

MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

Jukka Kohonen

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2022–2023
Periodi IV

Kurssin järjestelyt

Luennot ma ja pe klo 10–12

- Luennoitsija: Yliop.leht. **Jukka Kohonen**
Vastaanotto sopimuksen mukaan



Harjoitukset viikoittain 2 x 2h +
STACK-tehtävät

- Pääassistentti: **Joonas Vättö**
joonas.vatto@aalto.fi
- Muista ilmoittautua kurssille
- Käy **omassa ryhmässäsi**, ellet ole sopinut toisin
- Harjoitusten käytännön kysymyksissä kysy omalta assarilta tai pääassistentilta

<https://mycourses.aalto.fi/course/view.php?id=36208>

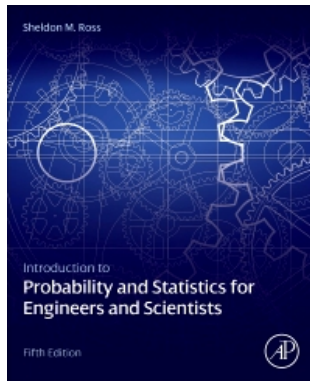
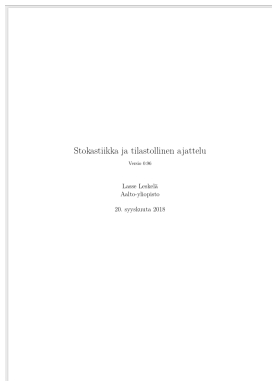
Suorittaminen

Kaksi vaihtoehtoa:

- (a) Tentti 60% + harjoitukset 40%
- (b) Tentti 100%

- Jokaiselle lasketaan arvosana molemmilla tavoilla ja valitaan parempi. Tätä ei tarvitse erikseen pyytää eikä valita.
- Kummassakin tavassa 50% pisteistä riittää kurssin läpäisyyn.
- Harjoituspisteet ovat voimassa kurssitentissä ja seuraavassa tentissä (uusintatenteistä ks. Sisu).

Oppimateriaalit



Luentomoniste sähköisenä (saattaa päivittyä kurssin aikana):

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

Oheiskirja sähköisenä (Aallon verkosta):

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

Www-sivut:

<https://mycourses.aalto.fi/course/view.php?id=36208>

Osaamistavoitteet

Kurssin suorittanut:

- osaa laskea yhdistelmä tapahtumien 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 posteriorijakaumia ja tehdä niistä päätelmiä
- 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 harjoitukseen 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

Itsenäinen opiskelu on matematiikassa olennaista. Siis:

- luentomonisteen lukemista ajatuksen kanssa
- tehtävien ahkeraa tekemistä
- myös tehtävien ratkaisuihin tutustumista
- kysymysten esittämistä opettajille

Luentorunko

L1A Todennäköisyyden käsite ja laskusäännöt

L1B Satunnaismuuttujat ja jakaumat

L2A Odotusarvo ja muunnokset

L2B Keskihajonta ja korrelaatio

L3A Satunnaismuuttujien summa ja keskiarvo

L3B Tilastolliset datajoukot

L4A Parametrien estimointi

L4B Tilastolliset luottamusvälit

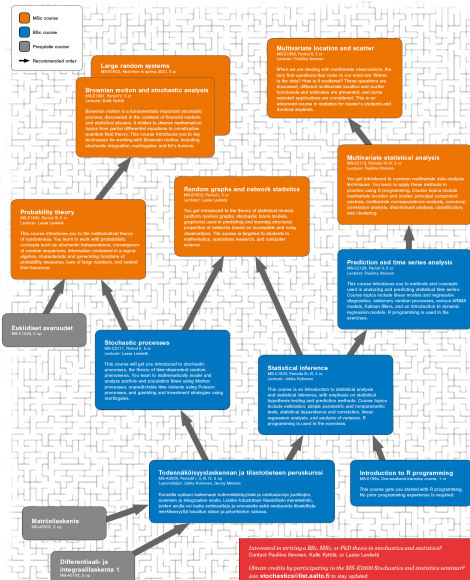
L5A Bayesläinen tilastollinen päättely

L5B Lisää Bayes-päätteleystä

L6A Tilastollisen merkitsevyyden testaaminen

L6B Kertaus ja yhteenveto

Mitä kurssin jälkeen?



MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

1A Todennäköisyyden käsite ja laskusäännöt

Jukka Kohonen

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2022–2023
Periodi IV

Sisältö

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

Sisältö

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

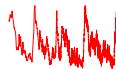
Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

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:

- Objektiiivinen / Frekvenssi (pitkän aikavälin esiintymistiheys)
- Subjektiiivinen (uskottavuus tietyn toimijan näkökulmasta)

Todennäköisyyden matemaattiset lait ovat samat tulkinnasta riippumatta, ja tulkinnoilla on muutenkin paljon yhteistä.

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

Sisältö

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

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 $S = \{1, 2, \dots, 6\}$
- 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

$$S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), \\ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$



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.

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

Tapahtumien yhdisteleminen

Perusjoukon tapahtumista voidaan muodostaa uusia tapahtumia loogisin päättelysäännöin:

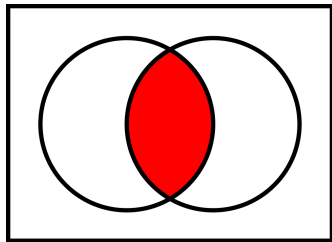
- " A ja B toteutuvat"
- " A tai B toteutuu"
- " A ei toteudu"
- " B toteutuu mutta A ei"

Todennäköisyyslaskentaa varten tapahtumat tulee ilmaista joukko-opin kielellä.

Tapahntumien leikkaus

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

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



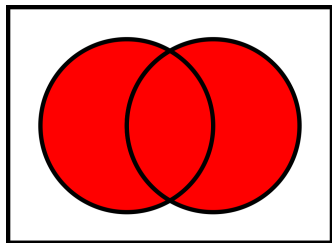
Esim (Noppa)

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

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)

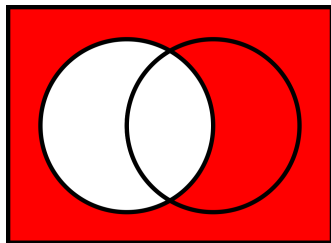
- $A = \text{"Silmäluku on } > 3\text{"} = \{4, 5, 6\}$
- $B = \text{"Silmäluku on parillinen"} = \{2, 4, 6\}$
- $A \cup B = \text{"Silmäluku on } > 3 \text{ tai parillinen"} = \{2, 4, 5, 6\}$

Huom. Matematiikassa ja logiikassa "tai" ymmärretään yleensä inklusiivisena, eli se sallii myös molempien vaihtoehtojen toteutumisen.

Tapahtuman vastakohta (komplementti)

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

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



Esim (Noppa)

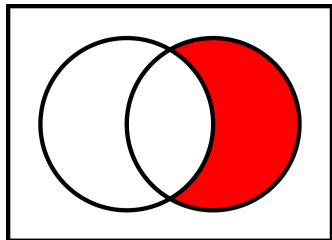
- $A =$ "Silmäluku on suurempi kuin 3" $= \{4, 5, 6\}$
- $A^c =$ "Silmäluku **ei** ole suurempi kuin 3"
 $=$ "Silmäluku on korkeintaan 3" $= \{1, 2, 3\}$

Huom. Tarkkana epäyhtälöiden kanssa. Tapahtuman "suurempi kuin 3" vastakohta ei ole "pienempi kuin 3" vaan "enintään 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)

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

Poissulkevat tapahtumat

Kaksi tapahtumaa A ja B **poissulkevat toisensa** (eli ovat *erilliset*), jos vain toinen niistä voi toteutua, eli

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











Monta tapahtumaa A_1, A_2, \dots **poissulkevat toisensa**, jos vain yksi niistä voi toteutua, eli

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

Esim (Nopanheitto)

- $A = \{1, 2\}$ ja $B = \{3, 4\}$ poissulkevat toisensa.
- $A = \{1, 2, 3\}$ ja $B = \{2, 4, 6\}$ eivät poissulje toisiaan.
- $A = \{1, 2\}$, $B = \{3, 4\}$ ja $C = \{5, 6\}$ poissulkevat toisensa.

Taphtumien 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

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

Todennäköisyyden aksioomat

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

- (i) Varman tapahtuman S todennäköisyys on $\mathbb{P}(S) = 1$.
- (ii) Jokaiselle tapahtumalle A pätee $0 \leq \mathbb{P}(A) \leq 1$.
- (iii) Mille tahansa äärelliselle tai äärettömälle jonolle toisensa poissulkevia tapahtumia A_1, A_2, \dots pätee

$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$$

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

Todennäköisyyden perussäännöt

- Yleinen summasääntö:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- Poissulkevien summasääntö:

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

- Vastakohtan ja erotuksen todennäköisyys:

$$\begin{aligned}\mathbb{P}(A^c) &= 1 - \mathbb{P}(A), \\ \mathbb{P}(B \setminus A) &= \mathbb{P}(B) - \mathbb{P}(A \cap B).\end{aligned}$$

- Monotonisuus:

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

Nämä säännöt voidaan johtaa todennäköisyyden aksioomista.

Sisältö

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

Ehdollinen todennäköisyys

Tapahtuman B ehdollinen todennäköisyys tapahtuman A toteutuessa määritellään kaavalla

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

- Luetaan “ B :n todennäköisyys ehdolla A ”
- Tarkoittaa B :n todennäköisyyttä *siinä tapauksessa*, että A toteutuu. Jos A ei toteudu, B :llä voi silloin olla eri todennäköisyys.
- Huomaa, että toisiaan muistuttavat merkinnät $\mathbb{P}(A \cap B)$, $\mathbb{P}(A|B)$ ja $\mathbb{P}(B|A)$ tarkoittavat aivan eri asioita!
- Mikäli $\mathbb{P}(A) = 0$, jätetään $\mathbb{P}(B|A)$ määrittelemättä.

Yleinen tulosääntö

Ehdollisen todennäköisyyden määritelmästä voidaan johtaa:

Laskusääntö

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

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

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 $\mathbb{P}(A_1 \cap \dots \cap A_{k-1}) \neq 0$, pätee *yleinen tulosääntö*

$$\begin{aligned}\mathbb{P}(A_1 \cap \dots \cap A_k) \\ = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_k|A_1 \cap \dots \cap A_{k-1}).\end{aligned}$$

Tulkinta

Yhteistapahtuman ”jokainen tapahtumista A_1, \dots, 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, \dots, 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

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

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

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

Stokastinen riippuvuus ja riippumattomuus

Tapahtumat A ja B ovat toisistaan **riippumattomat**, jos

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

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

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k})$$

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

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

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

Tulkinta

Jos $\mathbb{P}(A|B) \neq \mathbb{P}(A)$, niin informaatio B :n toteutumisesta vaikuttaa A :n todennäköisyyteen (joko suurentavasti tai pienentävästi).

Esimerkki: Korttipakka

Nostetaan pakasta satunnainen kortti.

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

Ovatko A ja B riippuvat vai riippumattomat?



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

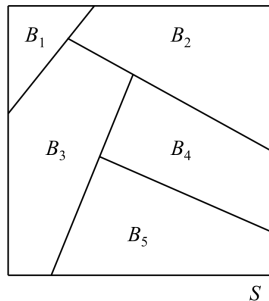
- $\mathbb{P}(A) = \frac{13}{52} = \frac{1}{4}$.
- $\mathbb{P}(B) = \frac{4}{52} = \frac{1}{13}$.
- $\mathbb{P}(A \cap B) = \mathbb{P}(\text{"kortti on pataässä"}) = \frac{1}{52}$.

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

Osituskaava

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

(Samoin voidaan määritellä minkä tahansa joukon ositus.)



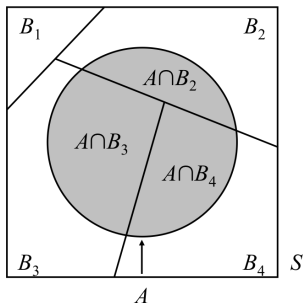
Laskusääntö

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

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

Todistus.

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



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

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n \mathbb{P}(C_i) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) \\ &= \sum_{i=1}^n \mathbb{P}(B_i)\mathbb{P}(A|B_i).\end{aligned}$$

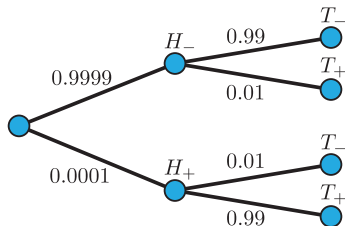


Esim. Harvinainen tauti

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

H_- = "henkilö ei sairasta tautia" T_- = "testi on negatiivinen"
 H_+ = "henkilö sairastaa tautia" T_+ = "testi on positiivinen"

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



Bayesin kaava

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

Laskusääntö (Bayesin kaava)

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

Todistus.

Ehdollisen todennäköisyyden määritelmästä

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



Esim. Harvinainen tauti

Erästä tautia esiintyy väestössä suhteessa $1/10000$. Taudin toteamiseen on testi, joka tuottaa väriä positiivisia ja väriä negatiivisia tnl:llä 1%. Millä tnl 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 $\mathbb{P}(T_+) = 0.010098$. Bayesin kaava \implies

$$\mathbb{P}(H_+ | T_+) = \frac{\mathbb{P}(H_+)\mathbb{P}(T_+ | H_+)}{\mathbb{P}(T_+)} = \frac{0.0001 \cdot 0.99}{0.010098} \approx 0.0098.$$

Onko tässä jotain outoa?

Esiintyvyysharha:

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

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

Tapahtumat L_1 ja L_2 muodostavat perusjoukon osituksen, joten osituskaavalla

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

ja Bayesin kaavalla

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

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

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

Tulosääntö

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

Osituskaava

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

Bayesin kaava

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

Sisältö

Johdanto

Satunnaisilmiön tapahtumat

Todennäköisyyden laskusäännöt

Ehdollinen todennäköisyys

Todennäköisyys ja kombinatoriikka

Todennäköisyys ja kombinatoriikka

Jos äärellisessä perusjoukossa S on n eri toteumaa ja ne ovat **yhtä todennäköiset**, niin kunkin toteuman todennäköisyys on $1/n$.

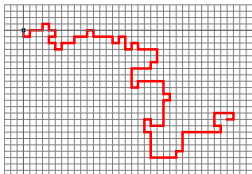
Jos sitten tapahtumaan A kuuluu k toteumaa, niin summasäännön mukaisesti

$$\mathbb{P}(A) = \frac{k}{n} = \frac{\#A}{\#S} = \frac{\text{tapahtuman } A \text{ toteumien lkm}}{\text{kaikkien toteumien lkm}}.$$

Tällaisessa **symmetrisessä** tapauksessa todennäköisyyslaskenta on periaatteessa hyvin helppoa, mutta ...

- Suuria lukumääriä on vaikea laskea yksitellen (“count”).
- Usein lukumääriä voi laskea (“calculate”) tehokkailla kaavoilla.
- Olennaista on ymmärtää, että molemmilla tavoilla tulee periaatteessa sama tulos.
- **Kombinatoriikka** on tämän tyyppisiin ongelmiin keskittynyt matematiikan osa-alue. Aina ei kyllä ole helppoja kaavojakaan.

Kombinatoriikkakaan ei ole aina helppoa



Esim (Vaikea kombinatorinen ongelma)

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

Kombinatoriikan perusperiaatteita

Vaiheittain eteneminen ja **summaperiaate**.

- Tutki, millä eri tavoilla prosessin 1. vaihe voi toteutua.
- Kussakin tapauksessa tutki erikseen, miten prosessi voi jatkua.
Huomioi: vaikuttaako 1. vaiheessa tehty valinta toiseen vaiheeseen vai ei (esim. sulkemalla pois joitakin vaihtoehtoja)?
- Lopuksi laske vaihtoehdot yhteen (summaperiaate).

Tuloperiaate. Jos ensimmäisessä vaiheessa on n vaihtoehtoa, ja jokainen niistä johtaa toisessa vaiheessa m vaihtoehtoon, niin lopputuloksia on $m + m + \dots + m = n \cdot m$.

Kombinatoriikan perusongelmia

- **Listojen** eli järjestettyjen jonojen lukumäärän laskeminen (annetuista alkioista)
 - kun alkion toistaminen sallitaan
 - kun alkion toistamista ei sallita
- **Järjestysten** laskeminen (annetusta listasta)
- **Osajoukkojen** lukumäärän laskeminen (alkion toistamista ei sallita ja eri järjestyksiä pidetään samoina)

Nämä kaikki voidaan järkeillä tuloperiaatteen avulla. Tuloksena on hyvin tunnettuja kaavoja kuten potenssi, laskeva tulo, kertoma ja binomikerroin.

Opettele sekä kaavat (perustilanteisiin) että taustalla oleva periaate (vaikeampiin tilanteisiin).

Listojen lukumäärä (toistojen kanssa)

Montako PIN-koodia voidaan numeroista $\{0, 1, 2, \dots, 9\}$ 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

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

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

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

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

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

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

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

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

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 (ns. laskeva tulo).

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 jääkiekkjoukkueen 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. \implies 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.

Tätä kutsutaan **binomikertoimeksi** (lue: “ n yli k : n ”).

Binomikerrointa tullaan tarvitsemaan myöhemmin mm. toistuvien tapahtumien käsittelyssä.

Matlab `nchoosek`, R `choose`

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

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

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

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

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

$$\implies \mathbb{P}(\text{"kolmoset"}) = \frac{13 \binom{12}{2} \binom{4}{3} \binom{4}{1} \binom{4}{1}}{\binom{52}{5}} = \frac{54\,912}{2\,598\,960} \approx 2.1\%.$$

Seuraavalla luennolla puhutaan satunnaismuuttujista ja niiden jakaumista. . .