitun annoksen miehelle, niin tapahtuman $A =$ (nuudeli,possu,satay) todennäköisyys on

$$\mathsf{P}(A) = \frac{1}{24}.$$

Tarkastellaan järjestettyä otantaa, jossa erittelemme samat alkiot sisältävät otokset toisistaan riippuen alkioiden järjestyksestä toisiinsa nähden.

Lause 8. Olkoon $\Omega = \{\omega_1, \dots, \omega_n\}$ joukko, jossa on n erilaista alkiota. Joukosta voidaan valita k:n ($k \leq n$) alkion järjestetty otos ilman palautusta $n \cdot (n-1) \cdots (n-k+1)$ tavalla. Näitä kutsutaan k-permutaatioiksi.

Todistus. Koska ensimmäiselle alkiolle on n vaihtoehtoa, toiselle alkiolle on $n-1$ vaihtoehtoa, ..., k. alkiolle on $(n-k+1)$ vaihtoehtoa, niin tulosäännön mukaisesti vaihtoehtoja on $n \cdot (n-1) \cdots (n-k+1)$. □

Käyttämällä **n-kertomaa**

$$n! := n \cdot (n-1) \cdots 2 \cdot 1,$$

voidaan k-permutaatio ilmaista muodossa

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1).$$

Erityisesti n-permutaatioiden määrä, kun alkioita on n kpl, on n-kertoma $n!$. Suurilla n arvoilla voidaan $n!$ laskea likimääräisesti Stirlingin kaavalla

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

jossa $e \approx 2.718282$ on Neperin luku.

Esimerkki 16. Opiskelijalla on 10 oppikirjaa luettavana, kuinka monessa järjestyksessä opiskelija voi kirjat lukea?

Eri järjestyksiä kirjojen lukemiselle on $10! = 3628800$ kappaletta. Stirlingin approksimaatio antaa järjestyksiä likimääräisesti

$$\sqrt{2\pi \cdot 10}\left(\frac{10}{e}\right)^{10} = 3598696.$$

Jos kirjat ovat numeroitu $1, \dots, 10$, niin todennäköisyys tapahtumalle $A =$ kirjat luetaan oikeassa järjestyksessä, on

$$\mathsf{P}(A) = \frac{1}{3628800}.$$

Jos valitaankin alkiot palauttaen, mutta pitämällä kirjaa järjestyksestä on kyseessä järjestetty otanta palautten.

Lause 9. Olkoon $\Omega = \{\omega_1, \dots, \omega_n\}$ joukko, ja olkoon $k \leq n$ luonnollinen luku. Joukosta voidaan valita k:n alkion järjestetty otos palauttaen n^k tavalla.

Todistus. Koska 1. alkio voidaan valita n tavalla, 2. alkio voidaan valita n tavalla, ..., k. alkio voidaan valita n tavalla, niin tulosäännön nojalla vaihtoehtoja on $\underbrace{n \cdot n \cdots n}_{k \text{ kpl}} = n^k$ kappaletta. □

Esimerkki 17. Vakioveikkauksessa veikataan 13 kohdeottelun lopputulosta (kotivoitto, tasapeli, vierasvoitto). Erilaisia vakioveikkauksen rivejä on siis $3^{13} = 1594323$ kappaletta. Jos pelataan satunnaisesti yhtä riviä, niin tapahtuman $A = 13$ oikein todennäköisyys on

$$\mathsf{P}(A) = \frac{1}{1594323} \approx 6 \cdot 10^{-7}.$$

Lause 10. Olkoon Ω joukko, jossa on on n erilaista alkiota, ja olkoon k luonnollinen luku siten, että $k \leq n$. Joukosta voidaan muodostaa k kokoinen järjestämätön otos ilman palautusta

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

tavalla. (Tätä lukua sanotaan binomikertoimeksi ja luetaan *n yli k:n*.)

Todistus. Koska erilaisia jonoja on $n!/(n-k)!$ kappaletta ja jokainen k:n pituinen joukko voidaan laittaa $k!$ eri järjestykseen, niin k-kombinaatioiden lukumäärä on

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

□

Esimerkki 18. Tilastotieteen peruskurssi a:lle on ilmoittautunut 109 opiskelijaa. Opiskelijoista voidaan muodostaa 3 hengen ryhmiä

$$\binom{109}{3} = \frac{109!}{3!106!} = 209934$$

tavalla. Jos 3 hengen ryhmä muodostetaan täysin satunnaisesti, niin tapahtuman $A =$ Pekka, Timoteus ja Anselmi pääsevät samaan ryhmään todennäköisyys on

$$\mathsf{P}(A) = \frac{1}{209934} \approx 4 \cdot 10^{-6}.$$

Esimerkki 19. Lotossa on 39 palloa, joista saadaan mahdollisia 7 numeron pituisia rivejä

$$\binom{39}{7} = 15380937.$$

Tapahtuman $A = 7$ oikein on

$$\mathsf{P}(A) = \frac{1}{15380937} \approx 6.5 \cdot 10^{-8} = 0.000000065.$$

Erilaisia 4 oikein rivejä on

$$\binom{7}{4} \cdot \binom{32}{3} = 35 \cdot 4960 = 173600.$$

Ensimmäisessä vaiheessa on laskettu monellako eri tapaa voi olla valittu neljä oikein seitsemästä. Toisessa vaiheessa on laskettu monellako eri tapaa 3 on valittu 32 luvusta, joita ei esiinny oikeassa rivissä. Todennäköisyys tapahtumalle $B = 4$ oikein on

$$\mathsf{P}(B) = \frac{173600}{15380937} \approx 0.011.$$

Lause 11. Olkoon Ω joukko, jossa on on n erilaista alkiota, ja olkoon k luonnollinen luku siten, että $k \leq n$. Joukosta voidaan muodostaa k kokoinen järjestämätön otos palauttaen

$$\binom{n+k-1}{k}$$

tavalla.

Todistus. Lause jätetään todistamatta. □

	Palauttaen	Palauttamatta
Järjestetty	n^k	$\frac{n!}{(n-k)!}$
Järjestämätön	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Taulukko 1.1: Eri tapausten määrä otannassa.

Multinomikerroin

Tarkastellaan n alkioista joukkoa. Ollaan kiinnostuttu siitä kuinka monella eri tavalla alkiot voidaan järjestää k kappaleeksi ryhmiä siten, että i. ryhmässä on n_i alkiota. Alkioiden järjestyksellä ryhmissä ei ole merkitystä. Kaikki alkiot jaetaan johonkin ryhmään, joten $n_1 + n_2 + \cdots + n_k = n$.

- Ryhmään 1 voidaan valita n alkion joukosta alkiot $\binom{n}{n_1}$ tavalla.
- Ryhmään 2 voidaan valita $n - n_1$ alkion joukosta alkiot $\binom{n-n_1}{n_2}$ tavalla.
- Ryhmään 3 voidaan valita $n - n_1 - n_2$ alkion joukosta alkiot $\binom{n-n_1-n_2}{n_3}$ tavalla.

  $\vdots$

- Ryhmään k voidaan valita $n - n_1 - n_2 - \cdots - n_{k-1} = n_k$ alkion joukosta alkiot $\binom{n_k}{n_k} = 1$ tavalla.

Kaikkien eri kombinaatioiden määrä löytyy tulosäännön avulla.

$$\begin{aligned}
&\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdots 1 \\
&= \frac{n!}{(n-n_1)!n_1!} \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \frac{(n-n_1-n_2)!}{(n-n_1-n_2-n_3)!n_3!} \cdots 1 \\
&= \frac{n!}{n_1!n_2!\cdots n_k!}.
\end{aligned}$$

Kombinaatioiden määrä määritellään multinomikertoimeksi.

$$\binom{n}{n_1, n_2, \ldots, n_k} := \frac{n!}{n_1!n_2!\cdots n_k!}$$

Binomikerroin on multinomikertoimen erikoistapaus, kun ryhmien määrä $k = 2$.

Esimerkki 20.

Multinomikertoimella on yllä esitetyn *lokerotulkinnan* lisäksi *järjestystulkinta*. Ajatellaa, että ryhmä koostuu n alkiota, joita on k erilaista alkiota. Erilaisia alkiota on kutakin $n_1, n_2, \ldots, n_k$ kpl. Kaikki ryhmän alkiot voidaan järjestää $n!$ erilaiseen järjestykseen. Nyt i. ryhmän alkiot, joita on n_i kpl, ovat keskenään identtisiä. Tällöin niiden keskinäisellä järjestyksellä ei ole merkitystä, kuten esimerkiksi sanojen erilaisten anagrammien määrää laskiessa. Kuitenkin jokainen näistä $n_i!$ järjestyksistä on laskettu omaksi järjestyksekseen permutaatioiden $n!$ määrässä. Poistamalla samanlaiset järjestykset pois saadaan yhteensä erilaisia järjestyksiä

$$\frac{n!}{n_1!n_2!\cdots n_k!},$$

joka on aiemmin määritelty multinomikerroin.

Esimerkki 21. Tarkastellaan sanaa 'anagrammi'. Yhdeksästä kirjaimesta koostuvassa sanassa esiintyy kirjaimet a (3), g (1), i, (1) m (2), n (1), r (1). Erilaisia anagrammeja sanalle on yhteensä

$$\binom{9}{3,1,1,2,1,1} = 30240.$$

1.4.2 Geometrinen todennäköisyys

Jos otosavaruudessa on ylinumeroituva määrä alkeistapauksia, ei klassisen todennäköisyyden käsitettä pystytä hyödyntämään todennäköisyyksien määräämisessä. Klassisen todennäköisyyden kanssa tietyssä mielessä yhtenevä todennäköisyyden käsite on geometrinen todennäköisyys. Tarkastellaan esimerkkinä tilannetta, jossa katkaistaan 50cm pitkä naru satunnaisesta kohtaa. Oletetaan, että katkaisukohta on missä tahansa narulla. Määritellään mielenkiintoinen tapahtuma $A =$ naru katkaistaan kohdasta $[20\text{cm}, 30\text{cm}]$. Tapahtuman A todennäköisyys on geometrisen todennäköisyyden avulla

$$\mathsf{P}(A) = \frac{\text{Suotuisan katkaisuvälin pituus}}{\text{Koko narun pituus}} = \frac{10}{50} = \frac{1}{5}$$

Esimerkki 22. Kaksi yliopistolla työskentelevää henkilöä käyvät molemmat kahvilla samassa kahvilla satunnaisesti 14:00-15:00 välillä tietämättä milloin toinen ehtii kahville. Molemmat odottavat toista korkeintaan 5 minuuttia kahviseuraksi, jonka jälkeen lähtevät omaan toimistoon juomaan kahvinsa. Molemmat lähtevät pois kahvilasta viimeistään $15:00$. Mikä on todennäköisyys, että henkilöt tapaavat toisensa?

Olkoon A tapahtuma, että henkilöt tapaavat toisensa. Aikaväli $14:00-15:00$ on 60 minuuttia. Olkoon henkilön 1 saapumisaika kahvilaan t_1 ja henkilön 2 saapumisaika t_2. Jos henkilö 1 saa ennen henkilöä 2 saa henkilö 2 saapua korkeintaan 5 minuuttia henkilön 1 jälkeen, eli

$$t_2 \leq t_1 + 5.$$

Jos henkilö 2 saa ennen henkilöä 1 saa henkilö 1 saapua korkeintaan 5 minuuttia henkilön 2 jälkeen, eli

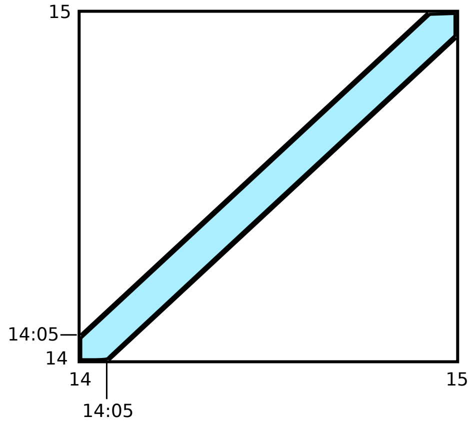
$$t_1 \leq t_2 + 5 \quad \Leftrightarrow \quad t_2 \geq t_1 - 5.$$

Yhteensä saadaan todennäköisyys tapahtumalle A

$$t_1 - 5 \leq t_2 \leq t_1 + 5.$$

Tapahtuma A on piirretty Kuvaan 1.10 väritettynä alueena. Nyt väritetyn alueen ala on suhteessa koko alaan

$$\mathsf{P}(A) = \frac{60 \cdot 60 - \frac{1}{2} \cdot 55 \cdot 55 - \frac{1}{2} \cdot 55 \cdot 55}{60 \cdot 60} = \frac{23}{144} \approx 0.160.$$



Kuva 1.10: Esimerkin 22 otosavaruus ja suotuisa tapahtuma A sinisellä väritettynä.

1.4.3 Frekventistinen todennäköisyys

Frekventistinen todennäköisyyden tulkinta perustuu oletukseen, että toistettaessa satunnaiskoetta tapahtuman esiintymisfrekvenssi stabiloituu kohti tiettyä arvoa, kun kokeiden määrä kasvaa. Tässä mielessä kyseessä on tilastollinen todennäköisyys. Tässä tulkinnassa tapahtuman todennäköisyys on esiintymisfrekvenssin raja-arvo toistojen määrän kasvaessa rajatta. Käytännössä käytetään hypoteettista frekvenssitulkintaa, eli jos koetta toistettaisiin äärettömän monta kertaa, niin mihin arvoon suhteellinen frekvenssi suppenisi. Esimerkiksi heitettäessä tasapainoista kolikkoa klaavojen suhteellinen osuus heitoista stabiloituu kohti arvoa 0.5. Frekvenssitulkinnan ongelmana on, että periaatteessa voimme puhua todennäköisyyksistä ainoastaan toistettavissa olevien kokeiden tai ilmiöiden yhteydessä. Eli ainutkertaisten tapahtumien todennäköisyyksistä ei voida puhua.

Matemaattisesti frekventisissä todennäköisyydessä määritellään suhteellinen frekvenssi tapahtumalle A, kun kokeita on n kpl

$$\mathsf{P}_n(A) = \frac{f_n(A)}{n},$$

jossa $f_n(A)$ on niiden tapausten lukumäärä, joissa esiintyy tapahtuma A. Todennäköisyys määritellään raja-arvona, kun kokeiden määrä n lähestyy ääretöntä

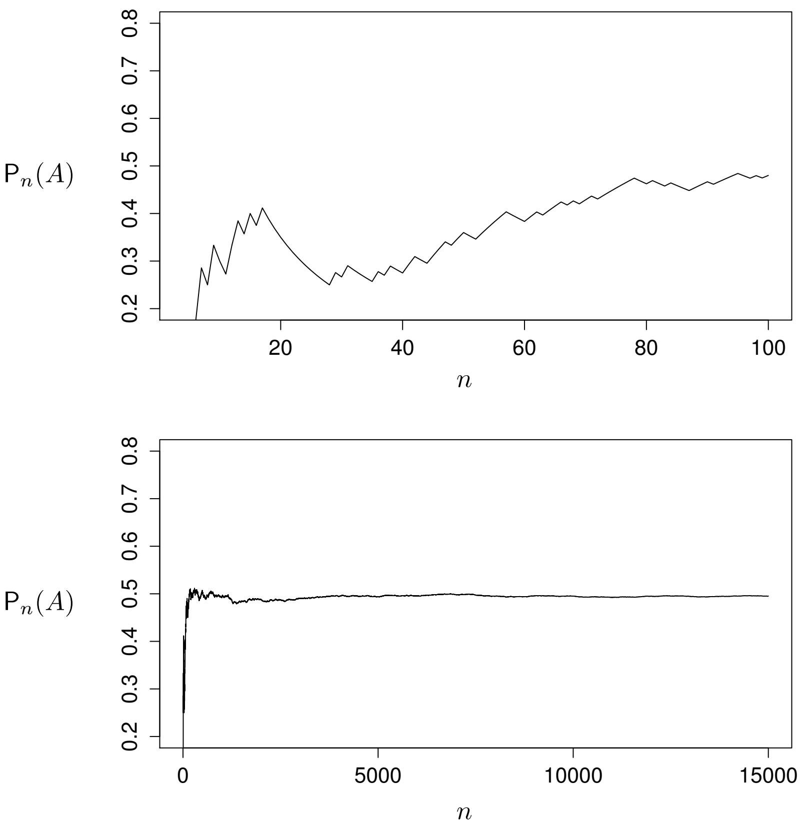
$$\mathsf{P}\,(A) = \lim_{n\to\infty} \mathsf{P}_n(A).$$

Esimerkki 23. Heitellään tavallista yhden euron kolikkoa ja kirjataan ylös saadaanko kruuna (H) vai klaava (T). Olemme kiinnostuneita tapahtuman A ='saadaan kruuna' todennäköisyydestä. Lasketaan jokaisen heiton jälkeen kruunien suhde kaikkiin heittoihin

$$P_n(A) = \frac{f_n(A)}{n}.$$

Kruunan suhteellinen frekvenssi 100 sekä 15000 heittoon asti on piirrettynä Kuvaan 1.11. Kuvasta näemme, että kruunan suhteellinen osuus stabiloituu kohti lukua 0.5. Tämän perusteella reilussa kolikon heitossa

$$\mathsf{P}\,(A) = \lim_{n\to\infty} \mathsf{P}_n(A) = 0.5.$$



Kuva 1.11: Kruunan suhteellinen frekvenssi Esimerkissä 23.

1.4.4 Bayesiläinen todennäköisyys

Bayesiläisen, eli subjektiivisen tulkinnan mukaan todennäköisyys on **uskomuksemme** tai **epävarmuutemme** mitta tapahtumista. Filosofisesti tämä on sopusoinnissa oletuksen kanssa, ettei ole olemassa puhtaasti satunnaisia tapahtumia, vaan todennäköisyys mallintaa epävarmuuttamme mahdollisista tapahtumista. Tämän todennäköisyyden tulkinnan mukaan voimme puhua ainutkertaisten tapahtumien todennäköisyyksistä, kuten esimerkiksi todennäköisyydestä että kolmas maailmansota syttyy ennen vuotta 2025. Subjektiivisen todennäköisyyden tulkinnan ongelmana voidaan pitää subjektiivisuutta. Koska todennäköisyys on epävarmuutemme mitta, tapahtumien todennäköisyydet saattavat vaihdella henkilöittäin.

1.5 Ehdollinen todennäköisyys ja tapahtumien riippumattomuus

Tähän mennessä olemme tarkastelleet todennäköisyyksiä määrittelemällä otosavaruuden ja tarkastelemalla tämän otosavaruuden tapahtumien todennäköisyyksiä. Usein kuitenkin olemme tilanteessa, jossa tarkastellessamme tapahtumaa A tiedämme jonkin muun tapahtuman B esiintyneen. Tämän kaltaisessa tilanteessa voimme ajatella tapahtuman B otosavaruudeksi, jossa tarkastelemme tapahtuman A esiintymisen todennäköisyyttä.

Esimerkki 24. Tarkastellaan kahden kolikon heittoa. Otosavaruus on $\{\mathsf{HH}, \mathsf{HT}, \mathsf{TH}, \mathsf{TT}\}$. Määritellään tapahtumat

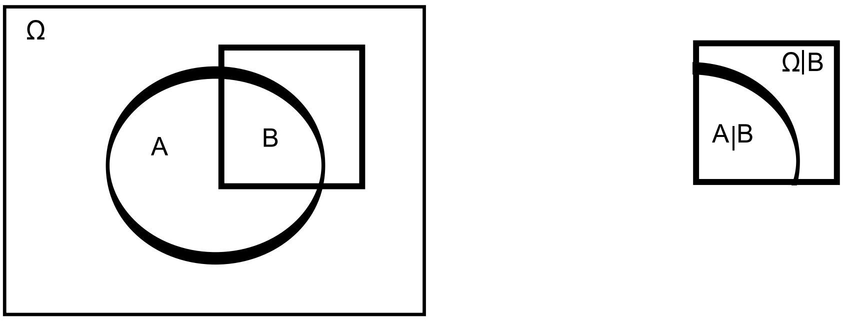
$$\begin{aligned} A &= \text{Molemmat kolikot ovat klaavoja} = \{\mathsf{TT}\} \\ B &= \text{Toinen kolikko on klaava} = \{\mathsf{HT}, \mathsf{TH}, \mathsf{TT}\}. \end{aligned}$$

Tapahtuman A todennäköisyys on $\mathsf{P}(A) = \frac{1}{4}$. Ajatellaan seuraavaksi, että B on tapahtunut. Nyt voidaan ajatella, että otosavaruus koostuu ainoastaan tapauksista, joissa esiinty B, eli $\{\mathsf{HT}, \mathsf{TH}, \mathsf{TT}\}$, ja tässä otosavaruudessa tapahtuman A todennäköisyys on $\frac{1}{3}$.

Määritelmä 9 (Ehdollinen todennäköisyys)**.** Kun tapahtuman B todennäköisyys $\mathsf{P}(B) > 0$, niin tapahtuman A ehdollinen todennäköisyys ehdolla, että B esiintyy on

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

Huomaa ehdollisen todennäköisyyden määrittelystä, että $\mathsf{P}(B \mid B) = 1$. Lisäksi jos tapahtumat A ja B ovat toisensa poissulkevat, niin $\mathsf{P}(A \mid B) = 0$, joka on intuitiivisesti selvää, sillä B:n esiintyminen *poissulkee* mahdollisuuden, että A esiintyisi. Ehdollista todennäköisyyttä on havainnollistettu Venn-diagrammin avulla Kuvassa 1.12



Kuva 1.12: Ehdollinen todennäköisyys havainnollistettuna. Alkuperäinen otosavaruus on Kuvassa oikealla ja vasemmalla on tapahtumalla B ehdollistettu otosavaruus.

Esimerkki 25. Lasketaan ehdollisen todennäköisyyden kaavalla Esimerkin 24 todennäköisyys.

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(\{TT\})}{\mathsf{P}(\{HT, TH, TT\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Esimerkki 26. Populaatiosta, josta 48% on miehiä ja 52% on naisia, valitaan satunnaisesti yksi henkilö. Miehistä 10% ja naisista 1% on värisokeita. Millä todennäköisyydellä valittu henkilö on värisokea mies?

Olkoon $A =$ Henkilö on värisokea ja $B =$ Henkilö on mies. Ehdollisen todennäköisyyden kaavan avulla saamme

$$\mathsf{P}(A \cap B) = \mathsf{P}(A \mid B)\,\mathsf{P}(B) = 0.48 \cdot 0.1 = 0.048.$$

Lause 12 (Kokonaistodennäköisyyden kaava)**.** Jos $B_1, \ldots, B_k$ muodostavat partition, niin tapahtuman A todennäköisyys saadaan nyt laskettua kokonaistodennäköisyyden kaavalla

$$\mathsf{P}(A) = \mathsf{P}(A \mid B_1)\,\mathsf{P}(B_1) + \cdots + \mathsf{P}(A \mid B_k)\,\mathsf{P}(B_k).$$

Todistus. Harjoitustehtävä. □

Esimerkki 27. Laatikossa H on yhdeksän korttia, joihin on kirjoitettu numerot $1, \ldots, 9$, ja laatikossa T on viisi korttia, joihin on kirjoitettu numerot $1, \ldots, 5$. Heitetään reilua kolikkoa. Jos tulee kruuna (heads), otetaan yksi kortti satunnaisesti laatikosta H, ja jos tulee klaava (tails), otetaan yksi kortti satunnaisesti laatikosta T. Mikä on on todennäköisyys, että otetun kortin numero on parillinen?

Määritellään tapahtumat $A =$ kortin numero on parillinen, $H =$ saadaan kruuna ja $T =$ saadaan klaava. Käytetään kokonaistodennäköisyyden kaavaa

$$\mathsf{P}(A) = \mathsf{P}(A \mid H)\,\mathsf{P}(H) + \mathsf{P}(A \mid T)\,\mathsf{P}(T) = \frac{4}{9} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2} = \frac{19}{45}.$$

Esimerkki 28. Avainnipussa on n avainta. Avaimia kokeillaan peräjälkeen, kunnes oikea avain löytyy. Kokeillut avaimet pidetään erillään kokeilemattomista. Millä todennäköisyydellä k. avain on oikea?

Määritelmä 10 (Bayesin teoreema)**.** Kun tapahtuman B todennäköisyys $\mathsf{P}(B) > 0$, niin Bayesin teoreema liittää tapahtumien A ja B, ja ehdollisten tapahtumien $A \mid B$ ja $B \mid A$ todennäköisyydet toisiinsa

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(B \mid A)\,\mathsf{P}(A)}{\mathsf{P}(B)}.$$

Esimerkki 29. Laatikossa on 10 punaista, 30 valkoista, 20 sinistä ja 15 mustaa marmorikuulaa. Kuulat ovat väriä lukuunottamatta samanlaisia. Laatikosta otetaan sokkona peräkkäin kaksi kuulaa siten, että ensin otettua kuulaa **ei palauteta** laatikkoon.

Määritellään tapahtumat

$$\begin{aligned} M_i &= i. \text{ kuula on musta} \\ V_i &= i. \text{ kuula on valkoinen} \\ P_i &= i. \text{ kuula on punainen.} \end{aligned}$$

Nyt voidaan laskea Bayesin teoreeman avulla esimerkiksi tapahtumat kumpikaan nostetuista kuulista ei ole musta

$$\mathsf{P}(M_1^c \cap M_2^c) = \mathsf{P}(M_1^c)\,\mathsf{P}(M_2^c \mid M_1^c) = \frac{60}{75} \cdot \frac{59}{74} = \frac{3540}{5550} = \frac{118}{185} \approx 0.64,$$

ja ensimmäinen kuula on valkoinen ja toinen ei ole

$$\mathsf{P}(V_1 \cap V_2^c) = \mathsf{P}(V_1)\,\mathsf{P}(V_2^c \mid V_1) = \frac{30}{75} \cdot \frac{45}{74} = \frac{1350}{5550} = \frac{9}{37} \approx 0.24.$$

Esimerkki 30. Tarkastellaan tietyn sairauden toteamiseksi kehitettyä testiä. Testin tulos on positiivinen tai negativinen, riippuen siitä onko henkilöllä testin mukaan sairautta. Väestöstä noin 0.5%:lla on kyseinen sairaus. Testistä tiedetään, että 5%:lle henkilöistä joilla ei ole sairautta, testin tulos on positiivinen ja 1%:lle sairaista testi antaa negatiivisen tuloksen.
a) Mikä on todennäköisyys saada positiivinen testitulos?

b) Mikä on todennäköisyys, että testattava henkilö on sairas vaikka testin tulos on negatiivinen?
c) Mikä on todennäköisyys, että testattava henkilö on ei ole sairas vaikka testin tulos on positiivinen?

Olkoon

$$T^+ = \text{Testin tulos on positiivinen}, \quad T^- = \text{Testin tulos on negatiivinen}$$
$$S = \text{Henkilöllä on sairaus}, \quad S^c = \text{Henkilöllä ei ole sairautta}$$

Tehtävänannossa ollaan annettu todennäköisyydet

$$\mathsf{P}(S) = \frac{5}{1000}, \quad \mathsf{P}\left(T^+ \mid S^c\right) = \frac{5}{100}, \quad \mathsf{P}\left(T^- \mid S\right) = \frac{1}{100}.$$

a) Käytetään kokonaistodennäköisyyden kaavaa

$$\begin{aligned}\mathsf{P}\left(T^+\right) &= \mathsf{P}\left(T^+ \mid S\right)\mathsf{P}(S) + \mathsf{P}\left(T^+ \mid S^c\right)\mathsf{P}(S^c) = (1 - \mathsf{P}\left(T^- \mid S\right))\mathsf{P}(S) + \mathsf{P}\left(T^+ \mid S^c\right)\mathsf{P}(S^c)\\ &= \left(1 - \frac{1}{100}\right)\cdot\frac{5}{1000} + \frac{5}{100}\cdot\left(1 - \frac{5}{1000}\right) = \frac{99\cdot 5 + 5\cdot 995}{100\cdot 1000} = \frac{547}{10000} = 0.0547.\end{aligned}$$

b) Käytetään Bayesin teoreemaa ja a)-kohdan tulosta

$$\mathsf{P}\left(S \mid T^-\right) = \frac{\mathsf{P}(T^- \mid S)\,\mathsf{P}(S)}{\mathsf{P}(T^-)} = \frac{\mathsf{P}(T^- \mid S)\,\mathsf{P}(S)}{1 - \mathsf{P}(T^+)} = \frac{\frac{1}{100}\cdot\frac{5}{1000}}{1 - \frac{547}{10000}} = \frac{5}{9453} \approx 0.00005$$

c) Käytetään Bayesin teoreemaa

$$\mathsf{P}\left(S^c \mid T^+\right) = \frac{\mathsf{P}(T^+ \mid S^c)\,\mathsf{P}(S^c)}{\mathsf{P}(T^+)} = \frac{\frac{5}{100}\left(1 - \frac{5}{1000}\right)}{\frac{547}{10000}} = \frac{4975}{5470} = \frac{995}{1094} \approx 0.910.$$

Lause 13. Lauseen 5 laskusäännöt pätevät, kun ehdollistetaan tapahtumat tapahtumalla C, jolle pätee $\mathsf{P}(C) > 0$.

1. $\mathsf{P}(A^c \mid C) = 1 - \mathsf{P}(A \mid C)$
2. $\mathsf{P}(\emptyset \mid C) = 0$
3. Jos $A \subset B$, niin $\mathsf{P}(A \mid C) \le \mathsf{P}(B \mid C)$
4. $0 \le \mathsf{P}(A \mid C) \le 1$
5. $\mathsf{P}(A \cup B \mid C) = \mathsf{P}(A \mid C) + \mathsf{P}(B \mid C) - \mathsf{P}(A \cap B \mid C)$
6. Jos tapahtumat $B_1, \ldots, B_k$ muodostavat partition, niin
$$\mathsf{P}(A \mid C) = \mathsf{P}(A \cap B_1 \mid C) + \cdots + \mathsf{P}(A \cap B_k \mid C).$$

Todistus. Harjoitustehtävä. □

Lause 14. Olkoon $A_1, A_2, \ldots, A_n$ tapahtumia. Tällöin pätee

$$\mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathsf{P}(A_1)\,\mathsf{P}(A_2 \mid A_1)\,\mathsf{P}(A_3 \mid A_1 \cap A_2)\cdots\mathsf{P}(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}),$$

kun kaikki ehdolliset todennäköisyydet ovat määriteltyjä.

Todistus. Harjoitustehtävä. □

Tapahtumien riippumattomuus

Tärkeä ominaisuus tapahtumilla on, että riippuvatko niiden esiintyminen toisista tapahtumista. Jos tapahtuman B esiintymisellä ei ole vaikutusta tapahtuman A esiintymiseen, sanotaan että tapahtumat ovat riippumattomia.

Määritelmä 11 (Tapahtumien riippumattomuus)**.** Tapahtumat A ja B ovat riippumattomia, jos ja vain jos

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)\,.$$

Huomaa, että on eri asia ovatko tapahtumat riippumattomia vai toisensa poissulkevia. Itse asiassa toisensa poissulkevat tapahtumat ovat aina riippuvia toisistaan, joka on intuitiivisesti selvää, sillä jos tapahtumat poissulkevat toisensa niin toisen tapahtuman esiintyminen tarkoittaa, että toinen ei voi esiintyä. Eli tapahtumien esiintyminen ei voi olla toisistaan täysin riippumatonta.

Lause 15. Jos tapahtumat A ja B toisistaan riippumattomia, niin tällöin

1. A ja B^c ovat toisistaan riippumattomia
2. A^c ja B ovat toisistaan riippumattomia
3. A^c ja B^c ovat toisistaan riippumattomia

Todistus. Harjoitustehtävä. □

Lause 16. Tapahtumat A ja B ovat riippumattomia ($\mathsf{P}(B) > 0$), jos ja vain jos

$$\mathsf{P}(A \mid B) = \mathsf{P}(A)\,.$$

Todistus. Harjoitustehtävä. □

Esimerkki 31. Tarkastellaan tavallista korttipakkaa, josta valitaan yksi kortti satunnaisesti. Olkoon $A =$ kortti on kuvakortti ja $B =$ kortti on hertta. Näytetään, että tapahtumat A ja B ovat toisistaan riippumattomat. Käyttäen ehdollisen todennäköisyyden kaavaa saamme

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{B} = \frac{4}{13}.$$

Lasketaan vielä todennäköisyys

$$\mathsf{P}(A) = \frac{16}{52} = \frac{4}{13}.$$

Koska $\mathsf{P}(A \mid B) = \mathsf{P}(A)$, niin tapahtumat A ja B ovat toisistaan riippumattomat.

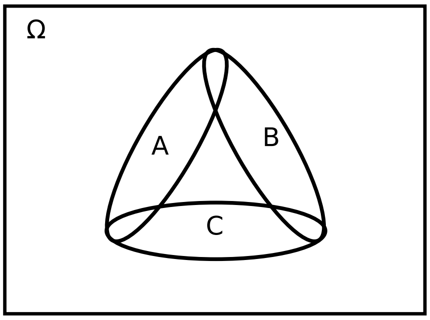
Määritelmä 12 (Tapahtumien riippumattomuus)**.** Tapahtumat $A_1, \ldots, A_n$ ovat toisistaan riippumattomia, jos ja vain jos $\forall i, j, k \in \{1, 2, \ldots, n\}$

$$\begin{aligned}
\mathsf{P}(A_i \cap A_j) &= \mathsf{P}(A_i)\,\mathsf{P}(A_j)\,, \quad i < j \\
\mathsf{P}(A_i \cap A_j \cap A_k) &= \mathsf{P}(A_i)\,\mathsf{P}(A_j)\,\mathsf{P}(A_k)\,, \quad i < j < k \\
&\vdots \\
\mathsf{P}(A_1 \cap \ldots \cap A_n) &= \mathsf{P}(A_1) \cdots \mathsf{P}(A_n)\,.
\end{aligned}$$

Useamman tapahtuman parittaisesta riippumattomuudesta ei seuraa välttämättä kaikkien tapahtumien yhteinen riippumattomuus.

Esimerkki 32. Tarkastellaan Kuvan 1.13 tilannetta. Olkoon $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2}$. Kaikkien kahden tapahtumien leikkauksien todennäköisyydet ovat $\mathsf{P}(A \cap B) = \mathsf{P}(A \cap C) = \mathsf{P}(B \cap C) = \frac{1}{4}$. Nyt tapahtumat ovat pareittan riippumattomat, sillä

$$\begin{aligned}
\mathsf{P}(A \cap B) &= \mathsf{P}(A)\,\mathsf{P}(B) = \frac{1}{4} \\
\mathsf{P}(A \cap C) &= \mathsf{P}(A)\,\mathsf{P}(C) = \frac{1}{4} \\
\mathsf{P}(B \cap C) &= \mathsf{P}(B)\,\mathsf{P}(C) = \frac{1}{4},
\end{aligned}$$



Kuva 1.13: Esimerkin 32 tilanne.

mutta kuitenkaan kaikki tapahtumat eivät ole riippumattomia sillä

$$\mathsf{P}(A \cap B \cap C) = \mathsf{P}(\emptyset) = 0 \neq \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C) = \frac{1}{8}.$$

Lause 17. Jos tapahtumat A, B ja C ovat toisistaan riippumattomat, niin A on riippumaton kaikista tapahtumista, jotka ovat muodostettu tapahtumista B ja C.

Todistus. Harjoitustehtävä. □

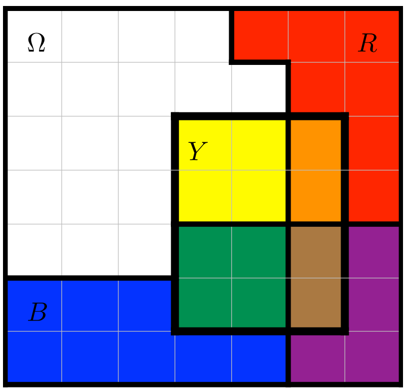
Määritelmä 13 (Ehdollinen riippumattomuus)**.** Tapahtumat A ja B ovat ehdollisesti riippumattomat ehdolla C, jos

$$\mathsf{P}(A \cap B \mid C) = \mathsf{P}(A \mid C)\,\mathsf{P}(B \mid C)\,.$$

Esimerkki 33. Määritellään tapahtumiin A, B ja Y liittyvät todennäköisyydet

$$\begin{aligned} \mathsf{P}(A) &= \frac{15}{49}, \quad \mathsf{P}(A \cap B) = \frac{6}{49} \\ \mathsf{P}(B) &= \frac{18}{49}, \quad \mathsf{P}(A \cap Y) = \frac{4}{49} \\ \mathsf{P}(Y) &= \frac{12}{49}, \quad \mathsf{P}(B \cap Y) = \frac{6}{49} \\ \mathsf{P}(A \cap B \cap Y) &= \frac{2}{49}. \end{aligned}$$

Tilannetta ollaan havainnollistettu Kuvassa 1.14. Nyt A ja B eivät ole riippumattomia toisistaan sillä,



Kuva 1.14: Esimerkin 33 tilanne kuvitettuna.

$$\mathsf{P}(A \cap B) = \frac{6}{49} \neq \mathsf{P}(A)\,\mathsf{P}(B) = \frac{15}{49}\frac{18}{49} = \frac{285}{2401},$$

mutta A ja B ovat toisistaan ehdollisesti riippumattomia ehdolla Y

$$\mathsf{P}(A \cap B \mid Y) = \frac{\mathsf{P}(A \cap B \cap Y)}{\mathsf{P}(Y)} = \frac{\frac{2}{49}}{\frac{12}{49}} = \frac{1}{6} = \mathsf{P}(A \mid Y)\,\mathsf{P}(B \mid Y) = \frac{\mathsf{P}(A \cap Y)}{\mathsf{P}(Y)}\frac{\mathsf{P}(B \cap Y)}{\mathsf{P}(Y)} = \frac{\frac{4}{49}}{\frac{12}{49}} \cdot \frac{\frac{6}{49}}{\frac{12}{49}} = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

Luku 2

Satunnaismuuttujat ja todennäköisyysjakaumat

2.1 Satunnaismuuttuja

Usein satunnaisilmiöiden käsittely määrittelemässämme todennäköisyysavaruudessa saattaa olla hankalaa. Ajatellaan esimerkkinä kolikon heittoa 10 kertaa peräkkäin, jossa olemme kiinnostuneita klaavojen lukumäärästä. Kolikonheittosarjaa satunnaisilmiönä voidaan pitää järjestettynä otantana palauttaen. Mahdollisia lopputuloksia kokeella on yhteensä $2^{10} = 1024$ kappaletta, joista jokainen on yhtä todennäköinen. Jos kuitenkin olemme kiinnostuneita ainoastaan klaavojen lukumäärästä heittosarjassa, niin on mielekästä määritellä muuttuja, jonka mahdolliset arvot ovat $0, 1, \ldots, 10$.

Yleisestikin satunnaisilmiöiden mallintaminen helpottuu määräämällä numeeriset arvot satunnaisilmiöiden mahdollisille tapahtumille. Tämä tapahtuu määrittelemällä **satunnaismuuttujaksi** kutsuttu funktio, joka liittää satunnaisilmiön lopputulokseen jonkin lukuarvon.

Määritelmä 14 (Satunnaismuuttuja)**.** Olkoon $(\Omega, \mathcal{F}, \mathsf{P})$ todennäköisyysavaruus. Satunnaismuuttuja X on otosavaruuden Ω reaaliarvoinen funktio $X : \Omega \to \mathbb{R}$, joka liittää jokaiseen otosavaruuden alkeistapahtumaan ω reaaliluvun $x = X(\omega)$.

Satunnaismuuttuja X on siis sääntö, joka määrää satunnaisilmiön jokaiseen tulokseen reaaliluvun, eikä kyseiseen sääntöön liity satunnaisuutta. Satunnaismuuttujan X tuloavaruutta voidaan kutsua X:n otosavaruudeksi Ω_X. Satunnaismuuttuja on satunnainen siinä mielessä, että ilmiötä tarkastellessa **ennen tuloksen realisoitumista** ei voida varmuudella sanoa kyseisen funktion arvoa. Tässä luentomonisteessa merkitsemme satunnaismuuttujia isoilla kirjaimilla $X, Y, Z, \ldots$, ja näiden realisoituneita arvoja vastaavilla pienillä kirjaimilla $x, y, z, \ldots$

Esimerkki 34. Heitetään kolikkoa kolme kertaa ja voittosumma X (euroissa) on saatujen klaavojen lukumäärä. Merkitään $H =$ kruuna, $T =$ klaava. Otosavaruus on

$$\Omega = \{\text{HHH, HHT, HTH, THH, HTT, THT, TTH ,TTT}\}.$$

Satunnaismuuttuja X liittää jokaiseen alkeistapahtumaan reaaliluvun klaavojen määrän mukaan, eli

$$X(\text{HHH}) = 0, \quad X(\text{HHT}) = 1, \quad X(\text{HTH}) = 1, \quad X(\text{THH}) = 1$$
$$X(\text{HTT}) = 2, \quad X(\text{THT}) = 2, \quad X(\text{TTH}) = 2, \quad X(\text{TTT}) = 3.$$

Satunnaismuuttujan $X(\omega)$ otosavaruus on siis $\Omega_X = \{0, 1, 2, 3\}$.

Esimerkki 35. Heitetään kahta kuusisivuista noppaa ja lasketaan yhteen saadut silmäluvut muuttujaksi X. Alkeistapahtumat $(\omega^{\mathsf{A}}, \omega^{\mathsf{B}})$ ja niitä vastaavat satunnaismuuttujan $X(\omega^{\mathsf{A}}, \omega^{\mathsf{B}})$ arvot ovat merkitty alla olevaan taulukkoon.

$X(\omega^{\mathsf{A}},\omega^{\mathsf{B}})$		Nopan A silmäluku (ω^{A}) 1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
Nopan B	3	4	5	6	7	8	9
silmäluku (ω^{B})	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Satunnaismuuttujan X otosavaruus $\Omega_X = \{2, 3, \ldots, 12\}$.

Satunnaismuuttujan todennäköisyysjakauma kertoo miten todennäköisyys on jakautunut satunnaismuuttujan mahdollisille arvoille. Todennäköisyysjakauma voidaan aina määritellä **kertymäfunktion** (cumulative distribution function) avulla.

Määritelmä 15 (Kertymäfunktio)**.** Satunnaismuuttujan X kertymäfunktio $F_X : \mathbb{R} \to [0, 1]$ määritellään

$$F_X(x) = \mathsf{P}\left(X \leq x\right).$$

Lause 18. Kertymäfunktio on kasvava funktio, jolla on mm. ominaisuudet

$$\begin{array}{ll}
\lim_{a\to-\infty} F_X(a) = 0, & \lim_{a\to\infty} F_X(a) = 1 \\
\mathsf{P}\left(X \leq a\right) = F_X(a), & \mathsf{P}\left(a < X \leq b\right) = F_X(b) - F_X(a) \\
\mathsf{P}\left(X < a\right) = F_X(a^-), & \mathsf{P}\left(a \leq X \leq b\right) = F_X(b) - F_X(a^-) \\
\mathsf{P}\left(X > a\right) = 1 - F_X(a), & \mathsf{P}\left(a \leq X < b\right) = F_X(b^-) - F_X(a^-) \\
\mathsf{P}\left(X \geq a\right) = 1 - F_X(a^-), & \mathsf{P}\left(a < X < b\right) = F_X(b^-) - F_X(a).
\end{array}$$

Todistus. Ei todisteta. □

Lauseen 18 merkintä $F_X(a^-)$ tarkoittaa funktion $F_X(x)$ raja-arvoa, kun x lähenee lukua a sen vasemmalta puolelta. Vastaavaa merkintä on $F_X(a^+)$, jos lähestytään oikealta puolelta.

Esimerkki 36. Olkoon satunnaismuuttujan X kertymäfunktio F muotoa

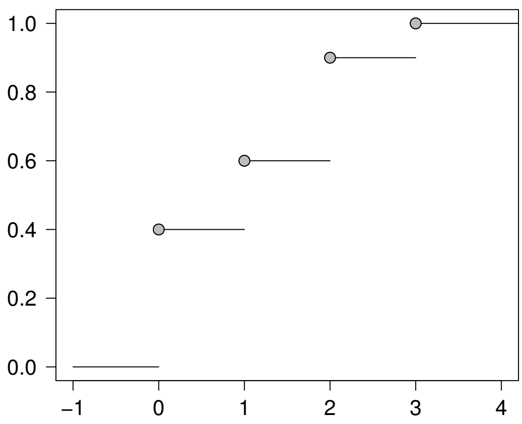
$$F(x) = \begin{cases} 0, & \text{kun } x < 0 \\ 0.4, & \text{kun } 0 \leq x < 1 \\ 0.6, & \text{kun } 1 \leq x < 2 \\ 0.9, & \text{kun } 2 \leq x < 3 \\ 1, & \text{kun } x \geq 3, \end{cases}$$

joka piirretty Kuvaan 2.1. Kertymäfunktiosta voidaan laskea todennäköisyyksiä satunnaismuuttujan arvoille käyttämällä Lauseen 18 ominaisuuksia

$$\begin{aligned}
\mathsf{P}\left(X < 1\right) &= F_X(1^-) = 0.4. \\
\mathsf{P}\left(X = 1\right) &= F_X(1) - F_X(1^-) = 0.6 - 0.4 = 0.2 \\
\mathsf{P}\left(X > 1\right) &= 1 - F_X(1^-) = 1 - 0.4 = 0.6 \\
\mathsf{P}\left(1 < X \leq 3\right) &= F_X(3) - F_X(1) = 1 - 0.6 = 0.4.
\end{aligned}$$

2.1.1 Diskreetti satunnaismuuttuja

Määritelmä 16 (Diskreetti satunnaismuuttuja)**.** Satunnaismuuttuja X on diskreetti, jos sen mahdolliset arvot Ω_X muodostuvat äärellisestä tai numeroituvasti äärettömästä määrästä reaalilukuja.



Kuva 2.1: Esimerkin 36 kertymäfunktio.

Diskreetillä satunnaismuuttujalla X on numeroituva määrä mahdollisia arvoja $\Omega_X = \{x_1, x_2, \ldots, \}$. Tällöin X:n arvojen todennäköisyydet voidaan määritellä todennäköisyyksien

$$p_X(x_i) = \mathsf{P}\left(\{\omega_i \mid X(\omega_i) = x_i\}\right)$$

avulla. Eli todennäköisyys, että satunnaismuuttuja X saa arvon x_i on niiden alkeistapahtumien ω_i yhteenlaskettu todennäköisyys, joilla $X(\omega_i) = x_i$. Diskreetin satunnaismuuttujan X **todennäköisyysjakauma** muodostuu X:n erisuurista arvoista ja niihin liittyvistä todennäköisyyksistä. Nämä voidaan ilmaista **pistetodennäköisyysfunktion** avulla.

Määritelmä 17 (Pistetodennäköisyysfunktio)**.** Diskreetin satunnaismuuttujan pistetodennäköisyysfunktio $p_X : \mathbb{R} \to [0, 1]$ on muotoa

$$p_X(x) = \begin{cases} \mathsf{P}\left(X = x\right), & \text{kun } x \in \Omega_X \\ 0, & \text{muulloin.} \end{cases}$$

Usein kirjoitamme pistetodennäköisyysfunktion lyhyemmin

$$p_X(x) = \mathsf{P}\left(X = x\right), \quad x \in \Omega_X.$$

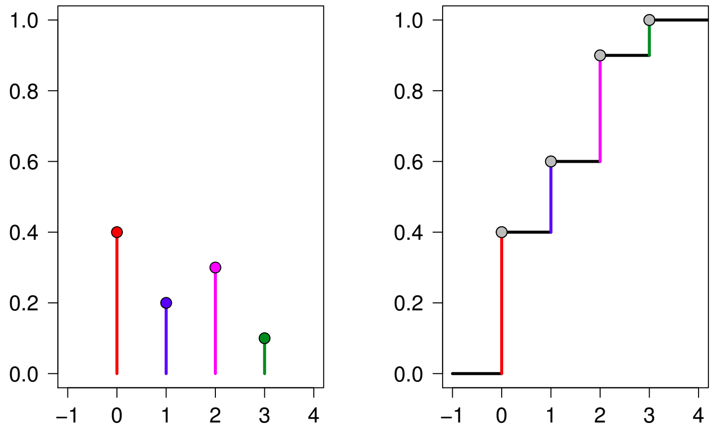
Pistetodennäköisyysfunktiolla on ominaisuudet

$$0 \le p_X(x) \le 1, \quad \forall x \in \Omega_X$$
$$\sum_{x \in \Omega_X} p_X(x) = 1,$$

jossa merkintä $\sum_{x \in \Omega_X}$ tarkoittaa, että summa on yli kaikkien otosavaruuden Ω_X arvojen.
Diskreetin satunnaismuuttujan kertymäfunktio on muotoa

$$F_X(x) = \sum_{x_i \le x} \mathsf{P}\left(X = x_i\right) = \sum_{x_i \le x} p_X(x_i).$$

Diskreetin satunnaismuuttujan pistetodennäköisyysfunktion ja kertymäfunktion yhtyettä on selvitetty Kuvassa 2.2, jossa on piirretty Esimerkin 36 satunnaismuuttujaan X liittyvät pistetodennäköisyysfunktio sekä kertymäfunktio.



Kuva 2.2: Esimerkin 36 satunnaismuuttujan pistetodennäköisyys- ja kertymäfunktio.

Esimerkki 37. Diskreettin satunnaismuuttujan X pistetodennäköisyysfunktio on

$$p_X(0) = p, \quad p_X(1) = 2p, \quad p_X(2) = \frac{p}{2}.$$

Haluamme määrätä luvun p. Koska p_X on satunnaismuuttujan X todennäköisyysjakauma, tiedämme että X:n mahdollisten arvojen todennäköisyydet summautuvat arvoon 1. Voimme siis ratkaista

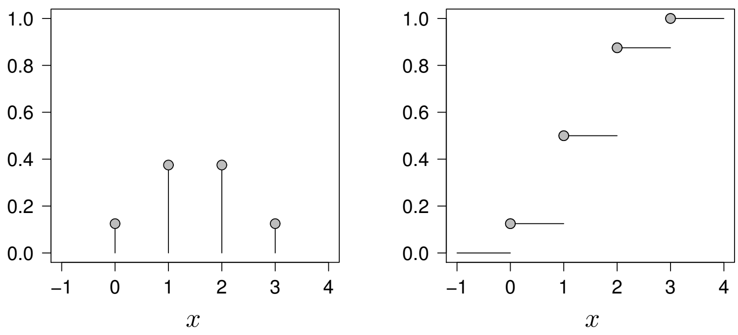
$$\begin{aligned} & p_X(0) + p_X(1) + p_X(2) = 1 \\ \Leftrightarrow \quad & p + 2p + \frac{p}{2} = 1 \\ \Leftrightarrow \quad & p = \frac{2}{7}. \end{aligned}$$

Esimerkki 38. Jatketaan Esimerkkiä 34. Satunnaismuuttujan X mahdolliset arvot ovat siis $\Omega_X = \{0, 1, 2, 3\}$. Laskemalla yhteen alkeistapahtumien todennäköisyydet, jotka tuottavat $X(\omega_i)$:n mahdolliset arvot, saamme todennäköisyyksiksi

$$\mathsf{P}(X = 0) = \frac{1}{8}, \mathsf{P}(X = 1) = \frac{3}{8}, \mathsf{P}(X = 2) = \frac{3}{8}, \mathsf{P}(X = 3) = \frac{1}{8}.$$

Satunnaismuuttujan X pistetodennäköisyysfunktio sekä kertymäfunktio ovat piirrettyinä Kuvaan 2.3.

Esimerkki 39. Neljä kirjettä, jotka ovat suunnattuja henkilöille A,B,C ja D laitetaan satunnaisesti neljään kirjekuoreen, joissa on kunkin henkilön osoite. Olkoon satunnaismuuttuja X niiden kirjeiden lukumäärä, jotka menevät oikeisiin osoitteisiin. Haetaan satunnaismuuttujan X pistetodennäköisyysfunktio listaamalla mahdolliset tapahtumat ja laskemalla montako kirjettä lähtee kullakin tapahtumalla oikeaan osoitteeseen. Huomaa, että kyseessä on *järjestetty otanta ilman palautusta*, joten alkeistapahtumia on $4! = 24$ kpl.



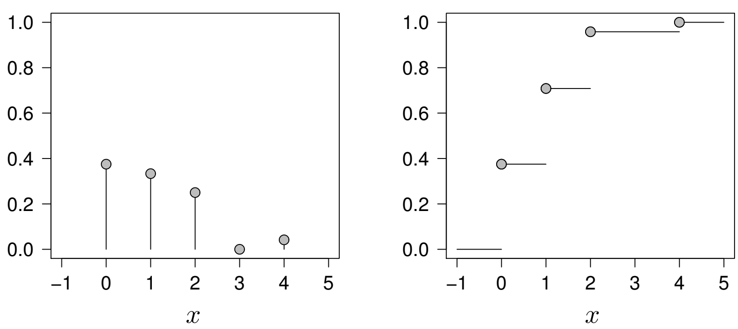
Kuva 2.3: Pistetodennäköisyysfunktio ja kertymäfunktio Esimerkin 38 satunnaismuuttujalle.

Oikea järjestys	ω_i																							
A	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	C	D	D	D	D	D	D
B	B	B	C	C	D	D	A	A	C	C	D	D	A	A	B	B	D	D	A	A	B	B	C	C
C	C	D	B	D	B	C	C	D	A	D	A	C	B	D	A	D	A	B	B	C	A	C	A	B
D	D	C	D	B	C	B	D	C	D	A	C	A	D	B	D	A	B	A	C	B	C	A	B	A
$X(\omega_i) =$	4	2	2	1	1	2	2	0	1	0	0	1	1	0	2	1	0	0	0	1	1	2	0	0

Oletetaan, että jokainen alkeistapahtumista on yhtä todennäköisiä. Tällöin pistetodennäköisyysfunktio satunnaismuuttujalle X on

x	0	1	2	3	4
$p_X(x)$	$\frac{9}{24}$	$\frac{8}{24}$	$\frac{6}{24}$	$\frac{0}{24}$	$\frac{1}{24}$,

jonka pistetodennäköisyysfunktio sekä kertymäfunktio ovat piirretty Kuvaan 2.4.



Kuva 2.4: Pistetodennäköisyysfunktio ja kertymäfunktio Esimerkin 39 satunnaismuuttujalle.

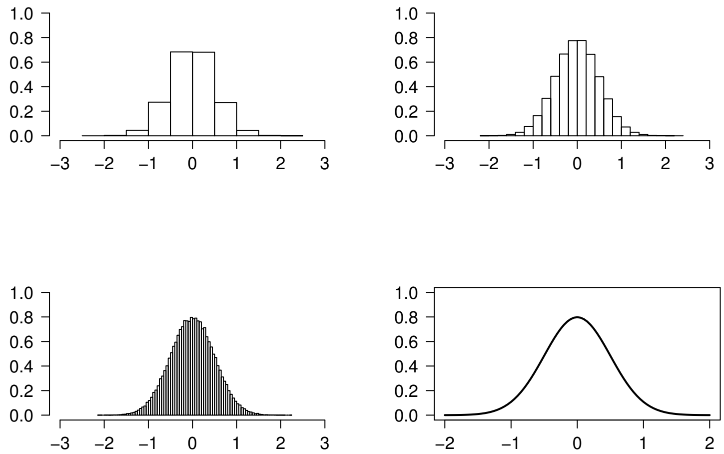
2.1.2 Jatkuva satunnaismuuttuja

Määritelmä 18 (Jatkuva satunnaismuuttuja). Satunnaismuuttuja X on jatkuva, jos sen mahdolliset arvot Ω_X muodostuvat reaaliakselin välistä.

Jatkuva satunnaismuuttuja saa arvokseen reaalilukuarvoja tietyltä lukuväliltä. Koska mahdollisia arvoja on ylinumeroituva määrä, niin jatkuvan satunnaismuuttujan todennäköisyysjakaumaa ei voida määritellä pistetodennäköisyysfunktion avulla. Jatkuvalle satunnaismuuttujalle X jokaisen yksittäisen arvon todennäköisyys on 0, eli

$$X \text{ on jatkuva} \Rightarrow \mathsf{P}\left(X = x\right) = 0, \forall x \in \Omega_X.$$

Pistetodennäköisyysfunktion sijaan käytetään määrittelyssä joko kertymäfunktiota tai **tiheysfunktiota** (probability density function). Tiheysfunktio kuvaa todennäköisyysmassan 'tiheyttä' satunnaismuuttujan arvojoukossa. Tätä voi ajatella esimerkiksi siten, että satunnaismuuttujan arvojoukko on ositettu numeroituvaan määrään välejä. Jokaiselle välille voidaan piirtää palkki, siten että palkin ala vastaa todennäköisyyttä, että satunnaismuuttujan arvo kuuluu vastaavaan väliin. Kun välejä pienennetään äärimmäisen lähelle nollaa, saadaan raja-arvona äärettömän kapeiden palkkien yläosista satunnaismuuttujan tiheysfunktion muoto. Tätä ajattelua selvennetään Kuvassa 2.5.



Kuva 2.5: Jatkuvan satunnaismuuttujan tiheysfunktion havainnollistus.

Satunnaismuuttuja on jatkuva, jos sen kertymäfunktio on jatkuva ja derivoituva. Kertymäfunktio F_X ja tiheysfunktio f_X voidaan laskea toistensa avulla

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u$$
$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x).$$

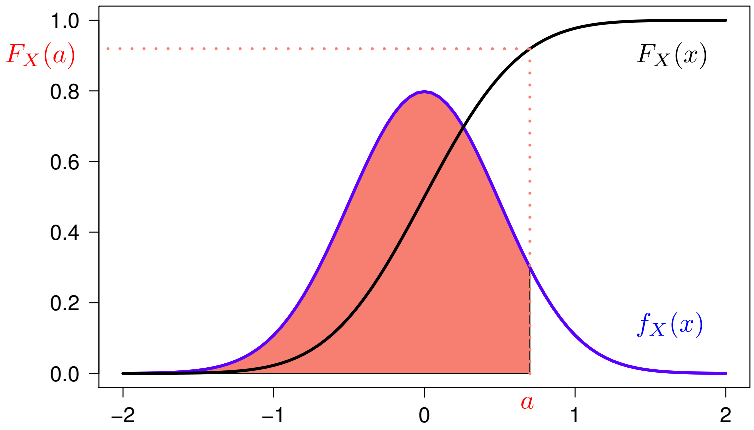
Eli käytännössä tiheysfunktion arvo saadaan derivoimalla kertymäfunktiosta, ja kertymäfunktion arvo pisteessä a on tiheysfunktion käyrän, x-akselin ja pisteessä a sijaitsevan pystysuoran rajoittaman alueen pinta-ala. Jatkuvan satunnaismuuttujan tiheysfunktion ja kertymäfunktion yhteyttä on selvitetty Kuvassa 2.6. Tiheysfunktiolla on ominaisuudet

$$f_X(x) \geq 0, \quad \forall x \in \Omega_X$$
$$\int_{\Omega_X} f_X(u)\mathrm{d}u = 1,$$

jossa merkintä Ω_X tarkoittaa integraalia otosavaruuden Ω_X yli. Huomaa, ettei tiheysfunktion arvoja ole rajoitettu ylhäältä välille $[0, 1]$.

Satunnaismuuttujan X todennäköisyysjakauma määritellään satunnaismuuttan otosavaruudessa Ω_X. Käytännössä tämä tarkoittaa, että jatkuvan satunnaismuuttujan tiheysfunktio $f_X(x) = 0$, kun $x \notin \Omega_X$. Tätä ei usein kirjoiteta erikseen tiheysfunktion määrittelyssä. Toisin sanoen merkintä

$$f_X(x) = f_X^+(x), \text{ kun } x \in \Omega_X$$



Kuva 2.6: Jatkuvan satunnaismuuttujan tiheys- ja kertymäfunktio.

tarkoittaa käytännössä samaa kuin

$$f_X(x) = \begin{cases} f_X^+(x), & \text{kun } x \in \Omega_X \\ 0, & \text{kun } x \notin \Omega_X. \end{cases}$$

Esimerkki 40. Satunnaismuuttujalla X on tiheysfunktio

$$f_X(x) = 1, \text{ kun } x \in [0,1].$$

Nyt X:n kertymäfunktio on muotoa

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u = \int_{-\infty}^{x} 1 \ \mathrm{d}x = x, \ \text{kun } x \in [0,1].$$

Esimerkki 41. Jatkuvan satunnaismuuttujan X, jonka otosavaruus on $\Omega_X = [5,10]$, tiheysfunktio on

$$f_X(x) = a(10-x), \quad x \in [5,10].$$

Halutaan määrätä tuntematon vakio a. Koska $f_X(x)$ on satunnaismuuttujan tiheysfunktio, niin tiedämme, että sen integraali satunnaismuuttujan otosavaruuden yli täytyy olla 1. Voimme siis ratkaista

$$\begin{aligned}
& \int_5^{10} f_X(x)\mathrm{d}x = \int_5^{10} a(10-x)\mathrm{d}x = 1 \\
\Leftrightarrow \quad & a\int_5^{10} (10-x)\mathrm{d}x = 1 \\
\Leftrightarrow \quad & a\left(10x - \frac{1}{2}x^2\right)\Big|_5^{10} = 1 \\
\Leftrightarrow \quad & a\left[\left(10\cdot 10 - \frac{1}{2}10^2\right) - \left(10\cdot 5 - \frac{1}{2}5^2\right)\right] = 1 \\
\Leftrightarrow \quad & a\cdot\frac{25}{2} = 1 \\
\Leftrightarrow \quad & a = \frac{2}{25}
\end{aligned}$$

Esimerkki 42. Tarkastellaan edellisessä tehtävässä määriteltyä tiheysfunktiota

$$f_X(x) = \frac{2}{25}(10-x), \quad x \in [5,10].$$

Lasketaan tiheysfunktiolle kertymäfunktio paloittain kolmelle eri alueelle. Arvoille $x \in (-\infty, 5]$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\mathrm{d}y = \int_{-\infty}^{x} 0\mathrm{d}y = 0.$$

Arvoille $x \in (5, 10]$

$$\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(y)\mathrm{d}y = \underbrace{\int_{-\infty}^{5} f_X(y)\mathrm{d}y}_{=F_X(5)=0} + \int_{5}^{x} f_X(y)\mathrm{d}y = \frac{2}{25}\left(10y - \frac{1}{2}y^2\right)\Big|_5^x \\
&= \frac{2}{25}\left[\left(10x - \frac{1}{2}x^2\right) - \left(10 \cdot 5 - \frac{1}{2} \cdot 5^2\right)\right] \\
&= \frac{20}{25}x - \frac{1}{25}x^2 - \frac{100}{25} + \frac{50}{50} = -\frac{1}{25}x^2 + \frac{4}{5}x - 3.
\end{aligned}$$

Arvoille $x \in (10, \infty]$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\mathrm{d}y = \underbrace{\int_{-\infty}^{10} f_X(y)\mathrm{d}y}_{=F_X(10)=1} + \underbrace{\int_{10}^{x} f_X(y)\mathrm{d}y}_{=0} = 1.$$

2.2 Satunnaismuuttujan funktiot

Satunnaismuuttujan määrittelynsä mukaisesti on funktio otosavaruudesta reaalilukujen joukkoon. Satunnaismuuttujan X reaaliarvoiset funktiotkin Y $= g(X)$ ovat satunnaismuuttujia, joilla on oma otosavaruutensa ja todennäköisyysjakaumansa.

Esimerkki 43. Olkoon X diskreetti satunnaismuuttuja, jolla on otosavaruus ja pistetodennäköisyys

x_i	-1	0	1	2
$p_X(x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{6}$.

Olkoon $Y = X^2$. Satunnaismuuttujan Y:n mahdolliset arvot ja näiden todennäköisyydet saadaan evaluoimalla X:n mahdolliset arvot funktiolla $g(x)^2$

x_i^2	$(-1)^2$	0^2	1^2	2^2
$p_X(x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{6}$,

joka on

x_i^2	1	0	1	4
$p_X(x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{6}$,

joten Y:n otosavaruus on $\Omega_Y = \{0, 1, 4\}$ ja todennäköisyysjakauma

y_i	0	1	4
$p_Y(y_i)$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$.

Yleisesti pätee, että jos satunnaismuuttuja X on diskreetti niin satunnaismuuttuja $Y = g(X)$ on myös diskreetti ja sen todennäköisyydet saadaan laskettua

$$\begin{aligned}
p_Y(y) &= \mathsf{P}\left(Y = y\right) = \mathsf{P}\left(g(X) = y\right) \\
&= \sum_{\{x_i : g(x_i) = y\}} \mathsf{P}\left(X = x_i\right) = \sum_{\{x_i : g(x_i) = y\}} p_X(x_i).
\end{aligned}$$

Silloin kun funktiolla $g(x)$ on olemassa käänteisfunktio, niin muunnoksen pistetodennäköisyysfunktion hakemisessa voidaan käyttää myös Lausetta 19.

Lause 19. Olkoon X diskreetti satunnaismuuttuja, jolla on pistetodennäköisyysfunktio $p_X(x)$. Olkoon $g(x)$ kääntyvä funktio, jonka käänteisfunktio on $g^{-1}(y)$. Tällöin diskreetillä satunnaismuuttujalla $Y = g(X)$ on pistetodennäköisyysfunktio

$$p_Y(y) = \sum_{\{x_i : g(x_i) = y\}} = \sum_{\{x_i : g(x_i) = y\}} p_X(x_i) = \sum_{\{x_i : x_i = g^{-1}(y)\}} p_X(x_i)$$

Jatkuvan satunnaismuuttujan funktiot ovat myös edelleen satunnaismuuttujia. Näiden käsittely on ei kuitenkaan ole aina yhtä suoraviivaista kuin diskreettien satunnaismuuttujien. Jos satunnaismuuttujan funktio on aidosti kasvava tai vähenevä, eli aidosti monotoninen, voidaan muunnetun satunnaismuuttujan tiheysfunktio löytää Lauseen 20 avulla. Aidosti kasvavalle funktiolle $g_1(x)$ pätee

$$x_1 < x_2 \Rightarrow g_1(x_1) < g_1(x_2),$$

ja vähenevälle funktiolle $g_2(x)$ pätee

$$x_1 < x_2 \Rightarrow g_2(x_1) > g_2(x_2).$$

Lause 20. Olkoon X jatkuva satunnaismuuttuja, jolla on tiheysfunktio $f_X(x)$. Olkoon $g(x)$ derivoituva ja aidosti monotoninen funktio, jonka käänteisfunktio on $g^{-1}(y)$. Tällöin jatkuvalla satunnaismuuttujalla $Y = g(X)$ on tiheysfunktio

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|.$$

Todistus. Oletetaan ensin, että lauseen mukaiset ehdot toteuttava $g_(x)$ on kasvava funktio. Nyt pätee

$$F_Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}(g(X) \le y) = \mathsf{P}\left(X \le g^{-1}(y)\right) = F_X(g^{-1}(y)),$$

jolloin saamme kertymäfunktiosta derivoimalla

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(g^{-1}(y)) \\ &\overset{(*)}{=} f_X(g^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y). \end{aligned} \tag{2.1}$$

Kohdassa (*) käytetään derivoinnin ketjusääntöä. Jos oletetaan toisaalta, että $g_(x)$ on vähenevä funktio, niin tällöin pätee

$$F_Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}(g(X) \le y) = \mathsf{P}\left(X \ge g^{-1}(y)\right) = 1 - F_X(g^{-1}(y)).$$

Nyt jälleen derivoimalla ketjusääntöä käyttäen saadaan

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} 1 - F_X(g^{-1}(y)) \\ &= -f_X(g^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y). \end{aligned} \tag{2.2}$$

Koska kasvavalle funktiolle $\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0$ ja vähenevälle $\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) < 0$, niin yhteensä (2.1) ja (2.2) todistavat Lauseen väitteen. □

Esimerkki 44. Satunnaismuuttujalla X on tiheysfunktio

$$f_X(x) = \frac{2}{25}(10 - x), \quad x \in [5, 10].$$

Haetaan satunnaismuuttujan $Y = 3X - 2$ tiheysfunktio Lauseen 20 avulla. Ensin lasketaan käänteisfunktio

$$\begin{aligned} & y = g(x) = 3x - 2 \\ \Leftrightarrow \quad & x = g^{-1}(y) = \frac{1}{3} y + \frac{2}{3}, \end{aligned}$$

jonka derivaatta on

$$\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) = \frac{1}{3}.$$

Satunnaismuuttujan Y tiheysfunktio on siis

$$\begin{aligned} f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right| &= \frac{2}{25}\left(10 - \left(\frac{1}{3}y + \frac{2}{3}\right)\right)\left|\frac{1}{3}\right| \\ &= \frac{20}{75} - \frac{2}{225}y - \frac{4}{225} = -\frac{2}{225}y + \frac{56}{225}. \end{aligned}$$

Koska $g(x)$ on aidosti kasvava, niin muunnetun satunnaismuuttujan Y otosavaruus on

$$\Omega_Y = [g(5), g(10)] = [13, 28].$$

Jos satunnaismuuttuja X on jatkuva, mutta satunnaismuuttujan $Y = g(X)$ määrittelevä funktio $g(x)$ ei ole kääntyvä, niin Y jakauman etsimisessä täytyy olla varovainen. Tarkastellaan tilannetta $Y = X^2$, jossa $g(x) = x^2$ ei ole kääntyvä funktio. Haetaan kertymäfunktio

$$F_Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}\left(X^2 \le y\right) = \mathsf{P}(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

josta derivoimalla ketjusääntöä käyttäen saadaan

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\mathrm{d}}{\mathrm{d}y}F_X(\sqrt{y}) - \frac{\mathrm{d}}{\mathrm{d}y}F_X(-\sqrt{y}) \\ &= f_X(\sqrt{y})\frac{\mathrm{d}}{\mathrm{d}y}\sqrt{y} - f_X(-\sqrt{y})\frac{\mathrm{d}}{\mathrm{d}y}(-\sqrt{y}) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})). \end{aligned}$$

2.3 Todennäköisyysjakaumien tunnuslukuja

Todennäköisyysjakaumista voidaan laskea monenlaisia **tunnuslukuja**, joilla pystyy kuvailemaan esimerkiksi todennäköisyysjakauman sijaintia reaaliakselilla, satunnaismuuttujan arvojen vaihtelevuutta tai jakauman muotoa. Tunnuslukujen avulla on myös kätevää vertailla jakaumien ominaisuuksia toisiinsa.

2.3.1 Odotusarvo

Jakauman sijaintia reaaliakselilla voidaan kuvata satunnaismuuttujan **odotusarvon** avulla. Odotusarvo on satunnaismuuttujan kaikkien arvojen painotettu keskiarvo, jossa painoina käytetään arvojen todennäköisyyksiä. Voidaan ajatella, että se on jakauman keskimääräinen arvo. On myös mahdollista, että satunnaismuuttujan jakauma on muotoa, jolla ei ole olemassa odotusarvoa. Tällöin odotusarvon määritellyssä esiintyvät summa tai integraali eivät ole äärellisinä olemassa. Muun muassa ns. paksuhäntäisillä jakaumilla, jotka jakavat todennäköisyyttä tasaisesti suurelle osalle avaruutta, ei välttämättä ole olemassa odotusarvoa.

Diskreetit satunnaismuuttujat

Määritelmä 19 (Odotusarvo)**.** Olkoon X diskreetti satunnaismuuttuja, jolla on otosavaruus Ω_X. Satunnaismuuttujan odotusarvo määritellään

$$\mathsf{E}(X) = \sum_{x \in \Omega_X} x p_X(x),$$

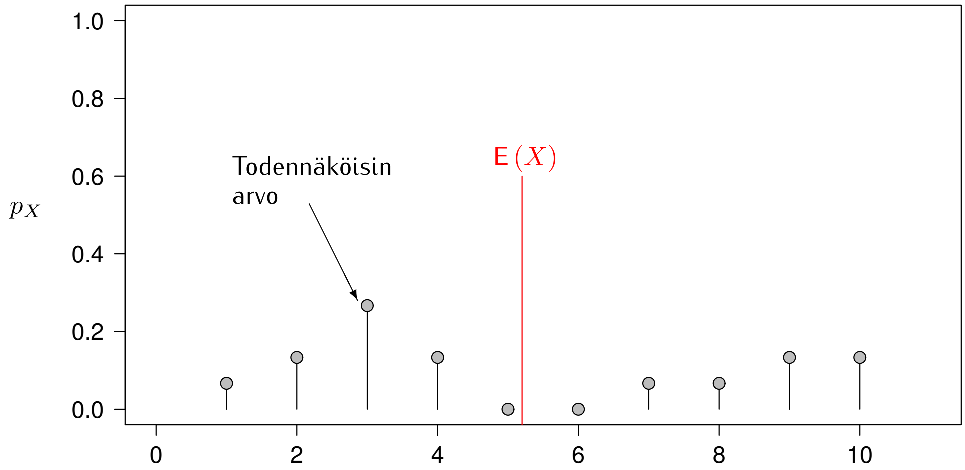
kun summa suppenee itseisesti

$$\sum_{x \in \Omega_X} |x|\, p_X(x) < \infty.$$

Huomaa, että satunnaismuuttujan odotusarvo ja satunnaismuuttujan **todennäköisin arvo** (jakauman moodi) ovat eri asioita, kuten Kuvassa 2.7 ollaan havainnollistettu pistetodennäköisyysfunktiolle

k	1	2	3	4	5	6	7	8	9	10
$p_X(k) =$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{2}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$.

Diskreetin satunnaismuuttujan odotusarvo ei välttämättä ole edes satunnaismuuttujan arvojoukossa.



Kuva 2.7: Todennäköisimmän arvon ja odotusarvon ero.

Esimerkki 45. Tarkastellaan tasapainoisen kuusisivuisen nopan heittoa. Olkoon X heitossa saatu silmäluku. Satunnaismuuttujan X odotusarvo on

$$\mathsf{E}\,(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Jatkuvat satunnaismuuttujat

Määritelmä 20 (Odotusarvo)**.** Olkoon X jatkuva satunnaismuuttuja. Satunnaismuuttujan odotusarvo määritellään

$$\mathsf{E}\,(X) = \int_{\Omega_X} x f_X(x) \mathrm{d}x,$$

kun integraali suppenee itseisesti

$$\int_{\Omega_X} |x|\, f_X(x) \mathrm{d}x < \infty.$$

Joskus odotusarvon integraalia merkitään ilman ala- ja ylärajaa, joka kuitenkin tarkoittaa samaa asiaa.

Esimerkki 46. Tarkastellaan satunnaismuuttujaa X, jonka tiheysfunktio on määritelty

$$f_X(x) = 1, \quad x \in [0, 1].$$

Lasketaan satunnaismuuttujan odotusarvo

$$\mathsf{E}\,(X) = \int_0^1 x f_X(x) \mathrm{d}x = \int_0^1 x \mathrm{d}x = \left(\frac{1}{2} x^2\right)\Big|_0^1 = \left(\frac{1}{2} \cdot 1^2\right) - \left(\frac{1}{2} \cdot 0^2\right) = \frac{1}{2}.$$

Odotusarvon ominaisuuksia

Joskus on tarpeen laskea satunnaismuuttujan X funktion $g(X)$ odotusarvo. Tämä on onnistuu suoraviivaisesti Lauseiden 21 ja 22 avulla.

Lause 21. Olkoon X diskreetti satunnaismuuttuja, joka on määritelty otosavaruudessa Ω_X, jolla on pistetodennäköisyysfunktio $p_X(x)$ ja määritellään $Y = g(X)$. Satunnaismuuttujan Y odotusarvo on

$$\mathsf{E}\,(Y) = \mathsf{E}\,(g(X)) = \sum_{x \in \Omega_X} g(x) p_X(x),$$

kun

$$\sum_{x \in \Omega_X} |g(x)| \, p_X(x) < \infty.$$

Lause 22. Olkoon X jatkuva satunnaismuuttuja, joka on määritelty otosavaruudessa Ω_X, jolla on tiheysfunktio $f_X(x)$ ja määritellään $Y = g(X)$. Satunnaismuuttujan Y odotusarvo on

$$\mathsf{E}\,(Y) = \mathsf{E}\,(g(X)) = \int_{\Omega_X} g(x) f_X(x) \mathrm{d}x,$$

kun

$$\int_{\Omega_X} |g(x)| \, f_X(x) \mathrm{d}x < \infty.$$

Erityisesti usein on tarpeellista laskea satunnaismuuttujan potenssien odotusarvoja.

Esimerkki 47. Tarkastellaan tasapainoisen kuusisivuisen nopan heittoa. Olkoon X heitossa saatu silmäluku. Lasketaan odotusarvo satunnaismuuttujalle X^2

$$\mathsf{E}\left(X^2\right) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}.$$

Esimerkki 48. Tarkastellaan satunnaismuuttujaa X, jonka tiheysfunktio on

$$f_X(x) = 1, \quad x \in [0, 1].$$

Lasketaan satunnaismuuttujan X^2 odotusarvo

$$\mathsf{E}\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f_X(x) \mathrm{d}x = \int_0^1 x^2 \mathrm{d}x = \left(\frac{1}{3}x^3\right)\Big|_0^1 = \left(\frac{1}{3} \cdot 1^2\right) - \left(\frac{1}{3} \cdot 0^2\right) = \frac{1}{3}.$$

Lause 23. Olkoon X satunnaismuuttuja, jolla on olemassa odotusarvo äärellisenä. Nyt jokaiselle vakiolle a ja b pätee

$$\mathsf{E}\,(aX + b) = a\mathsf{E}\,(X) + b.$$

Todistus. Harjoitustehtävä. □

Esimerkki 49. Olkoon X satunnaismuuttuja, jolla on olemassa odotusarvo äärellisenä. Satunnaismuuttujan $Y = X - \mu$ odotusarvo on

$$\mathsf{E}\,(Y) = \mathsf{E}\,(X - \mu) = \mathsf{E}\,(1 \cdot X + (-\mu)) = 1 \cdot \mathsf{E}\,(X) + (-\mu) = \mu - \mu = 0.$$

Lause 24. Olkoon X (diskreetti tai jatkuva) satunnaismuuttuja, joka on määritelty otosavaruudessa Ω_X, ja määritellään $Y = g_1(X)$ ja $Z = g_2(X)$. Nyt

$$\mathsf{E}\,(Y + Z) = \mathsf{E}\,(g_1(X) + g_2(X)) = \mathsf{E}\,(g_1(X)) + \mathsf{E}\,(g_2(X))\,.$$

Todistus. Harjoitustehtävä. □

2.3.2 Varianssi ja keskihajonta

Satunnaismuuttujan vaihtelua voidaan kuvata varianssin tai keskihajonnan avulla. Käytännössä nämä tunnusluvut kuvaavat kuinka kaukana satunnaismuuttujan mahdolliset arvot ovat keskimäärin odotusarvosta. Mitä pienempi satunnaismuuttujan keskipoikkeama on, niin sitä vähemmän jakaumasta havaitut arvot poikkeavat odotusarvosta.

Määritelmä 21 (Varianssi). Olkoon $\mathsf{E}\,(X) = \mu$. Satunnaismuuttujan varianssi määritellään odotusarvo-operaattoreiden avulla

$$\mathsf{Var}\,(X) = \mathsf{E}\left((X - \mu)^2\right).$$

Varianssin kaava on aukikirjoitettuna diskreetille satunnaismuuttujalle muotoa

$$\mathsf{Var}\,(X) = \sum_{x\in\Omega_X} (x-\mu)^2 p_X(x)$$

ja jatkuvalle satunnaismuuttujalle

$$\mathsf{Var}\,(X) = \int_{\Omega_X} (x-\mu)^2 f_X(x)\mathrm{d}x,$$

joissa Ω_X on satunnaismuuttujan otosavaruus.

Määritelmä 22 (Keskihajonta)**.** Olkoon $\sigma^2 = \mathsf{Var}\,(X)$. Satunnaismuuttujan X keskihajonta määritellään

$$\mathsf{Sd}\,(X) = \sigma.$$

Varianssin ominaisuuksia

Lause 25. Satunnaismuuttujan X varianssille pätee

1. $\mathsf{Var}\,(X) \geq 0$
2. $\mathsf{Var}\,(X) = 0$, jos ja vain jos $\mathsf{P}\,(X=a) = 1$, jollakin vakiolla a
3. $\mathsf{E}\left((X-a)^2\right)$ saa pienimmän arvonsa, kun $a = \mathsf{E}\,(X)$.

Todistus. Harjoitustehtävä. □

Usein varianssin laskemiseen on hyvin kätevää käyttää Lauseen 27 tulosta, jonka käyttämiseen meidän tarvitsee löytää satunnaismuuttujan X^2 odotusarvo.

Lause 26. Olkoon satunnaismuuttujan X odotusarvo μ. Satunnaismuuttujan X varianssi voidaan ilmaista muodossa

$$\mathsf{Var}\,(X) = \mathsf{E}\left(X^2\right) - \mu^2.$$

Todistus. Harjoitustehtävä. □

Lause 27. Olkoon satunnaismuuttujan X odotusarvo μ. Satunnaismuuttujan X varianssi voidaan ilmaista muodossa

$$\mathsf{Var}\,(X) = \mathsf{E}\,(X(X-1)) + \mu - \mu^2.$$

Todistus. Harjoitustehtävä. □

Esimerkki 50. Tarkastellaan tasapainoisen kuusisivuisen nopan heittoa. Olkoon X heitossa saatu silmäluku. Esimerkissä 45 laskettiin $\mathsf{E}\,(X) = \frac{7}{2}$, ja Esimerkissä 47 laskettiin odotusarvo $\mathsf{E}\left(X^2\right) = \frac{91}{6}$ Käyttämällä Lausetta 27 saadaan laskettua varianssi

$$\mathsf{Var}\,(X) = \mathsf{E}\left(X^2\right) - \mathsf{E}\,(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

josta lasketaan keskipoikkeama

$$\mathsf{Sd}\,(X) = \sqrt{\frac{35}{12}}.$$

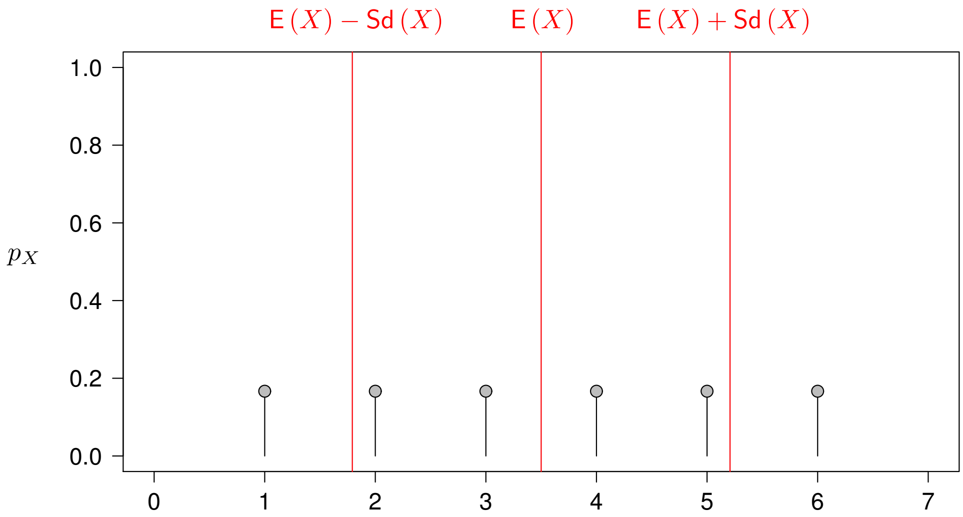
Tilannetta ollaan havainnollistettu Kuvassa 2.8.

Esimerkki 51. Tarkastellaan satunnaismuuttujaa X, jonka tiheysfunktio on määritelty

$$f_X(x) = 1, \quad x \in [0,1].$$

Aiemmin laskettiin satunnaismuuttujien X sekä X^2 odotusarvot $\mathsf{E}\,(X) = 1/2, \quad \mathsf{E}\left(X^2\right) = 1/3$. Lasketaan varianssi

$$\mathsf{Var}\,(X) = \mathsf{E}\left(X^2\right) - \mathsf{E}\,(X)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{4}{12} - \frac{3}{12} = \frac{1}{12}.$$



Kuva 2.8: Esimerkin 50 satunnaismuuttujan X pistetodennäköisyysfunktio, odotusarvo sekä keskipoikkeama.

Lause 28. Olkoon X satunnaismuuttuja, jolla on äärellinen varianssi. Jokaiselle vakiolle a ja b pätee

$$\mathrm{Var}\,(aX+b) = a^2\mathrm{Var}\,(X)\,.$$

Todistus. Harjoitustehtävä. □

Esimerkki 52. Pelataan peliä, jossa heitetään kuusisivuista tasapainoista noppaa, ja pelaaja saa heitosta rahasumman $Y = 10X - 30$ euroa, jossa X on nopanheitossa saatu silmäluku. Esimerkissä 50 laskettiin satunnaismuuttujan X varianssi. Pelin voittosumman varianssi on

$$\mathrm{Var}\,(Y) = \mathrm{Var}\,(10X-30) = 10^2\mathrm{Var}\,(X) = \frac{3500}{12} \approx 291.67.$$

Määritelmä 23 (Z-muunnos)**.** Olkoon $\mathrm{E}\,(X) = \mu$ ja $\mathrm{Sd}\,(X) = \sigma$. Määritellään muunnettu satunnaismuuttuja

$$Z = \frac{X-\mu}{\sigma}.$$

Lause 29. Z-muunnetulle satunnaismuuttujalle pätee

$$\mathrm{E}\,(Z) = 0 \text{ ja } \mathrm{Var}\,(Z) = 1.$$

Todistus. Harjoitustehtävä. □

2.3.3 Moodi, mediaani ja kvantiili

Satunnaismuuttujan odotusarvon lisäksi jakauman sijaintia voidaan kuvailla myös monilla muilla tunnusluvuilla, kuten moodi, mediaani ja kvantiilit.

Määritelmä 24 (Todennäköisyysjakauman moodi)**.** Todennäköisyysjakauman moodi on diskreetin satunnaismuuttujan tapauksessa pistetodennäköisyysfunktion ja jatkuvan satunnaismuuttujan tapauksessa tiheysfunktion maksimiarvo.

Diskreetin satunnaismuuttujan tapauksessa moodi on siis satunnaismuuttujan todennäköisin arvo, ja jatkuvan satunnaismuuttujan tapauksessa moodin välittömässä läheisyydesä on eniten todennäköisyyttä. Todennäköisyysjakauman moodi ei ole aina yksikäsitteisesti määrätty.

Todennäköisyysjakauman mediaani on sellainen arvo, joka jakaa todennäköisyyden kahteen yhtäsuureen osaan. Jakauman mediaani ei ole aina yksikäsitteisesti määrätty.

Määritelmä 25 (Todennäköisyysjakauman mediaani)**.** Mediaani on luku m, jolle pätee

$$\mathsf{P}\left(X \leq m\right) \geq \frac{1}{2}$$
$$\mathsf{P}\left(X \geq m\right) \geq \frac{1}{2}$$

Jatkuvalle satunnaismuuttujalle X voidaan jakauman mediaani määritellä $F_X(m) = \frac{1}{2}$. Mediaani on yksi kvantiilien erikoistapaus, ja kvantiilit voidaan määritellä yleisesti seuraavalla tavalla.

Määritelmä 26 (Todennäköisyysjakauman p-kvantiili)**.** Jakauman p-kvantiilit on sellainen luku, joka jakaa todennäköisyyden kahteen osaan, joihin kuuluu todennäköisyysmassaa p ja $1-p$ verran. Eli p-kvantiilille q pätee

$$\mathsf{P}\left(X \leq q\right) \geq p$$
$$\mathsf{P}\left(X \geq q\right) \geq 1-p$$

Jatkuvalle satunnaismuuttujalle X voidaan p-kvantiili q määritellä $F_X(q) = p$.

2.4 Satunnaismuuttujien yhteisjakauma

Useamman satunnaismuuttujan $X_1, \ldots, X_n$ yhdistelmä $(X_1, \ldots, X_n)$ on edelleen satunnaismuuttuja, jota voidaan kutsua myös satunnaisvektoriksi. Jos kaikki satunnaismuuttujat ovat diskreettejä (jatkuvia), niin myös satunnaismuuttujien yhdistelmä on diskreetti (jatkuva). Yleisesti ottaen on mahdollista, että satunnaismuuttuja koostuu sekä diskreeteistä, että jatkuvista osista, mutta tämänkaltaisia satunnaismuuttujia ei tällä kurssilla käsitellä.

Määritelmä 27. Olkoon $(\Omega, \mathcal{F}, \mathsf{P})$ todennäköisyysavaruus. Satunnaismuuttuja $(X_1, \ldots, X_n)$ on otosavaruuden Ω funktio $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, joka liittää jokaiseen otosavaruuden alkeistapahtumaan ω reaaliarvoisen vektorin $(x_1 \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$.

Diskreetillä satunnaisvektorilla on numeroituva määrä mahdollisia arvoja

$$\Omega_{(X_1,\ldots,X_n)} = \{(x_{11}, \ldots, x_{n1}), (x_{12}, \ldots, x_{n2}), (x_{13}, \ldots, x_{n3}), \ldots\},$$

ja jatkuvalla satunnaisvektorilla on ylinumeroituva määrä.

Satunnaisvektorin todennäköisyysjakauma määritellään samankaltaisesti kuin yksiulotteisessa tapauksessa. Todennäköisyysjakauma voidaan määritellä kertymäfunktion avulla.

Määritelmä 28 (Satunnaisvektorin kertymäfunktio)**.** Satunnaismuuttujien $X_1, X_2, \ldots, X_n$ yhteisjakauman kertymäfunktio $F_X : \mathbb{R}^n \to [0,1]$ on muotoa

$$F_{(X_1,X_2,\ldots,X_n)} = \mathsf{P}\left(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\right).$$

Tilastotieteen peruskursseilla käsitellään useimmiten diskreettejä satunnaisvektoreita, ja lisäksi laskuissa rajoitumme kahteen ulottuvuuteen, eli tapaukseen $n = 2$.

Määritelmä 29. Diskreetin satunnaisvektorin (X, Y) yhteispistetodennäköisyysfunktio $p_{(X,Y)} : \mathbb{R}^2 \to [0,1]$ on muotoa

$$p_{(X,Y)}(x,y) = \begin{cases} \mathsf{P}\left(X = x, Y = y\right), & \text{kun } (x,y) \in \Omega_{(X,Y)} \\ 0, & \text{muuten.} \end{cases}$$

Määritelmä 30 (Marginaalijakauma)**.** Olkoon $p_{(X,Y)}(x,y)$ diskreetin satunnaisvektorin (X, Y) pistetodennäköisyysfunktio. Satunnaismuuttujan X marginaalijakauma (reunajakauma) on

$$p_X(x) = \mathsf{P}\left(X = x\right) = \sum_{y \in \Omega_Y} \mathsf{P}\left(X = x, Y = y\right).$$

Määritelmä 31 (Marginaalijakauma). Olkoon $f_{(X,Y)}(x,y)$ jatkuvan satunnaisvektorin (X,Y) tiheysfunktio. Satunnaismuuttujan X marginaalijakauma (reunajakauma) on

$$f_X(x) = \int_{y \in \Omega_Y} f_{(X,Y)}(x,y)\mathrm{d}y.$$

Kahden diskreetin satunnaismuuttujan X ja Y yhteispistetodennäköisyysfunktio voidaan kirjoittaa kaksiulotteisen frekvenssitaulun avulla muodossa

		Y					
		y_1	$\cdots$	y_j	$\cdots$	y_s	p_X
	x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1s}	$p_{1\cdot}$
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
X	x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{is}	$p_{i\cdot}$
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_r	p_{r1}	$\cdots$	p_{rj}	$\cdots$	p_{rs}	$p_{r\cdot}$
	p_Y	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot s}$	$1,$

johon on kirjattu todennäköisyydet

$$\begin{aligned}
p_{ij} &:= \mathsf{P}\left(X = x_i, Y = y_j\right), \\
p_{i\cdot} &:= \mathsf{P}\left(X = x_i\right) = \sum_{j=1}^{s} p_{ij}, \\
p_{\cdot j} &:= \mathsf{P}\left(Y = y_j\right) = \sum_{i=1}^{r} p_{ij}.
\end{aligned}$$

Todennäköisyydet $p_{i\cdot},\ i = 1,\ldots,r$ muodostavat X:n marginaalijakauman ja $p_{\cdot j},\ j = 1,\ldots,s$ muodostavat Y:n marginaalijakauman.

Esimerkki 53. Olkoon satunnaismuuttujilla X ja Y yhteisjakauma

		Y		
		0	1	2
X	0	0.10	0.30	0.05
	1	0.20	0.25	0.10.

Yhteisjakaumasta voidaan määrätä X:n ja Y:n marginaalijakaumat, jotka on kirjoitettu taulukkoon

			Y			X:n marginaali
			0	1	2	p_X
	X	0	0.10	0.30	0.05	0.45
		1	0.20	0.25	0.10	0.55
Y:n marginaali		p_Y	0.30	0.55	0.15	1.

Marginaalijakaumat ovat laskettu

$$\begin{aligned}
\mathsf{P}\left(X = 0\right) &= p_{(X,Y)}(0,0) + p_{(X,Y)}(0,1) + p_{(X,Y)}(0,2) = 0.10 + 0.30 + 0.05 = 0.45 \\
\mathsf{P}\left(X = 1\right) &= p_{(X,Y)}(1,0) + p_{(X,Y)}(1,1) + p_{(X,Y)}(1,2) = 0.20 + 0.25 + 0.10 = 0.55 \\
\mathsf{P}\left(Y = 0\right) &= p_{(X,Y)}(0,0) + p_{(X,Y)}(1,0) = 0.10 + 0.20 = 0.30 \\
\mathsf{P}\left(Y = 1\right) &= p_{(X,Y)}(0,1) + p_{(X,Y)}(1,1) = 0.30 + 0.25 = 0.55 \\
\mathsf{P}\left(Y = 2\right) &= p_{(X,Y)}(0,2) + p_{(X,Y)}(1,2) = 0.05 + 0.10 = 0.15.
\end{aligned}$$

Yhteisjakaumasta voidaan laskea erilaisten tapahtumien todennäköisyyksiä, kuten tapahtuman $X < Y$ todennäköisyys. Todennäköisyys lasketaan etsimällä avaruuden $\Omega_{(X,Y)}$ tapaukset, joille tapahtuman ehto toteutuu, eli tapaukset $(0,1),(0,2),(1,2)$. Todennäköisyys on tällöin

$$\mathsf{P}(X < Y) = p_{(X,Y)}(0,1) + p_{(X,Y)}(0,2) + p_{(X,Y)}(1,2) = 0.30 + 0.05 + 0.10 = 0.45.$$

Muita tapahtumia ja niiden todennäköisyyksiä ovat mm.

$$\begin{aligned}\mathsf{P}(X+Y=2) &= p_{(X,Y)}(0,2) + p_{(X,Y)}(1,1) = 0.05 + 0.25 = 0.30\\ \mathsf{P}(\min(X,Y)=0) &= p_{(X,Y)}(0,0) + p_{(X,Y)}(0,1) + p_{(X,Y)}(0,2) + p_{(X,Y)}(1,0)\\ &= 0.10 + 0.30 + 0.05 + 0.20 = 0.65.\end{aligned}$$

Usean satunnaismuuttujan funktio on edelleen satunnaismuuttuja. Jos haluamme laskea satunnaismuuttujien funktion jakauman, tehdään tämä samoin kuin yksiulotteisen satunnaismuuttujan tapauksessa. Käytännössä haemme satunnaismuuttujan mahdolliset arvot ja laskemme niiden tapausten todennäköisyydet yhteen jotka tuottavat vastaavat arvot. Olkoon (X,Y) diskreetti kaksiulotteinen satunnaisvektori, ja $Z = g(X,Y)$ diskreetti yksiulotteinen satunnaismuuttuja, jonka todennäköisyydet saadaan laskettua

$$\begin{aligned}p_Z(z) &= \mathsf{P}(Z=z) = \mathsf{P}(g(X,Y)=z)\\ &= \sum_{\{(x_i,y_j):g(x_i,y_j)=z\}} \mathsf{P}(X=x_i, Y=y_j)\\ &= \sum_{\{(x_i,y_j):g((x_i,y_j))=z\}} p_{(X,Y)}(x_i,y_j).\end{aligned}$$

Esimerkki 54. Tarkastellaan Esimerkin 53 tilannetta. Lasketaan satunnaismuuttujan $Z = X + Y$ jakauma.

$$\begin{aligned}p_Z(0) &= p_{(X,Y)}(0,0) = 0.10\\ p_Z(1) &= p_{(X,Y)}(0,1) + p_{(X,Y)}(1,0) = 0.30 + 0.20 = 0.50\\ p_Z(2) &= p_{(X,Y)}(0,2) + p_{(X,Y)}(1,1) = 0.05 + 0.25 = 0.30\\ p_Z(3) &= p_{(X,Y)}(1,2) = 0.10.\end{aligned}$$

Samaantapaan kuin yksiulotteisen satunnaismuuttujan tapauksessa, joskus on tarpeen laskea satunnaismuuttujan (X,Y) funktion $g(X,Y)$ odotusarvo. Tämä on onnistuu suoraviivaisesti ilman muunnetun satunnaismuuttujan $g(X)$ todennäköisyysjakauman laskemista kaavoilla

$$\begin{aligned}\mathsf{E}(g(X,Y)) &= \sum_{i=1}^{r}\sum_{j=1}^{s} g(x_i,y_j)p_{(X,Y)}(x_i,y_j)\\ \mathsf{E}(g(X,Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p_{(X,Y)}(x,y)\mathrm{d}x\mathrm{d}y.\end{aligned}$$

Erityisesti usein on tarpeellista laskea satunnaismuuttujien tulon odotusarvo.

Esimerkki 55. Tarkastellaan Esimerkin 53 tilannetta. Satunnaismuuttujien X ja Y tulon odotusarvo on

$$\begin{aligned}\mathsf{E}(XY) = &\quad 0\cdot 0\cdot 0.10 + 0\cdot 1\cdot 0.30 + 0\cdot 2\cdot 0.05\\ &+ 1\cdot 0\cdot 0.20 + 1\cdot 1\cdot 0.25 + 1\cdot 2\cdot 0.10 = 0.45.\end{aligned}$$

Lause 30. Olkoon X ja Y satunnaismuuttujia, ja $a, b \in \mathbb{R}$ vakioita. Satunnaismuuttujien lineaarikombinaatioille pätee

$$\mathsf{E}(aX + bY) = a\mathsf{E}(X) + b\mathsf{E}(Y).$$

Todistus. Harjoitustehtävä. □

Kahden satunnaismuuttujan yhteisjakauman määräämä, näiden satunnaismuuttujan yhteyttä kuvaava tunnusluku on **kovarianssi**.

Määritelmä 32 (Kovarianssi)**.**

$$\mathsf{Cov}(X,Y) = \mathsf{E}((X - \mathsf{E}(X))(Y - \mathsf{E}(Y))).$$

Toisin kuin varianssi, kovarianssi voi saada positiivisia tai negatiivisia arvoja. Kovarianssi mittaa kuinka paljon X:n ja Y:n arvot 'muuttuvat' yhdessä, jos todennäköistä että molemmat muuttuvat samanaikaisesti pienemmiksi ja suuremmiksi, on kyseessä **positiivinen kovarianssi**. Jos taas toisen 'muuttuessa' suuremmaksi toinen muuttuu pienemmäksi, niin kyseessä on **negatiivinen kovarianssi**.

Lause 31. Kovarianssi voidaan ilmaista muodossa

$$\mathsf{Cov}(X,Y) = \mathsf{E}(XY) - \mathsf{E}(X)\,\mathsf{E}(Y).$$

Todistus. Harjoitustehtävä. □

Lause 32. Satunnaismuuttujan X kovarianssi itsensä kanssa on satunnaismuuttujan varianssi

$$\mathsf{Cov}(X,X) = \mathsf{Var}(X)$$

Todistus.

$$\mathsf{Cov}(X,X) = \mathsf{E}((X-\mathsf{E}(X))(X-\mathsf{E}(X))) = \mathsf{E}\left((X-\mathsf{E}(X))^2\right) = \mathsf{Var}(X).$$

□

Esimerkki 56. Olkoon satunnaismuuttujien X ja Y yhteisjakauma muotoa

		Y 3	Y 4
X	1	0.2	0.4
	2	0.1	0.3.

Lasketaan X:n ja Y:n kovarianssi kahdella tapaa. Ensin sijoittamalla todennäköisyydet suoraan kovarianssin kaavan, ja sen jälkeen käyttämällä Lausetta 31. Molemmissa tavoissa tarvitaan X:n ja Y:n odotusarvoja. Lasketaan siis ensin marginaalitodennäköisyydet

		Y 3	Y 4	p_X
X	1	0.2	0.4	0.6
	2	0.1	0.3	0.4
	p_Y	0.3	0.7	

joiden avulla lasketaan X:n ja Y:n odotusarvot

$$\begin{aligned}\mathsf{E}(X) &= 1\cdot 0.6 + 2\cdot 0.4 = 1.4\\ \mathsf{E}(Y) &= 3\cdot 0.3 + 4\cdot 0.7 = 3.7.\end{aligned}$$

Kovarianssin määritelmän avulla lasketaan

$$\begin{aligned}\mathsf{Cov}(X,Y) = &\ (1-1.4)\cdot(3-3.7)\cdot 0.2 + (1-1.4)\cdot(4-3.7)\cdot 0.4\\ &+ (2-1.4)\cdot(3-3.7)\cdot 0.1 + (2-1.4)\cdot(4-3.7)\cdot 0.3 = 0.02\end{aligned}$$

Lauseen 31 lasketaan

$$\mathsf{Cov}(X,Y) = \underbrace{1\cdot 3\cdot 0.2 + 1\cdot 4\cdot 0.4 + 2\cdot 3\cdot 0.1 + 2\cdot 4\cdot 0.3}_{\mathsf{E}(XY)} - \underbrace{1.4\cdot 3.7}_{\mathsf{E}(X)\mathsf{E}(Y)} = 0.02$$

Esimerkki 57. Olkoon satunnaismuuttujilla X ja Y seuraavanlainen yhteisjakauma

		Y			
		0	1	2	p_X
	0	0.05	0.10	0.25	0.40
X	1	0.05	0.25	0.10	0.40
	2	0.15	0.05	0	0.20
	p_Y	0.25	0.40	0.35	1,

johon ollaan valmiiksi laskettu marginaalijakaumat p_X ja p_Y. Lasketaan satunnaismuuttujien odotusarvot ja varianssit

$$
\begin{aligned}
\mathsf{E}\,(X) &= 0 \cdot 0.40 + 1 \cdot 0.40 + 2 \cdot 0.20 = 0.80 \\
\mathsf{E}\,(Y) &= 0 \cdot 0.25 + 1 \cdot 0.40 + 2 \cdot 0.35 = 1.10 \\
\mathsf{Var}\,(X) &= 0^2 \cdot 0.40 + 1^2 \cdot 0.40 + 2^2 \cdot 0.20 - 0.80^2 = 0.56 \\
\mathsf{Var}\,(Y) &= 0^2 \cdot 0.25 + 1^2 \cdot 0.40 + 2^2 \cdot 0.35 - 1.10^2 = 0.59 \\
\mathsf{E}\,(XY) &= \quad 0 \cdot 0 \cdot 0.05 + 0 \cdot 1 \cdot 0.10 + 0 \cdot 2 \cdot 0.25 \\
&\quad + 1 \cdot 0 \cdot 0.05 + 1 \cdot 1 \cdot 0.25 + 1 \cdot 2 \cdot 0.10 \\
&\quad + 2 \cdot 0 \cdot 0.15 + 2 \cdot 1 \cdot 0.05 + 2 \cdot 2 \cdot 0 = 0.55.
\end{aligned}
$$

Satunnaismuuttujien X ja Y kovarianssi on

$$
\mathsf{Cov}\,(X, Y) = \mathsf{E}\,(XY) - \mathsf{E}\,(x)\,\mathsf{E}\,(Y) = 0.55 - 0.80 \cdot 1.10 = -0.33
$$

Lisäksi voidaan hakea satunnaismuuttujan funktioiden jakaumia. Lasketaan satunnaismuuttujan $Z = X + Y$ jakauma

$$
\begin{aligned}
p_Z(0) &= p_{(X,Y)}(0,0) = 0.05 \\
p_Z(1) &= p_{(X,Y)}(0,1) + p_{(X,Y)}(1,0) = 0.10 + 0.05 = 0.15 \\
p_Z(2) &= p_{(X,Y)}(0,2) + p_{(X,Y)}(1,1) + p_{(X,Y)}(2,0) = 0.65 \\
p_Z(3) &= p_{(X,Y)}(1,2) + p_{(X,Y)}(2,1) = 0.15 \\
p_Z(4) &= p_{(X,Y)}(2,2) = 0,
\end{aligned}
$$

jolle voidaan laskea odotusarvo ja varianssi tavanomaiseen tapaan

$$
\begin{aligned}
\mathsf{E}\,(Z) &= 0 \cdot 0.05 + 1 \cdot 0.15 + 2 \cdot 0.65 + 3 \cdot 0.15 + 4 \cdot 0 = 1.9 \\
\mathsf{Var}\,(Z) &= 0^2 \cdot 0.05 + 1^2 \cdot 0.15 + 2^2 \cdot 0.65 + 3^2 \cdot 0.15 + 4^2 \cdot 0 - 1.9^2 = 0.49.
\end{aligned}
$$

Kovarianssin etumerkki antaa viitettä siitä liittyvätkö kahden satunnaismuuttujan arvojen todennäköisyydet toisiinsa, mutta kovarianssin suuruudella ei ole helppoa tulkintaa tämän yhteisvaikutuksen suuruudesta. Tätä varten voidaan laskea satunnaismuuttujien korrelaatiokerroin.

Määritelmä 33 (Korrelaatiokerroin).

$$
\mathsf{Corr}\,(X, Y) = \frac{\mathsf{Cov}\,(X, Y)}{\mathsf{Sd}\,(X)\,\mathsf{Sd}\,(Y)}
$$

Korrelaatiokerroin on satunnaismuuttujien lineaarisen yhteyden mitta.

Lause 33. Korrelaatiokertoimella on ominaisuudet

1. $-1 \leq \mathsf{Corr}\,(X, Y) \leq 1$
2. Jos $Y = aX + b$, niin $\mathsf{Corr}\,(X, Y) = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$

3. Jos Corr $(X, Y) = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$, niin $Y = aX + b$.

Esimerkki 58. Tarkastellaan Esimerkin 56 tilannetta. Lasketaan X:n ja Y:n varianssit

$$\begin{aligned}\mathsf{Var}\,(X) &= \mathsf{E}\left(X^2\right) - \mathsf{E}\,(X)^2 = 1^2 \cdot 0.6 + 2^2 \cdot 0.4 - 1.4^2 = 0.24 \\ \mathsf{Var}\,(Y) &= \mathsf{E}\left(Y^2\right) - \mathsf{E}\,(Y)^2 = 3^2 \cdot 0.3 + 4^2 \cdot 0.7 - 3.7^2 = 0.21,\end{aligned}$$

joiden avulla lasketaan X:n ja Y:n korrelaatiokerroin

$$\mathsf{Corr}\,(X, Y) = \frac{\mathsf{Cov}\,(X, Y)}{\mathsf{Sd}\,(X)\,\mathsf{Sd}\,(Y)} = \frac{0.02}{\sqrt{0.24}\sqrt{0.21}} \approx 0.089.$$

Satunnaismuuttujilla X ja Y on siis hyvin pieni positiivinen korrelaatio.

Esimerkki 59. Tarkastellaan Esimerkin 57 tilannetta. Satunnaismuuttjien X ja Y korrelaatiokerroin on

$$\mathsf{Corr}\,(X, Y) = \frac{\mathsf{Cov}\,(X, Y)}{\mathsf{Sd}\,(X)\,\mathsf{Sd}\,(Y)} = \frac{-0.33}{\sqrt{0.56}\sqrt{0.59}} \approx -0.574,$$

joten X ja Y ovat negatiivisesti korreloituneita.

Lause 34. Olkoon X ja Y satunnaismuuttujia sekä a ja b vakioita. Satunnaismuuttujien lineaarikombinaatioille pätee

$$\mathsf{Var}\,(aX + bY) = a^2\mathsf{Var}\,(X) + b^2\mathsf{Var}\,(Y) + 2ab\mathsf{Cov}\,(X, Y)\,.$$

Todistus. Harjoitustehtävä. □

Esimerkki 60. Tarkastellaan Esimerkissä 57 lasketun satunnaismuuttujan $Z = X + Y$ varianssia. Suoraan Z:n jakauman avulla saatiin varianssiksi Var $(Z) = 0.49$. Samaan tulokseen päästään käyttämällä Lausetta 34

$$\begin{aligned}\mathsf{Var}\,(Z) = \mathsf{Var}\,(X + Y) &= \mathsf{Var}\,(X) + \mathsf{Var}\,(Y) + 2\mathsf{Cov}\,(X, Y) \\ &= 0.56 + 0.59 + 2 \cdot (-0.33) = 0.49.\end{aligned}$$

Lause 34 on laajennettavissa n satunnaismuuttujan summalle.

Lause 35. Olkoon $X_1, \ldots, X_n$ satunnaismuuttujia ja $a_1 \ldots, a_n$ vakioita. Satunnaismuuttujien lineaarikombinaatiolle pätee

$$\mathsf{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathsf{Var}\,(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j \mathsf{Cov}\,(X_i, X_j)$$

Todistus. Ei todisteta. □

2.5 Ehdollinen jakauma ja satunnaismuuttujien riippumattomuus

Määritelmä 34 (Ehdollinen todennäköisyysjakauma)**.** Olkoon (X, Y) diskreetti tai jatkuva satunnaismuuttuja, jolla on yhteisjakauma $p_{(X,Y)}(x, y)$ (samoin $f_{(X,Y)}(x, y)$) ja marginaalijakaumat $p_X(x)$ ja $p_Y(y)$. Satunnaismuuttujan X ehdollinen todennäköisyysjakauma ehdolla $Y = y_0$, on

$$p_{X|Y}(x|y_0) = \frac{p_{(X,Y)}(x, y_0)}{p_Y(y_0)},$$

kun marginaalijakauman arvo pisteessä y_0 on positiivinen, $p_Y(y_0) > 0$.

Kahden satunnaismuuttujan ehdolliset pistetodennäköisyysfunktiot voidaan ilmoittaa taulukkomuodossa. Kun ehdollistetaan X satunnaismuuttujan Y arvoilla $y_1, \ldots, y_n$, voimme kerätä kaikki ehdolliset jakaumat taulukkoon

		Y					
		y_1	$\cdots$	y_j	$\cdots$	y_s	p_X
	x_1	$p_{1\|1}$	$\cdots$	$p_{1\|j}$	$\cdots$	$p_{1\|s}$	$p_{1\cdot}$
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
X	x_i	$p_{i\|1}$	$\cdots$	$p_{i\|j}$	$\cdots$	$p_{i\|s}$	$p_{i\cdot}$
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_r	$p_{r\|1}$	$\cdots$	$p_{r\|j}$	$\cdots$	$p_{r\|s}$	$p_{r\cdot}$
		1	$\cdots$	1	$\cdots$	1	$1,$

jossa

$$p_{i|j} = p_{X|Y}(x_i \mid y_j) = \frac{\mathsf{P}\left(X = x_i, Y = y_j\right)}{\mathsf{P}\left(Y = y_j\right)}.$$

Taulokossa on siis j. sarakkeessa X:n pistetodennäköisyysfunktio ehdolla $Y = y_j$. Vastaavasti haetaan Y:n ehdolliset jakaumat ehdoilla $X = x_i$ taulukkoon

		Y					
		y_1	$\cdots$	y_j	$\cdots$	y_s	
	x_1	$p_{1\|1}$	$\cdots$	$p_{j\|1}$	$\cdots$	$p_{s\|1}$	1
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
X	x_i	$p_{1\|i}$	$\cdots$	$p_{j\|i}$	$\cdots$	$p_{s\|i}$	1
		$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_r	$p_{1\|r}$	$\cdots$	$p_{j\|r}$	$\cdots$	$p_{s\|r}$	1
	p_Y	$p_{1\cdot}$	$\cdots$	$p_{j\cdot}$	$\cdots$	$p_{s\cdot}$	$1,$

jossa on tällä kertaa taulukoitu arvot

$$p_{i|j} = p_{Y|X}(y_i \mid x_j) = \frac{\mathsf{P}\left(Y = y_i, X = x_j\right)}{\mathsf{P}\left(X = x_j\right)}.$$

Ehdolliselle satunnaismuuttujalle voidaan laskea odotusarvo ja varianssi tavanomaisesti käyttäen ehdollista jakaumaa. Olkoon satunnaismuuttuja X diskreetti ja sen otosavaruus $\Omega_X = \{x_1, \ldots, x_r\}$. Merkitään satunnaismuuttujan X odotusarvoa ja varianssia ehdolla $Y = y$

$$\mathsf{E}\left(X \mid Y = y\right) := \sum_{i=1}^{r} x_i p_{X|Y}(x_i \mid y)$$
$$\mathsf{Var}\left(X \mid Y = y\right) := \mathsf{E}\left((X - \mu_{X|y})^2 \mid Y = y\right)$$

Esimerkki 61. Olkoon satunnaismuuttujilla X ja Y seuraavanlainen (Esimerkki 57) yhteisjakauma

		Y			
		0	1	2	p_X
	0	0.05	0.10	0.25	0.40
X	1	0.05	0.25	0.10	0.40
	2	0.15	0.05	0	0.20
	p_Y	0.25	0.40	0.35	1,

Lasketaan ensin X:n pistetodennäköisyysjakaumat ehdoilla $Y = 0$, $Y = 1$ ja $Y = 2$. Ehdollisen todennäköisyyden Määritelmän 34 kaavaan sijoittamalla saadaan

		Y 0	1	2	p_X
	0	$\frac{0.05}{0.25}$	$\frac{0.10}{0.40}$	$\frac{0.25}{0.35}$	0.40
X	1	$\frac{0.05}{0.25}$	$\frac{0.25}{0.40}$	$\frac{0.10}{0.35}$	0.40
	2	$\frac{0.15}{0.25}$	$\frac{0.05}{0.40}$	$\frac{0}{0.35}$	0.20
		1	1	1	1,

joka sievennettynä on

		Y 0	1	2	p_X
	0	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{5}{7}$	$\frac{2}{5}$
X	1	$\frac{1}{5}$	$\frac{5}{8}$	$\frac{2}{7}$	$\frac{2}{5}$
	2	$\frac{3}{5}$	$\frac{1}{8}$	0	$\frac{1}{5}$
		1	1	1	1.

Lasketaan X:n ehdolliset odotusarvot ehdoilla $Y = 0, 1, 2$

$$\begin{aligned}
\mathsf{E}(X \mid Y = 0) &= 0 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{3}{5} = \frac{7}{5} \\
\mathsf{E}(X \mid Y = 1) &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{5}{8} + 2 \cdot \frac{1}{8} = \frac{7}{8} \\
\mathsf{E}(X \mid Y = 2) &= 0 \cdot \frac{3}{5} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{3}{5} = \frac{53}{40}.
\end{aligned}$$

Määritelmä 35 (Satunnaismuuttujien riippumattomuus). Olkoon X, Y satunnaismuuttujia joilla on yhteisjakauma $p_{(X,Y)}(x, y)$ ja on marginaalijakaumat $p_X(x)$ ja $p_Y(y)$. Satunnaismuuttujat ovat riippumattomia jos ja vain jos

$$p_{(X,Y)}(x, y) = p_X(x)p_Y(y).$$

Lause 36. Olkoon X, Y satunnaismuuttujia joilla on yhteisjakauma $p_{(X,Y)}(x, y)$ ja jolle $p_Y(y_0) > 0$. Satunnaismuuttujat ovat riippumattomia jos ja vain jos kaikilla $x \in \mathbb{R}$ ja $y_0 \in \mathbb{R}$

$$p_{(X|Y)}(x|y_0) = p_X(x).$$

Todistus. Harjoitustehtävä. □

Esimerkki 62. Olkoon satunnaismuuttujilla X ja Y yhteisjakauma

		Y 0	1	2	p_X
	0	0.02	0.12	0.06	0.20
X	1	0.03	0.18	0.09	0.30
	2	0.05	0.30	0.15	0.50
		0.10	0.60	0.30	1,

johon ollaan laskettu marginaalijakaumat p_X ja p_Y. Lasketaan $p_{X|Y}(x \mid y_0)$, kun $y_0 = 0, 1, 2$, ja saadaan

		Y			
		0	1	2	p_X
	0	0.20	0.20	0.20	0.20
X	1	0.30	0.30	0.30	0.30
	2	0.50	0.50	0.50	0.50
		1	1	1	1.

Koska $p_{X|Y}(x \mid y_0)$ ei riipu arvosta y_0 tästä seuraa, että satunnaismuuttujat X ja Y ovat riippumattomia. Tämä oltaisiin nähty myös toteamalla, että yhteisjakauman todennäköisyydet $p_{(X,Y)}(x, y)$ ovat marginaalijakaumien tuloja $p_X(x)p_Y(y)$.

Lause 37. Olkoon satunnaismuuttujat X ja Y riippumattomia, ja olkoon $\mathsf{E}(X)$ ja $\mathsf{E}(Y)$ olemassa äärellisinä. Tällöin

$$\mathsf{E}(XY) = \mathsf{E}(X)\,\mathsf{E}(Y).$$

Todistus. Todistetaan tulos diskreeteille satunnaismuuttujille X ja Y.

$$\begin{aligned}
\mathsf{E}(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j p_{(X,Y)}(x_i, y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j p_X(x_i) p_Y(y_j) \\
&= \sum_{i=1}^{n} x_i p_X(x_i) \sum_{j=1}^{n} y_j p_Y(y_j) \\
&= \mathsf{E}(X)\,\mathsf{E}(Y).
\end{aligned}$$

□

On tärkeää huomata, että Lauseen 37 tulos pätee vain toiseen suuntaan. Tämä tarkoittaa, että vaikka pätisi $\mathsf{E}(XY) = \mathsf{E}(X)\,\mathsf{E}(Y)$, niin tästä **ei** seuraa, että satunnaismuuttujat X ja Y olisivat riippumattomia.

Lause 38. Olkoon satunnaismuuttujat X ja Y riippumattomia, ja olkoon $\mathsf{E}(X)$ ja $\mathsf{E}(Y)$ olemassa äärellisinä. Tällöin

$$\mathsf{Cov}(X, Y) = 0$$

ja

$$\mathsf{Corr}(X, Y) = 0.$$

Todistus. Harjoitustehtävä. □

Myös Lause 38 pätee vain toiseen suuntaan, eli satunnaismuuttujien riippumattomuudesta seuraa, että satunnaismuuttujien kovarianssi on nolla, mutta siitä että satunnaismuuttujien kovarianssi on nolla **ei** seuraa, että satunnaismuuttujat olisivat riippumattomia.

2.6 Diskreettejä todennäköisyysjakaumia

Käytetyimmät todennäköisyysjakaumat kuuluvat johonkin tiettyyn **jakauma-perheeseen**, johon kuuluvat jakaumat määritellään tiettyjen jakauman **parametrien** avulla.

Bernoullin jakauma, binomijakauma ja geometrinen jakauma määritellään Bernoullin kokeen (dikotomisen, eli kaksiarvoisen kokeen) avulla. Bernoullin kokeessa on kaksi mahdollista toisensa poissulkevaa lopputulosta, eli voimme ajatella että kokeessa jokin onnistuu tai ei onnistu. Satunnaismuuttujaksi X kokeen tulos on koodattu siten että $X = 1$ merkitsee onnistumista ja $X = 0$ epäonnistumista.

Määritelmä 36 (Bernoullin jakauma)**.** Olkoon $p \in (0,1)$ Bernoullin kokeen onnistumisen todennäköisyys. Satunnaismuuttujan X jakauma on Bernoullin jakauma, jota merkitään $X \sim \mathsf{Bernoulli}\,(p)$, kun

$$p_X(k) = p^k(1-p)^{1-k}, \quad k \in \{0,1\}$$

Lause 39. Bernoulli-jakautuneella satunnaismuuttujalla $X \sim \mathsf{Bernoulli}\,(p)$ on odotusarvo ja varianssi

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

Todistus. Harjoitustehtävä. □

Esimerkki 63. Heitetään tasapainoista kuusisivuista noppaa kerran. Tarkastellaan Bernoullin koetta, jossa onnistumme jos nopanheitossa saamme silmäluvun, jonka arvo on korkeintaan 2. Määritellään satunnaismuuttuja

$$X = \begin{cases} 1, & \text{kun nopan silmäluku on korkeintaan 2} \\ 0, & \text{kun nopan silmäluku on yli 2.} \end{cases}$$

Koska onnistumisen todennäköisyys on $\frac{1}{3}$, on X:n todennäköisyysjakauma muotoa

$$p_X(k) = \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{1-k}, \quad k \in \{0,1\}.$$

Geometrinen jakauma on ensimmäistä onnistumiseen vaadittavien riippumattomien Bernoullin kokeiden määrän jakauma. Jos Bernoullin kokeen onnistumisen todennäköisyys on p, niin saamme ensimmäisen onnistumisen k. kokeessa todennäköisyydellä $(1-p)^{k-1}p$, jossa $(1-p)^{k-1}$ on ensimmäisten $k-1$ kokeen epäonnistumisen ja p on k. kokeen onnistumisen todennäköisyys.

Määritelmä 37 (Geometrinen jakauma)**.** Olkoon $p \in (0,1)$ Bernoullin kokeen onnistumisen todennäköisyys. Satunnaismuuttujan X jakauma on Geometrinen jakauma, jota merkitään $X \sim \mathsf{Geometric}\,(p)$, kun

$$p_X(k) = (1-p)^{k-1}p, \quad k \in \{1,2,\ldots\}$$

Lause 40. Geometrisen jakauman mukaisesti jakautuneella satunnaismuuttujalla $X \sim \mathsf{Geometric}\,(p)$ on odotusarvo ja varianssi

$$\begin{aligned} \mathsf{E}\,(X) &= \frac{1}{p} \\ \mathsf{Var}\,(X) &= \frac{1-p}{p^2} \end{aligned}$$

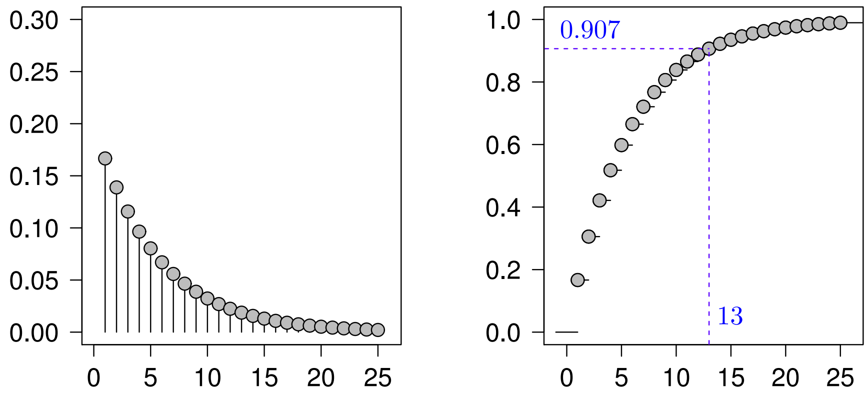
Todistus. Harjoitustehtävä. □

Esimerkki 64. Kimble-pelissä nappulan saa ottaa mukaan pelilaudalle kun saa 'nopanheitolla' silmäluvun kuusi. Oletetaan, että noppa on reilu. Olkoon X nopanheittojen määrä kunnes saadaan ensimmäinen silmäluku kuusi. Satunnaismuuttuja noudattaa todennäköisyysjakaumaa

$$p_X(k) = \left(1 - \frac{1}{6}\right)^{k-1} \frac{1}{6}, \quad k \in \{1,2,\ldots\}$$

Satunnaismuuttujan X pistetodennäköisyys- ja kertymäfunktiot ovat piirretty Kuvaa 2.9 Todennäköisyys, että tarvitaan korkeintaan 4 heittoa, jotta nappulan saa ottaa mukaan pelilaudalle on

$$\begin{aligned} \mathsf{P}\,(X \le 4) &= p_X(1) + p_X(2) + p_X(3) + p_X(4) \\ &= \left(1-\frac{1}{6}\right)^{1-1}\frac{1}{6} + \left(1-\frac{1}{6}\right)^{2-1}\frac{1}{6} + \left(1-\frac{1}{6}\right)^{3-1}\frac{1}{6} + \left(1-\frac{1}{6}\right)^{4-1}\frac{1}{6} \\ &= \frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} + \frac{5^3}{6^4} \\ &= \frac{671}{1296} \approx 0.518. \end{aligned}$$



Kuva 2.9: Esimerkin 64 pistetodennäköisyys- ja kertymäfunktiot.

Jos halutaan tietää, että kuinka monta heittoa tarvitaan, että vähintään yli 90% todennäköisyydellä saadaan silmäluku, voidaan tämä lukea kertymäfunktion pystysuoralta akselilta, ensimmäinen kertymäfunktion arvo joka on yli on $F_X(13) = 0.9065361$. Eli 13 heittoa vaaditaan. jotta yli 90% todennäköisyydellä pääsee aloittamaan pelin.

Binomijakauma on N riippumattoman Bernoullin kokeen onnistumisten määrän X todennäköisyysjakauma. Jakauman parametrit ovat Bernoullin kokeen onnistumisen todennäköisyys p, sekä riippumattomien Bernoullin kokeiden lukumäärä N.

Määritelmä 38 (Binomijakauma)**.** Olkoon $p \in (0, 1)$ Bernoullin kokeen onnistumisen todennäköisyys ja $N \in \mathrm{N}$ kokeiden lukumäärä. Satunnaismuuttujan X jakauma on binomijakauma, jota merkitään $X \sim \mathrm{Binomial}\,(N, p)$, kun

$$p_X(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k \in \{0, 1, \dots, N\}$$

Lause 41. Binomijakautuneella satunnaismuuttujalla $X \sim \mathrm{Binomial}\,(N, p)$ on odotusarvo ja varianssi

$$\begin{aligned} \mathrm{E}\,(X) &= Np \\ \mathrm{Var}\,(X) &= Np(1-p) \end{aligned}$$

Todistus. Harjoitustehtävä. □

Esimerkki 65. Yliopistoon sisäänkirjautuvan valmistumistodennäköisyys on 70%. Valitaan satunnaisesti 20 opiskelijaa syksyllä 2013 aloittavista opiskelijoista. Olkoon X valitusta kahdestakymmenestä opiskelijasta valmistuvien määrä. Satunnaismuuttuja X noudattaa todennäköisyysjakaumaa

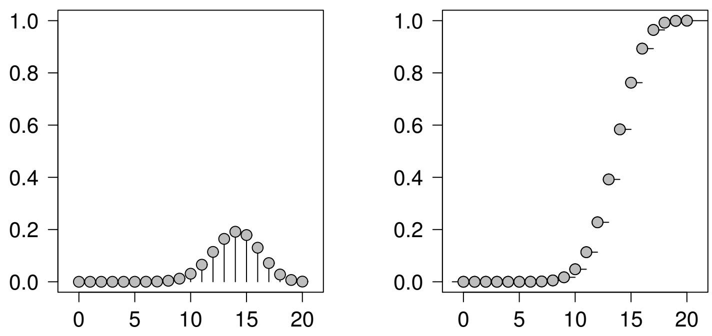
$$p_X(k) = \binom{20}{k} 0.70^k (1-0.70)^{N-k}, \quad k = 0, 1, \dots, 20$$

Todennäköisyysjakauman avulla voidaan laskea erilaisten tapahtumien todennäköisyyksiä, kuten mikä on todennäköisyys tapahtumalle kaikki valmistuvat

$$\mathrm{P}\,(X = 20) = p_X(20) = 8.0 \cdot 10^{-4}.$$

Satunnaismuuttujan X tiheysfunktio sekä kertymäfunktio on esitettynä Kuvassa 2.10.

Binomijakauman voidaan ajatella kuvaavan onnistumisten määrää yksinkertaisessa otannassa palauttaen, jossa p on onnistumisen todennäköisyys. Vastaava todennäköisyysjakauma yksinkertaiselle otannalle ilman palautusta on hypergeometrinen jakauma. Hypergeometrinen jakauma on onnistumisten määrä X n-kokoisessa poisto-otannassa N-kokoisesta populaatiosta, jossa jokainen populaation alkio tulee valituksi yhtä suurella todennäköisyydellä, ja onnistumista vastaavien suotuisien tapausten määrä on K. Jakauman parametrit ovat populaation koko N, suotuisten tapausten määrä K sekä otannan koko n.



Kuva 2.10: Esimerkin 65 pistetodennäköisyys- ja kertymäfunktiot.

Määritelmä 39 (Hypergeometrinen jakauma)**.** Olkoon $N \in \mathsf{N}$ populaation koko, $K \in \{0, \ldots, N\}$ suotuisten tapausten lukumäärä ja $n \in \{0, \ldots, N\}$ otannan koko. Satunnaismuuttujan jakauma on hypergeometrinen jakauma, jota merkitään $X \sim \mathsf{HyperGeometric}\,(N, K, n)$, kun

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \in \{0, 1, \ldots, n\}$$

Lause 42. Hypergeometrisen jakauman mukaisesti jakautuneella satunnaismuuttujalla $X \sim \mathsf{HyperGeometric}\,(N, K, n)$ on odotusarvo ja varianssi

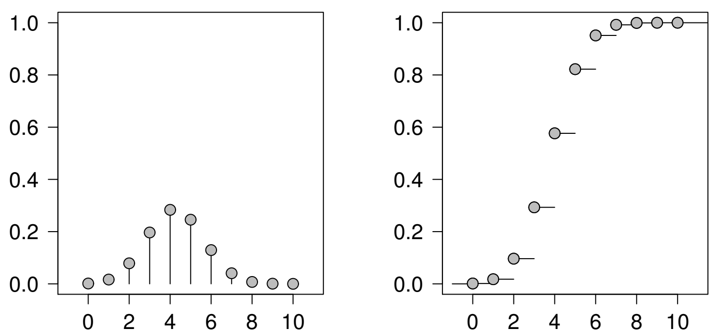
$$\mathsf{E}\,(X) = n\frac{K}{N}$$
$$\mathsf{Var}\,(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\left(1 - \frac{n-1}{N-1}\right)$$

Todistus. Harjoitustehtävä. □

Esimerkki 66. Toimistossa on 40 työntekijää, joista 17 ovat naisia. Työntekijät saavat virkistystarkoituksessa yhteensä 10 oopperalippua, jotka arvotaan kaikkien työntekijöiden kesken. Olkoon X naisten määrä jotka saavat lipun. Satunnaismuuttujalla X on pistetodennäköisyysfunktio

$$p_X(k) = \frac{\binom{17}{k}\binom{N-17}{10-k}}{\binom{N}{10}}, \quad k \in \{0, 1, \ldots, 10\},$$

jota on havainnollistettu Kuvassa 2.11.



Kuva 2.11: Esimerkin 66 pistetodennäköisyys- ja kertymäfunktiot.

Todennäköisyysjakaumasta voidaan laskea todennäköisyyksia tapahtumille kuten vähintään kaksi naista saa lipun

$$\mathsf{P}(X \geq 2) = 1 - \mathsf{P}(X < 2) = 1 - (p_X(0) + p_X(1)) = 0.982.$$

Jos tarkastellaan binomijakaumaa, jossa kokeiden määrä N lähestyy ääretöntä, mutta onnistumisen todennäköisyys p pienenee vastaavasti siten, että tulo $\lambda = Np$ pysyy vakiona, niin onnistumisten määrä lähestyy Poisson-jakaumaa parametrilla λ. Parametria λ voidaan ajatella keskimääräisenä onnistumisten määränä. Tätä on yksinkertaisinta tarkastella esimerkin kautta. Esimerkkinä tutkitaan pitkää suprajohdekerää, joka on täysin samanlaista koko pituudeltaan. Metallihiukkaset johteen pinnassa huonontavat sen ominaisuuksia, ja näiden määrä on suoraan suhteessa johteen laatuun. Ajatellaan, että odotusarvoisesti metrin (m) pituisella pätkällä on $\lambda = Np$ metallihiukkasta (mh), eikä täsmälleen samassa kohtaa voi olla useampaa metallihiukkasta. Tarkastellaan 1cm pituista osaa m-pituisesta johteesta, ja merkitään mh määrää m-pituisella osalla satunnaismuuttujalla X. Tämä voidaan kirjoittaa jakamalla osiin

$$\lambda \left[\frac{\text{mh}}{\text{m}}\right] = \underbrace{100}_{N} \left[\frac{\text{cm}}{\text{m}}\right] \underbrace{\frac{\lambda}{100}}_{p} \left[\frac{\text{mh}}{\text{cm}}\right].$$

Jos mallinnetaan mh/m määrää binomijakaumalla, voidaan todennäköisyysjakauma ilmaista muodossa

$$\mathsf{P}(X = k) = \binom{100}{k} \left(\frac{\lambda}{100}\right)^k \left(1 - \left(\frac{\lambda}{100}\right)\right)^{100-k}.$$

Jos tihennetään tarkasteluvälejä edelleen mm-pituisiksi, voidaan kirjoittaa

$$\lambda \left[\frac{\text{mh}}{\text{m}}\right] = \underbrace{1000}_{N} \left[\frac{\text{mm}}{\text{m}}\right] \underbrace{\frac{\lambda}{1000}}_{p} \left[\frac{\text{mh}}{\text{mm}}\right],$$

jota vastaa binomijakauma

$$\mathsf{P}(X = k) = \binom{1000}{k} \left(\frac{\lambda}{1000}\right)^k \left(1 - \left(\frac{\lambda}{1000}\right)\right)^{1000-k}.$$

Jos nyt pidetään metrin pituisen osan mh määrän odotusarvo vakiona ja osatarkasteluvälien määrää kasvatetaan, niin X:n jakauma lähenee Poisson-jakaumaa.

Määritelmä 40 (Poisson-jakauma)**.** Olkoon $\lambda > 0$. Satunnaismuuttujan X jakauma on Poisson-jakauma, jota merkitään $X \sim \mathsf{Poisson}(\lambda)$, kun

$$p_X(k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad k \in \mathsf{N}.$$

Poissonin-jakauman tiheysfunktiosta löytyvä eksponenttifunktio $\exp(x) = e^x$ on Neperin luvun ($e \approx 2.718282$) x. potenssi.

Lause 43. Poisson-jakautuneella satunnaismuuttujalla $X \sim \mathsf{Poisson}(\lambda)$ on odotusarvo ja varianssi

$$\mathsf{E}(X) = \lambda$$
$$\mathsf{Var}(X) = \lambda.$$

Todistus. Harjoitustehtävä. □

Esimerkki 67. Konekirjoittaja lyö satunnaisesti virhelyöntejä kirjoittaessaan dokumenttia. Oletetaan, että keskimäärin hän lyö 4 virhelyöntiä tunnissa. Ollaan kiinnostuneita seuraavan puolen tunnin sisällä tapahtuvista virhelyönneistä. Määritellään satunnaismuuttuja X =virhelyöntien määrä puolessa tunnissa. Koska tunnissa keskimäärin oli 4 virhelyöntiä, niin puolessa tunnissa virhelyöntejä on keskimäärin $\lambda = 2$. Todennäköisyys, että konekirjoittaja lyö yli 3 virhelyöntiä puolessa tunnissa on

$$\begin{aligned}
\mathsf{P}(X > 3) &= 1 - \mathsf{P}(X \leq 3) = 1 - (\mathsf{P}(X = 0) + \mathsf{P}(X = 1) + \mathsf{P}(X = 2) + \mathsf{P}(X = 3)) \\
&= 1 - \left(\frac{2^0}{0!}\exp(-2) + \frac{2^1}{1!}\exp(-2) + \frac{2^2}{2!}\exp(-2) + \frac{2^3}{3!}\exp(-2)\right) \\
&= 1 - \exp(-2)\left(1 + 2 + \frac{4}{2} + \frac{8}{6}\right) \\
&\approx 0.143.
\end{aligned}$$

Binomijakaumaa voidaan ajatella N riippumattoman kokeen lopputuloksena, kun kokeella on kaksi mahdollista lopputulosta ja tarkastelemme montako kertaa toinen lopputuloksista esiintyy. Jos tilanne yleistetään m mahdolliseen lopputulokseen ja tarkastellaan mikä on eri tuloskombinaatioiden todennäköisyys, niin päädymme tarkastelemaan multinomijakaumaa

Määritelmä 41 (Multinomijakauma). Olkoon $N \in \mathsf{N}$ ja $p_1, p_2, \ldots, p_m \in (0, 1)$ siten, että $p_1 + p_2 + \cdots + p_m = 1$. Satunnaisvektorin $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ jakauma on multinomijakauma, jota merkitään $\mathbf{X} \sim \text{Multinomial}\,(N, p_1, p_2, \ldots, p_m)$, kun

$$p_{\mathbf{X}}(k_1, k_2, \ldots, k_m) = \binom{N}{k_1, k_2, \ldots, k_m} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$
$$k_1, k_2, \ldots, k_m \in \mathsf{N}$$
$$N = k_1 + k_2 + \cdots + k_m$$

Lause 44. Olkoon $\mathbf{X} = (X_1, \ldots, X_m)$ multinomijakautunut satunnaisvektori $\mathbf{X} \sim \text{Multinomial}\,(N, p_1, p_2, \ldots, p_m)$. Satunnaismuuttujilla X_i on odotusarvo, varianssi ja kovarianssi

$$\mathsf{E}\,(X_i) = Np_i$$
$$\mathsf{Var}\,(X_i) = Np_i(1 - p_i)$$
$$\mathsf{Cov}\,(X_i, X_j) = -Np_ip_j, \quad i \neq j.$$

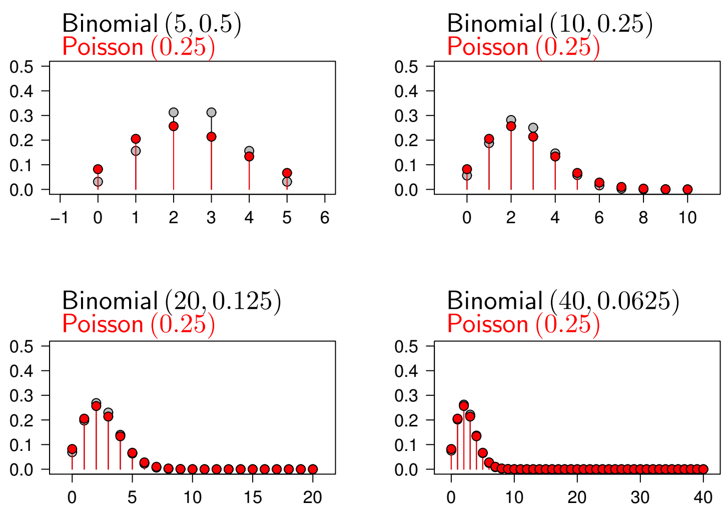
Todistus. Harjoitustehtävä. □

2.6.1 Binomijakauman approksimointi Poisson-jakaumalla

Poisson jakaumaa voidaan myös käyttää approksimoimaan binomijakaumaa. Tämä on hyödyllistä kun Bernoullin kokeiden määrä N on suuri ja onnistumisen todennäköisyyde on p on pieni, sillä tällöin binomijakauman binomikertoimien laskenta saattaa olla hankalaa. Poisson-jakauma approksimaatio on muotoa

$$\text{Binomial}\,(N, p) \approx \text{Poisson}\,(Np)\,.$$

Binomijakaumia ja vastaavia Poisson-jakauma-approksimaatioita on havainnollistettu Kuvassa 2.12.



Kuva 2.12: Binomijakauman approksimaatio Poisson-jakaumalla.

Esimerkki 68. Teollisuuskoneita valmistava yritys on saanut hyvin suuren lähetyksen koneiden valmistuksessa tarvittavia tietokoneprosessoreita, ja yritys haluaa tarkastella lähetyksen laatua. Oletetaan että joka tuhannes prosessori on viallinen. Yritys testaa prosessoreista 800 kappaletta, ja jos yli kolme viallista prosessoria löytyy, täytyy jokainen prosessori testata ennen koneeseen asentamista. Todennäköisyys, että kaikki prosessorit täytyy testata on

$$\mathsf{P}(X > 3) = 1 - \mathsf{P}(X \le 3) = 1 - \sum_{k=0}^{3} \binom{800}{k} \left(\frac{1}{1000}\right)^k \left(\frac{999}{1000}\right)^{800-k} = 0.009037691\ldots$$

Poisson approksimaatio ($\lambda = Np = \frac{800}{1000} = 0.8$) todennäköisyydelle on

$$\mathsf{P}(X > 3) = 1 - \exp(-0.8)\left(\frac{0.8^0}{0!} + \frac{0.8^1}{1!} + \frac{0.8^2}{2!} + \frac{0.8^3}{3!}\right) = 0.009079858.$$

2.7 Jatkuvia todennäköisyysjakaumia

Samoin kuin tiettyjä diskreettejä satunnaismuuttujia voidaan määritellä parametristen jakaumien avulla, myös jatkuville satunnaismuuttujille on olemassa tiettyjä parametrisoituja standardijakaumia. Yksinkertaisin jatkuva parametrinen jakauma on tasajakauma, joka asettaa todennäköisyyden tasaisesti reaalilukuvälille, jonka määrittelevinä parametreina on välin päätepisteet.

Määritelmä 42 (Tasajakauma)**.** Olkoon a ja b vakioita, joille $a < b$. Satunnaismuuttujan X jakauma on tasajakauma, jota merkitään $X \sim \mathsf{Uni}[a,b]$, kun

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

Lause 45. Tasajakautuneella satunnaismuuttujalla $X \sim \mathsf{Uni}[a,b]$ on odotusarvo ja varianssi

$$\mathsf{E}(X) = \frac{1}{2}(a+b)$$
$$\mathsf{Var}(X) = \frac{1}{12}(b-a)^2.$$

Todistus. Harjoitustehtävä. □

Esimerkki 69. Olkoon X tasajakautunut satunnaismuuttuja $X \sim \mathsf{Uni}[0,1]$. Satunnaismuuttujan tiheysfunktio on muotoa

$$f_X(x) = 1, \quad 0 \le x \le 1.$$

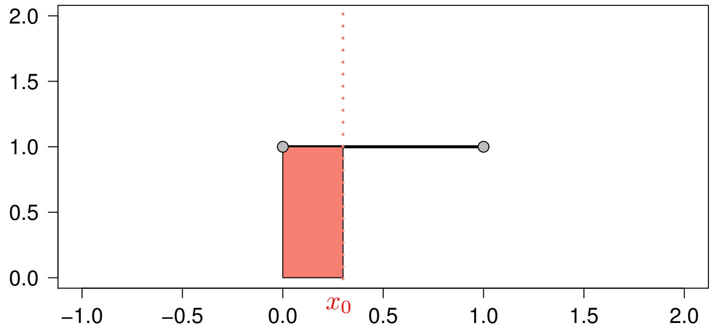
Satunnaismuuttujan kertymäfunktion arvon pisteessä x_0 voidaan ajatella olevan pisteen vasemmalle puolelle jäävän todennäköisyyden määrä. Jos nyt tarkastellaan paljonko todennäköisyysmassaa jää tiheysfunktiosta f_X pisteen x vasemmalle puolelle, kuten Kuvassa 2.13, saadaan kertymäfunktioksi

$$F_X(x) = \begin{cases} 0, & \text{kun } x \le 0 \\ x, & \text{kun } 0 < x < 1 \\ 1, & \text{kun } x \ge 1. \end{cases}$$

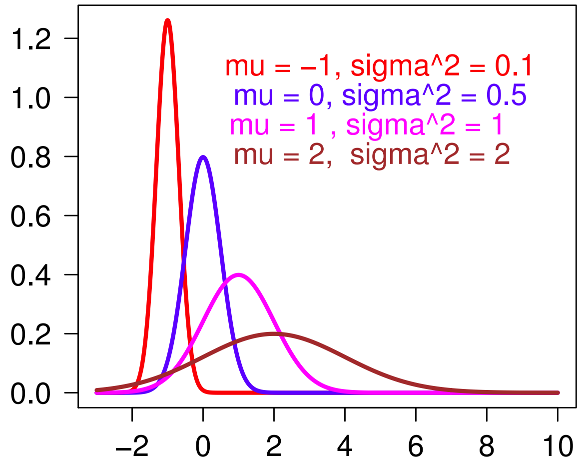
Normaalijakaumaa, eli Gaussin jakaumaa, voidaan pitää tietyssä mielessä tilastotieteen tärkeimpänä todennäköisyysjakaumana. Hyvin monet tilastotieteen menetelmät perustuvat oletukseen normaalijakautuneisuudesta, ja usein tämä oletus on myös perusteltavissa.

Määritelmä 43 (Normaalijakauma)**.** Olkoon μ ja $\sigma > 0$ vakioita. Satunnaismuuttujan X jakauma on normaalijakauma, jota merkitään $X \sim \mathsf{Normal}\left(\mu, \sigma^2\right)$, kun

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



Kuva 2.13: Esimerkin 69 tilanne.



Kuva 2.14: Normaalijakaumia muutamilla parametrien μ ja σ^2 arvoilla.

Lause 46. Normaalijakautuneella satunnaismuuttujalla $X \sim \mathsf{Normal}\left(\mu, \sigma^2\right)$ on odotusarvo ja varianssi

$$\mathsf{E}\left(X\right) = \mu$$
$$\mathsf{Var}\left(X\right) = \sigma^2.$$

Normaalijakauman määrittelevät yksikäsitteisesti sen parametrit odotusarvo ja varianssi. Normaalijakaumia eri parametrien arvoilla on havainnollistettu Kuvassa 2.14. Koska normaalijakauma määräytyy odotusarvonsa ja varianssinsa perusteella, niin jokaisen normaalijakautuneen satunnaismuuttujan Z-muunnos (Määritelmä 23) noudattaa myös normaalijakaumaa, jota kutsutaan **standardinormaalijakaumaksi**.

Määritelmä 44 (Standardinormaalijakauma)**.** Normaalijakaumaa $\mathsf{Normal}\left(\mu, \sigma^2\right)$, jolla $\mu = 0$ ja $\sigma^2 = 1$, kutsutaan standardinormaalijakaumaksi.

Normaalijakauman kertymäfunktio on muotoa

$$F(x) = \mathsf{P}\left(X \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(u-\mu)^2\right) \mathrm{d}u,$$

jonka laskemiseksi tarvitaan numeerisia menetelmiä. Perinteisesti kertymäfunktion arvon likimääräisen arvon hakemiseksi on käytetty standardinormaalijakauman kertymäfunktion

$$\Phi(z) := \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u$$

taulukoituja arvoja, jotka löytyvät luentomonisteen liittestä. Tämä on mahdollista, sillä jokaisen normaalijakautuneen satunnaismuuttujan X kertymäfunktion arvo vastaa tiettyä standardinormaalijakautuneen satunnaismuuttujan Z kertymäfunktion arvoa

$$\mathsf{P}\left(X \leq x\right) = \mathsf{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \mathsf{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Taulukosta löytyy kertymäfunktion $\Phi(z)$ arvot positiivisille z arvoille. Tämä on riittävää sillä normaalijakauma on odotusarvon suhteen symmetrinen jakauma.

Lause 47. Olkoon $\Phi(z)$ standardinormaalijakauman kertymäfunktio. Funktiolla on ominaisuudet

1. $\Phi(0) = 0.5$

2. $\Phi(z) > 0.5 \Leftrightarrow z > 0$

3. $\Phi(-z) = 1 - \Phi(z)$

Todistus. Harjoitustehtävä. □

Lauseen 47 tuloksia hyödyntäen on mahdollista arvioida tarpeelliset normaalijakauman kertymäfunktion arvot taulukoitujen arvojen avulla.

Esimerkki 70. Olkoon $X \sim$ Normal $\left(\mu, \sigma^2\right)$, ja F_X kertymäfunktio. Tällöin voimme laskea

$$\mathsf{P}\left(a < X \leq b\right) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Esimerkki 71. Olkoon $X \sim$ Normal $\left(50, 6^2\right)$. Nyt

$$\begin{aligned}
\mathsf{P}\left(X \leq 62\right) &= \mathsf{P}\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathsf{P}\left(Z \leq \frac{62-50}{\sqrt{36}}\right) = \mathsf{P}\left(Z \leq 2\right) = \Phi(2) = 0.9772 \\
\mathsf{P}\left(X > 57.2\right) &= \mathsf{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma}\right) \\
&= \mathsf{P}\left(Z > 1.2\right) = 1 - \Phi(1.2) = 0.1151 \\
\mathsf{P}\left(47 < X \leq 58\right) &= \mathsf{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \Phi(1.33) - \Phi(-0.50) \\
&= \Phi(1.33) - (1 - \Phi(0.50)) = 0.5997.
\end{aligned}$$

Taulukoituja kertymäfunktion arvoja voidaan käyttää myös sellaisten pisteiden hakemiseen, jotka rajaavat todennäköisyysjakaumasta tietyn osan. Nämä saadaan luettua kertymäfunktion käänteisfunktion arvoista etsimällä taulukosta lähinnä todennäköisyyttä p oleva arvo.

Esimerkki 72. Haetaan 0.9750-kvantiili taulukosta

$$\begin{aligned}
& \Phi(z) = 0.9750 \\
\Leftrightarrow \quad & z = \Phi^{-1}(0.9750) = 0.196.
\end{aligned}$$

Lause 48. Olkoon $\Phi(z)$ standardinormaalijakauman kertymäfunktio. Nyt p-kvantiilille sekä $(1-p)$-kvantiilille pätee

$$\Phi^{-1}(p) = -\Phi^{-1}(1-p).$$

Todistus. Merkitään p-kvantiilia z_p, nyt pätee

$$\begin{aligned} & z_p = \Phi^{-1}(p) \\ \Leftrightarrow\quad & \Phi(z_p) = p \\ \Leftrightarrow\quad & 1 - \Phi(-z_p) = p \\ \Leftrightarrow\quad & \Phi(-z_p) = 1 - p \\ \Leftrightarrow\quad & z_p = -\Phi^{-1}(1-p) \end{aligned}$$

□

Esimerkki 73. Erään mp3-soittimen keskimääräinen soittoaika ennen ensimmäistä vikaa on 100000min keskihajonnan ollessa 10000min. Valmistaja korjaa soittimet, jotka menevät rikki takuuaikana. Jos valmistaja varautuu korjaamaan 2% valmistetuista soittimista, niin kuinka pitkää takuuaikaa tämä vastaa? Voidaan olettaa ensimmäistä vikaa edeltävän soittoajan noudattavan normaalijakaumaa odotusarvolla 100000 ja keskihajonnalla 10000.

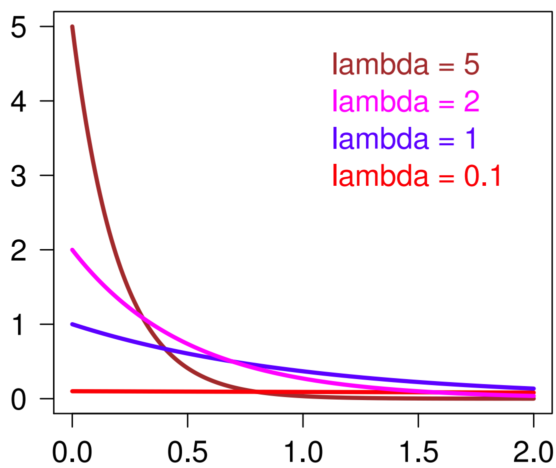
Olkoon X ensimmäistä vikaa edeltävä soittoaika minuuteissa, ja x takuuaika. Tällöin haluamme ratkaista takuuajan yhtälöstä

$$\begin{aligned} & \mathsf{P}\,(X \leq x) = 0.02 \\ \Leftrightarrow\quad & \mathsf{P}\left(Z \leq \frac{x - 100000}{10000}\right) = \Phi\left(\frac{x-100000}{10000}\right) = 0.02 \\ \Leftrightarrow\quad & \frac{x-100000}{10000} = \Phi^{-1}(0.02) = -\Phi^{-1}(0.98) \\ \Leftrightarrow\quad & x = 100000 - 10000 \cdot \Phi^{-1}(0.98) \end{aligned}$$

Poisson-jakaumalla kuvattiin tapahtumia, jotka esiintyvät prosessissa jatkuvasti, toisistaan riippumattomasti vakiolla esiintymistaajuudella λ. Poisson-prosessin peräkkäisten tapahtumien esiintymisajan eroa voidaan mallintaa eksponentiaalijakaumalla, joita on piirretty Kuvaan 2.15

Määritelmä 45 (Eksponentiaalijakauma)**.** Olkoon $\lambda > 0$ vakio. Satunnaismuuttujan X jakauma on eksponentiaalijakauma, jota merkitään $X \sim \mathsf{Exp}\,(\lambda)$, kun

$$f_X(x) = \mathsf{Exp}\,(\lambda) = \lambda \exp(-\lambda x), \quad x \geq 0$$



Kuva 2.15: Eksponentiaalifunktioita parametrin λ eri arvoilla.

Lause 49. Eksponentiaalijakautuneella satunnaismuuttujalla $X \sim \mathsf{Exp}(\lambda)$ on odotusarvo ja varianssi

$$\mathsf{E}(X) = \frac{1}{\lambda}$$
$$\mathsf{Var}(X) = \frac{1}{\lambda^2}$$

Todistus. Harjoitustehtävä. □

Lause 50. Eksponentiaalijakauman $\mathsf{Exp}(x)$ kertymäfunktio on muotoa

$$F_x(x) = 1 - \exp(-\lambda x).$$

Todistus. Harjoitustehtävä. □

2.7.1 Binomijakauman approksimointi normaalijakaumalla

Normaalijakaumaa käytetään joskus approksimoimaan muita jakaumia. Usein syynä normaaliapproksimaation käyttöön ovat laskennalliset syyt, esimerkiksi diskreettien jakaumien kertymäfunktiot saattavat olla hyvin raskaita laskettavia.

Yksinkertaisimmillaan satunnaismuuttujan X todennäköisyysjakaumaa voidaan approksimoida normaalijakaumalla $\mathsf{Normal}(\mathsf{E}(X), \mathsf{Var}(X))$. Esimerkiksi joskus normaalijakaumaa käytetään approksimoimaan binomijakaumaa. Binomijakautunut satunnaismuuttuja X on N riippumattoman ja identtisen Bernoullin kokeen X_i onnistumisten summa, jolla on odotusarvo Np ja varianssi $Np(1-p)$ Voimme siis arvioida binomijakautunutta satunnaismuuttujaa

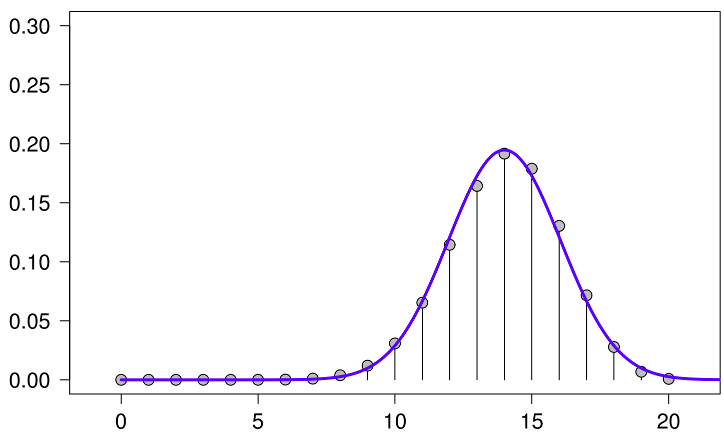
$$X \dot{\sim} \mathsf{Normal}(Np, Np(1-p)).$$

Binomijakauman normaaliapproksimaatio on sitä parempi mitä suurempi on kokeiden lukumäärä N ja mitä lähempänä parametrin p arvo on lukua 0.5. Konservatiivisena sääntönä pidetään joskus, että normaaliapproksimaatio on hyvä kun $\min(Np, N(1-p)) \geq 5$.

Esimerkki 74. Tarkastellaan Esimerkin 65 tilannetta. Satunnaismuuttuja X on binomijakautunut parametrein $N = 20, p = 0.70$. Normaaliapproksimaatio satunnaismuuttujalle on

$$X \dot{\sim} \mathsf{Normal}(20 \cdot 0.70, 20 \cdot 0.7 \cdot 0.3).$$

Approksimaatiota on havainnollistettu Kuvassa 2.16. Nyt esimerkiksi tapahtuman $X \leq 14$ todennäköisyys



Kuva 2.16: Esimerkin 74 tilanne.

on tarkasti

$$\mathsf{P}\left(X \leq 14\right) = 0.5836292.$$

Ja normaaliapproksimaatiolla laskettu likiarvo on

$$\mathsf{P}\left(Z \leq \frac{14 - Np}{\sqrt{Np(1-p)}}\right) = \Phi(0) = 0.50.$$

Kun käytämme jatkuvaa normaalijakaumaa approksimoimaan diskreettiä jakaumaa, kuten binomijakaumaa, niin saamme tarkempia tuloksia kun käytämme **jatkuvuuskorjausta**. Jatkuvuuskorjaus tehdään muuttamalla tapahtuman rajoittaviksi väleiksi reaalilukuväli, jolla olevat reaaliluvut pyöristettäisiin diskreetin tapahtuman rajoiksi. Jatkuvuuskorjaukset löytyvät Taulukosta 2.1.

Esimerkki 75. Tarkastellaan Esimerkin 74 tilannetta. Tehdään nyt jatkuvuuskorjaus normaaliapproksimaatioon. Eli nyt tapahtumaa $\mathsf{P}\left(X \leq 14\right)$ vastaa approksimaatio (taulukosta luettuna)

$$\mathsf{P}\left(Z \leq \frac{14.5 - Np}{\sqrt{Np(1-p)}}\right) = 0.5948.$$

$X \sim$ Binomial (N, p)	$X \sim$ Normal $(Np, Np(1-p))$
$\mathsf{P}\left(X \leq 10\right)$	$\mathsf{P}\left(X < 10.5\right)$
$\mathsf{P}\left(X < 10\right)$	$\mathsf{P}\left(X < 9.5\right)$
$\mathsf{P}\left(X \geq 10\right)$	$\mathsf{P}\left(X \geq 9.5\right)$
$\mathsf{P}\left(X > 10\right)$	$\mathsf{P}\left(X \geq 10.5\right)$

Taulukko 2.1: Esimerkkejä jatkuvuuskorjauksen käytöstä binomijakauman approksimoimisessa normaalijakauman avulla.

2.8 Normaalijakaumasta johdettuja todennäköisyysjakaumia

Normaalijakaumasta voidaan johtaa muutamia todennäköisyysjakaumia, joilla on erityisen tärkeä rooli klassisessa tilastollisessa päättelyssä.

2.8.1 Khiin neliö-jakauma

Tarkastellaan toisistaan riippumattomia satunnaismuuttujia $X_1, \ldots, X_n$, joilla kaikilla on todennäköisyysjakauma Normal $\left(\mu, \sigma^2\right)$. Z-muunnettujen satunnaismuuttujien neliösummalle pätee

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n),$$

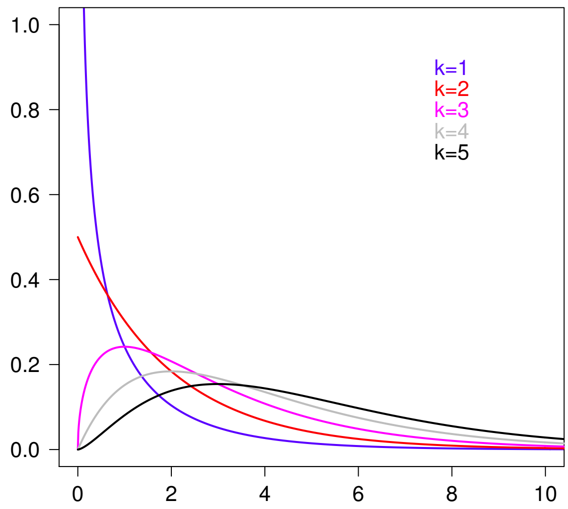
jossa $\chi^2(n)$ on parametrinen jakauma, jota kutsutaan **khiin neliö-jakaumaksi vapausastein** n. Khiin neliöjakaumia eri vapausastein on piirrettynä Kuvaan 2.17.

Määritelmä 46 (Khiin neliö-jakauma). Olkoon $Z_1, \ldots, Z_k$ ovat standardinormaalijakautuneita riippumattomia satunnaismuuttujia. Näiden neliösummalla on khiin neliö-jakauma ($\chi^2(k)$-jakauma) vapausastein $k > 0$

$$Q = Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi^2(k),$$

jossa $\chi^2(k)$-jakauman tiheysfunktio on

$$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k-2}{2}} \exp\left(-\frac{x}{2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}$$



Kuva 2.17: Khiin neliö-jakaumia eri vapausastein k.

Khiin neliö-jakauman tiheysfunktiosta löytyy usein todennäköisyysjakaumissa esiintyvä gamma-funktio

$$\Gamma(t) = \int_0^\infty x^{t-1} \exp(-x) \mathrm{d}x,$$

joka on tietyssä mielessä kertoman $n!$ yleistys kokonaislukujen ulkopuolelle ominaisuuksiensa $\Gamma(n+1) = n\Gamma(n)$ ja $\Gamma(1) = 1$ perusteella. Ominaisuuksista seuraa $\Gamma(n+1) = n!$, kun n on kokonaisluku.

Lause 51. Olkoon X khiin neliö-jakautunut satunnaismuuttuja $X \sim \chi^2(k)$. Tällöin

$$\begin{aligned} \mathsf{E}\,(X) &= k \\ \mathsf{Var}\,(X) &= 2k \end{aligned}$$

2.8.2 Studentin t-jakauma

Olkoon X_1 normaalijakautunut satunnaismuuttuja $X \sim \mathsf{Normal}\left(\mu, \sigma^2\right)$, ja X_2 khiin neliö-jakautunut satunnaismuuttuja vapausastein $k > 0$, eli $X_2 \sim \chi(k)$. Oletetaan, että X_1 ja X_2 ovat toisistaan riippumattomat. Tällöin satunnaismuuttujien suhteelle pätee

$$T = \frac{X_1}{\sqrt{X_2/k}} \sim \mathsf{t}(k),$$

jossa $\mathsf{t}(n)$ on Studentin t-jakauma vapausastein k.

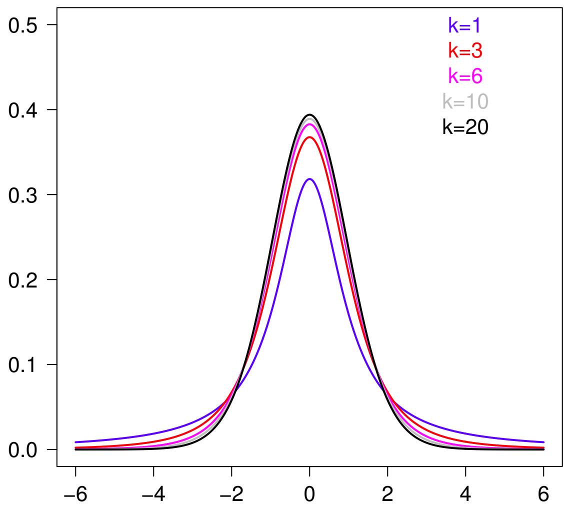
Studentin t-jakaumia eri vapausasteilla on piirretty Kuvaan 2.18. Kun Studentin t-jakauman vapausasteet kasvavat, lähestyy t-jakauman muoto normaalijakauman muotoa, mutta etenkin pienillä vapausasteilla t-jakaumalla on paksut hännät verrattuna normaalijakaumaan.

Määritelmä 47 (t-jakauma)**.** Olkoon X_1 ja X_2 toisistaan riippumattomia satunnaismuuttujia

$$\begin{aligned} X_1 &\sim \mathsf{Normal}\,(0, 1) \\ X_2 &\sim \chi^2(k). \end{aligned}$$

Tällöin satunnaismuuttujalla T on Studentin t-jakauma vapausastein $k > 0$, $T \sim \mathsf{t}(k)$ kun

$$T = \frac{X_1}{\sqrt{X_2/k}}.$$



Kuva 2.18: Studentin t-jakaumia eri vapausastein k.

Jakauman $\mathsf{t}(k)$ tiheysfunktio on

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}.$$

Lause 52. Olkoon X t-jakautunut satunnaismuuttuja $X \sim \mathsf{t}(k)$. Tällöin

$$\mathsf{E}\left(X\right) = 0, \quad k > 0$$
$$\mathsf{Var}\left(X\right) = \frac{k}{k-2}, \quad k \geq 2$$

2.8.3 F-jakauma

Tarkastellaan kahden khiin neliö-jakautuneen satunnaismuuttujan $X_1 \sim \chi^2(k_1)$ ja $X_2 \sim \chi^2(k_2)$ suhdetta. Nyt voidaan osoittaa, että satunnaismuuttujien suhteen

$$F = \frac{X_1/k_1}{X_2/k_2} = \frac{k_2}{k_1}\frac{X_1}{X_2}$$

jakauma on parametrista muotoa oleva F-jakauma vapausastein k_1 ja k_2. F-jakaumia eri vapausastein on piirretty Kuvaan 2.19 Jos satunnaismuuttujien X_1 ja X_2 jakaumilla on yhtä monta vapausastetta, niin satunnaismuuttuja supistuu muotoon

$$F_1 = \frac{X_1}{X_2},$$

jolla on sama F-jakauma kuin satunnaismuuttujalla

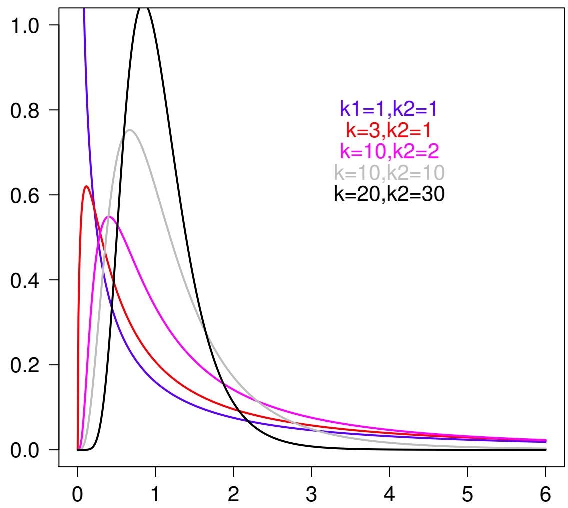
$$F_2 = \frac{X_2}{X_1}.$$

Määritelmä 48 (F-jakauma). Olkoon X_1 ja X_2 toisistaan riippumattomia satunnaismuuttujia

$$X_1 \sim \chi^2(k_1)$$
$$X_2 \sim \chi^2(k_2).$$

Tällöin satunnaismuuttujalla F on F-jakauma vapausastein $k_1 > 0$ ja $k_2 > 0$, $F \sim \mathsf{F}(k_1, k_2)$ kun

$$F = \frac{X_1/k_1}{X_2/k_2}.$$



Kuva 2.19: F-jakaumia eri vapausastein k_1 ja k_2.

Jakauman $\mathsf{F}(k_1, k_2)$ tiheysfunktio on

$$f(x) = \begin{cases} \frac{1}{B(\frac{k_1}{2}, \frac{k_2}{2})} \left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}} x^{\frac{k_1}{2}-1} \left(1 + \frac{k_1}{k_2} x\right)^{-\frac{k_1+k_2}{2}}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

F-jakauman määrittelyssä esiintyy beta-funktio

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Lause 53. Olkoon X F-jakautunut satunnaismuuttuja $X \sim \mathsf{F}(k_1, k_2)$. Tällöin

$$\mathsf{E}\,(X) = \frac{k_2}{k_2 - 2}, \quad k_2 > 2$$
$$\mathsf{Var}\,(X) = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}, \quad k_2 > 4$$

Lähteet

[1] T. W. Anderson, *An Introduction to the Statistical Analysis of Data*, Houghton Mifflin Company, 1978

[2] R. B. Ash, *Basic Probability Theory*, Dover Publications, Inc., Mineola, New York

[3] M. Grönroos, *Johdatus tilastotieteeseen: Kuvailu, mallit ja päättely*, Oy Finn Lectura Ab, 2003, Helsinki.

[4] G. Casella, R. L. Berger, *Statistical Inference*, 2ed.

[5] G. J. Kerns, *Introduction to Probability and Statistics Using R*, 2010.

[6] I. Mellin, *Todennäköisyyslaskenta: Todennäköisyys ja sen laskusäännöt*, Luentomoniste, TKK.

[7] H. Ruskeepää, *Todennäköisyyslaskenta I*, Luentomoniste, TY.

[8] B. Shahbaba, *Biostatistics with R: An Introduction to Statistics Through Biological Data*, Springer 2012.

[9] A. N. Shiryaev, *Probability*, Springer 1995.

Liite A

TAULUKOITA

Standardoidun normaalijakauman kertymäfunktion Φ arvoja.

$\mathsf{P}\,(Z \le z) = \Phi(z)\,,\quad \Phi(-z) = 1 - \Phi(z),\quad \Phi^{-1}(p) = -\Phi^{-1}(1-p)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Jakaumien $\mathtt{t}(k)$ p-yläkvantiileja $t_p^{(k)}$ eri vapausasteilla k. $P(T \geq t_p^{(k)}) = p$, kun $T \sim \mathtt{t}(k)$

	p								
k	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
31	0.682	1.309	1.696	2.040	2.453	2.744	3.022	3.375	3.633
32	0.682	1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33	0.682	1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34	0.682	1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601
36	0.681	1.306	1.688	2.028	2.434	2.719	2.990	3.333	3.582
38	0.681	1.304	1.686	2.024	2.429	2.712	2.980	3.319	3.566
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
45	0.680	1.301	1.679	2.014	2.412	2.690	2.952	3.281	3.520
50	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
55	0.679	1.297	1.673	2.004	2.396	2.668	2.925	3.245	3.476
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
150	0.676	1.287	1.655	1.976	2.351	2.609	2.849	3.145	3.357
200	0.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340
300	0.675	1.284	1.650	1.968	2.339	2.592	2.828	3.118	3.323
500	0.675	1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310
∞	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Khin-neliö(k)-jakaumien p-yläkvantiileja $q_p^{(k)}$ eri vapausasteilla k.

$$P(Q \geq q_p^{(k)}) = p\ , \quad \text{kun} \ \ Q \sim \chi^2(k)$$

	p								
k	0.99	0.975	0.95	0.1	0.05	0.025	0.01	0.005	0.001
1	0.000	0.001	0.004	2.706	3.841	5.024	6.635	7.879	10.828
2	0.020	0.051	0.103	4.605	5.991	7.378	9.210	10.597	13.816
3	0.115	0.216	0.352	6.251	7.815	9.348	11.345	12.838	16.266
4	0.297	0.484	0.711	7.779	9.488	11.143	13.277	14.860	18.467
5	0.554	0.831	1.145	9.236	11.070	12.833	15.086	16.750	20.515
6	0.872	1.237	1.635	10.645	12.592	14.449	16.812	18.548	22.458
7	1.239	1.690	2.167	12.017	14.067	16.013	18.475	20.278	24.322
8	1.646	2.180	2.733	13.362	15.507	17.535	20.090	21.955	26.124
9	2.088	2.700	3.325	14.684	16.919	19.023	21.666	23.589	27.877
10	2.558	3.247	3.940	15.987	18.307	20.483	23.209	25.188	29.588
11	3.053	3.816	4.575	17.275	19.675	21.920	24.725	26.757	31.264
12	3.571	4.404	5.226	18.549	21.026	23.337	26.217	28.300	32.909
13	4.107	5.009	5.892	19.812	22.362	24.736	27.688	29.819	34.528
14	4.660	5.629	6.571	21.064	23.685	26.119	29.141	31.319	36.123
15	5.229	6.262	7.261	22.307	24.996	27.488	30.578	32.801	37.697
16	5.812	6.908	7.962	23.542	26.296	28.845	32.000	34.267	39.252
17	6.408	7.564	8.672	24.769	27.587	30.191	33.409	35.718	40.790
18	7.015	8.231	9.390	25.989	28.869	31.526	34.805	37.156	42.312
19	7.633	8.907	10.117	27.204	30.144	32.852	36.191	38.582	43.820
20	8.260	9.591	10.851	28.412	31.410	34.170	37.566	39.997	45.315
21	8.897	10.283	11.591	29.615	32.671	35.479	38.932	41.401	46.797
22	9.542	10.982	12.338	30.813	33.924	36.781	40.289	42.796	48.268
23	10.196	11.689	13.091	32.007	35.172	38.076	41.638	44.181	49.728
24	10.856	12.401	13.848	33.196	36.415	39.364	42.980	45.559	51.179
25	11.524	13.120	14.611	34.382	37.652	40.646	44.314	46.928	52.620
26	12.198	13.844	15.379	35.563	38.885	41.923	45.642	48.290	54.052
27	12.879	14.573	16.151	36.741	40.113	43.195	46.963	49.645	55.476
28	13.565	15.308	16.928	37.916	41.337	44.461	48.278	50.993	56.892
29	14.256	16.047	17.708	39.087	42.557	45.722	49.588	52.336	58.301
30	14.953	16.791	18.493	40.256	43.773	46.979	50.892	53.672	59.703
31	15.655	17.539	19.281	41.422	44.985	48.232	52.191	55.003	61.098
32	16.362	18.291	20.072	42.585	46.194	49.480	53.486	56.328	62.487
33	17.074	19.047	20.867	43.745	47.400	50.725	54.776	57.648	63.870
34	17.789	19.806	21.664	44.903	48.602	51.966	56.061	58.964	65.247
35	18.509	20.569	22.465	46.059	49.802	53.203	57.342	60.275	66.619
36	19.233	21.336	23.269	47.212	50.998	54.437	58.619	61.581	67.985
37	19.960	22.106	24.075	48.363	52.192	55.668	59.893	62.883	69.346
38	20.691	22.878	24.884	49.513	53.384	56.896	61.162	64.181	70.703
39	21.426	23.654	25.695	50.660	54.572	58.120	62.428	65.476	72.055
40	22.164	24.433	26.509	51.805	55.758	59.342	63.691	66.766	73.402

Jakaumien $\mathsf{F}(k_1, k_2)$ 0.05-yläkvantiileja eri vapausasteilla k_1 ja k_2 .

	k_1											
k_2	1	2	3	4	5	6	7	8	9	10	11	12
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	2.09	2.06
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03
37	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14	2.10	2.06	2.02
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
39	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13	2.08	2.04	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00