

Todennäköisyyslaskennan peruskäsitteet

Todennäköisyyslaskennassa tarkastelun kohteena ovat *satunnaisilmiöt*. Esimerkkejä tällaisista ovat tavallisen kuusitahkoisen nopan heitto ja vaikkapa tikan heitto tauluun.

Kaikki ilmiön mahdolliset tulokset muodostavat joukon Ω , jota kutsutaan *otosavaruudeksi*. Nopan tapauksessa tämä on joukko $\{1, 2, 3, 4, 5, 6\}$ vastaten heiton tuloksena saatavia silmälukuja. Tikanheitossa tulokset ovat sopivassa koordinaatistossa ilmoitettuja koordinaattipareja (x, y) , jotka ilmaisevat, mihin pisteeseen tikka osuu. Luontevaa on valita tikkataulun keskipiste origoksi ja yksiköksi senttimetri. Otosavaruudeksi voidaan tällöin ottaa xy -taso, vaikka kaikki sen pisteet eivät varmasti olekaan mahdollisia heittotuloksia.

■ joukko

Otosavaruuden alkioita kutsutaan *alkeistapauksiksi*.

Tapahtumalla tarkoitetaan otosavaruuden Ω osajoukkoa. Nopanheitossa se voi olla esimerkiksi tulos 'vähintään 3', ts. joukko $\{3, 4, 5, 6\}$; tikanheitossa 'alle 5 senttimetrin päähän keskipisteestä' eli $\{(x, y) \mid x^2 + y^2 < 25\}$.

■ osajoukko

Jokaiseen tapahtumaan liitetään *todennäköisyys* P , so. funktio, jonka argumentteina ovat em. osajoukot. Tällä tulee olla seuraavat ominaisuudet:

■ funktio
■ unioni
■ leikkaus

- $0 \leq P(A) \leq 1$, olipa A mikä tahansa tapahtuma.
- $P(\Omega) = 1$ eli ns. varman tapahtuman todennäköisyys on $= 1$.
- $P(A \cup B) = P(A) + P(B)$, jos $A \cap B = \emptyset$, ts. jos A ja B ovat erillisiä.

Funktio P kuvaa tapahtuman todennäköisyyttä: mitä lähempänä $P(A)$ on arvoa 1, sitä todennäköisempi on tapahtuma A .

Eo. ominaisuuksien seurauksena on mm. *komplementtitapahtuman* $\bar{A} = \Omega \setminus A$ todennäköisyyttä koskeva sääntö: Koska $P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\Omega) = 1$, on $P(\bar{A}) = 1 - P(A)$. (Todennäköisyyslaskennassa käytetään yleensä merkintää \bar{A} joukko-opillisen komplementin merkinnän $\complement A$ sijasta.)

■ komplementti (joukon)

Eo. ominaisuudet eivät määrää funktiota P , vaan ainoastaan asettavat sille tietyt järkevyyssvaatimukset. Funktion määrittäminen kussakin yksityistapauksessa edellyttää jotakin lisätietoa ilmiöstä.

Hieman täsmällisempään muotoon kirjoitettuina funktiolta P vaadittavia ominaisuuksia kutsutaan *Kolmogorovin* aksioomiksi todennäköisyyslaskennan perusteita tutkineen neuvostoliittolaisen matemaatikon Andrei Nikolajevitš Kolmogorovin (1903 – 1987) mukaan.

■ aksiooma
■ aksiooma
■ Kolmogorov

Todennäköisyysfunktio P

Monissa alkeellisissa todennäköisyyslaskennan probleemoissa otosavaruus on luonnollista valita siten, että alkeistapauksia on äärellinen määrä ja niitä on perusteltua pitää yhtä todennäköisinä. Jos alkeistapausten lukumäärä on n , on jokaisen alkeistapausten e_k todennäköisyys tällöin $P(e_k) = 1/n$, $k = 1, 2, \dots, n$.

■ funktio

Esimerkiksi heitettäessä virheetöntä noppaa tarkoittaa 'virheettömyys' sitä, että kaikki silmäluvut ovat keskenään yhtä todennäköisiä. Alkeistapauksia on siten kuusi: 'heiton tuloksena on yksi', 'heiton tuloksena on kaksi', jne. ja jokaisen todennäköisyys on $1/6$, esimerkiksi $P(\{2\}) = 1/6$. Erillisten tapahtumien unionia koskeva sääntö antaa tällöin esimerkiksi tapahtuman 'heitto antaa vähintään 3' todennäköisyydeksi

$$P(\{3, 4, 5, 6\}) = P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 4 \cdot \frac{1}{6} = \frac{4}{6} = \frac{2}{3}.$$

Tämäntyyppisissä esimerkeissä todennäköisyys voidaan luonnehtia tapahtuman kannalta suotuisien alkeistapausten ja kaikkien alkeistapausten lukumäärien suhteeksi. Tätä luonnehdintaa kutsutaan *klassiseksi* todennäköisyydeksi.

Alkeistapausten symmetrisyys ei kuitenkaan välttämättä merkitse sitä, että ne olisivat yhtä todennäköisiä. Jossakin mielessä esimerkiksi poikien ja tyttöjen syntyminen on symmetristä, mutta tilastojen mukaan poikia kuitenkin syntyy hieman enemmän kuin tyttöjä. Tällöin on luontevaa määrittää todennäköisyysfunktio P tilastollisen frekvenssin perusteella, so. määrittämällä riittävän pitkältä ajanjaksolta syntyvien poikien ja tyttöjen prosenttiosuudet. Suomalaisten tilastojen perusteella saadaan tällöin $P(\text{poika}) \approx 0.51$ ja $P(\text{tyttö}) \approx 0.49$.

■ frekvenssi
■ tilastotiede (matemaattinen)

Tällaisessa tilanteessa puhutaan todennäköisyyden *frekvenssitulkinnasta*.

Tikanheittoon liittyvien todennäköisyyksien määrittäminen ei ole yhtä helppoa. Ne ilmeisesti riippuvat myös heittäjästä. Määrittäminen voisi tapahtua frekvenssitulkinnan mukaisesti pitkän heittosarjan pohjalta. Tällöin jollakin tavalla tilastoitaisiin tietyn heittäjän saamat tulokset ja näiden avulla luonnehdittaisiin hänen todennäköisyysjakaumansa. Luonnehdinnan tulisi olla sellainen, että sen avulla voidaan ainakin jollakin tarkkuudella määrittää esimerkiksi todennäköisyys, että tikka osuu kahdeksikkoon taulussa ylöspäin osoittavassa 45 asteen sektorissa.

Esimerkkejä kombinatorisesta todennäköisyyslaskennasta

Mikäli alkeistapauksia on äärellinen määrä ja on perusteltua pitää näitä yhtä todennäköisinä, johtaa todennäköisyyksien laskeminen usein erilaisten kombinaatioiden lukumäärien laskemiseen.

1) Arpajaisissa on 100 arpaa, joista kymmenen on voittoarpoja. Henkilö ostaa viisi arpaa. Millä todennäköisyydellä hän saa ainakin yhden voiton?

Alkeistapauksia olkoot 100 arvan joukosta poimitut viiden arvan kombinaatiot. Näitä on kaikkiaan $\binom{100}{5} = 75287520$. Kombinaatioita on perusteltua pitää yhtä todennäköisinä, jolloin tietyn kombinaation todennäköisyys on $1/\binom{100}{5} = 1/75287520$.

Tapahtuma, jonka todennäköisyyttä kysytään, on $A =$ 'ainakin yksi voitto'. Tämän komplementtitapahtuma on $\bar{A} =$ 'ei yhtään voittoa'. Tapahtumaan \bar{A} kuuluvia kombinaatioita on $\binom{90}{5} = 43949268$ kappaletta ja siis

$$P(\bar{A}) = 43949268 \cdot \frac{1}{75287520} \approx 0.584.$$

Todennäköisyys saada ainakin yksi voitto on siis $P(A) = 1 - P(\bar{A}) \approx 0.416$.

2) Hyvin sekoitetusta korttipakasta vedetään neljä korttia. Millä todennäköisyydellä ne ovat eri maita?

Keskenään yhtä todennäköisiä alkeistapauksia ovat 52 kortin pakasta otetut neljän kortin kombinaatiot, joiden lukumäärä on $\binom{52}{4}$. Jokaisen kombinaation todennäköisyys on siis $1/\binom{52}{4}$. Eri maita olevien kombinaatioiden määrä on 13^4 , koska jokaisesta maasta kortti voidaan valita 13 tavalla. Todennäköisyys on siten

$$13^4 \cdot \frac{1}{\binom{52}{4}} \approx 0.105.$$

■ lukumäärä
(merkkijonojen)

■ lukumäärä
(merkkijonojen)

■ lukumäärä
(osajoukkojen)

■ lukumäärä
(permutaatioiden)

■ lukumäärä
(osajonojen)

■ lukumäärä
(kombinaatioiden)

■ lukumäärä
(leikkaavien
joukkojen
alkioiden)

Ehdollinen todennäköisyys

Tiettyyn ilmiöön liittyvän todennäköisyysfunktion P määrittämisen ohjenuorana voidaan todennäköisyyden frekvenssitulkinnan mukaisesti pitää tilastollisia kokeita. Haluttaessa määrittää todennäköisyys tapahtumalle B , kun tapahtuma A on sattunut, on luonnollista suorittaa koesarja ja laskea tapaukset, joissa sattuu A ja joissa sattuu sekä A että B . Jos näiden lukumäärät ovat $n(A)$ ja $n(A \cap B)$, todennäköisyyttä approksimoi suhde

■ funktio

■ tilastotiede (matemaattinen)

$$\frac{n(A \cap B)}{n(A)} = \frac{n(A \cap B)/N}{n(A)/N},$$

missä N tarkoittaa sarjan kokeiden kokonaismäärää. Lausekkeen oikean puolen voidaan tulkita approksimoivan tapahtumien $A \cap B$ ja A todennäköisyyksien suhdetta $P(A \cap B)/P(A)$.

Ehdollinen todennäköisyys $P(B|A)$ tapahtumalle B , kun tapahtuma A on sattunut, määritellään tämän johdosta asettamalla

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

Esimerkkinä olkoon seuraava:

Hyvin sekoitetusta korttipakasta vedetään peräkkäin kaksi korttia. Ensimmäinen osoittautuu ässäksi. Millä todennäköisyydellä toinenkin on ässä?

Kahden kortin variaatioita on $52 \cdot 51 = 2652$ kappaletta. On luonnollista pitää näitä yhtä todennäköisinä, jolloin jokaisen todennäköisyys on $1/2652$. Tarkoitakoon A tapahtumaa 'ensimmäinen kortti on ässä, toinen mikä tahansa' sekä B tapahtumaa 'toinen kortti on ässä, ensimmäinen mikä tahansa'. Tapahtumia A ja $A \cap B$ vastaavien korttiparien lukumäärät ovat $4 \cdot 51 = 204$ ja $4 \cdot 3 = 12$, jolloin vastaavat todennäköisyydet ovat

■ variaatio

$$P(A) = 204 \cdot \frac{1}{2652} = \frac{1}{13} \quad \text{ja} \quad P(A \cap B) = 12 \cdot \frac{1}{2652} = \frac{1}{221}.$$

Kysytty ehdollinen todennäköisyys on tällöin

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{12}{204} = \frac{1}{17} \approx 0.059.$$

Tapahtumien riippumattomuus

Tapahtumat A ja B ovat *riippumattomat*, jos $P(B|A) = P(B)$, ts. jos tapahtuman B todennäköisyys ei riipu siitä, onko A sattunut vai ei.

Yhdistämällä tähän ehdollisen todennäköisyyden määritelmä, saadaan riippumattomille tapahtumille kaava $P(A \cap B) = P(A)P(B)$.

Esimerkkinä olkoon ehdollisen todennäköisyyden laskeminen sille, että toisessa nopanheitossa saadaan kuutonen sen jälkeen, kun ensimmäisen heiton tuloksena on ollut kuutonen:

Alkeistapauksia ovat kahdessa heitossa saatavat mahdolliset silmälukuparit. Näitä on $6 \cdot 6 = 36$ kappaletta ja niitä on perusteltua pitää yhtä todennäköisinä jokaisen todennäköisyyden ollessa $1/36$. Jos A tarkoittaa tapahtumaa 'ensimmäinen heitto antoi kuutosen (ja toinen mitä tahansa)' ja B tapahtumaa 'toinen heitto antoi kuutosen (ja ensimmäinen mitä tahansa)', on

$$P(A) = 6 \cdot \frac{1}{36} = \frac{1}{6}, \quad P(A \cap B) = \frac{1}{36}.$$

Ehdollinen todennäköisyys on siis

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{6}.$$

Tämä on sama kuin $P(B)$, kuten luonnollisesti pitää ollakin, koska perättäiset nopanheitot eivät millään tavoin riipu toisistaan.

Koska samaa koetta toistettaessa perättäiset toistot ovat toisistaan riippumattomia, voidaan niiden todennäköisyyksiä laskea helposti: Jos A tarkoittaa tapahtumaa 'ensimmäinen nopanheitto antaa enintään kolmosen' ja B tapahtumaa 'toinen heitto antaa vähintään kolmosen', on todennäköisyys saada ensimmäisellä heitolla 1, 2 tai 3 ja toisella 3, 4, 5 tai 6 kaavan $P(A \cap B) = P(A)P(B)$ mukaisesti

$$\frac{3}{6} \cdot \frac{4}{6} = \frac{1}{3}.$$

Stokastinen muuttuja

Otosavaruudessa Ω määritelty funktio on *stokastinen muuttuja* eli *satunnaismuuttuja*. Yleensä tämä on reaaliarvoinen.

■ funktio

Kahta noppaa heittäessä otosavaruus Ω muodostuu pareista (m, n) , missä m ja n ovat heiton tuloksena saadut silmäluvut, ts. m ja n saavat toisistaan riippumatta arvot 1, 2, 3, 4, 5, 6. Esimerkki stokastisesta muuttujasta on näiden summa: $X(m, n) = m + n$. Stokastinen muuttuja (siis funktio) X saa tällöin äärellisen monta arvoa: 2, 3, ..., 12.

Heittäessä tikkaa tauluun voidaan otosavaruudeksi Ω ottaa xy-taso, jonka origo sijaitsee taulun keskipisteessä. Alkeistapaukset eli otosavaruuden alkiot muodostuvat koordinaattipareista (x, y) , jotka kuvaavat tikan osumiskohtaa. Eräs stokastinen muuttuja on osumiskohdan etäisyys taulun keskipisteestä: $D(x, y) = \sqrt{x^2 + y^2}$. Muuttujan D arvoja ovat kaikki ei-negatiiviset reaalitylvut.

Edellisessä esimerkissä on kyseessä *diskreetti* stokastinen muuttuja, koska se saa vain erillisiä arvoja. Jälkimmäisessä tapauksessa muuttuja on *jatkuva*.

Stokastisten muuttujien argumentit on yleensä tapana jättää kirjoittamatta. Esimerkiksi merkintä $X = x$ tarkoittaa, että stokastinen muuttuja X saa arvon x .

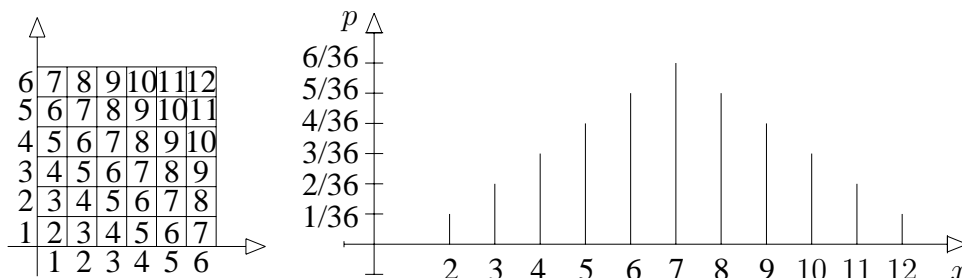
Noppaesimerkissä merkintä $\{X = 5\}$ tarkoittaa otosavaruuden osajoukkoa eli tapahtumaa $\{(m, n) \in \Omega \mid X(m, n) = 5\}$. Vastaavasti tikanheittoesimerkkiin liittyvä joukko $\{D \leq 5\} = \{(x, y) \in \Omega \mid D(x, y) \leq 5\}$ on tapahtuma.

Stokastisten muuttujien saamiin arvoihin liittyvien todennäköisyyksien tarkastelu johtaa tilastollisten jakaumien tarkasteluun.

Seuraavissa kuvissa on kahden nopan heittoon liittyvä otosavaruus ja funktion $X(m, n) = m + n$ sen eri osissa saamat arvot sekä todennäköisyysfunktion $p(x) = P(\{X = x\})$ kuvaaja.

■ kuvaaja

■ jakauma
(diskreetti)



Todennäköisyyslaskenta

ESITIEDOT: ■ joukko-oppi, ■ lukumäärän laskeminen, ■ funktiokäsite

KATSO MYÖS: ■ tilastomatematiikka, ■ todennäköisyysjakaumat

7/7

■ Sisältö

■ Hakemisto

Todennäköisyyslaskennan historiaa

Todennäköisyyslaskennan historia alkaa 1600-luvulta, jolloin ranskalaiset matemaatikot Pierre Fermat (1601 – 1655) ja Blaise Pascal (1623 – 1662) kävivät kirjeenvaihtoa uhkapeleistä. (Kumpikaan ei kuitenkaan itse ollut peluri.) Myöhempiä merkittäviä nimiä ovat sveitsiläinen Jakob Bernoulli (1654 – 1705), ranskalais-sveitsiläinen Abraham de Moivre (1667 – 1754) ja ranskalainen Pierre Simon de Laplace (1749 – 1827). Jokaisella on ansioita myös muilla matematiikan aloilla. Laplace kiteytti todennäköisyyslaskennan siihenastiset tulokset teokseensa *Théorie analytique des probabilités*.

■ Fermat

■ Pascal

■ Bernoulli

■ Bernoulli

■ de Moivre

■ Laplace

Todennäköisyys ymmärrettiin tällöin ns. klassisena todennäköisyytenä, jossa — esimerkiksi jonkin korttipelitalanteen — todennäköisyys määritellään suotuisien tapausten ja kaikkien tapausten lukumäärien suhteena. Tämä edellyttää, että pohjana ovat jotkin alkeistapaukset, joita on perusteltua pitää keskenään yhtä todennäköisinä. Suotuisien tapauksien määrän laskeminen johtaa tällöin yleensä kombinatorisiin tehtäviin, minkä johdosta todennäköisyyslaskenta usein vieläkin mielletään jonkinlaiseksi kombinaatio-opiksi.

Matematiikassa 1800-luku merkitsi voimakasta kehitystä ja erityisesti abstraktimpaan käsittelytapaan siirtymistä. Käsitteet pyrittiin määrittelemään aksiomaattisesti irrallaan sovelluksista, jolloin teoria muodosti oman itsenäisen kokonaisuutensa. Tällä saavutettiin se etu, että samaa teoriaa voitiin soveltaa hyvinkin erilaisiin, mutta periaatteelliselta rakenteeltaan samankaltaisiin ongelmiin.

Seurauksena oli, että klassisen todennäköisyyden käsitettä ei enää pidetty riittävänä eikä todennäköisyyslaskentaa aina edes matematiikkana. Tilanne muuttui 1900-luvun alkupuolella, jolloin myös todennäköisyyslaskenta sai aksiomaattisen perustansa. Todennäköisyysteorian perusteet kiteytti neuvostoliittolainen matemaatikko Andrei Nikolajevitš Kolmogorov (1903 – 1987) vuonna 1933 julkaistussa kirjoituksessaan.

■ aksioma

■ aksioma

■ Kolmogorov