Sisältö

Luku 1

Johdanto

1.1 Todennäköisyys ja tilastotiede

Tämä kurssi käsittelee sekä todennäköisyyslaskentaa että tilastotiedettä. Uhkapelurien ongelmat inspiroivat todennäköisyyslaskennan uranuurtajien ajattelua, mutta nykyisin todennäköisyyslaskennan sovellusalue on erittäin monipuolinen ja jatkuvasti laajeneva. Tilastotieteessä laaditaan satunnaisilmiöille todennäköisyysmalleja ja tutkitaan sitten havaintojen perusteella, miten hyvin mallit kuvaavat todellisuutta.

1.2 Havaitut frekvenssit ja empiiriset jakaumat

Jatkossa käytämme termiä *koe* tai *satunnaiskoe,* kun puhumme menettelystä tai prosessista, joka tuottaa (generoi) havaintoja. Esimerkkejä satunnaiskokeista ovat lantin heitto tai kännykkään tulevien viestien lukumäärä seuraavan tunnin aikana. Heitetään lanttia esimerkiksi 100 kertaa ja saadaan 56 klaavaa (L). Tapahtuman 'klaava' frekvenssi 100:n heiton sarjassa on tässä tapauksessa 56 ja suhteellinen frekvenssi $56/100 = 0.56$. Merkitään tapahtuman A lukumäärää eli frekvenssiä n:n kokeen sarjassa $N_n(A)$. Useimmissa sovelluksissa näyttää käyvän niin, että suhteellinen frekvenssi

$$\frac{N_n(A)}{n} \quad \text{lähenee jotain lukua} \quad P(A), \tag{1.2.1}$$

kun toistojen lukumäärä n kasvaa. On helppo todeta, että $0 \leq P(A) \leq 1$. Tätä lukua $P(A)$ kutsumme tapahtuman A todennäköisyydeksi.

Vaikka emme olekaan vielä määritelleet todennäköisyyttä, voimme todeta, että suhteellinen frekvenssi on ominaisuuksiltaan todennäköisyyden kaltainen ja antaa siksi hyvän intuitiivisen käsityksen todennäköisyydestä. Suhteellisen frekvenssin avulla voidaan myös arvioida todennäköisyyksiä numeerisesti. Näin tehdään esimerkiksi simulointikokeissa. Huomattakoon, että suhteellinen frekvenssi ei ole todennäköisyyden määritelmä vaan todennäköisyyden eräs tulkinta. Todennäköisyys määritellään aksiomaattisesti. Kun todennäköisyys on määritelty, seuraa tulos (1.2.1) näistä aksioomeista. Itse asiassa (1.2.1) voidaan perustella *vahvan suurten lukujen lain* avulla. Se on tilastotieteen kannalta yksi todennäköisyyslaskennan tärkeimpiä lauseita.

Olkoon $x_1, x_2, \dots, x_n$ jokin lukujono. Tavallisesti nämä luvut $x_1, x_2, \dots, x_n$ ovat jonkin suureen, kuten esimerkiksi pituuden tai painon, mittalukuja. Jos esimerkiksi n tilastoyksikköä on mitattu, niin silloin x_i on i. tilastoyksikön mittaluku ja luvut x_1, x_2, $\dots$, x_n muodostavat havaintoaineiston. Lukujen x_1, x_2, $\dots$, x_n (havaintoaineiston) *empiirinen kertymäfunktio* (ekf) reaalilukuakselilla $(-\infty, \infty)$ on

$$F_n(a) = \frac{1}{n}|\{\, i : 1 \le i \le n,\, x_i \le a \,\}|,$$

missä $-\infty < a < \infty$ ja $|\,.\,|$ on joukon alkioiden lukumäärä.

Lukujen $x_1, x_2, \dots, x_n$ *empiirinen jakaumafunktio* tai lyhyesti *empiirinen jakauma* (ej) on

$$P_n(a, b) = F_n(b) - F_n(a).$$

$P_n(a, b)$ on siis puoliavoimelle välille $(a, b]$ kuuluvien lukujen suhteellinen osuus lukujoukossa $\{x_1, x_2, \dots, x_n\}$:

$$P_n(a, b) = \frac{1}{n}|\{\, i : 1 \le i \le n,\, a < x_i \le b \,\}|.$$

Esimerkki 1.1 Olkoon hatussa n arpalippua ja i. lippuun on kirjoitettu luku x_i. Valitaan hatusta satunnaisesti yksi arpa. Silloin todennäköisyys, että arvan numero sattuu välille $(a, b]$ on $P_n(a, b)$. Tässä tilanteessa empiiriselle jakaumalle voidaan siis antaa todennäköisyystulkinta. □

Empiirisen jakauman kuvaajana käytetään tavallisesti histogrammia. Histogrammin piirtäminen aloitetaan valitsemalla ensin *jakopisteet* $b_1 < b_2 < \dots < b_m$ siten, että kaikki luvut x_i sisältyvät avoimelle välille (b_1, b_m) ja mikään jakopiste ei ole mittaluku. Jakopisteet määrittelevät $m - 1$ osaväliä (b_j, b_{j+1}), $1 \le j \le m - 1$. Histogrammi piirretään asettamalla vierekkäin $m - 1$ pylvästä (suorakaidetta) siten, että j. pylvään kannan (luokan) leveys on $b_{j+1} - b_j$ ja pylvään korkeus on

$$\frac{P_n(b_j, b_{j+1})}{b_{j+1} - b_j} = \frac{|\{\, i : 1 \le i \le n,\, b_j < x_i < b_{j+1} \,\}|}{n(b_{j+1} - b_j)}.$$

Korkeus on siis j. osaväliin kuuluvien *havaintojen suhteellinen osuus pituusyksikköä kohti.* Pylvään korkeutta kutsutaan *havaintotiheydeksi* tai lyhyesti *tiheydeksi.* Vastaavasti j. *pylvään pinta-ala* on $P_n(b_j, b_{j+1})$ ja kaikkien pylväiden yhteenlaskettu pinta-ala on 1.

Käytännön sovelluksissa mittaustarkkuus on aina äärellinen, sanokaamme Δx. Jokainen mittaluku on silloin muotoa *kokonaisluku*$\cdot \Delta x$. Kahden mittaluvun pienin mahdollinen erotus on Δx. Histogrammin jakopisteet valitaan siten, että ne ovat muotoa

$$\textit{kokonaisluku} \cdot \Delta x + \frac{\Delta x}{2}.$$

Silloin jakopiste ei voi olla mittaluku. Jakopisteet muodostavat aineistoon *luokituksen* ja puhumme silloin *luokitellusta aineistosta.* Jakopisteet b_j, b_{j+1} ovat silloin j. luokan ns. *todelliset luokkarajat* ja pisteet $b_j + \frac{\Delta x}{2}$, $b_{j+1} - \frac{\Delta x}{2}$ ovat ns. pyöristetyt luokkarajat.

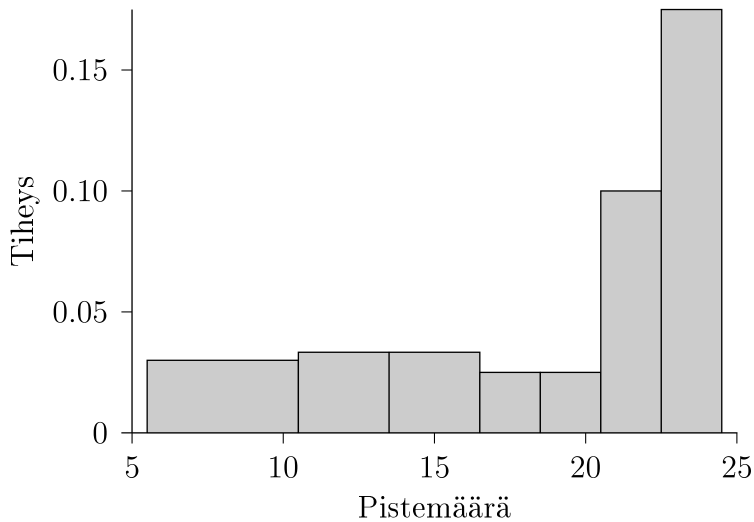
Esimerkki 1.2 Kurssin 1. välikokeen pistemäärät $x_i, 1 \leq i \leq 20$ olivat

$$18, 12, 14, 11, 24, 14, 24, 22, 24, 10, 8, 19, 21, 22, 24, 24, 24, 6, 24, 21.$$

Kokeeseen osallistui siis 20 opiskelijaa. Valitaan todellisiksi luokkarajoiksi

$$5.5, 10.5, 13.5, 16.5, 18.5, 20.5, 22.5, 24.5.$$

Nyt siis $b_1 = 5.5$ ja $b_8 = 24.5$. Luokkarajat määrittelevät 7 luokkaa.



Kuvio 1.1. Koepistemäärän histogrammi ($n = 20$).

Esimerkiksi $P_{20}(20.5, 22.5) = \frac{4}{20} = 0.2$ ja havaintotiheys luokassa $(20.5, 22.5)$ on

$$\frac{P_{20}(20.5, 22.5)}{22.5 - 20.5} = \frac{0.2}{2} = 0.1.$$

□

1.3 Todennäköisyysmallit

1.3.1 Satunnaiskoe

Todennäköisyyslaskenta on *satunnaisilmiöiden* matemaattista teoriaa. Kun tarkastelemme satunnaisilmiöitä, puhumme *satunnaiskokeista,* vaikka kyse on tavallisesti vain ajatelluista satunnaiskokeista. Se on siis matemaattinen abstraktio. Satunnaiskokeessa on oletuksena, että kokeen alkutila ei määritä tulosta deterministisesti, vaan väliintuleva tekijä, sattuma, vaikuttaa kokeen tulokseen. Satunnaiskokeen mahdolliset tulosvaihtoehdot tiedetään, mutta yksittäisen kokeen tulosta ei voida varmuudella ennustaa. Ainoa tapa saada tietoa satunnaisilmiöistä on tehdä satunnaiskokeita (eli havainnoida satunnaisilmiöitä).

Oletetaan nyt, että koe (ilmiö) on sellainen, että sen tulos ei ole varmuudella ennustettavissa, mutta kaikki mahdolliset tulosvaihtoehdot ovat tiedossa. Jos tällainen koe voidaan toistaa samoissa olosuhteissa, sitä kutsutaan satunnaiskokeeksi. Satunnaiskokeen kaikkien mahdollisten tulosten joukkoa kutsutaan *otosavaruudeksi* ja merkitään Ω:lla. Satunnaiskokeen yksittäistä mahdollista tulosta kutsutaan *alkeistapaukseksi* (satunnaiskokeeseen liittyvän otosavaruuden Ω yksi piste). Jos otosavaruus on äärellinen, merkitään

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\},$$

missä alkeistapaukset ovat ω_1, ω_2, $\ldots$, ω_n ja Ω:n alkeistapausten lukumäärä $|\Omega| = n$. Otosavaruus voi olla myös ääretön.

Tapahtuma on otosavaruuden Ω osajoukko. Otosavaruuden osajoukkoja merkitään isoilla kirjaimilla A, B, C, $\ldots$ Sanomme, että tapahtuma A sattuu, jos kokeen tulos ω kuuluu joukkoon A eli $\omega \in A$. Ω on ns. *varma tapahtuma,* koska jokin mahdollisista vaihtoehdoista sattuu varmasti.

Esimerkki 1.3 Heitetään lanttia. Tulosvaihtoehdot ovat kruuna (R) ja klaava (L), joten otosavaruus $\Omega = \{\mathrm{L}, \mathrm{R}\}$ ja $|\Omega| = 2$.

Heitetään lanttia, kunnes saadaan ensimmäinen kruuna. Silloin otosavaruus

$$\Omega = \{\mathrm{R}, \mathrm{LR}, \mathrm{LLR}, \mathrm{LLLR}, \ldots\}$$

ja $|\Omega| = \infty$. Jos tapahtuma A on 'enintään kaksi klaavaa ennen 1. kruunaa', niin $A = \{\mathrm{R}, \mathrm{LR}, \mathrm{LLR}\}$. □

Esimerkki 1.4 Tarkastellaan laitteen kestoa. Jokainen positiivinen reaaliluku voidaan tulkita kestoajaksi. Silloin

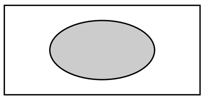
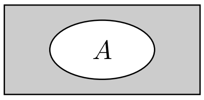
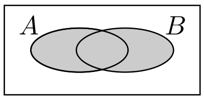
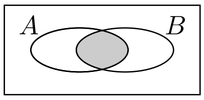
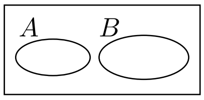
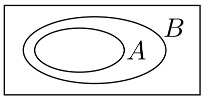
$$\Omega = \{\, \omega \in \mathbb{R} \mid \omega > 0 \,\}.$$

Esimerkiksi tapahtuma 'kestoaika ainakin 100 tuntia' on $[100, \infty)$ ja 'kestoikä yli 150, mutta korkeintaan 200 tuntia' on $(150, 200]$. □

1.3.2 Otosavaruudet, tapahtumat ja joukko-operaatiot

Oletetaan, että satunnaiskokeen $\mathcal{E}$ otosavaruus Ω on annettu. Kaikki tarkastelun kohteena olevat tapahtumat esitetään Ω:n osajoukkoina. Olkoon A tapahtuma. Jos A sattuu, se tarkoittaa, että kokeen $\mathcal{E}$ tulos ω kuuluu joukkoon A eli $\omega \in A$. Tulkitse Vennin diagrammi siten, että valitset suorakaiteesta

Taulukko 1.1. Joukko-opillisen ja todennäköisyyslaskennan terminologian vastaavuus.

Tapahtumat	Joukot	Joukkojen merkintä	Vennin diagrammi
otosavaruus	perusjoukko	Ω	
tapahtuma	Ω:n osajoukko	A, B, C jne.	
mahdoton tapahtuma	tyhjä joukko	$\emptyset$	
ei A, A ei satu	A:n komplementti	A^c	
joko A tai B tai molemmat	A:n ja B:n yhdiste	$A \cup B$	
sekä A että B	A:n ja B:n leikkaus	AB, $A \cap B$	
A ja B toisensa poissulkevat	A ja B pistevieraat	$A \cap B = \emptyset$	
jos A niin B	A on B:n osajoukko	$A \subset B$	

(Ω:sta) satunnaisesti pisteen. Jokainen suorakaiteen piste on alkeistapaus. Jokainen suorakaiteen osa-alue on tapahtuma.

Taulukossa 1.1 on esitetty joukko-opilliset operaatiot *komplementti, yhdiste* ja *leikkaus.* Nämä operaatiot toteuttavat monia käyttökelpoisia ominaisuuksia, kuten esimerkiksi

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

Yksinkertaista, mutta todennäköisyyslaskennassa hyödyllistä relaatiota $(A^c)^c = A$ kutsutaan *kaksinkertaisen komplementin säännöksi.* Keskeisiä joukko-opin laskusääntöjä ovat *vaihdantalait*

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

liitäntälait

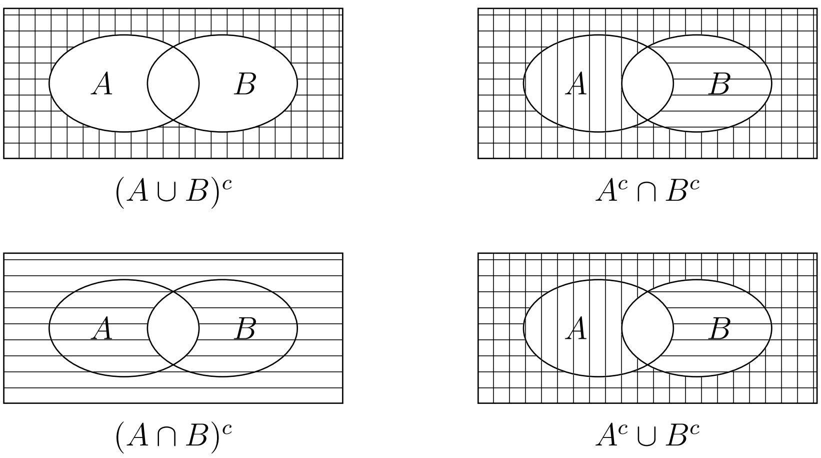
$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = A \cap (B \cap C)$$

osittelulait

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{ja}$$

De Morganin lait

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$



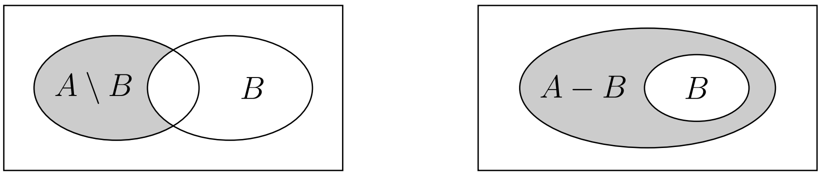
Kuvio 1.2. De Morganin lait.

Huomaa, että
$$A \cap (B \cup C) \neq (A \cap B) \cup C,$$
paitsi erikoistapauksissa. Lauseke $A \cap B \cup C$ ei ole siis hyvin määritelty, vaan tarvitaan sulut osoittamaan, kummasta tapahtumasta on kyse.

Joukkojen A ja B *erotukseen* $A \setminus B$ kuuluvat ne A:n pisteet, jotka eivät kuulu joukkoon B:

$$A \setminus B = A \cap B^c = \{\, \omega \mid \omega \in A \text{ ja } \omega \notin B \,\}.$$

Jos $B \subset A$, käytämme merkinnän $A \setminus B$ sijasta myös merkintää $A - B$. Tätä merkintää käyttäen
$$A \setminus B = A - (A \cap B)$$

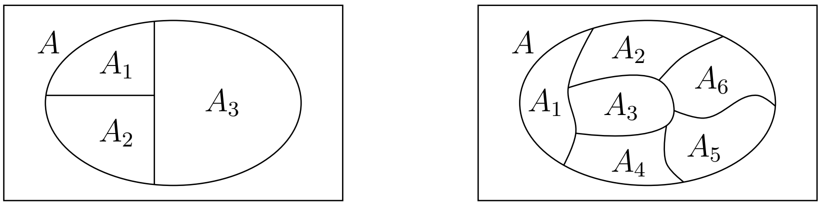


Kuvio 1.3. Joukkojen erotus.

ja

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

Sanomme, että tapahtumat A_1, A_2, $\ldots$, A_m muodostavat tapahtuman A *jaon* (tai osituksen), jos $A = A_1 \cup A_2 \cup \cdots \cup A_m$ ja tapahtumat A_1, A_2, $\ldots$, A_m ovat toisensa poissulkevat ($A_i \cap A_j = \emptyset$, kun $i \neq j$). Esimerkiksi A, A^c muodostaa otosavaruuden Ω jaon ja $A \setminus B$, $A \cap B$ muodostaa A:n jaon. Jos



Kuvio 1.4. Joukon A osituksia.

joukot A ja B ovat pistevieraat ($A \cap B = \emptyset$), niin voimme merkinnän $A \cup B$ sijasta käyttää merkintää $A + B$. Silloin esimerkiksi

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

Jos A_1, A_2, A_3 on A:n jako, niin

$$A = A_1 + A_2 + A_3.$$

Jos $A_1, A_2, \ldots, A_n$ on jono tapahtumia, niiden unioni on

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cdots \cup A_n$$

ja leikkaus

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

Kun tapahtumien jono $\{A_n\}, n \geq 1$ on ääretön, jonon tapahtumien unioni ja leikkaus voidaan määritellä äärellisten unionien ja leikkausten raja-arvona:

$$\bigcup_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcup_{n=1}^{m} A_n, \quad \bigcap_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcap_{n=1}^{m} A_n.$$

Kun tapahtumien jono on $\{A_n\}, n = 1, 2, \ldots$ on äärellinen tai ääretön, voimme merkitä myös

$$\begin{aligned}\bigcup_n A_n &= \{\omega \,|\, \omega \in A_n \text{ ainakin yhdellä } n\text{:n arvolla}\},\\ \bigcap_n A_n &= \{\omega \,|\, \omega \in A_n \text{ kaikilla } n\text{:n arvoilla}\}.\end{aligned}$$

Huomaa, että $\bigcup_{n=1}^{m} A_n$ ei vähene ja $\bigcap_{n=1}^{m} A_n$ ei kasva, kun m kasvaa. Jono $\{\bigcup_{n=1}^{m} A_n\}, m = 1, 2, \ldots$ paisuu kohti joukkoa $\bigcup_{n=1}^{\infty} A_n$ ja jono $\{\bigcap_{n=1}^{m} A_n\}, m = 1, 2, \ldots$ kutistuu kohti joukkoa $\bigcap_{n=1}^{\infty} A_n$. Ne ovat monotonisia jonoja. Jonoa $\{B_n\}, n = 1, 2, \ldots$ sanotaan *monotoniseksi*, jos $B_1 \subset B_2 \subset \ldots$ (*kasvava*) tai $B_1 \supset B_2 \supset \ldots$ (*vähenevä*). Monotonisille jonoille voidaan määritellä raja-arvo seuraavasti:

$$\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n, \quad \text{kun } \{B_n\}, n \geq 1 \text{ kasvava,}$$

ja

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n, \quad \text{kun } \{B_n\}, n \geq 1 \text{ vähenevä.}$$

Osittelulait ja De Morganin lait voidaan yleistää ilmeisellä tavalla koskemaan äärellisiä ja äärettömiä joukkojen jonoja. Esimerkiksi

$$\begin{aligned}B \cap (\bigcup_n A_n) &= \bigcup_n (B \cap A_n),\\ (\bigcap_n A_n)^c &= \bigcup_n A_n^c.\end{aligned}$$

1.3.3 Todennäköisyys

Oletetaan, että satunnaiskoe ja siihen liittyvä otosavaruus on annettu. Tarkastellaan nyt todennäköisyyden määrittelemistä. Oletamme aluksi, että otosavaruus on äärellinen. Silloin todennäköisyys voidaan määritellä alkeistapahtumien avulla.

Määritelmä 1.1 Olkoon $\mathcal{E}$ satunnaiskoe ja Ω sen äärellinen otosavaruus. Todennäköisyys on otosavaruudessa Ω määritelty reaaliarvoinen kuvaus

$$P \colon \Omega \to [0, 1],$$

jolla on seuraavat ominaisuudet:

1. $P(\{\omega\}) \geq 0$ kaikilla $\{\omega\} \in \Omega$, ja

2. $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$.

Sanomme, että $P(\{\omega\})$ on *alkeistapahtuman* $\{\omega\}$ *todennäköisyys.* Tapahtuman A eli Ω:n osajoukon todennäköisyys määritellään lukuna

$$P(A) = \sum_{\omega \in A} P(\{\omega\}). \tag{1.3.1}$$

Näin funktio P voidaan laajentaa joukkofunktioksi, joka liittää jokaiseen tapahtumaan $A \subset \Omega$ luvun $0 \leq P(A) \leq 1$. Ominaisuuksiensa nojalla todennäköisyyttä kutsutaan yleisessä teoriassa todennäköisyysmitaksi. Jos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, niin

$$\sum_{\omega_i \in \Omega} P(\{\omega_i\}) = \sum_{i=1}^{n} P(\{\omega_i\}) = 1.$$

Esimerkiksi tapahtuman $A = \{\omega_1, \omega_3, \omega_5\}$ todennäköisyys $P(A) = P(\{\omega_1\}) + P(\{\omega_3\}) + P(\{\omega_5\})$. Lisäksi määrittelemme *mahdottoman tapahtuman,* jota merkitään tyhjällä joukolla $\emptyset$, todennäköisyyden $P(\emptyset)$ nollaksi. Satunnaiskokeen todennäköisyysmalli määritellään antamalla kokeen otosavaruus Ω ja siihen liittyvä funktio P, joka toteuttaa Määritelmän 1.1 ehdot. *Todennäköisyysmalli* on siis pari (Ω, P).

Määritelmän mukaan $P(\emptyset) = 0$. Mahdoton tapahtuma $\emptyset$ on varman tapahtuman Ω komplementti eli $\Omega^c = \emptyset$. Tapahtuman A komplementti on joukko, johon kuuluvat kaikki ne alkeistapaukset, jotka eivät kuulu joukkoon A. Koska jokainen alkeistapaus ω kuuluu joukkoon A tai sen komplementtiin, mutta ei molempiin samanaikaisesti, niin

$$\sum_{\omega \in A} P(\{\omega\}) + \sum_{\omega \in A^c} P(\{\omega\}) = \sum_{\omega \in \Omega} P(\{\omega\}) = 1.$$

Tästä seuraa, että $P(A) + P(A^c) = 1$, joten

$$P(A^c) = 1 - P(A).$$

Määritelmän 1.1 oletukset toteuttava funktio määrittelee *todennäköisyysjakauman* Ω:ssa. Jos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, niin voimme esittää todennäköisyysjakauman muodossa

$$\begin{array}{cccc} \omega_1 & \omega_2 & \ldots & \omega_n \\ p_1 & p_2 & \ldots & p_n, \end{array}$$

missä $p_i = P(\{\omega_i\})$ ja $\sum_{i=1}^{n} p_i = 1$. Mikä tahansa Määritelmän 1.1 ehdot toteuttava reaalilukujoukko $\{\, p_i \mid p_i = P(\{\omega_i\}),\ 1 \leq i \leq n \,\}$ määrittelee todennäköisyysjakauman Ω:ssa.

Esimerkki 1.5 Heitetään harhatonta noppaa. Silloin silmälukujen muodostama otosavaruus on $\Omega = \{1, 2, 3, 4, 5, 6\}$. Jos jokainen silmäluku on yhtä mahdollinen, niin määritellään todennäköisyys P siten, että

$$P(i) = \frac{1}{6}, \qquad i = 1, \ldots, 6.$$

Tapahtuman 'silmäluku pariton' todennäköisyys on

$$P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}.$$

□

Olemme nyt määritelleet todennäköisyysmallin (Ω, P) äärellisessä otosavaruudessa siten, että jokaisen tapahtuman todennäköisyys voitiin määritellä. Olemme kiinnostuneita Ω :n tapahtumien todennäkäisyyksistä. Tapahtumista johdetaan uusia tapahtumia joukko-opin operaatioilla.

Määritelmä 1.2 Otosavaruuden Ω osajoukkojen kokoelma $\mathcal{A}$ on *algebra*, jos seuraavat kolme ehtoa toteutuvat:

a_1. $\Omega \in \mathcal{A}$.

a_2. Jos $A \in \mathcal{A}$, niin $A^c \in \mathcal{A}$.

a_3. Jos $A, B \in \mathcal{A}$, niin $A \cup B \in \mathcal{A}$.

Todennäköisyyslaskennassa tarkasteltavat joukkokokoelmat (tapahtumien kokoelmat) muodostavat aina algebran. Esimerkkejä joukkoalgebroista ovat:

(a) Suppein madollinen algebra $\{\Omega, \emptyset\}$, johon kuuluvat vain otosavaruus Ω ja tyhjä joukko $\emptyset$.

(b) Tapahtuman A generoima algebra $\{A, A^c, \Omega, \emptyset\}$.

(c) Otosavaruuden Ω kaikkien osajoukkojen kokoelma $\{A | A \subset \Omega\}$, joka sisältää myös tyhjän joukon $\emptyset$.

Todettakoon, että kaikki mainitut joukkoalgebrat liittyvät johonkin otosavaruuden Ω jakoon eli ositukseen. Olkoon

$$\mathcal{D} = \{D_1, \ldots, D_n\}$$

otosavaruuden jako. Silloin $D_1 + \cdots + D_n = \Omega$. Jos esimerkiksi $\Omega = \{\omega_1, \omega_2, \omega_3\}$, niin $\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$on Ω:n jako, koska $\Omega = \{\omega_1\} + \{\omega_2\} + \{\omega_3\}$. Tämän osituksen avulla voidaan määritellä 5 eri jakoa: $\mathcal{D}_1 = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{D}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$, $\mathcal{D}_3 = \{\{\omega_1, \omega_3\}, \{\omega_2\}\}$, $\mathcal{D}_4 = \{\{\omega_2, \omega_3\}, \{\omega_1\}\}$ $\mathcal{D}_5 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$. Jos muodostetaan jaon $\mathcal{D}$ joukkojen kaikki unionit, niin saadaan joukkokokoelma, joka on algebra. Mukaan otetaan myös $\emptyset$, joka aina voidaan ajatella olevan jaossa. Syntyvää joukkokokoelmaa sanotaan jaon $\mathcal{D}$ indusoimaksi joukkokokoelmaksi $\alpha(\mathcal{D})$. Myös käänteinen tulos pitää paikkansa. Jos $\mathcal{A}$ on äärellisen otosavaruuden Ω osajoukkojen muodostama algebra, niin on olemassa sellainen Ω:n yksikäsitteinen jako $\mathcal{D}$, että $\mathcal{A}$ on jaon $\mathcal{D}$ indusoima algebra eli $\mathcal{A} = \alpha(\mathcal{D})$.

Esimerkki 1.6 (*a*) Tarkastellaan otosavaruuden $\Omega = \{\omega_1, \omega_2, \omega_3\}$ jakoa $\mathcal{D}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$. Merkitään $A = \{\omega_1, \omega_2\}$, joten $A^c = \{\omega_3\}$. Silloin jaon $\mathcal{D}_2$ indusoima algebra on itse asiassa joukon A indusoima algebra.
(*b*) Olkoon otosavaruus $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_3\}$ ja sen jako $\mathcal{D} = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. Jaon $\mathcal{D}_2$ indusoima algebra on $\{\Omega, \emptyset, \{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_3, \omega_4\}\}$. Jos merkitään $D_1 = \{\omega_1, \omega_2\}, D_2 = \{\omega_3\}$ ja $D_3 = \{\omega_4\}$, niin jaon $\mathcal{D} = \{D_1, D_2, D_3\}$ indusoima algebra saadaan muodostamalla joukkojen D_1, D_2 ja D_3 kaikki mahdolliset unionit. Esimerkiksi $D_1 \cup D_3 = \{\omega_1, \omega_2, , \omega_4\}$ ja $D_2 \cup D_3 = \{\omega_3, \omega_4\}$. □

Kun satunnaiskokeelle määritellään todennäköisyysmalli, kiinnitetään ensin otosavaruus $\Omega = \{\omega_1, \ldots, \omega_n\}$. Sen jälkeen valitaan jokin sellainen osajoukkojen kokoelma $\mathcal{A}$, joka muodostaa algebran. Kokoelman $\mathcal{A}$ alkiot ovat tapahtumia. Kun Ω on äärellinen, valitaan joukkoalgebraksi $\mathcal{A}$ tavallisesti Ω:n kaikkien osajoukkojen kokoelma. Sitten jokaiseen alkeistapaukseen $\omega_i \in \Omega$, $1 \leq i \leq n$ liitetään Määritelmän 1.1 mukaisesti epänegatiivinen todennäköisyys. Tapahtuman $A \in \mathcal{A}$ todennäköisyys $P(A)$ määritellään kaavan (1.3.1) mukaisesti lukuna

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}).$$

Sanomme, että kolmikko

$$(\Omega, \mathcal{A}, P)$$

määrittelee *todennäköisyysmallin*, tai *todennäköisyysavaruuden*. Jos äärellisen otosavaruuden yhteydessä ei erikseen mainita joukkoalgebraa $\mathcal{A}$, tarkoitetaan Ω:n kaikkien osajoukkojen muodostamaa algebraa.

1.3.4 Äärettömät otosavaruudet

Edellä on käsitelty vain äärellisiä otosavaruuksia. Esimerkissä 1.3 esitettiin myös äärettömiä otosavaruuksia, jotka ovat sovelluksissa tavallisia. Jos Ω on numeroituvasti ääretön, niin

$$\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}.$$

Silloin todennäköisyysfunktiofunktio voidaan määritellä samalla tavalla kuin äärellisen otosavaruuden tapauksessa. Määritelmä 1.1 siis soveltuu myös numeroituvasti äärettömiin otosavaruuksiin. Silloin Määritelmän 1.1 2. ehdossa äärellinen summa korvataan äärettömällä summalla

$$\sum_{i=1}^{\infty} p_i = p_1 + p_2 + p_3 + \cdots = 1,$$

missä $P(\{\omega_i\}) = p_i$. Tapahtuman $A \in \Omega$ todennäköisyys on

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}), \tag{1.3.2}$$

mutta nyt summa voi olla ääretön. Jos Ω ei ole numeroituva (eli on ylinumeroituva), niin Määritelmä 1.1 ei sovellu tapahtumien todennäköisyyden määrittelemiseen, vaan tarvitaan uusia käsitteitä. Niihin palataan myöhemmin.

Esimerkki 1.7 Esimerkissä 1.3 tarkasteltiin satunnaiskoetta, jossa heitetään lanttia, kunnes saadaan ensimmäinen kruuna. Silloin otosavaruus

$$\Omega = \{\mathrm{R}, \mathrm{LR}, \mathrm{LLR}, \mathrm{LLLR}, \ldots\},$$

missä alkeistapaus

$$\omega_i = \underbrace{\mathrm{LL}\ldots\mathrm{L}}_{i-1}\mathrm{R}.$$

Jos kruunan todennäköisyys $P(\{R\}) = p$ ja klaavan todennäköisyys $P(\{L\}) = q$ $(p+q=1)$, ja määritellään $P(\{\omega_i\}) = q^{i-1}p$, niin silloin

$$\sum_{i=1}^{\infty} P(\{\omega_i\}) = \sum_{i=1}^{\infty} q^{i-1}p = \frac{p}{1-q} = 1.$$

Olemme näin määritelleet todennäköisyysjaukuman Ω:ssa, joka on numeroituvasti ääretön. $\square$

1.3.5 Todennäköisyyden tulkinnat

Todennäköisyyslaskenta ei ole riippuvainen todennäköisyyksien tulkinnoista eikä siitä, miten näitä lukuja mitataan tai arvioidaan. Todennäköisyyslaskenta on aksiomaattinen matemaattinen teoria. Esimerkiksi diskreetti todennäköisyyslaskenta perustuu Määritelmän 1.1 esittämiin todennäköisyyden ominaisuuksiin. Sovelluksissa tulkitsemme todennäköisyydet usein suureiksi, joita voidaan estimoida suhteellisilla frekvensseillä.

Mihin tahansa tapahtumaan A liittyy veto (odds), joka määritellään kaavalla

$$v(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}. \tag{1.3.3}$$

Tapahtuman A veto kertoo, kuinka monta kertaa todennäköisempää on, että A sattuu, verrattuna siihen, että A ei satu. Veto on todennäköisyyden muunnos. Olkoon p jokin todennäköisyys (tai suhteellinen frekvenssi), niin silloin todennäköisyyteen liittyvä veto on

$$v(p) = \frac{p}{1-p}.$$

Jos tapahtuman A veto $v(A)$ on annettu, niin A:n todennäköisyys on

$$P(A) = \frac{v(A)}{1+v(A)}.$$

Esimerkki 1.8 Jos 1000 henkilön populaatiossa on 600 naista ja 400 miestä, niin naisten suhteellinen osuus on

$$\frac{600}{600+400} = 0.6.$$

Jos tästä populaatista valitaan satunnaisesti yksi henkilö, niin naisen valitsemisen todennäköisyys on 0.6. Naisen mahdollisuus (veto) tulla valituksi on 6 vastaan 4. Mahdollisuus, että nainen ei tule valituksi on 4 vastaan 6. Jos $A = \{\text{nainen}\}$ ja $B = \{\text{mies}\}$, niin

$$v(A) = \frac{P(A)}{1-P(A)} = \frac{0.6}{0.4} = \frac{3}{2}.$$

□

Uhkapelurit ovat kiinnostuneita hieman erityyppisestä mahdollisuudesta, nimittäin *voiton mahdollisuudesta* (*payoff odds*). Pelikasinot ja vedonlyönnin välittäjät tarjoavat näitä mahdollisuuksia. Jos tapahtuman A mahdollisuus eli $v(A)$ (veto A) on 1 vastaan 10 ja lyöt euron vetoa tapahtuman puolesta, niin A:n sattuessa voitat 10 euroa. Jos A ei satu, häviät sen yhden euron. Kasinossa maksat pelimaksuna yhden euron eli panoksesi. Jos A sattuu, saat takaisin 11 euroa, joka on voittosi plus euron palautus. Jos A ei satu, kasino pitää maksamasi euron. *Panoksesi* on 1 euro, *kasinon panos* 10 euroa ja *kokonaispanos* 11 euroa.

Voiton mahdollisuuden ja tapahtuman mahdollisuuden välillä on yhteys, joka on ymmärretty uhkapelin yhteydessä paljon ennen varsinaisen todennäköisyyslaskennan syntyä. Puhutaan esimerkiksi ns. *reilun pelin säännöstä,* joka toteutuu silloin, kun tapahtumaa A koskevassa vedolyönnissä voiton mahdollisuus on sama kuin A:n mahdollisuus eli

$$\frac{\text{panos}}{\text{kasinon panos}} = v(A).$$

Reilun pelin säännön mukaan panoksen suhteellisen osuuden kokonasipanoksesta tulee olla $P(A)$.

Eivät ainoastaan tapahtumien mahdollisuudet vaan myös mahdollisuuksien suhteet ovat keskeisiä pelitilanteiden analysoinnissa. Ne ovat tärkeitä käsiteitä myös esimerkiksi frekvenssiaineistojen analyysissa ja logistisessa regressiossa. Olkoon A:n veto (mahdollisuus) $v(A)$ ja B:n veto $v(B)$. Silloin *vetosuhde* (*odds ratio*) $\theta(A,B)$ on

$$\theta(A,B) = \frac{v(A)}{v(B)} = \frac{P(A)/[1-P(A)]}{P(B)/[1-P(B)]}, \tag{1.3.4}$$

joka on jälleen vedonlyöntiterminologiaa. Todennäköisyyksien arviointi vedonlyönnissä perustuu pitkälti henkilökohtaisiin uskomuksiin ja kokemuksiin. Myös esimerkiksi liiketoiminnan päätöksenteossa henkilökohtaiset todennäköisyyden tulkinnat voivat olla käyttökelpoisia.

1.4 Ehdollinen todennäköisyys

Ehdollistaminen on varsin tehokas ja hyödyllinen tekniikka todennäköisyyslaskennassa ja tilastotieteessä. Käsittelemme tässä luvussa ensimmäisen kerran lyhyesti ehdollista todennäköisyyttä, joka tulee olemaan tärkeä käsite läpi koko kurssin.

Esimerkki 1.9 Heitetään harhatonta noppaa kuten Esimerkissä 1.5. Meille kerrotaan, että on saatu pariton silmäluku, mutta emme tiedä, mikä niistä. Mikä on silmäluvun 5 todennäköisyys? Olkoon B 'silmäluku pariton' ja A 'silmäluku 5'. Tiedämme siis, että silmäluku on 1, 3 tai 5. Nämä alkeistapaukset ovat yhtä todennäköisiä, joten silmäluvun 5 todennäköisyys on 1/3. Sanomme, että tapahtuman A *ehdollinen todennäköisyys* ehdolla B on 1/3. Tätä ehdollista todennäköisyyttä merkitään $P(A \mid B)$. Huomaamme, että ainakin tässä esimerkissä $P(A \mid B) \neq P(A) = 1/6$. □

Kun tarkastellaan tapahtuman A ehdollista todennäköisyyttä $P(A \mid B)$, rajoitutaan tarkastelemaan tapahtuman B alkeistapauksia. Sitten katsotaan, kuinka usein B:ssä sattuu myös A. Tämä on tapahtuma 'sekä A että B sattuvat', jota merkitään $A \cap B$. Edellisessä esimerkissä laskimme itse asiassa ehdollisen todennäköisyyden $P(A \mid B)$ kaavalla

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{1.4.1}$$

Todennäköisyys $P(A \mid B)$ on määritelty, kun $P(B) > 0$.

Esimerkki 1.10 Eloonjäämistaulukoissa esitetään eri ikäisenä elossa olevien odotettu lukumäärä 100000 elävänä syntynyttä kohti. Esimerkiksi seuraavassa taulukossa on annettu 20-, 45- ja 65-vuotiaana elossa olevien naisten lukumäärät eräässä väestössä 100000 elävänä syntynyttä tyttölasta kohti.

Ikä	20	45	65
Elossa	98040	95662	84483

Tässä voidaan ajatella, että alkuperäinen otosavaruus Ω on 100000 tyttölasta. Mikä on todennäköisyys, että 20-vuotias elää 45-vuotiaaksi (tarkoittaa itse asiassa, että elää ainakin 45-vuotiaaksi)? Olkoon A = 'elää 45-vuotiaaksi' ja B = 'elää 20-vuotiaaksi'. Koska 20-vuotiaaksi on elänyt 98040 naista ja näistä 45-vuotiaaksi 95662, niin kysytty todennäköisyys on $95662/98040 = 0.97574$. Laskettaessa ehdollista todennäköisyyttä valitaan perusjoukoksi B ja katsotaan kuinka moni näistä selviää 45-vuotiaaksi.

Nyt tapahtuma $A \cap B$ on 'elää 45-vuotiaaksi', koska 45-vuotiaksi eläneet ovat eläneet myös 20-vuotiaksi. Koska 20-vuotiaaksi elää 98040, niin $P(B) = 98040/100000 = 0.98040$. Vastaavasti $P(A \cap B) = 95662/100000 = 0.95662$. Ehdollinen todennäköisyys

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.95662}{0.98040} = 0.97574.$$

□

1.4.1 Ehdollisen todennäköisyyden frekvenssitulkinta

Olkoot A ja B jotkut satunnaiskokeen $\mathcal{E}$ otosavaruuteen Ω liittyvät tapahtumat ja $N_n(A \cap B)$ on tapahtuman $A \cap B$ frekvenssi ja $N_n(B)$ tapahtuman B frekvenssi, kun satunnaiskoe $\mathcal{E}$ toistetaan n kertaa. Voimme ajatella, että

$$P(A \mid B) \approx \frac{N_n(A \cap B)}{N_n(B)} = \frac{N_n(A \cap B)/n}{N_n(B)/n} \approx \frac{P(A \cap B)}{P(B)}, \tag{1.4.2}$$

kun toistojen lukumäärä n on suuri.

1.4.2 Kertolaskusääntö

Koska ehdollisen todennäköisyyden kaavassa (1.4.1) $P(B) > 0$, saadaan siitä kertolaskusääntö

$$P(A \cap B) = P(B)\, P(A \mid B) \tag{1.4.3}$$

tapahtuman $A \cap B$ todennäköisyyden laskemiseksi.

1.4.3 Riippumattomuus

Sanomme, että tapahtumat A ja B ovat *riippumattomat,* jos

$$P(A \cap B) = P(A)\, P(B). \tag{1.4.4}$$

Huomaa, että ehdollinen todennäköisyys (1.4.1) ei ole määritelty, jos $P(B) = 0$, mutta riippumattomuuden määritelmä (1.4.4) on silloinkin voimassa. Jos $P(B) \neq 0$ ja (1.4.4) pitää paikkansa, niin

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

Jos A ja B ovat riippumattomat, niin tieto B:n sattumisesta ei vaikuta A:n todennäköisyyteen. Jos $P(A) > 0$, niin myös $P(B \mid A) = P(A \cap B)/P(A) = P(B)$, kun A ja B ovat riippumattomat.

1.5 Odotetut frekvenssit

Kokeen $\mathcal{E}$ todennäköisyysmalli (Ω, P) on teoreettinen konstruktio. Mallin hyvyys käytännön sovelluksissa on tutkittava empiirisesti. Tämä tehdään vertailemalla kokeen (empiirisen ilmiön) havaittuja tuloksia mallin perusteella odotettavissa oleviin tuloksiin. Oletetaan, että koe toistetaan n kertaa. Jos tapahtuman A todennäköisyys on mallin mukaan p, niin silloin A:n *odotettu frekvenssi* eli *teoreettinen frekvenssi* on np. Jos A sattui suoritetussa toistokokeessa n_A kertaa, niin tätä *havaittua frekvenssiä* verrataan odotettuun frekvenssiin. Jos n_A poikkeaa "liian paljon" odotetusta frekvenssistä np, niin malli (teoria) joutuu kyseenalaiseksi. Havainnot eivät silloin tue teoriaa. Siihen, mikä on "liian suuri" poikkeama, pyrimme vastaamaan todennäköisyyslaskennan ja tilastotieteen avulla.

Johdanto: Yhteenveto

- Empiirinen kertymäfunktio. Lukujen $x_1, x_2, \ldots, x_n$ *empiirinen kertymäfunktio* on
$$F_n(a) = \frac{1}{n}|\{\, i : 1 \le i \le n,\, x_i \le a \,\}|,$$
missä $-\infty < a < \infty$ ja $|\,.\,|$ on joukon alkioiden lukumäärä.

- *Empiirinen jakaumafunktio* tai lyhyesti *empiirinen jakauma* on
$$P_n(a,b) = F_n(b) - F_n(a).$$

- Otosavaruus Ω on satunnaiskokeen (tai satunnaisilmiön) mahdollisten tulosten (alkeistapausten ω) joukko. Satunnaiskokeessa voi sattua yksi ja vain yksi alkeistapaus.

- Tapahtuma on otosavaruuden Ω osajoukko.

A ja B tapahtumia	$A \subset \Omega$ ja $B \subset \Omega$
Ω	varma tapahtuma
$\emptyset$	mahdoton tapahtuma
$A \subset B$	jos A sattuu, niin B sattuu
A^c	A ei satu
$A \cup B$	A tai B sattuu (tai molemmat)
$A \cap B$, AB	sekä A että B sattuvat
$A \setminus B = A \cap B^c$	A sattuu, mutta ei B
$A \cap B = \emptyset$	A ja B pistevieraat (toisensa poissulkevat)
A:n ositus	$A = A_1 \cup A_2 \cup \cdots \cup A_m$ ja $A_i \cap A_j = \emptyset$, $i \neq j$

- De Morganin lait
$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

- *Todennäköisyys* P on otosavaruudessa Ω (numeroituva) määritelty funktio $P\colon \Omega \to [0,1]$, jolla on seuraavat ominaisuudet:

  1. $P(\omega) \ge 0$ kaikilla $\omega \in \Omega$, ja
  2. $\sum_{\omega \in \Omega} P(\omega) = 1$.

- Tapahtuman A todennäköisyys $\quad P(A) = \sum_{\omega \in A} P(\omega)$.

- Tapahtuman A mahdollisuus
$$v(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

- Vedonlyöntisuhde

$$\theta(A, B) = \frac{v(A)}{v(B)}.$$

- A:n todennäköisyys ehdolla B

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

- Kertolaskusääntö $\quad P(A \cap B) = P(B)\,P(A \mid B)$.

- Riippumattomuus: A ja B ovat riippumattomat, jos $P(A \cap B) = P(A)\,P(B)$.

- Todennäköisyysmalli: Kokeen $\mathcal{E}$ todennäköisyysmalli on otosavaruuden Ω ja todennäköisyyden P muodostama kaksikko (Ω, P).

Harjoituksia

1. Aineistossa `kaivos_onn.dat` on aikajärjestyksessä pahojen (yli 10 kuollutta) peräkkäisten kaivosonnettomuuksien väliajat (päivinä) ajanjaksolta 6. 12. 1875 – 29. 5. 1951. Piirrä väliaikojen frekvenssihistogramma koko aineistosta ja erilliset histogrammat 56:sta ensimmäisestä ja 53:sta viimeisestä havainnosta. Kommentoi eroja ja yhtäläisyyksiä.

2. Oletetaan, että histogrammassa kahden vierekkäisen suorakaiteen kannan leveydet ovat k_1 ja k_2 sekä korkeudet h_1 ja h_2. Yhdistetään suorakaiteet yhdeksi suorakaiteeksi. Esitä uuden suorakaiteen korkeuden h lauseke ja osoita, että h on korkeuksien h_1 ja h_2 välissä.

3. Heitä harhatonta noppaa (R-ohjelma) 60, 120, 240, 480, 960 ja 2000 kertaa ja laske eri silmälukujen suhteelliset frekvenssit eri heittosarjoissa. Piirrä myös suhteellisten frekvenssien histogrammat. Miten heittojen lkm:n n kasvattaminen vaikuttaa suhteellisiin frekvensseihin?

4. Henkilöille X, Y, Z ja W on kullekin osoitettu kirje. Jokaiselle kirjeelle on varattu osoitteella varustettu kirjekuori. Kirjeet pannaan satunnaisesti kirjekuoriin.

(a) Mikä on tämän kokeen 24 alkeistapahtuman otosavaruus.

(b) Luettele seuraaviin tapahtumiin liittyvät alkaistapahtumat.

A: ”X:n kirje menee oikeaan kuoreen”;
B: ”Mikään kirje ei mene oikeaan kuoreen”;
C: ”Täsmälleen kaksi kirjettä menee oikeaan kuoreen”;
D: ”Täsmälleen kolme kirjettä menee oikeaan kuoreen”;

(c) Laske edellisessä kohdassa mainittujen tapahtumien todennäköisyydet, jos oletetaan, että kaikki alkeistapaukset ovat yhtä todennäköisiä. Määritä tapahtumien A, C ja D mahdollisuudet tapahtumaa B vastaan.

5. Kaksi joukkuetta pelaa paras seitsemästä sarjaa. Se joukkue voittaa, joka on ensiksi voittanut neljä peliä. Mikä on kokeen otosavaruus? Jos joukkueet ovat tasavahvoja (ja pelien tulokset toisistaan riippumattomia), niin mitkä ovat eri alkeistapahtumien todennäköisyydet? Mikä on todennäköisyys, että voittoon tarvitaan 7 peliä?

6. Tarkastellaan sellaista noppaa, että $p_1 = p_2 = p_3 = p_4 = p$ ja $p_5 = p_6 = q$. Kirjoitetaan tn p muodossa $p = \frac{1}{6} + \theta$.

(a) Lausu q θ:n avulla.

(b) Heitetään noppaa n kertaa ja saadaan silmälukujen 1, 2, 3, 4, 5, 6 lukumääriksi n_1, n_2, n_3, n_4, n_5, n_6. Miten estimoisit θ:n arvon?

(c) Heitettiin noppaa 30, 120, 600 ja 1200. Silmälukujen frekvenssit olivat.

	Silmäluvut					
n	1	2	3	4	5	6
30	6	10	6	5	0	3
120	29	17	35	25	9	5
600	126	119	141	124	50	40
1200	255	278	231	254	90	92

Laske θ:n, p:n ja q:n estimaatit.

7. (a) Mikä on tn-malli, kun heitetään samanaikaisesti kolmea harhatonta lanttia.

(b) Määritä tn saada x kruunaa.

(c) Heitettiin kolmea lanttia 80 kertaa ja saatiin seuraavat kruunien lukumäärät.

```
1 1 1 1 2 1 1 2 2 1 1 2 2 3 2 1 1 2 1 2 0 1 1 0 2 1 0
1 1 3 0 3 0 1 2 1 2 1 2 2 1 3 1 2 2 0 1 1 1 3 2 0 3 2
0 2 0 1 0 1 1 3 2 2 1 1 2 1 2 1 1 1 2 3 3 2 0 2 1 3
```

Määritä kruunien lukumäärän odotetut ja havaitut frekvenssit. Ovatko havainnot sopusoinnussa mallin kanssa (Heitot tiedostossa `H1.8_heitot.dat`)?

8. (a) Heitetään samanaikaisesti kahta noppaa ja olkoon tulos silmälukujen summa. Olkoot kaikki 36 alkeistapausta ovat yhtä todennäköisiä. Osoita, että tuloksen tn-jakauma on:

Tulos	2	3	4	5	6	7	8	9	10	11	12
$36 \times$ tn	1	2	3	4	5	6	5	4	3	2	1

(b) Heitä kahta noppaa 100 kertaa. Vertaa tuloksen havaittuja frekvenssejä odotettuihin frekvensseihin.

9. Vuoden 2003 jääkiekon pudotuspelijoukkueet olivat HPK (1/3), Jokerit (1/2), Kärpät (1/3), Espoon BLUES (1/6), Tappara (1/3), JYP (1/7), HIFK (1/6) ja TPS (1/9). Eräällä työpaikalla järjestettiin ennen pudotuspelien alkua vuoden mestaria koskeva vedonlyönti käyttäen suluissa ilmoitettuja voiton mahdollisuuksia. Jos veikkasit esimerkiksi Tapparaa mestariksi, niin voitit panoksesi kolminkertaisena.

(a) Laske annettujen voiton mahdollisuuksien (payoff odds) avulla joukkueiden voiton todennäköisyydet kaavalla (1.3.2). Laske todennäköisyyksien summa S.

(b) Skaalaa edellisessä kohdassa lasketut ”todennäköisyydet” jakamalla ne summalla S. Miksi skaalaus on tarpeellinen?

(c) Oleta, että skaalatut todennäköisyydet ovat ”oikeita”. Laske odotettu voittosi, jos veikkasit Tapparaa [voitto $\times P(A) +$ panoksesi $\times (1 - P(A))$]. Toteuttaako veikkaus reilun pelin säännön?

10. Eräässä kyselyssä tutkittiin suhtautumista lailliseen aborttiin ja saatiin oheisessa taulukossa esitetyt tulokset.

	Asenne		
Sukupuoli	Myönteinen	Kielteinen	Yhteensä
Nainen	309	191	500
Mies	319	281	600
Yhteensä	628	472	1100

Käytä todennäköisyyksien estimaatteina suhteellisia frekvenssejä.

(a) Laske todennäköisyys, että (i) nainen (ii) mies suhtautuu aborttiin positiivisesti (tarkasteltavassa otosavaruudessa).

(b) Laske mahdollisuudet (odds), että (i) nainen (ii) mies suhtautuu aborttiin positiivisesti.

(c) Laske mahdollisuuksien suhde (odds ratio, vedonlyöntisuhde).

11. Esimerkissä 1.2 (luennot) on annettu erään kurssin 1. välikokeen pistemäärät.

(a) Laske empiirisen kertymäfunktion (ekf) arvo pisteessä 15.3.

(b) Lausu empiirisen jakauman arvo $P_{20}(18.5, 20.5)$ ekf:n avulla.

(c) Laske histogrammissa luokkaa $[18.5, 20.5]$ kuvaavan pylvään korkeus.