MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

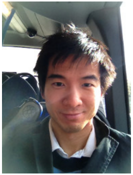
Kurssin järjestelyt

Luennot ma ja pe klo 10–12

- Luennoitsija: Prof Lasse Leskelä
  Vastaanotto ma 14–15 @ Y242a

Harjoitukset viikoittain 2 x 2h + STACK-tehtävät

- Pääassistentti: FM Hoa Ngo
  hoa.ngo@aalto.fi
- Muista ilmoittautua harjoitusryhmään



https://mycourses.aalto.fi/course/view.php?id=22035

Suorittaminen

Kaksi vaihtoehtoa:

(a) Tentti 60% + harjoitukset 40%

(b) Tentti 100%

Harjoituspisteet ovat voimassa 11.4.2019 tentissä

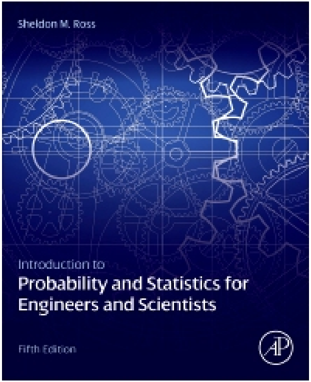
Oppimateriaalit

Stokastiikka ja tilastollinen ajattelu

Versio 0.96

Lasse Leskelä
Aalto-yliopisto

20. syyskuuta 2018



Luentomoniste sähköisenä (saattaa päivittyä kurssin aikana):

http://math.aalto.fi/~lleskela/LectureNotes003.html

Kurssikirja sähköisenä (Aallon verkosta):

http://www.sciencedirect.com.libproxy.aalto.fi/science/book/9780123948113

Www-sivut:

https://mycourses.aalto.fi/course/view.php?id=22035

Osaamistavoitteet

Kurssin suorittanut:

- osaa laskea yhdistelmätapahtumien todennäköisyyksiä hyödyntämällä joukko-opin operaatioita
- tuntee tärkeimmät diskreetit ja jatkuvat todennäköisyysjakaumat sekä tunnistaa tilanteita, joita niillä voi mallintaa
- osaa yhteisjakauman perusteella laskea satunnaisvektorin tunnuslukuja sekä tunnistaa, milloin kaksi satunnaismuuttujaa ovat stokastisesti riippumattomat
- tuntee menetelmiä tilastollisten mallien parametrien estimoimiseen
- osaa laskea yksinkertaisen mallin posteriorijakauman annetusta priorijakaumasta ja havaitusta datasta
- osaa selittää, millaisia johtopäätöksiä voi ja ei voi tehdä valittuun tilastolliseen testiasetelmaan liittyvän p-arvon pohjalta

Työmäärä

- Osallistuminen luennoille 22 h (4 h/viikko)
- Osallistuminen harjoituksiin 24 h (4 h/viikko)
- Viikottainen itsenäinen opiskelu 36–72 h (6–12 h/viikko)
- Osallistuminen ja valmistautuminen kokeisiin 4–40 h

Yhteensä 88–160 h $\approx$ 5 op

Luentorunko

L1A Todennäköisyyden käsite ja laskusäännöt
L1B Satunnaismuuttujat ja todennäköisyysjakaumat

L2A Satunnaismuuttujan odotusarvo ja muunnokset
L2B Keskihajonta ja korrelaatio

L3A Normaaliapproksimaatio
L3B Tilastolliset datajoukot

L4A Parametrien estimointi
L4B Tilastolliset luottamusvälit

L5A Bayesläinen tilastollinen päättely
L5B Frekventistiset vs. bayeslaiset menetelmät

L6A Tilastollisen merkitsevyyden testaaminen
L6B Kertaus (tarvittaessa)

Mitä kurssin jälkeen?

Aalto University

Stochastics and Statistics Courses 2018–2019

MSc course

BSc course

Prequisite course

Recommended order

How to lie with statistics
MS-E1992, Period II, 5 cr
Lecturer: Pauliina Ilmonen

The goal is to learn to spot if there is something fishy in a statistical analysis and to tell the truth, not lies, with statistics. You will work with various problematic data sets. The findings and ideas for improving data analyses are discussed during the lectures. You will also learn to defend your ideas and discoveries by participating in a debate.

Large random systems
MS-E1602, Period IV, 5 cr
Lecturer: Kalle Kytölä

Many important random systems are composed of a large number of simple interacting constituents. In this course you learn to work with such systems using generic mathematical techniques, tightness and weak convergence of probability measures. You also get introduced to probabilistic models including random walks, Brownian motion, percolation, and Ising model.

Multivariate statistical analysis
MS-E2112, Periods III–IV, 5 cr
Lecturer: Pauliina Ilmonen

You get introduced to common multivariate data analysis techniques. You learn to apply these methods in practice using R programming. Course topics include multivariate location and scatter, principal component analysis, multivariate correspondence analysis, canonical correlation analysis, discriminant analysis, classification, and clustering.

Random graphs and network statistics
MS-E1603, Period I, 5 cr
Lecturer: Lasse Leskelä

You get introduced to the theory of statistical models (uniform random graphs, stochastic block models, graphons) used in predicting and learning structural properties of networks based on incomplete and noisy observations. The course is targeted to students in mathematics, operations research, and computer science.

Probability theory
MS-E1600, Period III, 5 cr
Lecturer: Kalle Kytölä

This course introduces you to the mathematical theory of randomness. You learn to work with probabilistic concepts such as stochastic independence, convergence of random sequences, information contained in a sigma-algebra, characteristic and generating functions of probability measures, laws of large numbers, and central limit theorems.

Prediction and time series analysis
MS-C2128, Period II, 5 cr
Lecturer: Pauliina Ilmonen

This course introduces you to methods and concepts used in analyzing and predicting statistical time series. Course topics include linear models and regression diagnostics, stationary random processes, various ARIMA models, Kalman filters, and an introduction to dynamic regression models. R programming is used in the exercises.

Euklidiset avaruudet
MS-C1540, 5 op

Stochastic processes
MS-C2111, Period II, 5 cr
Lecturer: Lasse Leskelä

This course will get you introduced to stochastic processes, the theory of time-dependent random phenomena. You learn to mathematically model and analyze particle and population flows using Markov processes, unpredictable time instants using Poisson processes, and gambling and investment strategies using martingales.

Statistical inference
MS-C1620, Periods III–IV, 5 cr
Lecturer: Joni Virta

This course is an introduction to statistical analysis and statistical inference, with emphasis on statistical hypothesis testing and prediction methods. Course topics include estimation, simple parametric and nonparametric tests, statistical dependence and correlation, linear regression analysis, and analysis of variance. R programming is used in the exercises.

Todennäköisyyslaskennan ja tilastotieteen peruskurssi
MS-A050X, Periodit I, II, III, IV, 5 op
Luennoitsijat: Georg Metsalo, Joni Virta, Ragnar Freij, Lasse Leskelä

Kurssilla opitaan laskemaan todennäköisyyksiä ja odotusarvoja joukkojen, summien ja integraalien avulla. Lisäksi tutustutaan tilastollisiin menetelmiin, joiden avulla voi laatia estimaatteja ja ennusteita sekä analysoida tilastollista merkitsevyyttä havaitun datan ja prioritiedon valossa.

Introduction to R programming
MS-E1994, One-weekend intensive course, 1 cr
Lecturers: Sami Helander, Paavo Raittinen, Niko Lietzén

This course gets you started with R programming. No prior programming experience is required.

Matriisilaskenta
MS-A000X, 5 op

Differentiaali- ja integraalilaskenta 1
MS-A010X, 5 op

Interested in writing a BSc, MSc, or PhD thesis in stochastics and statistics?
Contact Pauliina Ilmonen, Kalle Kytölä, or Lasse Leskelä

Obtain credits by participating in the MS-E1609 Stochastics and statistics seminar?
Join stochastics@list.aalto.fi to stay updated

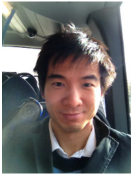
Tervetuloa kurssille!

Luennot ma ja pe klo 10–12

- Luennoitsija: Prof Lasse Leskelä
  Vastaanotto ma 14–15 @ Y242a

Harjoitukset viikoittain 2 x 2h + STACK-tehtävät

- Pääassistentti: FM Hoa Ngo
  hoa.ngo@aalto.fi



https://mycourses.aalto.fi/course/view.php?id=22035

MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

1A Todennäköisyyden käsite ja laskusäännöt

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

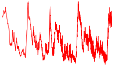
Sisältö

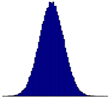
Sisältö

Tilastotiede ja stokastiikka

Tilastotiede on tieteenala, jonka tavoitteena on kehittää menetelmiä valistuneiden arvausten ja päätösten tekemiseen puutteellisen ja epävarman datan pohjalta.

Stokastiikka on sattuman ja todennäköisyyden lakeihin ja malleihin keskittynyt matematiikan osa-alue.





Laskennan ja visualisoinnin tietokonealgoritmit riittävät yksittäisen datajoukon ominaisuuksien tutkimiseen.

Stokastiikan matemaattiset mallit ovat välttämättömiä silloin, kun havaitun datan pohjalta halutaan laatia ennusteita ja yleistyksiä laajempaan kontekstiin.

Todennäköisyyden käsite

Todennäköisyys on tapa kvantifioida uskottavuuksia:

- Kolikkoa heittämällä saadaan kruuna todennäköisyydellä $\frac{1}{2}$

- Ensi ma Otaniemessä sataa todennäköisyydellä
  - 14% (Ilmatieteen laitos)
  - 19% (Foreca)

Todennäköisyyden tulkintoja:

- Objektiivinen (pitkän aikavälin esiintymistiheys)
- Subjektiivinen (uskottavuus tietyn toimijan näkökulmasta)

Todennäköisyyden matemaattiset lait ovat samat tulkinnasta riippumatta

http://www.stat.berkeley.edu/~aldous/Real-World/100.html

Sisältö

Satunnaisilmiö

Satunnaisilmiö on ilmiö, jonka toteumaa ei varmuudella tunneta

- Perusjoukko S sisältää satunnaisilmiön kaikki toteumat
- Toteuma = perusjoukon alkio $s \in S$
- Tapahtuma = perusjoukon osajoukko $A \subset S$

Tulkinta

- Tapahtuma A toteutuu, kun satunnaisilmiön toteuma $s \in A$
- Täysi osajoukko S on varma tapahtuma
- Tyhjä osajoukko $\emptyset$ on mahdoton tapahtuma

Esim. Noppa

- Toteuma i = nopanheiton tulos
- Perusjoukko $S = \{1, 2, \ldots, 6\}$
  - Tapahtumia ovat S:n osajoukot, esim.
    - $A =$ "tulos on parillinen" $= \{2, 4, 6\}$.
    - $B =$ "tulos on suurempi kuin neljä" $= \{5, 6\}$.

Esim. Kaksi nopanheittoa

- Toteuma on lukupari (i, j), jossa i on ensimmäisen ja j toisen heiton tulos
- Perusjoukko on

$$
\begin{aligned}
S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
&(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
&(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
&(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
&(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
&(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}
\end{aligned}
$$

Tapahtumia ovat esim.

- $A =$ "tulokset ovat samat" $= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$.
- $B =$ "ensimmäisen heiton tulos on 1" $= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}$.

Esim. Huomisen sademäärä Otaniemessä (mm)

- Toteumat ovat reaalilukuja $x \geq 0$.
- Perusjoukko $S = [0, \infty)$.

Tapahtumia ovat esim.

- $A =$ "huomenna sataa yli 10 mm" $= (10, \infty)$
- $B =$ "huomenna ei sada" $= \{0\}$

Tapahtumien yhdisteleminen

Perusjoukon tapahtumista voidaan muodostaa uusia tapahtumia loogisin päättelysäännöin:

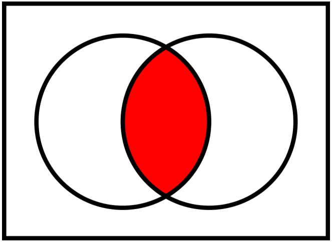
- "A ja B toteutuvat"
- "A tai B toteutuu"
- "A ei toteudu"
- "B toteutuu mutta A ei"

Todennäköisyyslaskentaa varten tapahtumat tulee ilmaista joukko-opin kielellä.

Tapahtumien leikkaus

Tapahtuma "A ja B toteutuvat" sisältää ne toteumat, jotka kuuluvat sekä joukkoon A että joukkoon B:

$$A \cap B = \{s \in S : s \in A \text{ ja } s \in B\}.$$



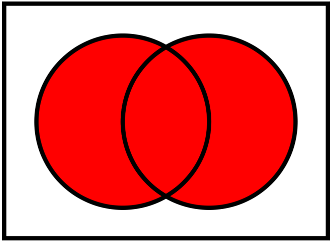
Esim (Noppa)

- $A =$ "Silmäluku on > 3" $= \{4, 5, 6\}$
- $B =$ "Silmäluku on parillinen" $= \{2, 4, 6\}$
- $A \cap B =$ "Silmäluku on > 3 ja parillinen" $= \{4, 6\}$

Tapahtumien yhdiste

Tapahtuma "A tai B toteutuu" sisältää ne toteumat, jotka kuuluvat joukkoon A tai joukkoon B:

$$A \cup B = \{s \in S : s \in A \text{ tai } s \in B\}.$$



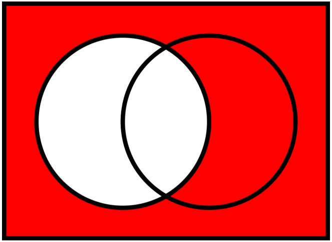
Esim (Noppa)

- $A =$ "Silmäluku on > 3" $= \{4, 5, 6\}$
- $B =$ "Silmäluku on parillinen" $= \{2, 4, 6\}$
- $A \cup B =$ "Silmäluku on > 3 tai parillinen" $= \{2, 4, 5, 6\}$

Tapahtuman vastakohta

Tapahtuma ”A ei toteudu” sisältää ne toteumat, jotka eivät kuulu joukkoon A:

$$A^c = \{s \in S : s \notin A\}.$$



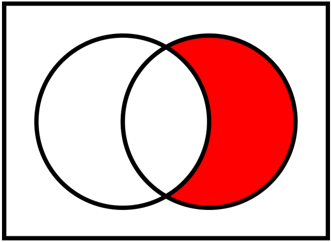
Esim (Noppa)

- $A =$ ”Silmäluku on suurempi kuin 3” $= \{4, 5, 6\}$
- $A^c =$ ”Silmäluku ei ole suurempi kuin 3”
  $=$ ”Silmäluku on korkeintaan 3” $= \{1, 2, 3\}$

Tapahtumien erotus

Tapahtuma "B toteutuu mutta A ei" sisältää ne toteumat, jotka kuuluvat joukkoon B mutta eivät joukkoon A:

$$B \setminus A = \{s \in S : s \in B \text{ ja } s \notin A\}.$$



Esim (Noppa)

- $A =$ "Silmäluku on > 3" $= \{4, 5, 6\}$
- $B =$ "Silmäluku on parillinen" $= \{2, 4, 6\}$
- $B \setminus A =$ "Silmäluku on parillinen ja ≤ 3" $= \{2\}$

Poissulkevat tapahtumat

Tapahtumat A ja B poissulkevat toisensa, jos vain toinen niistä voi toteutua, eli

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

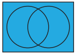
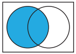
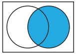
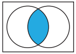
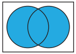
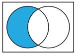
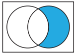
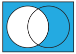
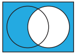
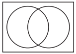
Tapahtumat $A_1, A_2, \ldots$ poissulkevat toisensa, jos vain yksi niistä voi toteutua, eli

$$A_i \cap A_j = \emptyset \quad \text{aina kun } i \neq j.$$

Esim (Noppa)

- $A =$ "Silmäluku on parillinen",
  $B =$ "Silmäluku on kolme tai viisi"
  Tapahtumat A ja B poissulkevat toisensa.
- $A_i =$ "Silmäluku on i".
  Tapahtumat $A_1, A_2, \ldots, A_6$ poissulkevat toisensa.

Tapahtumien yhdisteleminen — Yhteenveto

Termi	Merkintä	Määritelmä	Venn-kaavio	Tulkinta
Perusjoukko	S	$\{x \in S : x \in S\}$		Varma tapahtuma
Osajoukko	A	$\{x \in S : x \in A\}$		A toteutuu
Osajoukko	B	$\{x \in S : x \in B\}$		B toteutuu
Leikkaus	$A \cap B$	$\{x \in S : x \in A \text{ ja } x \in B\}$		A ja B toteutuvat
Yhdiste	$A \cup B$	$\{x \in S : x \in A \text{ tai } x \in B\}$		A tai B toteutuu
Erotus	$A \setminus B$	$\{x \in S : x \in A \text{ ja } x \notin B\}$		A toteutuu mutta B ei
Erotus	$B \setminus A$	$\{x \in S : x \in B \text{ ja } x \notin A\}$		B toteutuu mutta A ei
Komplementti	A^c	$\{x \in S : x \notin A\}$		A ei toteudu
Komplementti	B^c	$\{x \in S : x \notin B\}$		B ei toteudu
Tyhjä joukko	$\emptyset$	$\{x \in S : x \notin S\}$		Mahdoton tapahtuma

Sisältö

Todennäköisyyden aksioomat

Todennäköisyysjakauma eli todennäköisyysmitta perusjoukolla S on kuvaus, joka liittää jokaiseen tapahtumaan $A \subset S$ luvun $\mathrm{P}(A)$, ja toteuttaa:

(i) Varman tapahtuman S todennäköisyys on $\mathrm{P}(S) = 1$.

(ii) Jokaiselle tapahtumalle A pätee $0 \leq \mathrm{P}(A) \leq 1$.

(iii) Mille tahansa äärelliselle tai äärettömälle jonolle toisensa poissulkevia tapahtumia $A_1, A_2, \ldots$ pätee

$$\mathrm{P}(A_1 \cup A_2 \cup \cdots) \; = \; \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots$$

Näitä ominaisuuksia kutsutaan aksioomiksi, koska niistä voidaan johtaa muut todennäköisyyden laskusäännöt.

Todennäköisyyden perussäännöt

- Yleinen summasääntö:

$$P(A \cup B) \; = \; P(A) + P(B) - P(A \cap B).$$

- Poissulkevien summasääntö:

$$P(A \cup B) = P(A) + P(B), \quad \text{kun } A \cap B = \emptyset.$$

- Vastakohdan ja erotuksen todennäköisyys:

$$P(A^c) = 1 - P(A),$$
$$P(B \setminus A) = P(B) - P(A \cap B).$$

- Monotonisuus:

$$P(A) \le P(B), \quad \text{kun } A \subset B.$$

Nämä säännöt voidaan johtaa todennäköisyyden aksioomista.

Sisältö

Ehdollinen todennäköisyys

Tapahtuman A ehdollinen todennäköisyys tapahtuman B toteutuessa määritellään kaavalla

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \quad \mathrm{P}(B) \neq 0.$$

Mikäli $\mathrm{P}(B) = 0$, jätetään $\mathrm{P}(A|B)$ määrittelemättä.

Yleinen tulosääntö

Ehdollisen todennäköisyyden määritelmästä voidaan johtaa:

Laskusääntö

Aina kun $\mathrm{P}(A) \neq 0$, *pätee yleinen tulosääntö*

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

Tulkinta

Yhteistapahtuman "A ja B toteutuvat" todennäköisyys saadaan kertomalla tapahtuman A todennäköisyys tapahtuman B ehdollisella todennäköisyydellä A:n sattuessa.

Monen tapahtuman tulosääntö

Laskusääntö

Aina kun $\mathrm{P}(A_1 \cap \cdots \cap A_{k-1}) \neq 0$, *pätee yleinen tulosääntö*

$$\begin{aligned}&\mathrm{P}(A_1 \cap \cdots \cap A_k)\\ &\quad = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1 \cap A_2) \cdots \mathrm{P}(A_k|A_1 \cap \cdots \cap A_{k-1}).\end{aligned}$$

Tulkinta

Yhteistapahtuman "jokainen tapahtumista $A_1, \ldots, A_k$ sattuu" todennäköisyys saadaan kertomalla keskenään:

- A_1:n todennäköisyys,
- A_2:n ehdollinen tn tapahtuman A_1 sattuessa,
- A_3:n ehdollinen tn tapahtumien A_1 ja A_2 sattuessa,
- ...
- A_k:n ehdollinen tn tapahtumien $A_1, A_2, \ldots, A_{k-1}$ sattuessa.

Tulosääntö — Esimerkki

Nostetaan korttipakasta palauttamatta 3 korttia. Millä todennäköisyydellä kaikki ovat patoja?

- A_i = "i:s kortti on pata"
- $A = A_1 \cap A_2 \cap A_3$

Tapa 1. Yleisen tulosäännön perusteella

$$P(A) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 \cap A_2) = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \approx 0.013.$$

Tapa 2. Vaihtoehtoinen kombinatorinen tapa:

- S = "kolmen kortin järjestämättömät osajoukot", $\#S = \binom{52}{3}$.
- Tapahtuman A toteumat vastaavat kolmen kortin osajoukkoja patojen joukosta. Näitä on $\#A = \binom{13}{3}$ kpl.
- Symmetrian nojalla satunnaisilmiö on tasajakautunut, joten

$$P(A) = \frac{\#A}{\#S} = \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{13 \cdot 12 \cdot 11}{52 \cdot 51 \cdot 50} \approx 0.013.$$

Stokastinen riippuvuus ja riippumattomuus

Tapahtumat A ja B ovat toisistaan riippumattomat, jos

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

Kokoelma tapahtumia $\{A_i, i \in I\}$ on riippumaton, jos

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) \;=\; P(A_{i_1}) \cdots P(A_{i_k})$$

kaikilla $i_1, i_2, \ldots, i_k \in I$.

Esim

Tilanteita, joissa riippumattomuus on intuitiivisesti selvää:

- Perättäiset kolikonheitot, kunhan kolikkoa heitetään riittävän korkealle.
- Otanta palauttaen: nostetaan korista arpalippuja niin, että nostettu lippu palautetaan koriin ja sen jälkeen kori sekoitetaan hyvin.

Riippumattomuus ja ehdollinen todennäköisyys

Fakta

Kun $\mathsf{P}(A) \neq 0$ *ja* $\mathsf{P}(B) \neq 0$, *ovat seuraavat yhtäpitävät:*

- *A ja B ovat riippumattomat.*
- $\mathsf{P}(A|B) = \mathsf{P}(A)$.
- $\mathsf{P}(B|A) = \mathsf{P}(B)$.

Tulkinta

Jos $\mathsf{P}(A|B) \neq \mathsf{P}(A)$, niin informaatiota B:n toteutumisesta voidaan hyödyntää A:n todennäköisyyden määrittämiseen.

Esimerkki: Korttipakka

Nostetaan pakasta satunnainen kortti.

- A = "kortti on pata"
- B = "kortti on ässä"

Ovatko A ja B riippuvat vai riippumattomat?

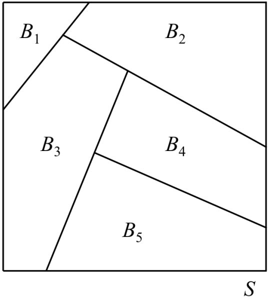
Tarkastetaan laskemalla, päteekö $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$.

- $\mathrm{P}(A) = \frac{13}{52} = \frac{1}{4}$.
- $\mathrm{P}(B) = \frac{4}{52} = \frac{1}{13}$.
- $\mathrm{P}(A \cap B) =$ P("kortti on pataässä") $= \frac{1}{52}$.

Koska $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$, ovat A ja B toisistaan riippumattomat.

Osituskaava

Perusjoukon S ositus on kokoelma toisensa poissulkevia tapahtumia $B_1, \dots, B_n$, joiden yhdiste on S.



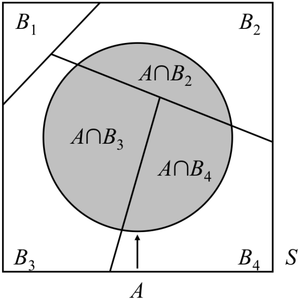
Laskusääntö

Jos $B_1, \dots, B_n$ muodostavat perusjoukon osituksen ja $\mathrm{P}(B_i) \neq 0$ *kaikilla i, niin*

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(B_i)\,\mathrm{P}(A|B_i).$$

Todistus.

Tapahtumat $C_i = A \cap B_i$ poissulkevat toisensa ja niiden yhdiste on A.



Poissulkevien summasäännöstä ja yleisestä tulosäännöstä $\mathrm{P}(A \cap B_i) = \mathrm{P}(B_i)\,\mathrm{P}(A|B_i)$ seuraa

$$\mathrm{P}(A) = \mathrm{P}\left(\bigcup_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} \mathrm{P}(C_i) = \sum_{i=1}^{n} \mathrm{P}(A \cap B_i)$$

$$= \sum_{i=1}^{n} \mathrm{P}(B_i)\,\mathrm{P}(A|B_i).$$

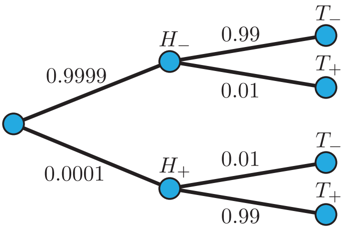
□

Esim. Harvinainen tauti

Erästä tautia esiintyy väestössä suhteessa 1/10000. Taudin toteamiseen on testi, joka tuottaa vääriä positiivisia ja vääriä negatiivisia tn:llä 1%. Millä tn satunnaisen henkilön testitulos on positiivinen?

$$\begin{aligned} H_- &= \text{“henkilö ei sairasta tautia”} & T_- &= \text{“testi on negatiivinen”} \\ H_+ &= \text{“henkilö sairastaa tautia”} & T_+ &= \text{“testi on positiivinen”} \end{aligned}$$

$$\begin{aligned} \text{Osituskaava} \implies \mathrm{P}(T_+) &= \mathrm{P}(H_-)\,\mathrm{P}(T_+ \mid H_-) + \mathrm{P}(H_+)\,\mathrm{P}(T_+ \mid H_+) \\ &= 0.9999 \cdot 0.01 + 0.0001 \cdot 0.99 \\ &= 0.010098. \end{aligned}$$



Bayesin kaava

Kun tunnetaan $\mathrm{P}(A|B)$ sekä $\mathrm{P}(A) \neq 0$ ja $\mathrm{P}(B) \neq 0$, voidaanko määrittää käänteinen ehdollinen todennäköisyys $\mathrm{P}(B|A)$?

Laskusääntö (Bayesin kaava)

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

Todistus.

Ehdollisen todennäköisyyden määritelmästä

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

$\square$

Esim. Harvinainen tauti

Erästä tautia esiintyy väestössä suhteessa 1/10000. Taudin toteamiseen on testi, joka tuottaa vääriä positiivisia ja vääriä negatiivisia tn:llä 1%. Millä tn positiivisen testituloksen saanut henkilö sairastaa tautia?

$$\begin{aligned} H_- &= \text{“henkilö ei sairasta tautia”} & T_- &= \text{“testi on negatiivinen”} \\ H_+ &= \text{“henkilö sairastaa tautia”} & T_+ &= \text{“testi on positiivinen”} \end{aligned}$$

Aiemmin laskettiin $\mathrm{P}(T_+) = 0.010098$. Bayesin kaava $\implies$

$$\mathrm{P}(H_+ \mid T_+) = \frac{\mathrm{P}(H_+)\,\mathrm{P}(T_+ \mid H_+)}{\mathrm{P}(T_+)} = \frac{0.0001 \cdot 0.99}{0.010098} \approx 0.0098.$$

Onko tässä jotain outoa?

Esiintyvyysharha:

- Kaikista testituloksista 99% on oikeita
- Positiivisista testituloksista yli 99% on vääriä

Esimerkki: Tehtaan laadunvalvonta

Samaa tuotetta valmistetaan kahdella tuotantolinjalla. Valmiit tuotteet sekoitetaan ja pakataan laatikoihin.

- Linjalta 1 valmistuu 3 tuotetta/min, joista 2% on viallisia.
- Linjalta 2 valmistuu 5 tuotetta/min, joista 9% on viallisia.

Tarkastetaan satunnaisesta laatikosta satunnaisesti valittu tuote.

- Millä todennäköisyydellä tarkastettava tuote on linjalta 1?
- Millä todennäköisyydellä vialliseksi havaittu tuote on linjalta 1?

Esimerkki: Tehtaan laadunvalvonta — Ratkaisu

- Linjalta 1 valmistuu 3 tuotetta/min, joista 2% on viallisia.
- Linjalta 2 valmistuu 5 tuotetta/min, joista 9% on viallisia.

Tunnetut todennäköisyydet:

- L_1 = "Tuote on linjalta 1', $\mathrm{P}(L_1) = 3/8$
- L_2 = "Tuote on linjalta 2", $\mathrm{P}(L_2) = 5/8$
- V = "Tuote on viallinen", $\mathrm{P}(V|L_1) = 0.02$, $\mathrm{P}(V|L_2) = 0.09$

Tapahtumat L_1 ja L_2 muodostavat perusjoukon osituksen, joten osituskaavalla

$$\begin{aligned}\mathrm{P}(V) &= \mathrm{P}(V \mid L_1)\,\mathrm{P}(L_1) + \mathrm{P}(V \mid L_2)\,\mathrm{P}(L_2) \\ &= 0.02 \times \frac{3}{8} + 0.09 \times \frac{5}{8} = 6.375\%,\end{aligned}$$

ja Bayesin kaavalla

$$\mathrm{P}(L_1|V) = \frac{\mathrm{P}(V|L_1)\,\mathrm{P}(L_1)}{\mathrm{P}(V)} = \frac{0.02 \times 3/8}{0.06375} \approx 11.8\%.$$

Esimerkki: Tehtaan laadunvalvonta — Yhteenveto

Samaa tuotetta valmistetaan kahdella tuotantolinjalla. Valmiit tuotteet sekoitetaan ja pakataan laatikoihin.

- Linjalta 1 valmistuu 3 tuotetta/min, joista 2% on viallisia.
- Linjalta 2 valmistuu 5 tuotetta/min, joista 9% on viallisia.

Tarkastettavan tuotteen alkuperän prioritodennäköisyydet ovat:

- Tuote on linjalta 1 tn:llä $3/8 = 37.5$ %
- Tuote on linjalta 2 tn:llä $5/8 = 62.5$ %

Tarkastettavan tuotteen alkuperän posterioritodennäköisyydet (sen jälkeen kun tuote on havaittu vialliseksi) ovat:

- Tuote on linjalta 1 tn:llä $\approx 11.8\%$
- Tuote on linjalta 2 tn:llä $\approx 88.2\%$

Todennäköisyyden laskusäännöt — Yhteenveto

Summasääntö

$$\begin{aligned} \mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) \\ &= \mathrm{P}(A) + \mathrm{P}(B) \quad \text{(jos A ja B poissulkevat toisensa)} \end{aligned}$$

Tulosääntö

$$\begin{aligned} \mathrm{P}(A \cap B) &= \mathrm{P}(A)\,\mathrm{P}(B|A) \\ &= \mathrm{P}(A)\,\mathrm{P}(B) \quad \text{(jos A ja B riippumattomat)} \end{aligned}$$

Osituskaava

$$\mathrm{P}(A) = \sum_i \mathrm{P}(B_i)\,\mathrm{P}(A|B_i) \quad \text{(jos B_i:t muodostavat osituksen)}$$

Bayesin kaava

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A|B)\,\mathrm{P}(B)}{\mathrm{P}(A)}$$

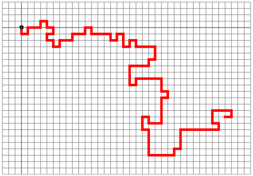
Sisältö

Todennäköisyys ja kombinatoriikka

Jos äärellisen perusjoukon S jokainen toteuma on yhtä todennäköinen, on tapahtuman $A \subset S$ todennäköisyys

$$\mathrm{P}(A) \;=\; \frac{\#A}{\#S} \;=\; \frac{\text{tapahtuman } A \text{ toteumien lkm}}{\text{kaikkien toteumien lkm}}.$$

- Suuressa perusjoukossa voi lukumäärien $\#A$ ja $\#S$ laskeminen olla vaikeaa, ellei jopa mahdotonta.
- Kombinatoriikka on tämäntyyppisiin ongelmiin keskittynyt matematiikan osa-alue.



Esim (Vaikea kombinatorinen ongelma)

Millä todennäköisyydellä 10^8 askeleen itseään välttävä satunnaiskulku päättyy etäisyydelle 10^6 lähtöpisteestään?

Listojen lukumäärä (toistojen kanssa)

Montako PIN-koodia voidaan numeroista $\{0, 1, 2, \ldots, 9\}$ muodostaa?

Muodostetaan kaikki PIN-koodit neljässä vaiheessa:

1. Valitaan ensimmäinen PIN-koodin numero: 10 tapaa
2. Valitaan seuraava PIN-koodin numero: 10 tapaa
3. Valitaan seuraava PIN-koodin numero: 10 tapaa
4. Valitaan seuraava PIN-koodin numero: 10 tapaa

$\implies$ Yhteensä $10^4 = 10000$ tapaa

0000, 0001, 0002, 0003, 0004, 0005, 0006, 0007, 0008, 0009, 0010, 0011, 0012, 0013, 0014, 0015, 0016, 0017, 0018, 0019, 0020, 0021, 0022, 0023, 0024, 0025, 0026, 0027, 0028, 0029, 0030, 0031, 0032, 0033, 0034, 0035, 0036, 0037, 0038, 0039, 0040, 0041, 0042, 0043, 0044, 0045, 0046, 0047, 0048, 0049, 0050, 0051, 0052, 0053, 0054, 0055, 0056, 0057, 0058, 0059, 0060, 0061, 0062, 0063, 0064, 0065, 0066, 0067, 0068, 0069, 0070, 0071, 0072, 0073, 0074, 0075, 0076, 0077, 0078, 0079, 0080, 0081, 0082, 0083, 0084, 0085, 0086, 0087, 0088, 0089, 0090, 0091, 0092, 0093, 0094, 0095, 0096, 0097, 0098, 0099, 0100, 0101, 0102, 0103, 0104, 0105, 0106, 0107, 0108, 0109, 0110, 0111, 0112, 0113, 0114, 0115, 0116, 0117, 0118, 0119, 0120, 0121, 0122, 0123, 0124, 0125, 0126, 0127, 0128, 0129, 0130, 0131, 0132, 0133, 0134, 0135, 0136, 0137, 0138, 0139, 0140, 0141, 0142, 0143, 0144, 0145, 0146, 0147, 0148, 0149, 0150, 0151, 0152, 0153, 0154, 0155, 0156, 0157, 0158, 0159, 0160, 0161, 0162, 0163, 0164, 0165, 0166, 0167, 0168, 0169, 0170, 0171, 0172, 0173, 0174, 0175, 0176, 0177, 0178, 0179, 0180, 0181, 0182, 0183, 0184, 0185, 0186, 0187, 0188, 0189, 0190, 0191, 0192, 0193, 0194, 0195, 0196, 0197, . . . , 9983, 9984, 9985, 9986, 9987, 9988, 9989, 9990, 9991, 9992, 9993, 9994, 9995, 9996, 9997, 9998, 9999

Listojen lukumäärä (ilman toistoja)

Monellako tapaa on mahdollista jakaa mitalisijat SM-liigassa pelaavien 15 joukkueen (HPK, IFK, ILV, JUK, JYP, KAL, KÄR, KOO, LUK, PEL, SAI, SPO, TAP, TPS, ÄSS) kesken?

Muodostetaan kaikki mitalisijakombinaatiot kolmessa vaiheessa:

1. Valitaan sijalle 1 jokin joukkue: 15 tapaa
2. Valitaan sijalle 2 jokin vielä sijoittamaton joukkue: 14 tapaa
3. Valitaan sijalle 3 jokin vielä sijoittamaton joukkue: 13 tapaa

$\Longrightarrow$ Yhteensä $15 \times 14 \times 13 = 2730$ tapaa

(HPK,IFK,ILV), (HPK,IFK,JUK), (HPK,IFK,JYP), (HPK,IFK,KAL), (HPK,IFK,KÄR), (HPK,IFK,KOO), (HPK,IFK,LUK),
(HPK,IFK,PEL), (HPK,IFK,SAI), (HPK,IFK,SPO), (HPK,IFK,TAP), (HPK,IFK,TPS), (HPK,IFK,ÄSS), (HPK,ILV,IFK),
(HPK,ILV,JUK), (HPK,ILV,JYP), (HPK,ILV,KAL), (HPK,ILV,KÄR), (HPK,ILV,KOO), (HPK,ILV,LUK), (HPK,ILV,PEL),
(HPK,ILV,SAI), (HPK,ILV,SPO), (HPK,ILV,TAP), (HPK,ILV,TPS), (HPK,ILV,ÄSS), (HPK,JUK,IFK), (HPK,JUK,ILV),
(HPK,JUK,JYP), (HPK,JUK,KAL), (HPK,JUK,KÄR), (HPK,JUK,KOO), (HPK,JUK,LUK), (HPK,JUK,PEL), (HPK,JUK,SAI),
(HPK,JUK,SPO), (HPK,JUK,TAP), (HPK,JUK,TPS), (HPK,JUK,ÄSS), (HPK,JYP,IFK), (HPK,JYP,ILV), (HPK,JYP,JUK),
(HPK,JYP,KAL), (HPK,JYP,KÄR), (HPK,JYP,KOO), (HPK,JYP,LUK), (HPK,JYP,PEL), (HPK,JYP,SAI), (HPK,JYP,SPO),
(HPK,JYP,TAP), (HPK,JYP,TPS), (HPK,JYP,ÄSS), (HPK,KAL,IFK), (HPK,KAL,ILV), (HPK,KAL,JUK), (HPK,KAL,JYP),
(HPK,KAL,KÄR), (HPK,KAL,KOO), (HPK,KAL,LUK), (HPK,KAL,PEL), (HPK,KAL,SAI), (HPK,KAL,SPO), (HPK,KAL,TAP),
(HPK,KAL,TPS), (HPK,KAL,ÄSS), (HPK,KÄR,IFK), (HPK,KÄR,ILV), (HPK,KÄR,JUK), (HPK,KÄR,JYP), (HPK,KÄR,KAL),
(HPK,KÄR,KOO), (HPK,KÄR,LUK), (HPK,KÄR,PEL), (HPK,KÄR,SAI), (HPK,KÄR,SPO), (HPK,KÄR,TAP), (HPK,KÄR,TPS),
(HPK,KÄR,ÄSS), (HPK,KOO,IFK), (HPK,KOO,ILV), (HPK,KOO,JUK), (HPK,KOO,JYP), (HPK,KOO,KAL), (HPK,KOO,KÄR),
(HPK,KOO,LUK), (HPK,KOO,PEL), (HPK,KOO,SAI), (HPK,KOO,SPO), (HPK,KOO,TAP), (HPK,KOO,TPS), (HPK,KOO,ÄSS),
(HPK,LUK,IFK), (HPK,LUK,ILV), (HPK,LUK,JUK), (HPK,LUK,JYP), (HPK,LUK,KAL), (HPK,LUK,KÄR), (HPK,LUK,KOO),
(HPK,LUK,PEL), (HPK,LUK,SAI), (HPK,LUK,SPO), (HPK,LUK,TAP), (HPK,LUK,TPS), (HPK,LUK,ÄSS), (HPK,PEL,IFK),
(HPK,PEL,ILV), (HPK,PEL,JUK), (HPK,PEL,JYP), (HPK,PEL,KAL), (HPK,PEL,KÄR), (HPK,PEL,KOO), (HPK,PEL,LUK),
(HPK,PEL,SAI), (HPK,PEL,SPO), (HPK,PEL,TAP), (HPK,PEL,TPS), (HPK,PEL,ÄSS), (HPK,SAI,IFK), ...

Listojen lukumäärä — Yhteenveto

Fakta

Järjestettyjä k alkion listoja voidaan n alkion joukosta muodostaa:

- *toistojen kanssa n^k kappaletta,*
- *ilman toistoja $n(n-1)\cdots(n-k+1)$ kappaletta.*

Esim (PIN-koodit)

Järjestettyjä $k = 4$ numeron listoja (toistojen kanssa) voidaan $n = 10$ numeron joukosta muodostaa 10^4 kappaletta

Esim (SM-liigan mitalisijat)

Järjestettyjä $k = 3$ joukkueen listoja (ilman toistoja) voidaan $n = 15$ joukkueen joukosta muodostaa
$15(15-1)\cdots(15-3+1) = 15 \times 14 \times 13 = 2730$ kappaletta

Järjestysten lukumäärä

Monellako tapaa voidaan SM-liigan kaikki 15 joukkuetta asettaa paremmuusjärjestykseen?

Järjestettyjä $k = 15$ joukkueen listoja voidaan $n = 15$ liigajoukkueen joukosta muodostaa $n(n-1)\cdots(n-k+1)$ eli $15 \times 14 \times \cdots \times 1$ kappaletta

Fakta

n alkiota voidaan järjestää listaan $n! = n(n-1)\cdots 1$ tavalla.

Esim (SM-liigan sarjataulukko)

SM-liigan 15 joukkuetta voidaan asettaa paremmuusjärjestykseen $15! \approx 1.3 \times 10^{12}$ tavalla.

Osajoukkojen lukumäärä

Kuinka monta eri pelaajaviisikkoa voidaan jääkiekkojoukkueen 20 kenttäpelaajan joukosta muodostaa?

20 pelaajan joukosta voidaan muodostaa 5 eri pelaajan järjestettyjä listoja $20 \times 19 \times 18 \times 17 \times 16 = 1\ 860\ 480$ kappaletta.

Pelaajaviisikko on sama huolimatta siitä, miten sen pelaajat järjestää. $\Longrightarrow$ Jokaista pelaajaviisikkoa vastaa $5! = 120$ järjestettyä 5 eri pelaajan listaa

Pelaajaviisikkojen lukumäärä on

$$\frac{20 \times 19 \times 18 \times 17 \times 16}{5!} = \frac{1\ 860\ 480}{120} = 15\ 504$$

Osajoukkojen lukumäärä — Yleistys

Fakta

Järjestämättömiä k alkion osajoukkoja voidaan n alkion joukosta muodostaa

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

kappaletta.

Esim. Lottoarvonta

Mikä on todennäköisyys saada 7 oikein yhdellä lottorivillä?

- Veikkaus Oy:n lottoarvonnan perusjoukko on

$$S = \text{“7:n alkion osajoukot joukosta } \{1, \dots, 40\}\text{”}$$

ja sen koko on $\#S = \binom{40}{7} = 18\,643\,560$.

- Tapahtuma

$$A = \text{“valitulla lottorivillä 7 oikein”}$$

sisältää täsmälleen yhden toteuman, joten $\#A = 1$.

- Symmetrian perusteella lottoarvonta on tasajakautunut, joten

$$\mathrm{P}(A) = \frac{\#A}{\#S} = \frac{1}{18\,643\,560}$$

Esim. Johtoryhmä

Viiden hengen johtoryhmään haki 6 miestä 10 naista. Jos johtoryhmä valittaisiin arpomalla, niin millä tn valituksi tulisi 3 miestä ja 2 naista?

Perusjoukko $S =$ "viisikot 16 henkilön hakijajoukosta"

$$\#S = \binom{16}{5}.$$

Tapahtumaa $A =$ "valitaan 3 miestä ja 2 naista" vastaavat henkilökombinaatiot voidaan muodostaa seuraavasti:

1. Valitaan ensin 3 miestä 6 miehen joukosta: $\binom{6}{3}$ tapaa
2. Valitaan 2 naista 10 naisen joukosta: $\binom{10}{2}$ tapaa

$$\#A = \binom{6}{3}\binom{10}{2}$$

Kysytty todennäköisyys on

$$\mathrm{P}(A) = \frac{\#A}{\#S} = \frac{\binom{6}{3}\binom{10}{2}}{\binom{16}{5}} = \frac{900}{4368} \approx 20.6\%.$$

Esim. Kolmoset pokerissa

Mikä on todennäköisyys saada “kolmoset” viidellä kortilla?
“kolmoset” $=$ 5 kortin (järjestämätön) joukko, jossa esiintyy 3 arvoa: yksi kolmesti ja muut kerran. Muodostetaan kaikki “kolmosia” vastaavat korttiviisikot:

1. Valitaan kolmesti esiintyvä arvo a kaikista arvoista: 13 tapaa
2. Valitaan kerran esiintyvien arvojen (järjestämätön) joukko käyttämättömistä arvoista: $\binom{12}{2} = 66$ tapaa
3. Valitaan kolmen kortin joukko arvon a korteista: $\binom{4}{3} = 4$ tapaa
4. Valitaan yhden kortin joukko pienempää kerran esiintyvää arvoa olevista korteista: $\binom{4}{1} = 4$ tapaa
5. Valitaan yhden kortin joukko suurempaa kerran esiintyvää arvoa olevista korteista: $\binom{4}{1} = 4$ tapaa

$$\Longrightarrow \; \mathrm{P}(\text{“kolmoset”}) \;=\; \frac{13\binom{12}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} \;=\; \frac{54\ 912}{2\ 598\ 960} \;\approx\; 2.1\%.$$

Seuraavalla luennolla puhutaan satunnaismuuttujista ja niiden jakaumista. . .