

## MAB8 TIIVISTELMÄ

### KOMBINATORIIKKA

- $n$  esinettä voidaan laittaa järjestykseen

$$n! := n \cdot (n-1)(n-2) \cdots 2 \cdot 1 \text{ tavalla.}$$

Esimerkiksi kirjaimet ABC voidaan järjestää  $3! = 3 \cdot 2 \cdot 1 = 6$ :lla tavalla

- Valinta takaisinpanolla: sama olio voidaan valita toisenkin kerran. Eri tapoja valita  $k$  kappaletta olioita  $n$ :stä takaisinpanolla on  $n^k$  kappaletta.

Esimerkiksi kirjaimista ABCDE voi muodostaa nelikirjaimisen sanan  $5^4 = 625$  tavalla.

- Valinta ilman takaisinpanoa: olio voidaan valita vain kerran. Eri tapoja valita *järjestyksessä*  $k$  kappaletta olioita  $n$ :stä on

$$n \cdot (n-1)(n-2) \cdots (n-k+2)(n-k+1) = \frac{n!}{(n-k)!}.$$

Esimerkiksi nelinumeroisen pin-koodi, jossa jokainen numero esiintyy vain kerran, voidaan valita  $10 \cdot 9 \cdot 8 \cdot 7$  tavalla.

- $n$ :stä oliosta voidaan valita  $k$ :n olion joukko (jossa järjestyksellä ei ole väliä)

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

Esimerkiksi kirjaimista ABCDE voidaan valita kolme

$$\binom{5}{2} = \frac{5!}{3!2!} = 10 \text{ tavalla.}$$

### TODENNÄKÖISYYSLASKENTAA

- Todennäköisyyksiä lasketaan tapahtumille. Jos tapahtuma ei koostu yksinkertaisemmista tapahtumista, se on alkeistapaus.

Esimerkiksi “nostettu kortti on pata” on tapahtuma, mutta ei alkeistapaus.

“Nostettu kortti on patanelonen” on alkeistapaus. Kaikki alkeistapaukset ovat tapahtumia.

- Jos tapahtuma  $A$  koostuu äärellisestä määrästä alkeistapauksia, sen todennäköisyys on näiden alkeistapauksien todennäköisyyksien summa.

Esimerkiksi

$$P(\text{Nostetaan } \spadesuit) = P(\spadesuit A) + P(\spadesuit 2) + P(\spadesuit 3) + \cdots + P(\spadesuit Q) + P(\spadesuit K).$$

- $P(\text{ei-}A) = q - P(A)$ .  
 $P(A \text{ tai } B) = P(A) + P(B) - P(A \text{ ja } B)$ .

Esimerkiksi

$$P(\text{nopanheiton tulos on } 6) = 1 - P(\text{nopanheiton tulos on } 1, 2, 3, 4 \text{ tai } 5).$$

$$P(\spadesuit \text{ tai } 2) = P(\spadesuit) + P(2) - P(\spadesuit 2).$$

- Peräkkäisten tapahtumien todennäköisyydet saadaan kertomalla. Joskus pitää huomioida aiempien tapahtumien vaikutus todennäköisyyksiin.

Esimerkiksi todennäköisyys heittää kolmella kolikonheitolla kolme kruunaa on  $0.5 \cdot 0.5 \cdot 0.5 = 0.125$ .

Todennäköisyys nostettaessa kolme korttia pakasta, että kaikki ovat herttaa on  $\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$

- Ehdollinen todennäköisyys  $A$  ehdolla  $B$  on

$$P(A|B) := \frac{P(A \text{ ja } B)}{P(B)}, \text{ kun } P(B) > 0.$$

Esimerkiksi nopanheitossa

$$P(6|\text{parillinen}) = \frac{P(6)}{P(2, 4 \text{ tai } 6)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Arkijärjellä: jos silmäluku on parillinen, se on 2, 4 tai 6, joten todennäköisyys kutoselle tällä ehdolla on  $1/3$ .

## TOISTOKOE JA BINOMITODENNÄKÖISYYDET

- Toistokokeessa tehdään sama koe monta kertaa ja lasketaan kunkin tuloksen lukumäärä.

Esimerkiksi kolikon heitto sata kertaa on toistokoe, ja sen tulos voisi olla vaikka 53 kruunaa.

- Jos yhdessä kokeessa onnistumisen todennäköisyys on  $p$  ja koe toistetaan  $n$  kertaa, on  $k$ :n onnistumisen todennäköisyys

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

LibreOfficella `=binom.dist(k,n,p,0)`

TI-Nspirellä `binomPdf(n,p,k)`.

Esimerkiksi, jos kolikko antaa kruunan todennäköisyydellä 0.4, ja kolikkoa heitetään sata kertaa, on tuloksen 53 kruunaa todennäköisyys

$$\binom{100}{53} 0.4^{53} 0.6^{47}.$$

LibreOfficella `=binom.dist(53, 100, 0.4, 0)`

TI-Nspirellä `binomPdf(100, 0.4, 53)`.

- Kertymäfunktio kertoo todennäköisyyden saada enintään  $k$  onnistumista. Käsin sen voi laskea laskemalla vain yhteen todennäköisyydet  $1, 2, \dots, k$ :lle onnistumiselle.

LibreOfficella `=binom.dist(k,n,p,1)` (Ero tasan  $k$  onnistumiseen on se, että lopussa on ykkönen eikä nolla).

TI-Nspirellä `binomCdf(n,p,k)`.

Esimerkiksi jos kolikko antaa kruunan todennäköisyydellä 0.4, saadaan enintään 45 kruunan todennäköisyys LibreOfficessa näin:

`=binom.dist(45, 100, 0.4, 1)`

TI-Nspirellä näin: `binomCdf(100, 0.4, 45)`.

- Normaalijakauma on hyvin yleinen todennäköisyysjakauma. Se eroaa binomijakaumasta siinä, että arvojen ei tarvitse olla kokonaislukuja. Kruunia ei voi saada 4.5 kappaletta kymmenellä heitolla, mutta ihmisen pituus voi olla 182.5 cm. Jos tiedetään normaalijakautuneen suureen keskiarvo  $\mu$  ja keskihajonta  $\sigma$ , voidaan laskea todennäköisyys suureen (esimerkiksi henkilön pituuden) sijoittumisesta annetulle välille.

TI-Nspire: `normCdf(alaraja, yläraja,  $\mu$ ,  $\sigma$ )` kertoo todennäköisyyden, että suure on välillä `[alaraja, yläraja]`.

LO Calc: `=norm.dist(yläraja,  $\mu$ ,  $\sigma$ , 1)` kertoo todennäköisyyden, että suure on enintään `yläraja`. Todennäköisyys olla välillä `[alaraja, yläraja]` saadaan vähentämällä todennäköisyys sille, että suure on enintään `alaraja`:

`=norm.dist(yläraja,  $\mu$ ,  $\sigma$ , 1) - norm.dist(alaraja,  $\mu$ ,  $\sigma$ , 1)`

Esimerkki: ihmisten pituus on normaalijakautunut keskiarvolla 176 cm ja keskihajonnalla 8.2 cm. Lasketaan todennäköisyys olla 156-168 cm pitkä. TI-Nspire: `normCdf(156, 168, 176, 8.2)`

LO Calc:

`=norm.dist(168, 176, 8.2, 1) - norm.dist(156, 176, 8.2, 1)`

- Nspiressä Cdf tarkoittaa aina kumulatiivista todennäköisyyttä, eli todennäköisyyttä, että suure on *alle* jonkin rajan. Pdf tarkoittaa todennäköisyyttä, että suure on *tasan* annettu luku. LO Calcissa tämä sama erottelu tehdään sillä, onko viimeinen parametri 1 vai 0. Esimerkiksi `binom.dist(3, 5, 0.50, 0)` on todennäköisyys tasan kolmelle kruunalle viidellä heitolla ja `binom.dist(3, 5, 0.50, 1)` enintään viidelle kruunalle.

Normaalijakaumassa ei ole järkeä laskea todennäköisyyttä olla tasan 180 cm (tms). Kone kyllä antaa erään luvun, mutta sillä ei tee mitään. Jos haluat laskea tällaisen todennäköisyyden, laske todennäköisyys esim. sentin tarkkuudella `normCdf(179.5, 180.5, 176, 8.2)`.

- *Populaatio* on koko tutkittava joukko. *Otos* on siitä satunnaisesti valittu osajoukko. Jos osajoukko on valittu epäsatunnaisesti, se on *näyte* eikä otos. Näytteestä ei voi tehdä johtopäätöksiä koko otoksen suhteen. Erityisen vaarallisia ovat näytteet, jotka perustuvat jäsenyyteen, asiakkuuteen tms., koska tällöin kuuluminen joukkoon voi tarkoittaa eroamista keskimääräisestä.

Klassinen esimerkki: <https://www.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html>

- Koko populaation keskiarvolle paras arvio on tutkitun otoksen keskiarvo. Koko populaation keskihajonnalle paras arvio on otoksen *otoskeskihajonta*. Tämä johtuu siitä, että ääripäät (esim. hyvin pitkät tai lyhyet ihmiset) voivat todennäköisemmin jäädä otoksen ulkopuolelle.

Otoskeskihajonta lasketaan Nspirellä listat-tilassa näin `stdevsamp(eka ruutu, vika ruutu)` ja LO Calcissa näin `=stdev(eka ruutu, vika ruutu)`.

Esimerkki ruuduissa A1-A12 on henkilöiden pituuden otoksesta. Otoskeskihajonta lasketaan Nspirellä näin `stdevsamp(A1, A12)` ja LO Calcissa näin `=stdev(A1, A12)`. Se on ennuste pituuden keskihajonnalle koko populaatiossa.

- Jos otoksen koko on  $n$ , keskiarvo otoksessa on  $\bar{x}$  ja otoskeskihajonta otoksessa on  $s$ , on 95% luottamusväli

$$\left[ \bar{x} - 1.96 \cdot \frac{s}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{s}{\sqrt{n}} \right].$$

Huomaa, että sama luku miinustetaan keskiarvosta vasemmalla ja plussataan oikealla. Luku 1.96 tulee siitä, että halutaan nimenomaan 95%:n luottamusväli. Muille yleisille luottamusväleille on omat lukunsa MAOLissa, ja monisteessa selitetään,

miten lasketaan yleisiä luottamusvälejä. (Ne ovat luultavasti lyhyen matematiikan lukijoille tarpeettomia).

Esimerkki: 56 henkilön otoksessa keskipituus on 175.4 cm ja keskihajonta 7.5 cm. Lasketaan ensin:

$$1.96 \cdot \frac{7.5}{\sqrt{56}} \approx 1.964370128$$

Sitten miinustetaan ja plussataan tämä keskipituuteen 175,4 95%:n luottamusvälin saamiseksi:

$$[173.4 \text{ cm} , 177.4 \text{ cm} ] .$$

Huomaa: jos lasket paloissa, käytä pyöristämättömiä lukuja. Jos lasket Nspirellä, voi olla helpompi lätkiä siihen vain suoraan luottamusvälin kaava.

- Kun lasketaan luottamusväliä luokittelulle eikä mitoille, esimerkiksi gallupissa, lasketaan ensin suhteellinen osuus  $p := \frac{k}{n}$ , missä  $k$  on jotain tyyppiä äänestävien lukumäärä ja  $n$  koko otoksen lukumäärä. Suhteellinen osuus on välillä  $[0, 1]$  ja sen keskihajonta on

$$s = \sqrt{p(1 - p)}.$$

Loppuosa menee samoin kuin pituuksienkin luottamusvälit.

Esimerkki: Gallupissa 143:sta vastaajasta 84 sanoo äänestävänsä Niinistöä. Niinistön suhteellinen osuus on

$$p = \frac{84}{143} \approx 0.587412587,$$

josta keskihajonnaksi saadaan

$$\begin{aligned} s &= \sqrt{0.587412587 \cdot (1 - 0.587412587)} \\ &= \sqrt{0.587412587 \cdot 0.412587412} \\ &\approx 0.492299745. \end{aligned}$$

Nyt 95%:n luottamusvälissä vähennettävä/lisättävä osa on

$$1.96 \cdot s\sqrt{n} = 1.96 \cdot 0.492299745\sqrt{143} \approx 0.080689618,$$

ja luottamusväli saadaan vähentämällä/lisäämällä tämä suhteelliseen osuuteen  $p$  ja pyöristämällä:

$$[0.506722968, 0.668102206] \approx [0.506, 0.669] = [50.6\%, 66.9\%].$$

**Tämä on pyöristetty tahallaan väärin siten, että välistä tulee suurempi kuin tarkemmilla arvoilla. En tiedä, onko tämä virallinen käytäntö, mutta ainakin se on hyvin perusteltu: on järkevämpää arvioida väli liian suureksi kuin liian pieneksi. Jos vastaisin esimerkiksi yokeissa kysymykseen luottamusväleistä, pyöristäisin ensin sääntöjen**

mukaan, ja sitten kirjoittaisin erikseen, että jos halutaan välin varmasti olevan tarpeeksi kattava, pyöristetään tällä toisella tavalla...

- Ajatellaan tilannetta, jossa lopputuloksia voivat olla  $x_1, x_2, \dots, x_n$ , niiden todennäköisyydet ovat  $P(x_1), P(x_2), \dots, P(x_n)$  ja niistä aiheutuva rahallinen hyöty on  $R(x_1), R(x_2), \dots, R(x_n)$ . Rahallinen hyöty voi olla myös negatiivinen (eli haitta). Tällöin rahallisen hyödyn odotusarvo on

$$P(x_1) \cdot R(x_1) + P(x_2) \cdot R(x_2) + \dots + P(x_n) \cdot R(x_n).$$

Positiivisella odotusarvolla tilanne on kannattava ja negatiivisella ei. Rahallisen hyödyn asemesta voidaan tietenkin laskea odotusarvoa muillekin asioille, kuten kruunien lukumäärälle, kulutetulle ajalle jne., raha on nostettu tässä esille vain konkretian takia.

Esimerkki: Pelissä maksetaan kaksi euroa siitä, että heitetään neljää noppaa. Jokaisesta saadusta kutosesta voittaa kaksi euroa. Kannattaako peliä pelata?

Lasketaan tämä taulukkolaskennan avulla. Kutosten lukumäärä noudattaa binomijakaumaa. Koska odotusarvo on negatiivinen, peliä ei kannata pelata.

<u>kutosten lkm</u>	<u>tn</u>	<u>voitto</u>	<u>tn*voitto</u>
0	0.482253086	-2	-0.964506173
1	0.385802469	0	0
2	0.115740741	2	0.231481481
3	0.015432099	4	0.061728395
4	0.000771605	6	0.00462963
		<u>Yhteensä:</u>	-0.666666667