Luku 3

Ehdollinen todennäköisyys ja riippumattomuus

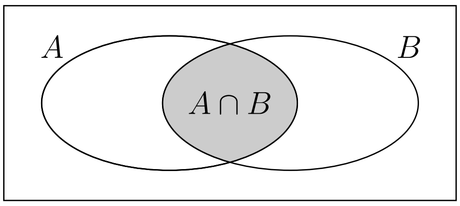
Tässä luvussa täydennetään ja laajennetaan edellisissä luvuissa esitettyjä todennäköisyyslaskennan tietoja. Ehdollistaminen ja riippumattomuus ovat todennäköisyyslaskennan ja tilastotieteen keskeisiä käsitteitä.

3.1 Ehdollinen todennäköisyys

Määritelmä 3.1 (Ehdollinen todennäköisyys) Olkoot A ja B otosavaruuden Ω tapahtumia. Jos $P(B) > 0$, niin tapahtuman A ehdollinen todennäköisyys ehdolla B on

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{3.1.1}$$

Lauseke $P(A \mid B)$ luetaan "A:n todennäköisyys ehdolla B".



Voidaan ajatella, että $P(B)$ on alueen B pinta-ala ja $P(A \cap B)$ alueen $A \cap B$ pinta-ala. Ehdollinen todennäköisyys $P(A \mid B)$ on siis alueen $A \cap B$ pinta-alan suhteellinen osuus B:n pinta-alasta.

Esimerkki 3.1 Mikä on todennäköisyys saada pokerissa kuninkaallinen värisuora K (samaa maata olevat kortit 10, 11, 12, 13 ja 14 = ässä)? Jos oletetaan, että kaikki 5 kortin kädet ovat yhtä todennäköisiä, niin

$$P(K) = \frac{4}{\binom{52}{5}} = \frac{1}{649740}.$$

Oletetaan, että jakaja jakaa 4 ensimmäistä korttia pöytään kuvapuoli alaspäin ja 5. kortin kuvapuoli ylöspäin. Viimenen korttisi on herttaässä (H_{14}). Millä todennäköisyydellä tämä käsi on kuninkaallinen värisuora? Ehdollisen todennäköisyyden (3.1.1) mukaan

$$P(K \mid H_{14}) = \frac{P(K \cap H_{14})}{P(H_{14})} = \frac{1/\binom{52}{5}}{\binom{51}{4}/\binom{52}{5}} = \frac{1}{\binom{51}{4}}.$$

Voimme nyt helposti todeta, että

$$P(K \mid H_{14}) = \frac{13}{5}\, P(K).$$

Kuninkaallisen värisuoran mahdollisuus siis yli kaksinkertaistuu, kun saat tietää, että viimeinen kortti on herttaässä. □

3.1.1 Tulosääntö, kokonaistodennäköisyys ja Bayesin kaava

Ehdollisen todennäköisyyden määritelmästä saadaan tulosääntö tapahtuman 'A ja B sattuvat' todennäköisyyden laskemiseksi. Kaavasta (3.1.1) saadaan todennäköisyyksien tulokaava

$$P(A \cap B) = P(A)\, P(B \mid A). \tag{3.1.2}$$

Tulokaavan perusteella saadaan vastaavasti $P(A^c \cap B) = P(A^c)\, P(B \mid A^c)$. Lauseen 2.3 perusteella

$$P(B) = P(A \cap B) + P(A^c \cap B),$$

josta tulokaavan avulla saadaan *kokonaistodennäköisyyden* lauseke

$$P(B) = P(A)\, P(B \mid A) + P(A^c)\, P(B \mid A^c). \tag{3.1.3}$$

Kun ehdollisen todennäköisyyden $P(A \mid B)$ lausekkeeseen (3.1.1) sijoitetaan $P(A \cap B)$:n paikalle yhtälön (3.1.2) oikea puoli ja $P(B)$:n paikalle vastaavasti kokonaistodennäköisyys (3.1.3), saadaan *Bayesin kaava*

$$P(A \mid B) = \frac{P(A)\, P(B \mid A)}{P(A)\, P(B \mid A) + P(A^c)\, P(B \mid A^c)}.$$

Jos siis tunnetaan todennäköisyydet $P(A)$, $P(B \mid A)$ ja $P(B \mid A^c)$, voidaan todennäköisyys $P(A \mid B)$ laskea Bayesin kaavan avulla.

Tulokaava (3.1.2) yleistyy myös useammalle kuin kahdelle tapahtumalle. Esimerkiksi

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

Tulokaavan, kokonaistodennäköisyyden ja Bayesin kaavan yleistykset käsitellään tämän luvun loppupuolella.

Taulukko 3.1. Kokonaistuotanto ja viallisten %-osuus eri maissa.

	Maa		
	Fahru	Russo	Swedla
Kokonaistuotanto	1000000	2000000	3000000
Viallisten %-osuus	20 %	10 %	5 %

Esimerkki 3.2 Suuri teollisuuskonserni valmistaa kännyköitä kolmessa eri maassa, jotka ovat nimeltään Fahru, Russo ja Swedla. Ostat kännykän, mutta et tiedä, missä se on valmistettu. Olkoon V tapahtuma, että tuote on viallinen. F on tapahtuma, että tuote on valmistettu Fahrussa. Vastaavasti R ja S viittaavat valmistusmaihin Russo ja Swedla. Viallisten %-osuus eri maissa on annettu oheisessa taulukossa. Lasketaan todennäköisyydet (a) $P(F \mid S^c)$, (b) $P(V \mid S^c)$, (c) $P(V)$, (d) $P(F \mid V)$. Oletetaan, että kaikki valmistetut 6000000 kännykkää ovat yhtä todennäköisiä.

Ratkaisu.

(a)
$$\begin{aligned} P(F \mid S^c) &= \frac{P(F \cap S^c)}{P(S^c)} \\ &= \frac{P(F)}{P(S^c)} \qquad (\text{koska } F \subseteq S^c) \\ &= \frac{1000000/6000000}{3000000/6000000} = \frac{1}{3}. \end{aligned}$$

(b)
$$\begin{aligned} P(V \mid S^c) &= \frac{V \cap S^c}{P(S^c)} \\ &= \frac{P[V \cap (F \cup R)]}{P(S^c)} \qquad (\text{koska } S^c = F \cup R) \\ &= \frac{P(V \cap F) + P(V \cap R)}{P(S^c)} \qquad (\text{koska } F \cap R = \emptyset) \\ &= \frac{P(V \mid F)\, P(F) + P(V \mid R)\, P(R)}{P(S^c)} \\ &= \frac{\frac{1}{5} \cdot \frac{1}{6} + \frac{1}{10} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{15}. \end{aligned}$$

Kohdat (c) ja (d) jätetään harjoitustehtäviksi. □

Esimerkki 3.3 (Väärä positiivinen) Oletetaan, että eräs verinäytteiden laboratoriotesti antaa kaksi ja vain kaksi tulosta: positiivisen ja negatiivisen. Tiedetään, että 95 % tautia A sairastavista saa testissä positiivisen tuloksen.

Myös 2 % niistä, joilla ei ole tautia A, saa positiivisen tuloksen (väärän positiivisen!). Oletetaan, että 1 % populaatiosta sairastaa tautia A. Jos satunnaisesti valitun henkilön testitulos on positiivinen, mikä on todennäköisyys, että hän sairastaa tautia A?

Olkoon nyt $T = \{\text{sairastaa tautia}\}$ ja $+$ tarkoittaa positiivista testitulosta. Tiedämme, että

$$P(+ \mid T) = 0.95, \quad P(+ \mid T^c) = 0.02, \quad P(T) = 0.01 \quad \text{ja} \quad P(T^c) = 0.99.$$

Soveltamalla Bayesin kaavaa (3.3.6) saadaan

$$\begin{aligned} P(T \mid +) &= \frac{P(T)\,P(+ \mid T)}{P(T)\,P(+ \mid T) + P(T^c)\,P(+ \mid T^c)} \\ &= \frac{0.01 \cdot 0.95}{0.01 \cdot 0.95 + 0.99 \cdot 0.02} = \frac{95}{293} \approx 0.32. \end{aligned}$$

Todennäköisyys vaikuttaa ensi näkemältä kovin pieneltä. Alhainen todennäköisyys selittyy sillä, että positiiviset tulevat joukosta, joka on pieni verrattuna siihen joukkoon, josta väärät positiiviset tulevat. □

3.1.2 Riippumattomuus

Milloin ehdollinen todennäköisyys $P(B \mid A)$ on sama kuin ehdollistamaton todennäköisyys $P(B)$? Silloin on voimassa identiteetti

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

Tämä kysymys johtaa riippumattomuuden määritelmään.

Määritelmä 3.2 Tapahtumat A ja B ovat *riippumattomat*, jos

$$P(A \cap B) = P(A)\,P(B) \tag{3.1.4}$$

Jos tapahtumat A ja B ovat riippumattomat, niin silloin identiteetit

$$P(A \mid B) = P(A) \qquad \text{ja} \qquad P(B \mid A) = P(B)$$

pitävät paikkansa. Tapahtumien A ja B riippumattomuudesta seuraa, että myös niiden komplementit ovat riippumattomat.

Lause 3.1 *Jos tapahtumat A ja B ovat riippumattomat, niin myös*

1. *A ja B^c,*
2. *A^c ja B,*

3. A^c *ja* B^c

ovat riippumattomat.

Todistus. Todistetaan 1. kohta. On siis näytettävä, että A:n ja B:n riippumattomuudesta seuraa identiteetti $P(A \cap B^c) = P(A)\,P(B^c)$. Seurauslauseen **??** mukaan

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)\,P(B) && [A \text{ ja } B \text{ riippumattomat}] \\ &= P(A)[1 - P(B)] \\ &= P(A)\,P(B^c) && [\text{Lause } 2.1], \end{aligned}$$

joten A ja B^c ovat riippumattomat. Muut kohdat todistetaan vastaavalla tavalla. □

Esimerkki 3.4 Gynekologisen irtosolunäytteen eli Papa-kokeen avulla voidaan todeta kohdun kaulaosan syöpää edeltävät kudosmuutokset. Oletetaan, että 30–65-vuotiaista naisista $100p$ %:lla on epänormaaleja (muuntuneita) soluja (kohdunsuussa ja kohdunkaulassa). Papa-kokeen suorittamiseen liittyvät seuraavat virheet:

1. Tapahtuma B: Kohdunkaulassa on epänormaaleja soluja, mutta ne *eivät osu* otokseen. Olkoon $P(B) = b$.

2. Tapahtuma C: Otoksessa on poikkeavia soluja, mutta niitä *ei havaita*. Olkoon $P(C) = c$.

3. Tapahtuma D: Pelkästään normaaleja soluja sisältävä otos *luokitellaan väärin* poikkeavaksi. Olkoon $P(D) = d$.

Oletetaan, että kaikki mainitut otanta- ja määritysvirheet ovat toisistaan riippumattomat. Jos satunnaisesti valitulle 30–65-vuotiaalle naiselle tehdään Papa-koe, niin

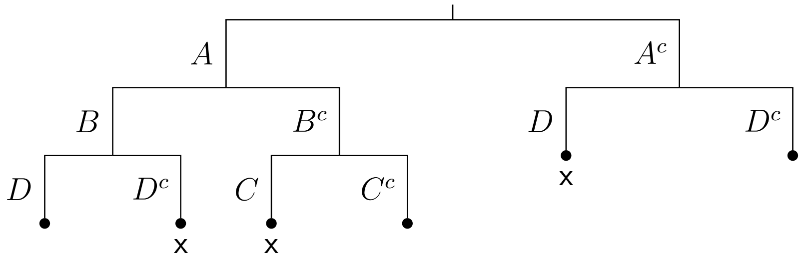
(a) millä todennäköisyydellä koe antaa väärän tuloksen?

(b) Jos testitulos osoittaa poikkeavia soluja löytyneen, millä todennäköisyydellä henkilöllä *ei ole* poikkeavia soluja?

Ratkaisu. (a) Tarkastellaan tapahtumia

V: Testi antaa virheellisen tuloksen,

A: Poikkeavia soluja on kohdunkaulassa



Kuvio 3.1. Kaaviokuva eri tulosvaihtoehdoista. Rastilla (x) merkityissä tilanteissa saadaan virheellinen testitulos.

ja tapahtumaa B (Poikkeavia soluja on, mutta ne eivät osu otokseen). Oletuksen mukaan $P(A) = p$, joten (Seurauslause 2.1)

$$\begin{aligned} P(V) &= P(A)\,P(V \mid A) + P(A^c)\,P(V \mid A^c) \\ &= p\,P(V \mid A) + (1-p)\,P(V \mid A^c). \end{aligned}$$

Virhetodennäköisyyden 3 mukaan $P(V \mid A^c) = d$. Toisaalta

$$P(V \mid A) = P(V \cap B \mid A) + P(V \cap B^c \mid A).$$

Virhetodennäköisyyksien 1 ja 3 mukaan

$$P(V \cap B \mid A) = (1-d)b$$

ja vastaavasti virheiden 1 ja 2 seurauksena

$$P(V \cap B^c \mid A) = c(1-b),$$

joten

$$P(V) = p[(1-d)b + c(1-b)] + (1-p)d.$$

(b) Jätetään harjoitustehtäväksi. □

Useamman kuin kahden tapahtuman riippumattomuuden määrittely vaatii hieman harkintaa. Milloin tapahtumat A, B ja C ovat riippumattomat? Ehdosta $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$ ei nimittäin seuraa, että tapahtumat ovat parittain riippumattomat.

Määritelmä 3.3 Tapahtumat A, B ja C ovat keskenään riippumattomat, jos

$$\begin{aligned} P(A \cap B) &= P(A)\,P(B), \\ P(A \cap C) &= P(A)\,P(C), \\ P(B \cap C) &= P(B)\,P(C) \\ \text{ja} \qquad P(A \cap B \cap C) &= P(A)\,P(B)\,P(C). \end{aligned}$$

Esimerkki 3.5 Keskinäinen riippumattomuus ei seuraa parittaisesta riippumattomuudesta. Olkoon Ω otosavaruus, jonka alkeistapahtumia ovat tavallisen korttipakan kortit. Valitaan pakasta satunnaisesti yksi kortti. Olkoon $A = \{\spadesuit, \heartsuit\}$ tapahtuma, että saadaan pata tai hertta. Vastaavasti määritellään $B = \{\spadesuit, \clubsuit\}$ ja $C = \{\spadesuit, \diamondsuit\}$. Tapahtumien todennäköisyydet ovat $P(A) = P(B) = P(C) = \frac{26}{52} = \frac{1}{2}$. Mutta $A \cap B = A \cap C = B \cap C = \{\spadesuit\}$, joten

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(\{\spadesuit\}) = \frac{13}{52} = \frac{1}{4}.$$

Nyt A, B ja C ovat parittain riippumattomat, sillä $P(A \cap B) = P(A)\,P(B)$, $P(A \cap C) = P(A)\,P(C)$ ja $P(B \cap C) = P(B)\,P(C)$. Koska $A \cap B \cap C = \{\spadesuit\}$ ja

$$P(A \cap B \cap C) = P(\{\spadesuit\}) = \frac{1}{4} \neq P(A)\,P(B)\,P(C) = \left(\frac{1}{2}\right)^3 = \frac{1}{8},$$

niin A, B ja C eivät ole keskenään riippumattomat. $\square$

Esimerkki 3.6 Valitaan korttipakasta satunnaisesti yksi kortti. Määritellään tapahtumat $A = \{$ässä tai punainen kuningas tai punainen kuningatar$\}$, $M = \{$musta$\}$ ja $R = \{$risti$\}$. Silloin $P(A) = \frac{8}{52}$, $P(M) = \frac{1}{2}$ ja $P(R) = \frac{1}{4}$. Tapahtuma $A \cap M \cap R = \{$ristiässä$\}$ ja

$$P(A \cap M \cap R) = P(A)\,P(M)\,P(R) = \frac{8}{52} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{52}.$$

Toisaalta

$$\begin{aligned} P(M \cap R) = P(R) = \frac{1}{4} &\neq P(M)\,P(R) = \frac{1}{8}, \\ P(A \cap M) = \frac{2}{52} &\neq P(A)\,P(M) = \frac{8}{52} \cdot \frac{1}{2} = \frac{4}{52}, \\ P(A \cap R) = \frac{1}{52} &\neq P(A)\,P(R) = \frac{8}{52} \cdot \frac{1}{4} = \frac{2}{52}, \end{aligned}$$

joten tapahtumat A, M ja R eivät ole parittain riippumattomia. Identiteetistä $P(A \cap M \cap R) = P(A)\,P(M)\,P(R)$ ei siis seuraa tapahtumien parittainen riippumattomuus. $\square$

Tapahtumien keskinäinen riippumattomuus vaatii toteutuakseen varsin voimakkaita ehtoja.

Määritelmä 3.4 Tapahtumat $A_1, \ldots, A_n$ ovat keskenään riippumattomat, jos jokainen tapahtumien osakokoelma $A_{i_1}, \ldots, A_{i_k}$ $(1 < k \leq n)$ toteuttaa ehdon

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Ehdollinen riippumattomuus. Tapahtumat A ja B ovat riippumattomat ehdolla C, jos $P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$.

3.1.3 Joukko-oppi ja todennäköisyys

Todennäköisyyslaskennan kannalta hyödylliset joukko-opin merkinnät esitettiin 1. luvussa. Tapahtumat A ja sen komplementti A^c eivät voi sattua samanaikaisesti, sillä $A \cap A^c = \emptyset$ ja $P(A \cap A^c) = P(\emptyset) = 0$. Toisaalta $\{A, A^c\}$ on otosavaruuden Ω ositus, sillä $A \cup A^c = \Omega$ ja . Tapahtuma "A tai A^c" sattuu varmasti eli $P(A \cup A^c) = P(A) + P(A^c) = 1$. Tästä seuraa erittäin käyttökelpoinen sääntö (Lause **??**, kohta 5)

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

De Morganin sääntö

$$(A \cap B)^c = A^c \cup B^c \tag{3.1.5}$$

on tärkeä apuväline todennäköisyyslaskennassa. Se pitää paikkansa myös mielivaltaisen monille tapahtumille. Tapahtuma-avaruuden kielellä luemme identiteetin (3.1.5) seuraavasti

Vasen puoli: Ei ole totta, että sekä A että B sattuvat.

Oikea puoli: Ainakin toinen tapahtumista A, B ei satu.

Soveltamalla kaksinkertaisen komplementin sääntöä $(A^c)^c = A$ saadaan De Morganin säännöstä (3.1.5) toinen vastaava sääntö

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

3.2 Satunnaismuuttujan ehdollinen jakauma

Olkoon X diskreetti satunnaismuuttuja ja $P(\cdot)$ on X:n todennäköisyysjakauma. Oletetaan, että tapahtuma $A \subset \Omega$, $P(A) > 0$, on sattunut. Määrittelemme nyt X:n ehdollisen jakauman käyttäen ehdollisen todennäköisyyden määritelmää.

Jokaista X:n arvoa $x_i \in \mathbb{R}$ kohti voimme määritellä tapahtuman

$$B_i = \{\, \omega \mid X(\omega) = x_i \,\}.$$

3.2.1 Ehdollinen todennäköisyysfunktio

Ehdollisen todennäköisyyden määritelmän mukaan

$$P(X(\omega) = x_i \mid A) = P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)} \geq 0, \text{ kun } P(A) > 0. \tag{3.2.1}$$

Koska $\bigcup_i B_i = \Omega$ ja $B_i \cap B_j = \emptyset$ kaikilla $i \neq j$, niin

$$\sum_i P(B_i \mid A) = \sum_i \frac{P(B_i \cap A)}{P(A)} = \frac{P(\Omega \cap A)}{P(A)} = 1. \tag{3.2.2}$$

Määritellään nyt funktio

$$f(x_i \mid A) = P(B_i \mid A) = P(X = x_i \mid A), \tag{3.2.3}$$

joka on (3.2.1):n ja (3.2.2):n perusteella todennäköisyysfunktio. Funktio (3.2.3) on X:n *ehdollinen todennäköisyysfunktio ehdolla* A.

Esimerkki 3.7 Jos X:n arvojoukko on $S_X = \{1, 2, \ldots, N\}$ ja $P(X = i) = 1/N$ kaikilla $i \in S_X$, niin sanomme, että X noudattaa diskreettiä tasajakaumaa $\mathrm{Tasd}(1, N)$. Määritellään tapahtuma $A = \{\,\omega \mid a \leq X \leq b\,\}$, missä a, b ja N, $1 \leq a < b \leq N$, ovat kokonaislukuja. Silloin

$$P(A) = \sum_{i=a}^{b} \frac{1}{N} = \frac{b-a+1}{N}$$

ja

$$P(\{X = k\} \cap A) = \begin{cases} 1/N; & a \leq k \leq b \\ 0; & \text{muutoin.} \end{cases}$$

Siksi X:n ehdollinen todennäköisyysfunktio ehdolla A on

$$f(x \mid A) = \begin{cases} \dfrac{1}{b-a+1}; & a \leq x \leq b \\ 0; & \text{muutoin.} \end{cases}$$

□

3.2.2 Yhteisjakauman todennäköisyysfunktio

Määritellään nyt kahden diskreetin satunnaismuuttujan X ja Y yhteisjakauman todennäköisyysfunktio $f_{X,Y}(\cdot)$. Olkoon X:n arvoalue $S_X = \{x_1, \ldots, x_n\}$ ja todennäköisyysfunktio $f_X(x_i) = p_i, 1 \leq i \leq n$ sekä Y:n arvoalue $S_Y = \{y_1, \ldots, y_m\}$ ja todennäköisyysfunktio $f_Y(y_j) = q_j, 1 \leq j \leq m$. Satunnaismuuttujienn X ja Y yhteisjakauman todennäköisyysfunktio on

$$f_{X,Y}(x_i, y_j) = P[(X = x_i) \cap (Y = y_j)], \quad 1 \leq i \leq n, 1 \leq j \leq m.$$

Merkitään lyhyesti

$$f_{X,Y}(x_i, y_j) = p_{ij} \quad \text{ja} \quad (X = x_i) \cap (Y = y_j) = (X = x_i, Y = y_j).$$

Apulause 3.1 *Olkoon $f_{X,Y}(x_i, y_j) = p_{ij}$, $1 \leq i \leq n, 1 \leq j \leq m$, satunnaismuuttujien X ja Y yhteisjakauman todennäköisyysfunktio. Silloin* *(i)* $\sum_{i=1}^{n} p_{ij} = q_j$ *kaikilla* $1 \leq j \leq m$.

(ii) $\sum_{j=1}^{m} p_{ij} = p_i$ *kaikilla* $1 \leq i \leq n$.

(iii) $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$.

Määrittelemme nyt satunnaismuuttujien X ja Y riippumattomuuden.

Määritelmä 3.5 Satunnaismuuttujat X ja Y ovat riippumattomat, jos

$$P[(X = x_i) \cap (Y = y_j)] = P(X = x_i)\, P(Y = y_j), \quad 1 \leq i \leq n,\ 1 \leq j \leq m.$$

Satunnaismuuttujien X ja Y riippumattomuutta merkitään $X \perp\!\!\!\perp Y$. Jos siis $p_{ij} = p_i q_j$ kaikilla $1 \leq i \leq n$ ja kaikilla $1 \leq j \leq m$, niin $X \perp\!\!\!\perp Y$.

Esimerkki 3.8 Sinun on arvattava satunnaisesti valittu kokonaisluku $\theta \in \{0, 1, \ldots, 9\}$ seuraavan kokeen tuloksen perusteella. Heitetään lanttia kahdesti (sinä et näe tuloksia) ja heittojen tulokset raportoidaan seuraavasti: Jos tulee kruuna, tulokseksi ilmoitetaan $\theta + 1$, muutoin ilmoitetaan tulokseksi $\theta - 1$. Olkoon X 1. heiton raportoitu tulos ja Y 2. heiton raportoitu tulos. Silloin $P(X = \theta + 1) = P(X = \theta - 1) = 0.5$ ja $P(Y = \theta + 1) = P(Y = \theta - 1) = 0.5$. Saadaan esimerkiksi rapotoidut havainnot $X = 5$ ja $Y = 5$.

Arvaussääntö

$$C(x, y) = \begin{cases} \frac{1}{2}(x + y), & \text{jos } x \neq y \\ x - 1, & \text{jos } x = y. \end{cases}$$

osuu 75%:n todennäköisyydellä oikeaan. Tavanomaisen luottamusajattelun mukaan edellä esitetyllä arvauksella on 75%:n luottamustaso. Jos $x \neq y$, meillä on kuitenkin 100%:n varmuus, että arvaus on oikea. Muutoin meillä on vain 50%:n varmuus. Jos havaitaan $x \neq y$, on järjetöntä väittää, että meillä on ainoastaan 75%:n luottamus arvaukseen $\frac{1}{2}(x + y)$. Kun havainto (x, y) on saatu, voidaan kertoa, mikä on parametria θ koskeva epävarmuus. Se on kuitenkin eri asia kuin arvauksen todennäköisyys osua oikeaan. □

3.3 Yleinen tulokaava ja Bayesin lause

Tietyssä mielessä kaikki todennäköisyydet ovat ehdollisia, mutta tavallisesti selvänä pidetyt ehdot jätetään mainitsematta. Rahanheitossa mainitsemme vain vaihtoehdot 'kruunu' ja 'klaava', vaikka lantti voi jäädä myös reunalleen. Presidenttiehdokkaasta tulee presidentti vain sillä ehdolla, että säilyy hengissä vaalikampanjan ajan. Valitsemistodennäköisyyttä laskettaessa ei hengissäpysymisen todennäköisyyttä tavallisesti oteta huomioon.

3.3.1 Yleinen tulokaava

Tässä alaluvussa leikkausta $A_1 \cap A_2$ merkitään kaavojen yksinkertaistamiseksi lyhyesti $A_1 A_2$.

Väittämä 3.1 (Tulokaava) *Olkoot $A_1, A_2, \ldots, A_n$ mitä tahansa tapahtumia. Silloin*

$$
(3.3.1) \quad P(A_1 A_2 \cdots A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 A_2) \cdots \cdot P(A_n \mid A_1 A_2 \cdots A_{n-1}),
$$

jos $P(A_1 A_2 \cdots A_{n-1}) > 0$.

Todistus. Jos $P(A_1 A_2 \cdots A_{n-1}) > 0$, niin kaavassa (3.3.1) esitetyt ehdolliset todennäköisyydet ovat hyvin määritellyt, koska

$$
P(A_1) \geq P(A_1 A_2) \geq \cdots \geq P(A_1 A_2 \cdots A_{n-1}) > 0.
$$

Kun yhtälön (3.3.1) oikea puoli kirjoitetaan auki ehdollisen todennäköisyyden kaavaa (3.1.1) soveltaen, saadaan

$$
\frac{P(A_1)}{P(\Omega)} \cdot \frac{P(A_1 A_2)}{P(A_1)} \cdot \frac{P(A_1 A_2 A_3)}{P(A_1 A_2)} \cdots \frac{P(A_1 A_2 \cdots A_n)}{P(A_1 A_2 \cdots A_{n-1})},
$$

joka supistuu todennäköisyydeksi $P(A_1 A_2 \cdots A_n)$. □

Kutsumme kaavaa (3.3.1) tapahtumien yhdisteen *yleiseksi tulokaavaksi.* Jos $A_1, A_2, \ldots, A_n$ ovat keskenään riippumattomat, niin saadaan

$$
P(A_1 A_2 \cdots A_n) = P(A_1)\, P(A_2) \cdots P(A_n).
$$

Olkoot $X_1, X_2, \ldots, X_n$ disrkreettejä satunnaismuuttujia ja $S_1, S_2, \ldots, S_n$ niiden arvoalueet. Määritellään tapahtumat

$$
A_i = \{X_i = x_i\}, \qquad i = 1, 2, \ldots,
$$

missä $x_i \in S_i$. Silloin voimme kirjoittaa kertolaskukaavan (3.3.1) avulla

$$
\begin{aligned}
(3.3.2) \quad & P(X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n) \\
&= P(X_1 = x_1)\, P(X_2 = x_2 \mid X_1 = x_1)\, P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \cdots \\
&\quad \cdot P(X_n = x_n \mid X_1 = x_1,\, \ldots,\, X_{n-1} = x_{n-1}).
\end{aligned}
$$

$P(X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n)$ on satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *yhteistodennäköisyys,* joka on lausuttu peräkkäisten ehdollisten todennäköisyyksien avulla.

Esimerkki 3.9 (Syntymäpäiväongelma uudelleen) Olemme jo aikaisemmin implisiittisesti soveltaneet yleistä tulokaavaa (3.3.1). Tarkastellaan uudelleen Esimerkin 2.5 syntymäpäiväongelmaa. Kutsuilla on r henkilöä. Millä todennäköisyydellä ainakin kahdella henkilöllä on sama syntymäpäivä? Käytössämme on osanottajalista (nimet satunnaisessa järjerstyksessä), johon syntymäpäivät on merkitty (karkausvuotta ei oteta huomioon). Käydään listaa läpi alusta lähtien järjestyksessä niin pitkälle, kunnes löydetään syntymäpäivä, joka jo oli listalla aikaisemmin. Silloin etsintä lopetetaan siihen ja todetaan, että ainakin kahdella vieraalla on sama syntymäpäivä. Jos lista päästään läpi löytämättä toistoa, kellään ei ole samaa syntymäpäivää.

Olkoon B_j tapahtuma, että tarkistus lopetetaan j. vieraaseen, koska hänen kohdallaan huomataan 1. toistuva syntymäpäivä. Olkoon A_j tapahtuma, että j:llä ensimmäisellä on eri syntymäpäivä. Silloin

$$A_r^c = B_2 \cup B_3 \cup \cdots \cup B_r$$

on tapahtuma, että ainakin kahdella on sama syntymäpäivä. Koska tapahtumat B_2, B_3, ..., B_r ovat toisensa poissulkevat, niin

$$P(A_r^c) = P(B_2) + P(B_3) + \cdots + P(B_r),$$

missä

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

Lasketaan kysytty todennäköisyys $P(A_r^c)$ todennäköisyyden $P(A_r)$ avulla.

Kuvataan tarkistusprosessi toistokokeena:

$$\begin{tikzcd}
B_2 & B_3 & B_4 & B_5 & & B_r & \\
A_1 \arrow[u, "\frac{1}{365}"] \arrow[r, "\frac{364}{365}"] & A_2 \arrow[u, "\frac{2}{365}"] \arrow[r, "\frac{363}{365}"] & A_3 \arrow[u, "\frac{3}{365}"] \arrow[r, "\frac{362}{365}"] & A_4 \arrow[u, "\frac{4}{365}"] & \cdots & A_{r-1} \arrow[u, "\frac{r-1}{365}"] \arrow[r, "\frac{365-(r-1)}{365}"] & A_r
\end{tikzcd}$$

Jotta tarkistusprosessi menee koko listan läpi, sattuu tapahtuma A_r, eli kaikilla vierailla on eri syntymäpäivä. Sitä ennen ovat sattuneet A_2, A_3, ..., A_{r-1}. Esimerkiksi A_2 on tapahtuma, että tarkistusprosessi ei pysähdy 2. vieraaseen, vaan hänellä on eri syntymäpäivä kuin 1. vieraalla. Todennäköisyys

$$P(A_2) = \frac{364}{365} = 1 - \frac{1}{365} = 1 - P(B_2).$$

koska valittavana on 364 päivää, jotka poikkeavat 1. vieraan syntymäpäiväpäivästä. Jos j:n ensimmäisen syntymäpäivän joukossa ei ole samoja, niin ei myöskään i:n ensimmäisen, jos $i < j$, jolloin $A_i \subset A_j$. Tästä seuraa, että $A_2 A_3 \cdots A_j = A_j$ ja

$$P(A_{j+1} \mid A_2 A_3 \cdots A_j) = P(A_{j+1} \mid A_j) = \frac{365 - j}{365} = 1 - \frac{j}{365}.$$

Soveltamalla tapahtumien yhdisteen tulokaavaa saadaan

$$\begin{aligned} P(A_r) &= P(A_2A_3A_4\cdots A_r) \\ &= P(A_2)\,P(A_3 \mid A_2)\,P(A_4 \mid A_2A_3)\cdots P(A_r \mid A_2\cdots A_{r-1}) \\ &= P(A_2)\,P(A_3 \mid A_2)\,P(A_4 \mid A_3)\cdots P(A_r \mid A_{r-1}) \\ &= \frac{364}{365}\cdot\frac{363}{365}\cdot\frac{362}{365}\cdots\frac{365-r+1}{365} = \frac{365^{(r)}}{365^r}. \end{aligned}$$

□

3.3.2 Bayesin lause

Pastori Thomas Bayesin (1763) mukaan nimetty lause seuraa suoraan ehdollisen todennäköisyyden määritelmästä. Bayesilainen lähestymistapa tilastotieteeseen perustuu tähän lauseeseen. Olkoot H_1, H_2, …, H_k sellaiset tapahtumat, että

$$H_iH_j = \emptyset \quad (i \neq j) \qquad \text{ja} \qquad \sum_{i=1}^{k} H_i = \Omega.$$

Nyt siis tapahtumajoukko H_1, H_2, …, H_k muodostaa otosavaruuden Ω osituksen. Tämä tarkoittaa sitä, että yksi ja vain yksi tapahtumista H_1, H_2, …, H_k sattuu, kun tehdään satunnaiskoe $\mathcal{E}$, jonka otosavaruus on Ω. Oletamme lisäksi, että $P(H_i) > 0$ kaikilla $i = 1, 2, \ldots, k$.

Lause 3.2 *Olkoon*

$$\Omega = \sum_i H_i$$

jokin otosavaruuden ositus. Silloin minkä tahansa tapahtuman $T \subset \Omega$ todennäköisyys voidaan lausua muodossa

(3.3.3) $$P(T) = \sum_i P(H_i)\,P(T \mid H_i).$$

Todistus. Joukko-opin sääntöjen nojalla saadaan

$$T = \Omega T = \Big(\sum_i H_i\Big)T = \sum_i H_iT,$$

josta todennäköisyyden P additiivisuuden (Määritelmä 2.3) perusteella seuraa kaava

$$P(T) = P\Big(\sum_i H_iT\Big) = \sum_i P(H_iT).$$

Kun kaavaan sijoitetaan

$$P(H_iT) = P(H_i)\,P(T \mid H_i),$$

saadaan (3.3.3). □

Jos kaavassa (3.3.3) jokin $P(H_i) = 0$, vastaava summan termi on 0, vaikka $P(T \mid H_i)$ ei olekaan määritelty. Kaavaa (3.3.3) kutsutaan *kokonaistodennäköisyyden kaavaksi.*

Olkoot X ja Y kokonaislukuarvoiset satunnaismuuttujat ja k jokin kokonaisluku. Soveltamalla kaavaa (3.3.3) tapahtumiin

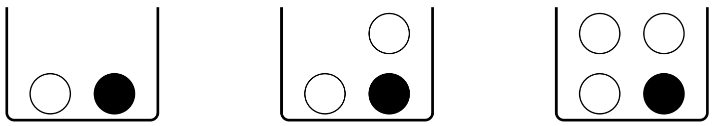
$$H_i = \{X = i\}, \qquad T = \{Y = k\}$$

saadaan

$$P(Y = k) = \sum_i P(X = i)\, P(Y = k \mid X = i), \tag{3.3.4}$$

missä summa käy yli kaikkien kokonaislukujen. Jos $P(X = i) = 0$, niin vastaava yhteenlaskettava summassa on 0. Kaava on helppo yleistää mille tahansa satunnaismuuttujalle X, jonka arvojoukko S_X on numeroituva. Y voi olla jokin yleisempi satunnaismuuttuja, ei välttämättä kokonaislukuarvoinen, ja tapahtuma $T = \{Y = k\}$ voidaan korvata vaikkapa tapahtumalla $T = \{Y > a\}$, $a \in \mathbb{R}$.

Esimerkki 3.10 (Mikä laatikko?) Meillä on 3 laatikkoa, joiden numerot ovat $i = 1, 2, 3$. Laatikossa i on i valkoista palloa ja yksi musta. Tilanne on siis oheisen kuvion kaltainen



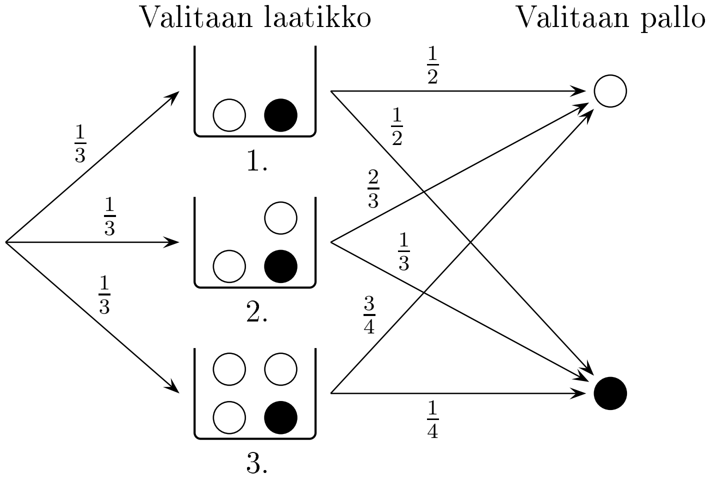
Laatikko 1 Laatikko 2 Laatikko 3

Yksi laatikko valitaan satunnaisesti harhattoman nopan heitolla. Jos silmäluku on k, valitaan laatikko, jonka numero $i = \lceil \frac{k}{2} \rceil$ on $\frac{k}{2}$ pyöristettynä lähimpään (suurempaan) kokonaislukuun. Jos esimerkiksi $k = 3$, niin $\lceil \frac{k}{2} \rceil = 2$ ja valitaan Laatikko 2. Kun laatikko on valittu, poimitaan valitusta laatikosta satunnaisesti yksi pallo. Kun valitun pallon väri on tiedossa, arvataan, mistä laatikosta pallo on valittu.

Mikä on arvauksesi, jos valittu pallo on valkoinen? Tuntuisi järkevältä arvata Laatikko 3, koska siellä on suhteellisesti eniten valkoisia. Olkoon $H_i = \{\text{Pallo Laatikosta } i\}$, $T = \{\text{Pallo valkoinen}\}$. Arvion varmentamiseksi lasketaan todennäköisyydet

$$P(H_i \mid T) = \frac{P(H_i T)}{P(T)}, \qquad i = 1, 2, 3. \tag{3.3.5}$$

Seuraavassa kuviossa on esitetty havainnollisesti tilanteeseen liittyvät todennäköisyydet.



Todennäköisyyksien tulokaavan mukaan

$$P(H_iT) = P(H_i)\,P(T \mid H_i) = \frac{1}{3}\cdot\frac{i}{i+1}, \qquad i = 1,2,3.$$

Koska $\sum_{i=1}^{3} H_iT = T$ ja T_1, T_2 ja T_3 muodostavat T:n osituksen, niin yhteenlaskulauseen perusteella

$$P(T) = \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{2}{3} + \frac{1}{3}\cdot\frac{3}{4} = \frac{23}{36}.$$

Kaavasta (3.3.5) saadaan

$$P(H_i \mid T) = \frac{\frac{1}{3}\cdot\frac{i}{i+1}}{\frac{23}{36}} = \frac{12}{23}\cdot\frac{i}{i+1}, \qquad i = 1,2,3.$$

Jos veikkaat Laatikkoa 3, todennäköisyys osua oikeaan on $\frac{9}{23}$. Laatikolla 1 vastaava todennäköisyys on $\frac{6}{23}$ ja Laatikolla 2 se on $\frac{8}{23}$. Intuitiivisesti oikealta tuntunut Laatikon 3 valinta on siis paras arvaus. □

Lause 3.3 (Bayesin lause) *Olkoon H_1, H_2, ..., H_k otosavaruuden Ω ositus ja T sellainen tapahtuma, että $P(T) > 0$. Silloin*

(3.3.6) $$P(H_i \mid T) = \frac{P(H_i)\,P(T \mid H_i)}{\sum_{j=1}^{k} P(H_j)\,P(T \mid H_j)}.$$

Todistus. Todennäköisyyksien tulokaavan nojalla sadaan

$$P(H_iT) = P(H_i)\,P(T \mid H_i) = P(T)\,P(H_i \mid T),$$

mistä seuraa

$$P(H_i \mid T) = \frac{P(H_i)\,P(T \mid H_i)}{P(T)}.$$

Lauseen 3.2 mukaan $P(T) = \sum_j P(H_j)\,P(T \mid H_j)$, joten kaava (3.3.6) on todistettu. □

Kaavaa (3.3.6) kutsutaan Bayesin säännöksi. Tapahtumat H_1, H_2, $\ldots$, H_k voidaan usein ajatella *hypoteeseiksi*, joista täsmälleen yksi on tosi. T on taas jokin tunnettu tieto satunnaiskokeen tuloksesta: tiedämme, että tapahtuma T on sattunut. Todennäköisyydet $P(H_i)$, $i = 1, 2, \ldots, k$ ovat hypoteeseja koskevia ns. *prioritodennäköisyyksiä,* jotka voivat kuvastaa uskoa tai luottamusta kyseisiin hypoteeseihin. Ehdollista todennäköisyyttä $P(H_i \mid T)$ kutsutaan hypoteesin H_i *posterioritodennäköisyydeksi* tai posterioriluottamukseksi hypoteesiin H_i. Tapahtuman T todennäköisyys $P(T \mid H_i)$ ehdolla, että hypoteesi H_i on tosi, on tapahtuman T *uskottavuus* (likelihood) ehdolla H_i.

3.3.3 Peräkkäisotanta

Populaatiossa on N henkilöä, joista Np $(0 \le p \le 1)$ henkilöä kannattaa puoluetta A ja loput $N - Np$ eivät kannata A:ta (ts. kannattavat jotain muuta puoluetta, eivät kannata mitään puoluetta, eivät ota kantaa yms.). Haluamme estimoida kannattajien suhteellisen osuuden p, joka on tuntematon parametri. Haastattelija kysyy n:n satunnaisesti valitun henkilön mielipiteen (otanta palauttamatta). Määritellään

$$X_i = \begin{cases} 1, & \text{jos } i. \text{ haastateltava kannattaa } A\text{:ta;} \\ 0 & \text{muutoin,} \end{cases}$$

missä $1 \le i \le n$ ja $1 \le n \le N$. Tarkastellaan siis satunnaismuuttujien jonoa $\{X_1, X_2, \ldots, X_n\}$ tai lyhyesti $\{\, X_i \mid 1 \le i \le n \,\}$. Tällaista jonoa kutsutaan *stokastiseksi prosessiksi,* mikä on satunnaismuuttujien perheestä käytetty nimitys.

Merkitään nyt $A_i = \{X_i = 1\}$ ja $A_i^c = \{X_i = 0\}$. Silloin kokonaistodennäköisyyden kaavan mukaan

$$P(A_2) = P(A_1)\, P(A_2 \mid A_1) + P(A_1^c)\, P(A_2 \mid A_1^c). \tag{3.3.7}$$

Helposti nähdään, että

$$P(A_1) = \frac{Np}{N} = p \qquad \text{ja} \qquad P(A_1^c) = \frac{N - Np}{N} = 1 - p.$$

Toisaalta

$$P(A_2 \mid A_1) = \frac{Np - 1}{N - 1} \qquad \text{ja} \qquad P(A_2 \mid A_1^c) = \frac{Np}{N - 1},$$

koska 1. haastatellun jälkeen jäljellä on $N - 1$ haastateltavaa, joiden joukossa on $Np - 1$ A:n kannattajaa, jos 1. haastateltava oli A:n kannattaja. Jos 1. haastateltava ei ollut A:n kannattaja, niin jäljellä on vielä Np A:n kannattajaa. Kun nämä todennäköisyydet sijoitetaan kaavaan (3.3.7), saadaan

$$P(A_2) = p\frac{Np - 1}{N - 1} + (1 - p)\frac{Np}{N - 1} = p.$$

Näin olemme osoittaneet, että $P(A_1) = P(A_2)$. Mutta tämä tulos pitää paikkansa yleisesti:

$$P(A_i) = p, \qquad i = 1, 2, \ldots, n; \quad 1 \leq n \leq N. \tag{3.3.8}$$

Näytämme nyt, että tämä yleinen tulos pitää paikkansa. Voimme ajatella, että A:n kannattajat on numeroitu 1, 2, ..., Np ja muut $Np+1$, $Np+2$, ..., N. Kysymyksessä on otanta palauttamatta, kun järjestys otetaan huomioon. Tarkastellaan tapahtumaa A_{i+1}, että $(i+1)$. haastateltava on A:n kannattaja. Kaikkien $(i+1)$:n kokoisten järjestettyjen jonojen (otosten) lukumäärä on $N^{(i+1)}$. Sellaisia jonoja, joissa $(i+1)$. alkio on 1 (A:n kannattaja) on $Np(N-1)^{(i)}$ kappaletta, koska A:n kannattaja voidaan valita Np tavalla ja loput i otosalkiota $(N-1)^{(i)}$ tavalla. Tuloperiaatteen mukaan suotuisia otoksia on siis $Np(N-1)^{(i)}$ kappaletta. Tästä seuraa, että

$$\begin{aligned} P(A_{i+1}) = \frac{Np(N-1)^{(i)}}{N^{(i+1)}} &= \frac{pN(N-1)\cdots(N-1-i+1)}{N^{(i+1)}} \\ &= \frac{pN^{(i+1)}}{N^{(i+1)}} = p. \end{aligned}$$

Olemme näin todistaneet tuloksen (3.3.8)

Määritellään nyt satunnaismuuttuja

$$X = X_1 + X_2 + \cdots + X_n,$$

joka on A:n kannattajien lukumäärä otoksessa. Tiedämme alaluvun 2.7.1 tarkastelujen perusteella, että X noudattaa hypergeometrista jakaumaa. Satunnaismuuttujan odotusarvoa käsitellään systemaattisesti vasta seuraavassa luvussa, mutta esitämme jo tässä satunnaismuuttujan X odotusarvon. Satunnaismuuttujan X_i odotusarvo on

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p, \qquad i = 1, 2, \ldots, n.$$

Jatkossa tullaan todistamaan, että satunnaismuuttujien summan odotusarvo on yhteenlaskettavien odotusarvojen summa. Odotusarvo on lineaarinen operaattori. Tästä ominaisuudesta seuraa, että

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

3.3.4 Usean tapahtuman unionin todennäköisyys

Lauseessa 2.5 esitettiin kolmen tapahtuman A_1, A_2 ja A_3 unionin todennäköisyyden $P(A_1 \cup A_2 \cup A_3)$ lauseke. Yleistetään nyt tämä tulos n:n tapahtuman A_1, A_2, ..., A_n unionin tapaukseen. Alaluvussa 2.5 määriteltiin tapahtuman indikaattorifunctio (Määritelmä 2.5), joka on kaksiarvoinen satunnaismuuttuja. Tapahtuman A todennäköisyys on sen indikaattorifunktion I_A arvon 1 todennäköisyys, eli $P(A) = P(I_A = 1)$. Indikaattorifunktion I_A odotusarvo

$$E(I_A) = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) = P(I_A = 1) = P(A).$$

tapahtuman A todennäköisyys. Tätä indikaattorifunktion ominaisuutta käytetään seuraavan lauseen todistuksessa.

Lause 3.4 *Otosavaruudessa Ω määriteltyjen tapahtumien $A_1, A_2, \ldots, A_n$ unionin $\bigcup_{i=1}^n A_i$ todennäköisyys on*

$$P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n P(A_i) - \sum_{j>i}^n P(A_iA_j) + \sum_{k>j>i}^n P(A_iA_jA_k) - + \cdots + (-1)^{n-1} P(A_1A_2\cdots A_n). \tag{3.3.9}$$

Todistus. Olkoon $\alpha_i = I_{A_i}$ tapahtuman A_i indikaattorifunktio, eli

$$\alpha_i(\omega) = \begin{cases} 1, & \text{kun } \omega \in A_i \\ 0, & \text{kun } \omega \in A_i^c. \end{cases}$$

Huomaa, että tapahtuman A_i^c indikaattorifunktio on $1 - \alpha_i$ ja tapahtuman $A_1^cA_2^c\cdots A_n^c$ indikaattorifunktio $\prod_{i=1}^n(1-\alpha_i)$. Koska $\bigcup_{i=1}^n A_i = (A_1^cA_2^c\cdots A_n^c)^c$, niin sen indikaattorifunktio on

$$\begin{aligned} I_{\bigcup A_i} &= 1 - \prod_{i=1}^n (1 - \alpha_i) \\ &= \sum_{i=1}^n \alpha_i - \sum_{j>i}^n \alpha_i\alpha_j + \sum_{k>j>i}^n \alpha_i\alpha_k\alpha_k \\ &\qquad - + \cdots + (-1)^{n-1}\alpha_1\alpha_2\cdots\alpha_n. \end{aligned} \tag{3.3.10}$$

Kun nyt yhtälössä (3.3.10) otetaan odotusarvot puolittain ja käytetään hyväksi odotusarvon lineaarisuutta, saadaan tulos (3.3.9), sillä

$$E(I_{\bigcup A_i}) = P\big(\bigcup_{i=1}^n A_i\big), \; E(\alpha_i) = P(A_i)$$

ja

$$E(\alpha_i\alpha_j) = P(A_iA_j), \ldots, E(\alpha_1\alpha_2\cdots\alpha_n) = P(A_1A_2\cdots A_n).$$

□

Esimerkki 3.11 (Yhteensopivuusongelma) Meillä on kaksi n:n kortin korttipakkaa, joiden kortit on numeroitu juoksevasti 1:stä n:ään. Asetetaan 1. pakan kortit pöydälle riviin numerojärjestyksessä 1, 2, ..., n. Sekoitetaan

2. pakka ja asetetaan kortit riviin pöydälle saadussa satunnaisjärjestyksessä. Mikä on todennäköisyys, että i. kortin numero on i? Silloin molemmissa riveissä i. kortti on i eli on saatu i-pari. Mikä on todennäköisyys, että saadaan ainakin yksi pari?

Ratkaisu. Olkoon A_i tapahtuma, että saadaan i-pari. Pakan 2 kortit voidaan asettaa $n!$ erilaiseen järjestykseen. Jos numero i kiinnitetään i. paikalle, niin loput kortit voidaan asettaa $(n-1)!$ erilaiseen järjestykseen, joten

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}. \tag{3.3.11}$$

Jos kiinnitetään i-pari ja j-pari ($i \neq j$), niin loput $(n-2)$ korttia voidaan permutoida $(n-2)!$ tavalla. Silloin

$$P(A_iA_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}. \tag{3.3.12}$$

Vastaavalla tavalla voidaan laskea todennäköisyys, että saadaan i-pari, j-pari ja k-pari ($i \neq j \neq k$):

$$P(A_iA_jA_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)} \tag{3.3.13}$$

ja yleisesti

$$P(A_{i_1}A_{i_2}\dots A_{i_m}) = \frac{(n-m)!}{n!} = \frac{1}{n(n-1)\cdots(n-m+1)}, \qquad 1 \leq m \leq n.$$

Todennäköisyys, että saadaan ainakin yksi pari on siis Lauseen 3.4 mukaan

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-2)} \\ &\qquad - + \cdots + (-1)^{n-1}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - + \cdots + (-1)^{n-1}\frac{1}{n!}. \end{aligned}$$

Huomaa, että

$$1 - \frac{1}{2!} + \frac{1}{3!} - + \cdots + (-1)^{n-1}\frac{1}{n!} + \cdots = \sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i!} = 1 - \mathrm{e}^{-1} = 0.632\dots$$

Kun siis n on suuri, niin

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \approx 1 - \mathrm{e}^{-1} = 0.632\dots$$

Suurilla n:n arvoilla todennäköisyys saada ainakin yksi pari on hyvin lähellä lukua 0.632. □

Ehdollinen todennäköisyys ja riippumattomuus: Yhteenveto

Todennäköisyys

- Ehdollinen todennäköisyys

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

- Tulosääntö $\quad P(AB) = P(A)\,P(B \mid A).$
- Yleinen tulokaava

$$P(A_1A_2A_3\cdots A_{n-1}A_n) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1A_2)\cdots \cdot P(A_n \mid A_1A_2\cdots A_{n-1}).$$

- Riippumattomuus. A ja B ovat riippumattomat, jos $P(AB) = P(A)\,P(B)$.
- $P(A_1 \text{ tai } A_2 \text{ tai } A_3)$

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3).$$

Bayesin lause

- Kokonaistodennäköisyys

$$P(T) = \sum_{i=1}^{k} P(H_i)\,P(T \mid H_i),$$

  missä $T \subset \Omega$ ja $H_1, H_2, \ldots, H_k$ on Ω:n ositus.
- Bayesin kaava

$$P(H_i \mid T) = \frac{P(H_i)\,P(T \mid H_i)}{\sum_{j=1}^{k} P(H_j)\,P(T \mid H_j)}.$$

- Prioritodennäköisyydet $\quad P(H_i)$.
- Posterioritodennäköisyydet $\quad P(H_i \mid T)$, $i = 1, 2, \ldots, n$.
- Uskottavuus. $P(T \mid H_i)$ on tapahtuman T uskottavuus ehdolla, että H_i on tosi.

Harjoituksia

1. Populaatiossa on M miestä ja N naista. Miehissä on m ja naisissa n tupakoitsijaa. Populaatiosta valitaan satunnaisesti yksi. A on tapahtuma, että valittu on mies ja B tapahtuma, että on valittu tupakoitsija. Mitkä ehdot lukumäärien M, N, m ja n on toteutettava, jotta A ja B ovat toisistaan riippumattomat?

2. Lennolla Havannasta Helsinkiin laukkuni eivät olleet perillä Helsingissä samaan aikaan kuin minä. Laukkuja on reitillä siirretty koneesta toiseen 3 kertaa ja todennäköisyydet, että siirtoa ei ole tehty ajoissa tai oikein, ovat siirtojärjestyksessä 0.4, 0.2 ja 0.1. Mikä on todennäköisyys, että virhe sattui jo ensimmäisessä siirrossa?

3. Tarkastellaan kaksilapsisia perheitä. Oletetaan pojat ja tytöt yhtä todennäköisiksi ja 2. lapsen sukupuoli on riippumaton 1. lapsen sukupuolesta. Tarkastellaan neljää tapahtumaa:

$$\begin{aligned}
A &= \text{1. lapsi on poika,}\\
B &= \text{lapset ovat eri sukupuolta,}\\
C &= \text{1. lapsi on tyttö,}\\
D &= \text{2. lapsi on poika.}
\end{aligned}$$

(a) Mitkä tapahtumaparit $\{A, B\}$, $\{A, C\}$, $\{B, C\}$ ovat keskenään riippumattomat?

(b) Ovatko tapahtumat A, B ja D keskenään eli täydellisesti riippumattomat?

4. Liukuhihnalta tulevat pullot ovat vikaantuneita, toisistaan riippumatta, todennäköisyydellä 0.2. Hihnalta tulevat pullot tarkistetaan, vikaantuneet poistetaan ja loput pakataan 12 pullon laatikoihin.

(a) Millä todennäköisyydellä on tutkittava täsmälleen 17 pulloa, kunnes laatikko saadaan täyteen?

(b) Ainakin 17 pulloa, kunnes laatikko saadaan täyteen?

5. Lääkärillä oli oheisessa taulukossa esitetty uuden hoidon vaikutusta koskeva potilasaineisto:

	Asuu kaupungissa		Asuu maaseudulla	
	Saanut hoidon	Ei hoitoa	Saanut hoidon	Ei hoitoa
Elossa	1000	50	95	5000
Kuollut	9000	950	5	5000

Tarkastellaan tapahtumia $A =$ 'potilas elossa', $B =$ 'saanut hoidon' ja $C =$ 'asuu kaupungissa'. Estimoi tarvittavat todennäköisyydet taulukon frekvenssien avulla ja laske

(a) $P(A \mid B)$ ja $P(A \mid B^c)$ sekä

(b) $P(A \mid BC)$, $P(A \mid B^cC)$, $P(A \mid BC^c)$ ja $P(A \mid B^cC^c)$.

(c) Oliko hoidosta apua?