Todennäköisyys

Todennäköisyys on ”epävarman matematiikkaa”. Matemaattinen todennäköisyys mallintaa *satunnaisia ilmiöitä*, kuten esimerkiksi nopan- tai lantinheitto. Todennäköisyyttä voi lähestyä mm. tilastollisesti tutkimalla toistokokeessa tietyn lopputuloksen tai havaintoarvojen esiintymiskertojen lukumäärää.

Alunperin todennäköisyyslaskenta kehittyi uhkapelien (esim. pokeri) teoriana.

Määritelmiä

Satunnaisilmiön **perusjoukko** Ω koostuu kaikista mahdollisista **alkeistapauksista** $\omega_1, \omega_2, \ldots, \omega_n$. Tällöin perusjoukon Ω koko eli sen alkioiden lukumäärä $\sharp\Omega = n$. Oletetaan toistaiseksi, että $n < \infty$.

Esim. Lantinheitossa $\omega_1 =$"saadaan kruuna", $\omega_2 =$"saadaan klaava" ja siis $\Omega = \{\omega_1, \omega_2\}$, jolloin $\sharp\Omega = 2$.
Nopanheitossa taas $\omega_1 =$"saadaan 1", $\omega_2 =$"saadaan 2", ..., $\omega_6 =$"saadaan 6" ja koska $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$ on $\sharp\Omega = 6$.

Alkeistapausten yhdistelminä saadaan **tapahtumat**, jotka ovat siis perusjoukon Ω osajoukkoja.

Esim. Nopanheitossa voidaan asettaa tapahtumaksi
$A =$"saadaan parillinen" $=$"saadaan $2, 4$ tai 6" $= \{\omega_2, \omega_4, \omega_6\}$.

Symmetrinen todennäköisyys

Tarkastellaan seuraavaksi vain **symmetristä todennäköisyyttä**, toisin sanoen oletetaan, että perusjoukko Ω on äärellinen ja kaikilla alkeistapahtumilla on sama todennäköisyys.

Symmetriseen satunnaisilmiöön liittyvän tapahtuman A todennäköisyys $P(A)$ on luku

$$P(A) = \frac{\sharp A}{\sharp \Omega}$$

ja tulkinta: A tapahtuu $100 \cdot P(A)\%$ varmuudella.
Huomaa, että $P(\emptyset) = 0$ ja $P(\Omega) = 1$ eli "jotain tapahtuu varmasti".
Selvästi $P(A) \in [0, 1]$.

Esim. Nopanheitossa $\sharp A = 3$, kun $A =$"saadaan parillinen" ja tapahtuman todennäköisyys $P(A) = \frac{\sharp A}{\sharp \Omega} = \frac{3}{6} = 0,5$. Siis parillinen saadaan 50% varmuudella.

Esim. Heitetään kahta noppaa. Millä tn:llä ainakin toinen on kakkonen?

Vastatapahtuma

Tapahtuman A **vastatapahtuma** on $A^{\complement} =$"A ei tapahdu" ja sen todennäköisyys
$$P(A^{\complement}) = 1 - P(A).$$

Esim. Nopanheitossa $A =$"saadaan parillinen", jolloin $A^{\complement} =$"ei saada parillista" ja siten

$$P(A^{\complement}) = 1 - P(A) = 1 - 0,5 = 0,5.$$

Samaan tulokseen oltaisiin päädytty laskemalla suoraan tapahtuman $A^{\complement} =$"ei saada parillista"="saadaan 1, 2 tai 3" todennäköisyys.

Ongelma - Biologit ja matemaatikot

Oletetaan, että meillä on b kpl biologeja ja m kpl matemaatikkoja. Näiden joukosta valitaan satunnaisesti 5 henkilön joukko. Millä todennäköisyydellä tässä viiden joukossa on tasan 3 biologia ja 2 matemaatikkoa?

Tehtävän ratkaisemiseksi pitäisi tietää, kuinka monta erilaista 5 henkilön joukkoa voidaan valita $b + m$ henkilön joukosta (eli perusjoukon alkioiden lukumäärä) ja kuinka monta erilaista, täsmälleen halutunlaista viisikkoa on näistä kaikista (tapahtuman alkioiden lukumäärä).

Työkaluja tähän ongelmaan saamme *kombinatoriikasta*! (Yo. ongelman ratkaisu harjoitustehtävänä.)

Kombinatoriikka

Kombinatoriikassa tarkastellaan mm. äärellisiä joukkoja, niiden relaatioita ja osajoukkoja sekä näiden välisiä kuvauksia. Kombinatoriikan klassisimmassa osassa keskeinen ongelma on joukon alkioiden lukumäärän laskeminen ja se liittyykin läheisesti todennäköisyyslaskentaan. Monialainen tiedemies Blaise Pascal oli eräs kombinatoriikan kehitykseen vaikuttaneista. Hänen mukaansa on nimetty ns. binomikertoimet sisältävä Pascalin kolmio, johon tutustumme hetken kuluttua.

Blaise Pascal (1623-1662)

Tuloperiaate

Tarkastellaan toimintaa, joka voidaan suorittaa n:ssä **toisistaan riippumattomassa** (peräkkäisessä) vaiheessa.
Oletetaan, että
vaihe 1 voidaan suorittaa α_1 eri tavalla
vaihe 2 voidaan suorittaa α_2 eri tavalla
$\vdots$
vaihe n voidaan suorittaa α_n eri tavalla.
Yleisesti: k:s vaihe ($1 \leq k \leq n$) voidaan suorittaa α_k eri tavalla riippumatta siitä miten muut vaiheet suoritetaan. Tällöin koko toiminta voidaan suorittaa

$$\alpha_1 \alpha_2 \cdots \alpha_n \text{ eri tavalla.}$$

Esim. Ravintola tarjoilee kolmea erilaista alkupalaa, kuutta erilaista pääruokaa ja viittä erilaista jälkiruokaa. Kolmen ruokalajin ateria voidaan valita $3 \cdot 6 \cdot 5 = 90$ eri tavalla.

Summaperiaate

Tarkastellaan toimintaa, joka voidaan suorittaa n:llä **toisensa poissulkevalla** vaihtoehtoisella menetelmällä. Oletetaan, että
menetelmä 1 voidaan toteuttaa α_1 eri tavalla
menetelmä 2 voidaan toteuttaa α_2 eri tavalla
$\vdots$
menetelmä n voidaan toteuttaa α_n eri tavalla.
Yleisesti: k:s menetelmä voidaan toteuttaa α_k eri tavalla.
Eri keinoja suorittaa toiminta on tällöin

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n \text{ kpl.}$$

Esim. Montako kahden ruokalajin (pääruoan sisältävää) ateriaa voidaan tällöin valita edellisen esimerkin ravintolassa?
Voidaan valita joko alkupala ja pääruoka, tai pääruoka ja jälkiruoka.
Alkupala ja pääruoka voidaan tuloperiaatteen mukaan valita $3 \cdot 6 = 18$ eri tavalla, kun taas pääruoka ja jälkiruoka voidaan valita $6 \cdot 5 = 30$ eri tavalla. Summaperiaatteen mukaan erilaisia kahden ruokalajin (pääruoan sisältäviä) aterioita on $18 + 30 = 48$ kappaletta.

Esimerkki - Syntymäpäivä

Millä todennäköisyydellä n:n henkilön joukossa ainakin kahdella on sama syntymäpäivä?
Ratkaisu: Nyt alkeistapauksina ovat kaikki jonot, joissa on n kappaletta päivämääriä eli kaikki mahdolliset n henkilön syntymäpäivät. Näistä n:stä päivämäärästä jokainen voi olla mikä tahansa 365:stä päivämäärästä. Siis perusjoukko Ω sisältää nämä kaikki $365^n (= \sharp\Omega)$ jonoa.
Tapahtuman $A =$"ainakin kahdella on sama syntymäpäivä" vastatapahtuma on $A^{\complement} =$"kaikilla eri syntymäpäivä". Tuloperiaatteen nojalla $\sharp A^{\complement} = 365 \cdot 364 \cdots (365 - n + 1)$. Siis

$$P(A) = 1 - P(A^{\complement}) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

Kun $n \geq 23$, niin $P(A) > 0,5$
Kun $n \geq 41$, niin $P(A) > 0,9$
Kun $n \geq 57$, niin $P(A) > 0,99$.

Kertoma

Neljä henkilöä A, B, C ja D muodostavat komitean. Heistä yhden on oltava puheenjohtaja, yhden sihteeri, yhden rahastonhoitaja ja yhden PR-henkilö. Kuinka monella eri tavalla tehtävät voidaan jakaa?
Ratkaisu:
Puheenjohtajaksi voidaan valita kuka tahansa neljästä. Jäljelle jäävistä kolmesta valitaan sitten sihteeri , ja loput kaksi jakavat rahastonhoitajan ja PR-henkilön tehtävät. Vaihtoehtoja on siis $4 \cdot 3 \cdot 2 \cdot 1 = 24$ kappaletta. Lukua sanotaan neljän kertomaksi ja sille käytetään merkintää 4!.

Määritellään luonnollisen luvun n **kertoma**

$$n! = 1 \cdot 2 \cdot 3 \cdots n = \prod_{k=1}^{n} k$$

ja sovitaan, että $0! = 1$.
Yleisesti, n objektia voidaan järjestää jonoon $n!$ eri tavalla.

Kertoman $n!$ arvoja

Kertoman $n!$ arvot on helppo laskea pienillä luvuilla n:
$1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, $6! = 720$, $7! = 5040$, $8! = 40320$, $9! = 362880$, $10! = 3628800$, ...
Huomataan, että kertoma kasvaa kuitenkin hyvin nopeasti. Sen suuruusluokkaa voidaan arvioida Stirlingin kaavalla:

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

Tässä merkinnällä $\sim$ tarkoitetaan sitä, että

$$\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{n!} \to 1, \quad \text{kun} \quad n \to \infty,$$

eli toisin sanoen yo. osamäärä lähestyy lukua 1, kun n kasvaa rajatta.
Huomaa: Olipa $a > 0$ mikä hyvänsä kantaluku , niin $n! > a^n$ kaikilla tarpeeksi suurilla n.

Täydellisten graafien kierroksista

Tarkastellaan täydellistä graafia K_p. Monellako eri tavalla voimme tehdä kierroksen K_p:ssä? Eri pisteistä alkavat, mutta samoja viivoja kulkevat kierrokset on mielekästä samaistaa. Siten voimme valita kierroksen alkupisteen vapaasti. Alkupisteestä kierros voi jatkua $p-1$ pisteeseen , josta edelleen $p-2$ pisteeseen. Jatkamalla tällä tavoin ollaan tilanteessa, jossa kaikki graafin pisteet on käyty läpi ja viimeisestä pisteestä palataan alkupisteeseen. Tuloperiaatteen mukaan eri vaihtoehtoja on siten $(p-1)!$. Kutakin kierrosta kohti graafissa on samat viivat käänteisessä järjestyksessä kulkeva kierros. Kauppamatkustajan ongelmassa on yhdentekevää kuljetaanko kierros etu- vai takaperin. Siten tuota ongelmaa (p:n kaupungin tapauksessa) tarkasteltaessa on huomioitava $\frac{1}{2}(p-1)!$ kierrosta.

Kertoma kasvaa kuitenkin niin nopeasti, että lyhimmän (tai halvimman) reitin löytäminen on tietokoneellekin mahdotonta, kun kaupunkeja on tarpeeksi monta.

Jonojen muodostus isommasta valikoimasta objekteja

Aiemmin esillä olleeseen neljän hengen komiteaan pyrkii 20 ihmistä. Montako erilaista komiteaa voidaan muodostaa?
Ratkaisu:
Puheenjohtaja voidaan ensin valita kahdestakymmenestä pyrkijästä , sitten sihteeri yhdeksästätoista , jonka jälkeen rahastonhoitaja kahdeksastatoista ja lopulta PR-henkilö seitsemästätoista. Vaihtoehtoja on siis $20 \cdot 19 \cdot 18 \cdot 17 = 116280$ kappaletta.

Yleisesti, n:stä objektista voidaan muodostaa k:n mittaisia jonoja

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

Osajoukkojen muodostus isommasta valikoimasta objekteja

Tarkastellaan taas tilannetta, jossa neljän hengen komiteaan pyrkii 20 ihmistä, mutta kiinnittämättä huomiota eri tehtäviin. Montako eri kokoonpanoa voidaan pyrkijöistä valita?

Ratkaisu:

Eri komiteoita on edellisen esimerkin mukaan $\frac{20!}{16!}$ kappaletta. Osassa niistä on kuitenkin samat ihmiset, koska kussakin kokoonpanossa tehtävät voidaan jakaa 4! eri tavalla, kuten aikaisemmin todettiin. Erilaisia kokoonpanoja (huomioimatta tehtävänjakoa) on siten

$$\frac{20!}{4!16!} = 4845 \text{ kappaletta.}$$

Yleisesti, n:stä objektista voidaan muodostaa erilaisia k:n objektin (järjestämättömiä) joukkoja

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

kappaletta. Merkintä $\binom{n}{k}$ luetaan 'n yli k:n'.

Esimerkki - Lottorivit

Montako erilaista lottoriviä on?
Ratkaisu:
Lottorivi on 7 lottonumeron osajoukko 39 lottonumeron joukosta.
Erilaisia lottorivejä on siis

$$\binom{39}{7} = \frac{39!}{7! \cdot (39-7)!} = \frac{39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33 \cdot 32!}{7! \cdot 32!} = 15380937$$

kappaletta.

Millä todennäköisyydellä lottoaja voittaa täysosuman valitsemillaan seitsemällä numerolla? *Vastaus:* $\frac{1}{15380937} \approx 0,000000065$ eli $0,000007\%$ varmuudella!

Esimerkki - Jalkapallo

Jalkapalloturnaukseen osallistuu 8 joukkuetta. Montako peliä on pelattava, jos kaikkien joukkueiden halutaan kohtaavan toisensa täsmälleen kerran?

Ratkaisu:

Otteluita vaaditaan yhtä monta kuin on tapoja valita 2 joukkuetta 8:n joukosta , eli

$$\binom{8}{2} = \frac{8!}{2! \cdot (8-2)!} = \frac{8 \cdot 7 \cdot 6!}{2! \cdot 6!} = \frac{8 \cdot 7}{2} = 7 \cdot 4 = 28.$$

Potenssijoukon alkioiden lukumäärä

Muistetaan, että potenssijoukolla $\mathcal{P}(X)$ tarkoitettiin kaikkien joukon X osajoukkojen muodostamaa joukkoa. Esimerkiksi kun $X = \{1, 2, 3\}$, niin

$$\mathcal{P}(\{1,2,3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

Esimerkissämme $\sharp X = 3$ ja $\sharp\mathcal{P}(X) = 8 = 2^3$. Myös yleisesti pätee $\sharp\mathcal{P}(X) = 2^n$, kun $\sharp X = n < \infty$.

Tämä voidaan nähdä esimerkiksi tuloperiaatteen avulla: Muodostetaan mielivaltainen joukon X osajoukko. Tämä voidaan suorittaa niin, että jokaisen alkion kohdalla päätetään, otetaanko se osajoukkoon vai ei. Näin jokaisen alkion kohdalla on tasan 2 vaihtoehtoa (otetaan - ei oteta). Koska valinnat eri alkioiden välillä ovat toisistaan riippumattomia, niin erilaisia joukon X osajoukkoja on $\underbrace{2 \cdot 2 \cdots 2}_{n \text{ kpl}} = 2^n$ kpl.

Potenssijoukon alkioiden lukumäärä - jatkuu

Toisaalta, tilannetta voidaan tarkastella jakamalla joukon X osajoukot luokkiin niiden alkioiden lukumäärän mukaan. Tällaisten osajoukkojen lukumääräthän me osataankin laskea! Nyt esimerkiksi joukon X 3-alkioisia osajoukkoja on $\binom{n}{3}$ kpl. Summaamalla yhteen kaikkien k-alkioisten ($0 \leq k \leq n$) osajoukkojen lukumäärät saadaan kaikkien (erityisesti kaiken kokoisten) joukon X osajoukkojen lukumäärä yhteensä.

Siis $\sharp\mathcal{P}(X) = \sum_{k=0}^{n} \binom{n}{k}$. Tällä *kombinatorisella päättelyllä* on saatu identiteetti

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

(Täsmällinen todistus induktiolla harjoitustehtävänä.)

Symmetrisyys

Olkoot $n \geq 1$ ja $0 \leq k \leq n$. Tällöin

$$\binom{n}{k} = \binom{n}{n-k}.$$

Tämä on selvää, sillä kutakin k objektin valintaa n objektista vastaa $n-k$ objektin (jäljelle jääneet) valinta. Siis esim.

$$\binom{7}{2} = \binom{7}{5} \quad \text{ja} \quad \binom{13}{4} = \binom{13}{9}.$$

Pascalin identiteetti

Olkoot $n \geq 1$ ja $0 < k < n$. Tällöin

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

Lähdetään liikkeelle yhtälön vasemmasta puolesta ja lasketaan määritelmää käyttäen

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

Laventamalla ensimmäistä termiä k:lla ja toista $n-k$:lla saadaan

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

Yhdistämällä termit, ottamalla $(n-1)!$ yhteiseksi tekijäksi ja sieventämällä saadaan lopulta

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

Binomikertoimet

Lukuja $\binom{n}{k}$ kutsutaan **binomikertoimiksi**, koska $\binom{n}{k}$ on termin $x^{n-k}y^k$ kertoimena aukikerrotussa binomin $x+y$ potenssissa $(x+y)^n$.
Esim.

$$(x+y)^2 = x^2 + 2xy + y^2 = \binom{2}{0}x^2 + \binom{2}{1}xy + \binom{2}{2}y^2$$

$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3 = \binom{3}{0}x^3 + \binom{3}{1}x^2y + \binom{3}{2}xy^2 + \binom{3}{3}y^3$$

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

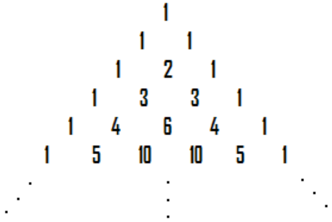
$$= \binom{4}{0}x^4 + \binom{4}{1}x^3y + \binom{4}{2}x^2y^2 + \binom{4}{3}xy^3 + \binom{4}{4}y^4$$

Yleisesti:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k.$$

Pascalin kolmio

Binomikertoimet muodostavat ns. Pascalin kolmion. Kerroin $\binom{n}{k}$ on kolmion $n+1$:nnen rivin $k+1$:s luku. Pascalin identiteetin mukaan kolmiossa oleva luku on saatu sen yläpuolella olevien vierekkäisten lukujen summana.



Ongelma - Tikkataulu

Heitetään umpimähkäisesti tikkaa (kuitenkin niin, että tikka osuu tauluun). Millä todennäköisyydellä tikka osuu kymppiin?

Otetaan perusjoukoksi tikkataulun määrittämä pistejoukko. Mutta aiemmilta luennoiltahan tiedetään, että kymppiympyrässä on (numeroituvasti) äärettömän verran pisteitä - samoin kuin tikkataulussa!

Lisäksi todennäköisyydet tapahtumien $A_k =$"saaadaan k", $1 \leq k \leq 10$, välillä eivät tuntuisi olevan samat. Miksi?

Todennäköisyysfunktio

Määritellään todennäköisyys funktiona $p : \mathcal{P}(\Omega) \longrightarrow [0, 1]$, missä funktio p liittää jokaiseen tapahtumaan $A \in \mathcal{P}(\Omega)$ sen todennäköisyyden $p(A)$. Todennäköisyysfunktiolle pätee

- $p(A) \geq 0$ kaikilla $A \in \mathcal{P}(\Omega)$
- $p(\Omega) = 1$ ja $p(\emptyset) = 0$
- Jos $A, B \in \mathcal{P}(\Omega)$ ja $A \cap B = \emptyset$, niin $p(A \cup B) = p(A) + p(B)$.

Äärettömän perusjoukon tapauksessa alkeistapausten todennäköisyydet *eivät välttämättä* määrää muiden tapahtumien todennäköisyyksiä (tikkatauluesimerkki!). Geometrisessä todennäköisyysmallissa todennäköisyysmitta voidaan rakentaa minkä tahansa äärellisen *mitta-avaruuden* pohjalta. Lyhyesti: jos perusjoukkomme Ω on jollakin tapaa *äärellismitallinen* (lukumäärältään, pituudeltaan, pinta-alaltaan,...), niin

$$p(A) = \frac{\mu(A)}{\mu(\Omega)}, \text{ jossa } \mu \text{ on mitta.}$$

Ratkaisu - Tikkataulu

Kuten huomattiin, klassinen (symmetrinen) todennäköisyysmalli ei tässä riitä, vaan mallia täytyy laajentaa geometriseen todennäköisyysmalliin. Luonteva (geometrinen) mitta tapahtuman $A =$"tikka osuu kymppiin" todennäköisyydelle on joukon A pinta-ala jaettuna koko taulun pinta-alalla. Olkoot tikkataulun säde R ja kymppiympyrän r $(0 < r < R)$. Nyt koko tikkataulun ala on πR^2 ja kymppiympyrän ala πr^2. Siten todennäköisyys sille, että tikka osuu kymppiin on

$$p(A) := P(A) = \frac{\pi r^2}{\pi R^2} = \left(\frac{r}{R}\right)^2.$$

Tässä alkeistapaukset vastaavat yhden pisteen joukkoja, jotka ovat nollamitallisia, joten tapahtuman todennäköisyyttä ei voi laskea siihen sisältyvien alkeistapausten todennäköisyyksien summana!