

---

**Ilkka Mellin**

**Todennäköisyyslaskenta**

**Osa 1: Todennäköisyys ja sen laskusäännöt**

**Johdanto: Deterministisyys ja satunnaisuus**

**Todennäköisyyden määritteleminen**

**Todennäköisyyslaskennan peruskäsitteet**

# Todennäköisyys ja sen laskusäännöt

---

- >> **Johdanto: Deterministisyys ja satunnaisuus**
- Todennäköisyyden määrittely**
- Todennäköisyyden perusominaisuudet**

## Deterministisyys ja satunnaisuus

# Deterministiset ilmiöt

---

- Reaalimaailman ilmiö on **deterministinen**, jos *ilmiön alkutilan perusteella voidaan tarkasti ennustaa ilmiön lopputila eli tulos.*
- Deterministisen ilmiön *alkutila määrää tarkasti ilmiön lopputilan eli tuloksen.*
- Deterministisiä ilmiöitä kutsutaan usein **eksakteiksi** tai **kausaalisiksi**.

## Deterministisyys ja satunnaisuus

# Deterministiset ilmiöt:

## Esimerkkejä

---

- Monia *fysiikan*, kuten *klassisen mekaniikan*, tutkimia ilmiöitä pidetään tavallisesti deterministisinä.
- Esimerkiksi kappaleen lentorata voidaan ennustaa hyvin tarkasti, jos tunnetaan kappaleen paino, lähtönopeus, lähtökulma, lähtösuunta, ilmanvastus jne.
- Huomautuksia:
  - Deterministisistä ilmiöistä tehtäviin *havaintoihin* liittyy hyvin usein luonteeltaan *satunnaisia havaintovirheitä*.
  - Deterministisiin ilmiöihin saattaa liittyä *ennustamattomuutta*, jota kutsutaan *kaaokseksi*.

## Deterministisyys ja satunnaisuus

# Satunnaisilmiöt

---

- Reaalimaailman ilmiö on **stokastinen ilmiö** eli **satunnaisilmiö**, jos sillä on seuraavat ominaisuudet:
  - (i) Ilmiö voi päätyä *alkutilastaan* useisiin erilaisiin *lopputiloihin* eli ilmiöllä on useita erilaisia *vaihtoehtoisia tuloksia*.
  - (ii) Ilmiön alkutilan perusteella *ei voida* tarkasti *ennustaa* ilmiön lopputilaa eli sitä, mikä mahdollisista *tulosvaihtoehtoista realisoituu* eli *toteutuu*.
  - (iii) Vaikka ilmiön lopputilaa ei voida ennustaa tarkasti, tulosvaihtoehtojen *suhteellisten frekvenssien* eli *osuuksien nähdään* ilmiön toistuesssa *käyttäytyvän säännönmukaisesti*.

## Satunnaisilmiöt:

### Esimerkkejä 1/2

---

- Biologiset ilmiöt
    - sukupuolen määräytyminen
    - perinnöllisyys
  - Havaintovirheiden syntyminen empiirisissä tutkimuksissa
  - Ihmisen ominaisuuksien periytyminen
    - fyysiset tai henkiset ominaisuudet
    - fyysinen tai henkinen suorituskyky
  - Kvanttimekaniikan ilmiöt
    - radioaktiivinen hajoaminen
    - hiukkasfysiikan ilmiöt
- **Tilastollisen tutkimusaineiston keruu**
    - *otoksen* poiminta
    - *satunnaistus* empiirisissä kokeissa
  - Uhkapelit
    - rahanheitto
    - korttipelit
    - lotto
    - ruletti
    - arpajaiset
  - Yhteiskunnalliset ilmiöt
    - sosiologiset ilmiöt
    - taloustieteelliset ilmiöt
-

## Satunnaisilmiöt:

### Esimerkkejä 2/2

---

- *Satunnaisilmiöiden tulosta ei voida ennustaa tarkasti, mutta ilmiön toistuessa mahdollisten tulosvaihtoehtojen suhteellisten frekvenssien eli osuuksien havaitaan käyttäytyvän säännönmukaisesti.*
- Esimerkkejä säännönmukaisuuksista satunnaisilmiöissä:
  - Satunnaisesti valitun ihmisen älykkyydosamäärää ei tiedetä, mutta älykkyydosamäärät *jakautuvat* suurissa ihmisjoukoissa ns. *normaalijakauman mukaan*.
  - Havaintovirheen suuruutta ei voida ennustaa yksittäiselle havainnolle, mutta havaintovirheet *jakautuvat* suurissa havaintomäärissä usein ns. *normaalijakauman mukaan*.
  - Radioaktiivisen aineen yksittäisen atomin hajoamishetkeä ei voida ennustaa, mutta *puoliintumisaika* on jokaiselle radioaktiiviselle aineelle ominainen *vakio*.

## Satunnaisuus *ei ole* mielivaltaisuutta

---

- Ilmiön satunnaisuudella tarkoitetaan sitä, että ilmiön tulos *vaihtelee* ilmiön toistuessa tavalla, jota *ei voida ennustaa tarkasti*.
- Satunnaisilmiön tulos *ei kuitenkaan saa ilmiön toistuessa vaihdella mielivaltaisella tavalla*.
- Satunnaisilmiön *säännönmukaisten piirteiden on tultava esille ilmiön toistuessa*.



## Deterministisyys ja satunnaisuus

# Tilastollinen stabiliteetti

---

- Satunnaisilmiön toistuessa ilmenevää *säännönmukaisuutta* kutsutaan tilastotieteessä **tilastolliseksi stabiliteetiksi**.
- Jos satunnaisilmiö *ei ole tilastollisesti stabiili*, sitä *ei voida mallintaa* tilastollisilla malleilla.
- Huomautus:

Tilastollisen stabiliteetin idea saa matemaattisesti täsmällisen muotoilun ns. *suurten lukujen laissa*; ks. lukua **Stokastiikan konvergenssikäsitteet ja raja-arvolauseet**.

## Deterministisyys ja satunnaisuus

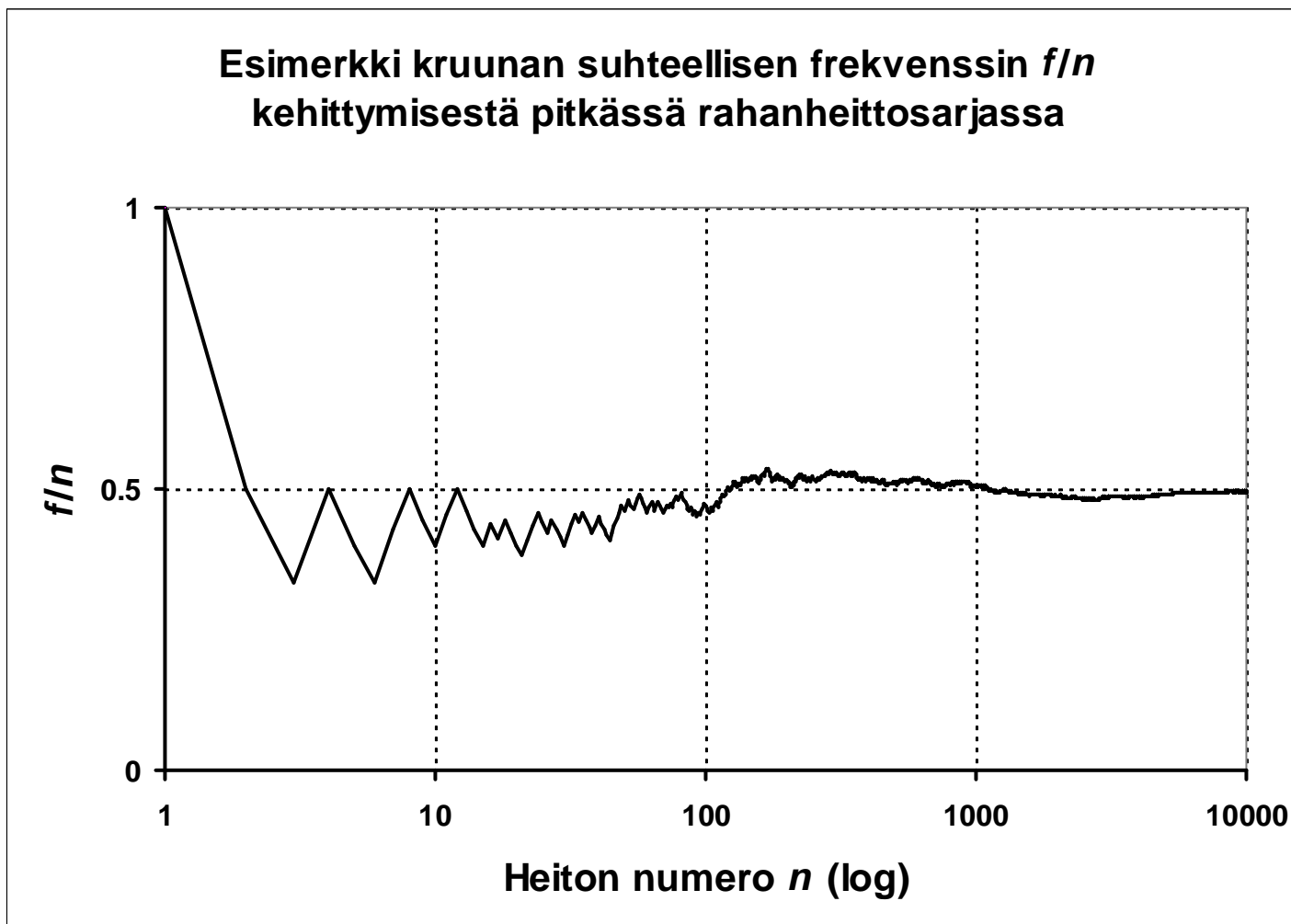
# Tilastollinen stabiliteetti:

### Esimerkki 1/2

---

- Heitetään *virheetöntä* eli *harhatonta* rahaa toistuvasti ja pidetään kirjaa kruunien *suhteellisesta osuudesta* eli *frekvenssistä*.
- *Yksittäisen heiton tulosta* ei voida ennustaa.
- Kruunien suhteellinen frekvenssi *vaihtelee* heittoja toistettaessa, mutta *lähestyy* virheettömän rahan tapauksessa lukua  $1/2$  siten, että *suuret poikkeamat* luvusta  $1/2$  tulevat yhä *epätodennäköisemmiksi* eli *harvinaisemmiksi*.
- Huomautuksia:
  - Luku  $1/2$  *ei ole* kruunien suhteellisen frekvenssin *raja-arvo* tavanomaisessa mielessä.
  - Tapa, jolla kruunien suhteellinen frekvenssi lähestyy lukua  $1/2$ , on esimerkki ns. *stokastisesta konvergenssista*.

Deterministisyys ja satunnaisuus  
**Tilastollinen stabiliteetti:**  
**Esimerkki 2/2**



## Satunnaisilmiöt, tilastolliset mallit ja todennäköisyyslaskenta 1/2

---

- Tilastotieteen tehtävänä on rakentaa (*tilastollisia*) malleja, joiden avulla voidaan *kuvata* ja *selittää mekanismit*, jotka *tuottavat tiedot* tutkimuksen kohteena olevasta reaali-maailman ilmiöstä.
- Koska tilastollisissa tutkimusasetelmissä ilmiötä koskeviin tietoihin sisältyy *satunnaisuutta* ja *epävarmuutta*, tilastollisia malleja rakennettaessa sovelletaan *todennäköisyyslaskentaa*.

## Satunnaisilmiöt, tilastolliset mallit ja todennäköisyyslaskenta 2/2

---

- Satunnaisilmiölle voidaan rakentaa tilastollisia malleja vain, jos *ilmiöiden tulokset eivät vaihtele mielivaltaisella tavalla.*
- *Ei-mielivaltaisuuudella* tarkoitetaan sitä, että ilmiön toistussa tulosvaihtoehtojen *suhteelliset frekvenssit eli osuudet* käyttäytyvät *tilastollisesti stabiilisti.*

## Satunnaiskokeet ja koetoistot: Määritelmät

---

- Kutsumme *satunnaisilmiötä* **satunnaiskokeeksi**.  
Esimerkkejä:
  - Lapsen sukupuolen *määräytymismekanismi* munasolun hedelmöityessä on satunnaiskoe.
  - Nopanheitto on satunnaiskoe.
- Kutsumme *satunnaisilmiön esiintymiskertaa* **koetoistoksi**.  
Esimerkkejä:
  - Yksittäisen lapsen sukupuolen *määräytyminen* on koetoisto.
  - Yksittäinen nopanheitto on koetoisto.

## Satunnaiskokeet ja koetoistot, tilastollinen stabiliteetti ja reilun pelin vaatimus

---

- Satunnaiskokeen toistaminen *samoissa olosuhteissa* – ts. kokeen *olosuhteiden vakiointi* – takaa tavallisesti sen, että satunnaiskokeen tulokset käyttäytyvät *tilastollisesti stabiilisti*.
- Vaatimus satunnaiskokeen tulosten käyttäytymisen tilastollisesta stabiliteetista voidaan tulkita vaatimukseksi **reilusta pelistä**.

## Deterministisyys ja satunnaisuus

# Reilu vs epäreilu peli:

### Esimerkki 1/3

---

- Pelat Mr. Ebenezer Scroogea vastaan *peliiä*, jolla on seuraavat *säännöt*:
  - (i) Mr. Scroogella on hallussaan useita *erilaisia* noppia, joiden silmälukuja *et tiedä*.
  - (ii) Mr. Scrooge valitsee nopistaan *yhden*.
  - (iii) Et saa ottaa Mr. Scroogen valitsemaa noppaa käteesi, mutta Mr. Scroogen on heitettävä valitsemaansa noppaa *niin monta kertaa kuin haluat*.
  - (iv) Jokaisen heiton jälkeen Mr. Scroogen on näytettävä sinulle heiton *tulos* eli silmäluku, joka on nopan ylösjääneellä tahkolla.
  - (v) Voitat ennalta sovitun rahasumman, jos saat selville Mr. Scroogen heittämän nopan silmäluvut.



## Deterministisyys ja satunnaisuus

# Reilu vs epäreilu peli:

### Esimerkki 2/3

---

- Saat Mr. Scroogen heittämän nopan silmäluvut selville, jos *toistatat nopanheittoa* riittävän monta kertaa ja *tarkkailet* heittojen tuloksena esiintyvien silmälukujen *suhteellisia frekvenssejä*.
- Oletetaan esimerkiksi, että Mr. Scroogen valitseman nopan silmäluvut ovat

1, 1, 1, 2, 2, 3

- Tällöin tuntuu ilmeiseltä, että silmälukujen

1, 2, 3

suhteellisten frekvenssien on *jakauduttava* pitkässä heittosarjassa suunnilleen suhteessa

3 : 2 : 1

## Deterministisyys ja satunnaisuus

# Reilu vs epäreilu peli:

### Esimerkki 3/3

---

- Oletetaan, että Mr. Scrooge *vaihtaa jatkuvasti salaa* noppaansa pelin aikana.
- Tällöin *et voi voittaa* peliä, koska Mr. Scrooge *rikkoo tietämättäsi pelin sääntöä (ii) vastaan*.
- *Vaatimus reilusta pelistä* tarkoittaa sitä, että tällaista sääntöjen rikkomista *ei saa* tapahtua.

## Tilastollinen tutkimus on peliä luontoa vastaan 1/3

---

- *Tilastollista tutkimusta* voidaan kuvata **peliksi luontoa vastaan**.
- Tilastollisessa tutkimuksessa pyritään *tekemään luonnon tilaa koskevia johtopäätöksiä havaintojen perusteella*.
- **Luonnon tila** on sitä, että luonnolla on kädessään joukko ”pelikortteja”.
- Tutkijan tavoitteena on *ottaa selville* luonnon kädessä olevat ”kortit”.
- Luonnon tavoitteena on *salata* kädessään olevat ”kortit” tutkijalta.

## Tilastollinen tutkimus on peliä luontoa vastaan 2/3

---

- Peli koostuu eristä, joissa jokaisessa tutkija voi katsoa yhden *satunnaisesti valitsemansa* ”kortin” luonnon kädessä – tämä on *havaintojen keräämistä*.
- Tutkija voi saada selville luonnon tilan eli luonnon kädessä olevat ”kortit” *pelaamalla riittävän monta erää* eli *keräämällä riittävästi havaintoja*.

## Tilastollinen tutkimus on peliä luontoa vastaan 3/3

---

- *Tilastollisessa tutkimuksessa* pyritään satunnaisilmiötä koskevien *havaintojen* perusteella *päättelemään*, millainen on *havainnot tuottanut mekanismi*.
- Päätely *ei onnistu*, jos *havainnot tuottanut mekanismi* ei ole jossakin mielessä *pysyvä* eli *tilastollisesti stabiili*.
- Oletus *havainnot tuottaneen mekanismin* pysyvyydestä voidaan tulkita oletukseksi siitä, että luonto *pelaa reilusti* eikä *riko pelin sääntöjä vastaan* vaihtamalla pelin aikana *salaa* kädessään olevia ”kortteja”.

# Todennäköisyys ja sen laskusäännöt

---

**Johdanto: Deterministisyys ja satunnaisuus**

**>> Todennäköisyyden määrittely**

**Todennäköisyyden perusominaisuudet**

## Todennäköisyyden *naiivit* määritelmät

---

- Satunnaisilmiöiden tapahtumien todennäköisyydelle voidaan esittää seuraavat *naiivit* **määritelmät**:
  - (i) Tapahtuman (**klassinen**) **todennäköisyys** on tapahtumalle suotuisien tulosvaihtoehtojen suhteellinen frekvenssi.
  - (ii) Tapahtuman (**empiirinen**) **todennäköisyys** on tapahtuman suhteellinen frekvenssi ilmiön toistokertojen joukossa.
  - (iii) Tapahtuman **todennäköisyys on** tapahtuman sattumisen mahdollisuuden **mitta**.

## Klassinen todennäköisyys: Määritelmä

---

- Tarkastellaan *satunnaisilmiötä*, johon liittyy  $n$  yhtä *todennäköistä tulosvaihtoehtoa*.
- Tarkastellaan satunnaisilmiön puitteissa *tapahtumaa*, johon liittyy  $k$  yhtä *todennäköistä tulosvaihtoehtoa*, joita sanotaan ko. tapahtumalle *suotuisiksi*.
- Ko. tapahtuman **klassinen todennäköisyys**  $p$  on tapahtumalle *suotuisien tulosvaihtoehtojen suhteellinen frekvenssi* eli tapahtumalle *suotuisien tulosvaihtoehtojen osuus* satunnaisilmiön *kaikista tulosvaihtoehdoista*:

$$p = \frac{k}{n}$$



## Klassinen todennäköisyys: Kommentteja

---

- Klassisen todennäköisyyden käsite sopii erityisesti *uhkapelien* analysointiin.
- Uhkapeleissä pelitapahtumien todennäköisyydet voidaan tavallisesti määrätä *päättelemällä* ne pelin säännöistä.
- *Historiallisesti* todennäköisyyslaskenta sai alkunsa 1600-luvulla juuri uhkapeleihin liittyvien ongelmien ratkaisuyrityksistä.
- Tulosvaihtoehtojen *lukumäärien laskeminen* on usein epätriviaali tehtävä ja apuna tarvitaan *kombinatoriikaksi* kutsuttua matematiikan osa-aluetta; ks. lukua **Klassinen todennäköisyys ja kombinatoriikka**.

## Klassinen todennäköisyys: Ongelmat määritelmässä

---

- Klassisen todennäköisyyden määritelmä *ei anna* mahdollisuutta puhua sellaisten tapahtumien todennäköisyyksistä, *joihin liittyvät tulosvaihtoehdot eivät ole yhtä todennäköisiä.*
- Klassisen todennäköisyyden määritelmä *ei anna* mahdollisuutta puhua sellaisten tapahtumien todennäköisyyksistä, *joihin liittyy äärettömän monta tulosvaihtoehtoa.*

Todennäköisyyden määrittely

## Klassinen todennäköisyys ja todennäköisyyden frekvenssitulkinta

---

- Oletetaan, että *toistamme* jotakin satunnaiskoetta ja tarkkailemme jonkin *tulosvaihtoehdon suhteellista frekvenssiä* koetoistojen aikana.
- **Todennäköisyyden frekvenssitulkinnan** mukaan ko. *tulosvaihtoehdon suhteellinen frekvenssi vaihtelee satunnaisesti koetoistosta toiseen, mutta saa keskimäärin tulosvaihtoehdon todennäköisyyttä lähellä olevia arvoja.*
- *Vahvistavatko havainnot tämän on empiirinen kysymys.*

## Todennäköisyyden määrittelemisen

# Klassinen todennäköisyys:

## Esimerkki nopanheitosta 1/2

---

- Heitetään noppaa.
- Tällöin tulosvaihtoehtoja on 6 kpl:  
Silmäluvut 1, 2, 3, 4, 5, 6
- Tarkastellaan tapahtumia  
 $A = \text{”Silmäluku on parillinen”}$   
 $B = \text{”Silmäluku} < 3\text{”}$
- Tapahtumalle  $A$  *suotuisia* tulosvaihtoehtoja on 3 kpl:  
Silmäluvut 2, 4, 6
- Tapahtumalle  $B$  *suotuisia* tulosvaihtoehtoja on 2 kpl:  
Silmäluvut 1, 2

## Todennäköisyyden määrittelemisen

# Klassinen todennäköisyys: Esimerkki nopanheitosta 2/2

---

- Tapahtuman  $A =$  ”Silmäluku on parillinen” todennäköisyys on

$$p = \frac{3}{6} = \frac{1}{2} = 0.5$$

- Tapahtuman  $B =$  ”Silmäluku  $< 3$ ” todennäköisyys on

$$p = \frac{2}{6} = \frac{1}{3} \approx 0.333$$

- Siten tapahtuma  $A$  on *todennäköisempi* kuin tapahtuma  $B$ .
- Oletetaan, että heität noppaa useita kertoja.
- *Todennäköisyyden frekvenssitulkinnan* mukaan *on odotettavissa*, että keskimäärin  $1/3$  heitoista antaa tulokseksi tapahtuman  $B$  ja tapahtuma  $A$  esiintyy heittojen tuloksena useammin kuin tapahtuma  $B$ .

## Empiirinen todennäköisyys: Määritelmä

---

- Tarkastellaan *satunnaiskoetta*, jota voidaan *toistaa* siten, että seuraavat ehdot pätevät:
  - (i) Kokeen olosuhteet *säilyvät muuttumattomina* koetoistosta toiseen.
  - (ii) Koetoistot *ovat riippumattomia*.
- Tarkkaillaan jonkin *tulosvaihtoehdon* esiintymistä koetoistojen aikana.
- Jos tulosvaihtoehdon *suhteellinen frekvenssi* eli *osuus* lähestyy jotakin kiinteätä lukua koetoistojen lukumäärän kasvaessa rajatta, lukua kutsutaan tulosvaihtoehdon **empiiriseksi todennäköisyydeksi**.

## Todennäköisyyden määritteleminen

# Empiirinen todennäköisyys ja suhteellinen frekvenssi 1/2

---

- *Toistetaan* satunnaiskoetta  $n$  kertaa.
- Tarkkaillaan jonkin *tulosvaihtoehdon* esiintymistä koetoistojen aikana.
- Olkoon  $f$  ko. tulosvaihtoehdon **frekvenssi** eli *lukumäärä* koetoistojen joukossa.

- Tällöin

$$\frac{f}{n}$$

on ko. tulosvaihtoehdon **suhteellinen frekvenssi** eli *suhteellinen osuus* koetoistojen joukossa.

## Empiirinen todennäköisyys ja suhteellinen frekvenssi 2/2

---

- Annetaan koetoistojen lukumäärän  $n$  kasvaa rajatta.
- Oletetaan, että (jossakin mielessä)

$$\frac{f}{n} \rightarrow p, \text{ kun } n \rightarrow +\infty$$

- Tällöin luku  $p$  on ko. tulosvaihtoehdon **empiirinen todennäköisyys**.
- Huomautus:

Suhteellisen frekvenssin  $f/n$  rajakäyttäytyminen koetoistojen lukumäärän  $n$  kasvaessa ei ole tavanomaista lukujonokongruenssia.

Todennäköisyyslaskennan konvergenssikäsitteitä käsitellään luvussa **Stokastiikan konvergenssikäsitteet ja raja-arvolauseet**.



## Empiirinen todennäköisyys: Kommentteja

---

- Tulosvaihtoehdon empiirinen todennäköisyys on tulosvaihtoehdon suhteellinen frekvenssi ”*pitkässä juoksussa*”.
- Empiirisen todennäköisyyden määritelmä edellyttää tulosvaihtoehtojen suhteellisilta frekvensseiltä *tilastollista stabiliteettia*:  
*Tulosvaihtoehdon empiirisestä todennäköisyydestä ei ole mielekästä puhua, ellei tulosvaihtoehdon suhteellinen frekvenssi käyttyädy satunnaiskoetta toistettaessa tilastollisesti stabiilisti.*
- *Matemaattista todennäköisyyttä* voidaan pitää empiirisen todennäköisyyden käsitteen *idealisointina*.

## Empiirinen todennäköisyys: Ongelmat määritelmässä 1/2

---

- Empiirinen todennäköisyys on *empiirinen käsite* siinä mielessä, että tulosvaihtoehdon suhteellisen frekvenssin  $f/n$  määrittäminen vaatii satunnaiskokeen *toistamista* ja *havaintojen keräämistä* satunnaiskokeen tuloksista.
- Tulosvaihtoehdon *empiiristä todennäköisyyttä ei kuitenkaan voida* – nimestään huolimatta – *määrittää kokeellisesti*, koska suhteellisen frekvenssin tilastollisen stabiliteetin empiirinen todentaminen vaatisi satunnaiskokeen toistamista *äärettömän* monta kertaa.
- Empiirisen todennäköisyyden käsite *ei anna* mahdollisuutta puhua sellaisten tapahtumien todennäköisyyksistä, *joista ei ole havaintoja*.

## Empiirinen todennäköisyys: Ongelmat määritelmässä 2/2

---

- Empiirisen todennäköisyyden määritelmässä esiintyvä suhteellisen frekvenssin raja-arvo *ei ole hyvin määritelty*: Mikään ei takaa, että määritelmässä esiintyvä raja-arvo *on olemassa*.
- Empiiristä todennäköisyyttä voidaan pikemminkin pitää *tilastollisesti stabiilisti käyttäytyvän suhteellisen frekvenssin ominaisuutena* kuin *todennäköisyyden määritelmänä*.

Todennäköisyyden määrittely

## Empiirinen todennäköisyys ja todennäköisyyden frekvenssitulkinta

---

- Oletetaan, että *toistamme* jotakin satunnaiskoetta ja tarkkailemme jonkin *tulosvaihtoehdon suhteellista frekvenssiä* koetoistojen aikana.
- **Todennäköisyyden frekvenssitulkinnan** mukaan ko. *tulosvaihtoehdon suhteellinen frekvenssi vaihtelee satunnaisesti koetoistosta toiseen, mutta saa keskimäärin tulosvaihtoehdon todennäköisyyttä lähellä olevia arvoja.*
- *Vahvistavatko havainnot tämän on empiirinen kysymys.*

Todennäköisyyden määrittelemisen

## Empiirinen todennäköisyys: Esimerkki laadunvalvonnasta 1/3

---

- Tehdas valmistaa erästä sähkölaitetta 300 kpl päivässä.
- Osa laitteista ei täytä ankaria laatukriteereitä.
- Merkitään:
  - $K$  = Laite on kelvollinen
  - $V$  = Laite on viallinen
- Oletetaan, että vialliset laitteet syntyvät tuotannossa *täysin satunnaisesti*.
- Eräänä päivänä valmistettujen laitteiden joukossa on 6 viallista laitetta.

## Todennäköisyyden määrittelemine

# Empiirinen todennäköisyys: Esimerkki laadunvalvonnasta 2/3

---

- Satunnaisilmiö: Laitteen laatu
- Koetoisto: Valmistetaan 1 laite
- Koetoistojen lkm  $n$ : 300
- Tulostavaihtoehto  $V$ : Laite on viallinen
- Viallisten laitteiden *frekvenssi*:

$$f = 6$$

- Viallisten laitteiden *suhteellinen frekvenssi*:

$$\frac{f}{n} = \frac{6}{300} = \frac{1}{50} = 0.02$$

## Todennäköisyyden määritteleminen

# Empiirinen todennäköisyys: Esimerkki laadunvalvonnasta 3/3

---

- Oletetaan, että viallisten laitteiden suhteellinen osuus pysyy päivästä toiseen *suunnilleen samana* eli, että viallisten laitteiden suhteellinen osuus käyttäytyy *tilastollisesti stabiilisti*.
- Tällöin suhteellista frekvenssiä

$$p = \frac{f}{n} = \frac{6}{300} = \frac{1}{50} = 0.02$$

on järkevää kutsua *todennäköisyydeksi* saada viallinen laite, jos tehtaalla valmistettujen laitteiden joukosta *poimitaan satunnaisesti* 1 laite tarkastettavaksi.

## Todennäköisyyden määrittelemisen

# Empiirinen todennäköisyys:

## Esimerkki otannasta 1/4

---

- Väestötilaston mukaan Suomen väestö jakautui vuoden 1998 lopussa miehiin ja naisiin seuraavasti:

Miehet	2 516 000
Naiset	2 643 600
<b>Yhteensä</b>	<b>5 159 600</b>

- Tyypillisessä *otantatutkimuksessa* tutkimuksen kohteet valitaan *poimimalla satunnaisotos* kaikkien suomalaisten joukosta.
- Satunnaisotoksen poimintaa voidaan kuvata *arvontana*, jossa jokaista suomalaista vastaa yksi arpalippu.
- *Todennäköisyys* poimia tietty henkilö on

$$\frac{1}{5159600}$$



## Todennäköisyyden määritteleminen

# Empiirinen todennäköisyys:

### Esimerkki otannasta 2/4

---

- Suomalaisien lukumäärä: 5 159 600
- Miesten *lukumäärä* eli *frekvenssi*: 2 516 000
- Miesten *suhteellinen osuus* eli *suhteellinen frekvenssi* kaikkien suomalaisten joukosta:

$$\frac{2516000}{5159600} = 0.488$$

- Naisten *lukumäärä* eli *frekvenssi*: 2 643 600
- Naisten *suhteellinen osuus* eli *suhteellinen frekvenssi* kaikkien suomalaisten joukosta:

$$\frac{2643600}{5159600} = 0.512$$

## Todennäköisyyden määrittelyminen

# Empiirinen todennäköisyys:

### Esimerkki otannasta 3/4

---

- Koska suomalaisia on näinkin paljon, miesten ja naisten suhteelliset frekvenssit voidaan *tulkita* empiirisen todennäköisyyden määritelmän mukaan todennäköisyyksiksi.
- Siten *todennäköisyys*, että satunnaisesti suomalaisten joukosta poimittu henkilö on *mies*, on
$$\frac{2516000}{5159600} = 0.488$$
- Siten *todennäköisyys*, että satunnaisesti suomalaisten joukosta poimittu henkilö on *nainen*, on
$$\frac{2643600}{5159600} = 0.512$$
- Todennäköisyys poimia suomalaisten joukosta nainen on *suurempi kuin* todennäköisyys poimia mies, koska naisia on *enemmän*.

## Todennäköisyyden määrittelyminen

# Empiirinen todennäköisyys:

### Esimerkki otannasta 4/4

---

- Oletetaan, että suomalaisten joukosta poimitaan *arpomalla* yhä uusia 1 000 henkilön *satunnaisotoksia*.
- Tällöin otokseen poimittujen miesten ja naisten suhteelliset osuudet *vaihtelevat otoksesta toiseen*, mutta otokseen poimituista henkilöistä *keskimäärin*

$$100 \times \frac{2516000}{5159600} = 48.8 \%$$

on miehiä ja *keskimäärin*

$$100 \times \frac{2643600}{5159600} = 51.2 \%$$

on naisia.

- *Tilastollinen stabiliteetti* on sitä, että nämä suhdeluvut pysyvät otoksesta toiseen *suunnilleen samoina*.

Todennäköisyyden määrittelemisen  
**Todennäköisyys *mittana*:**  
**Määritelmä**

---

- Hyödyllisen *mielikuvan* todennäköisyyden luonteesta antaa seuraava *naiivi määritelmä*:

**Todennäköisyys on mitta**, jolla mitataan satunnaisilmiön tapahtumavaihtoehtojen *sattumisen mahdollisuutta*.

## Todennäköisyyden määrittelemisen

# Todennäköisyys *mittana*:

## Kommentteja 1/2

---

- Määritelmä *ei täytä hyvän määritelmän tunnusmerkkejä*, koska se on *kehämääritelmä*:  
Sattumisen mahdollisuus ja todennäköisyys *tarkoittavat suunnilleen samaa*.
- Kuitenkin on totta, että ns. **Kolmogorovin aksioomien** mukaan todennäköisyys on *mitta matemaattisen mitta-teorian tarkoittamassa mielessä*; ks. lukua **Todennäköisyyden aksioomat**.
- Kolmogorovin aksioomien mukaan *todennäköisyysmitta* käyttäytyy samalla tavalla kuin *pinta-alamitta* paitsi, että todennäköisyysmitalla on ylärajana ns. varman tapahtuman todennäköisyys.

## Todennäköisyyden määrittelemisen

# Todennäköisyys *mittana*:

## Kommentteja 2/2

---

- Todennäköisyyden laskusääntöjä voidaan havainnollistaa joukko-opissa käytettävien **Venn-diagrammien** avulla.
- Venn-diagrammien idea:
  - (i) Tapahtumia kuvataan *tasoalueilla*.
  - (ii) Tapahtumien todennäköisyyksiä kuvataan *tasoalueiden pinta-aloilla*.
- Venn-diagrammien käyttö todennäköisyyslaskennan laskusääntöjen havainnollistamisessa perustuu siihen, että *todennäköisyydellä on mittana (lähes kaikki) samat ominaisuudet kuin pinta-alalla*.

# Todennäköisyys ja sen laskusäännöt

---

**Johdanto: Deterministisyys ja satunnaisuus**

**Todennäköisyyden määrittely**

**>> Todennäköisyyden perusominaisuudet**

## Satunnaisilmiöt ja niiden tilastolliset mallit

---

- Tilastotieteen tehtävänä on kehittää satunnaisilmiöille **tilastollisia malleja**, joiden avulla pyritään tekemään satunnaisilmiöitä koskevia *johtopäätöksiä*.
- Satunnaisilmiöiden tilastolliset mallit perustuvat **todennäköisyyslaskentaan** ja siksi niitä kutsutaan usein myös *stokastisiksi malleiksi* eli *todennäköisyysmalleiksi*.



## Todennäköisyysmalli satunnaisilmiön tilastollisena mallina

---

- Satunnaisilmiön **tilastollisessa mallissa** eli **todennäköisyysmallissa** eli **stokastisessa mallissa** on kaksi osaa:
  - (i) Satunnaisilmiön *kaikkien mahdollisten tulosvaihtoehtojen* kuvaus.
  - (ii) Tulosvaihtoehtojen **todennäköisyyksien** kuvaus.
- Satunnaisilmiön tilastollinen malli esitetään tavallisesti satunnaisilmiön tulosvaihtoja numeerisessa muodossa kuvaavaan *satunnaismuuttujan* ja sen *todennäköisyysjakauman* avulla; ks. lukua **Satunnaismuuttajat ja todennäköisyysjakaumat**.

## Tilastollisten mallien rakentaminen ja tilastollisen tutkimuksen tavoitteet

---

- *Tilastollisen tutkimuksen päätavoitteena on tilastollisen mallin rakentaminen tutkimuksen kohteena olevalle satunnaisilmiöille.*
- Tilastollisen mallin rakentamisen työvaiheet:
  - (1) *Mallin muodostaminen ilmiölle.*
  - (2) *Ilmiötä koskevien havaintojen kerääminen.*
  - (3) *Mallin parametrien estimointi.*
  - (4) *Mallin ja havaintojen yhteensopivuuden testaaminen.*  
Jos mallissa havaitaan puutteita vaiheessa (4), palataan vaiheeseen (1).

## Todennäköisyyslaskenta ja joukko-oppi

---

- Todennäköisyyslaskennan historian tärkeimpiä teoreettisia oivalluksia on ollut se, että satunnaisilmiön tapahtumia voidaan käsitellä **joukkoina**.
- Siksi seuraavassa palautetaan mieleen **joukko-opin perusmääritelmät**.

## Joukko-opin perusmääritelmät: Joukko ja sen alkiot

---

- **Joukko** on *kokoelma olioita*, joita kutsutaan joukon **alkioiksi**.
- Joukko on **hyvin määritelty**, jos *sen alkiot tunnetaan* eli voimme sanoa *jokaisesta* oliosta onko se joukon alkio vai ei.
- Merkitään joukon ja sen alkioiden välistä relaatiota seuraavasti:
  - (i)  $s$  on joukon  $A$  alkio eli  $s$  kuuluu joukkoon  $A$ :
$$s \in A$$
  - (ii)  $s$  ei ole joukon  $A$  alkio eli  $s$  ei kuulu joukkoon  $A$ :
$$s \notin A$$

## Joukko-opin perusmääritelmät: Osajoukko

---

- Olkoot  $A$  ja  $B$  kaksi joukkoa.
- Jos jokaiselle joukon  $B$  alkiolle  $s$  pätee, että

$$s \in B \Rightarrow s \in A$$

niin sanomme, että *joukko  $B$  on joukon  $A$  osajoukko* tai, että *joukko  $B$  sisältyy joukkoon  $A$ .*

- Merkintä:

$$B \subset A \text{ tai } A \supset B$$

## Joukko-opin perusmääritelmät:

### Tyhjä joukko

---

- Joukko on **tyhjä**, jos siihen ei kuulu yhtään alkiota.
- Tyhjää joukkoa merkitään symbolilla  
 $\emptyset$
- Jos joukko  $\emptyset$  on tyhjä, *ei ole olemassa* oliota  $s$ , jolle  
 $s \in \emptyset$
- Tyhjä joukko  $\emptyset$  on *jokaisen* joukon osajoukko eli mieltävaltaiselle joukolle  $A$  pätee:

$$\emptyset \subset A$$

## Otosavaruus ja alkeistapahtumat

---

- Satunnaisilmiön *kaikkien mahdollisten tulosvaihtoehtojen joukkoa* kutsutaan **otosavaruudeksi**.
- Otosavaruuden *alkioita* kutsutaan **alkeistapahtumiksi**.
- Merkinnät:
  - (i) *Otosavaruutta* (engl. *sample space*) merkitään tavallisesti isolla kirjaimella  $S$ .
  - (ii) *Otosavaruuden  $S$  alkiota* merkitään usein vastaavalla pienellä kirjaimella  $s$ .
- Jos siis alkeistapahtuma  $s$  kuuluu otosavaruuteen  $S$ , merkitään:

$$s \in S$$

## Otosavaruus ja alkeistapahtumat: Kommentteja

---

- Otosavaruus muodostaa *perusjoukon*, jossa satunnaisilmiön tulosvaihtoehtoja tarkastellaan.
- Satunnaisilmiötä ei voida “purkaa” otosavaruuden alkeistapahtumia *alkeellisempiin* tulosvaihtoehtoihin.



## Tapahtumat 1/2

---

- Olkoon  $S$  otosavaruus eli tarkasteltavan satunnaisilmiön kaikkien mahdollisten tulosvaihtoehtojen joukko.
- Tarkasteltavan satunnaisilmiön **tapahtumat** ovat otosavaruuden  $S$  alkeistapahtumien muodostamia joukkoja.
- Siten tapahtumat ovat tarkasteltavaan satunnaisilmiöön liittyvän otosavaruuden  $S$  osajoukkoja.

## Tapahtumat 2/2

---

- Jos siis  $A$  on jokin otosavaruuden  $S$  *tapahtuma*, niin

$$A \subset S$$

eli

$$s \in A \Rightarrow s \in S$$

jossa  $s$  on tapahtumaan  $A$  kuuluva *alkeistapahtuma*.

- **Kun sanomme, että tapahtuma  $A$  sattuu, tarkoitamme aina sitä, että jokin tapahtumaan  $A$  kuuluva alkeistapahtuma  $s$  sattuu.**

Todennäköisyyden perusominaisuudet

## Otosavaruus, alkeistapahtumat ja tapahtumat: Esimerkki sukupuolen määräytymisestä

---

- Satunnaisilmiö:  
Lapsen sukupuolen määräytyminen
- Otosavaruus:  
 $S = \{\text{Tyttö, Poika}\}$
- Alkeistapahtumat:  
 $s_1 = \text{Tyttö}$   
 $s_2 = \text{Poika}$

Todennäköisyyden perusominaisuudet

## Otosavaruus, alkeistapahtumat ja tapahtumat: Esimerkki 1 nopanheitosta

---

- Satunnaisilmiö:  
Nopanheiton tulos
- Otosavaruus:  
Silmälukujen joukko  $S = \{1, 2, 3, 4, 5, 6\}$
- Alkeistapahtumat:  
Silmäluvut 1, 2, 3, 4, 5, 6
- Esimerkki tapahtumasta:  
 $A = \text{”Silmäluku on parillinen”} = \{2, 4, 6\}$

## Otosavaruus, alkeistapahtumat ja tapahtumat: Esimerkki 2 nopanheitosta 1/3

---

- Satunnaisilmiö:

Tulokset kahdesta nopanheitosta

- Otosavaruus  $S$ :

Silmälukuparien  $(i, j)$  (36 kpl) joukko, jossa

$i = 1$ . nopanheiton silmäluku,  $i = 1, 2, 3, 4, 5, 6$

$j = 2$ . nopanheiton silmäluku,  $j = 1, 2, 3, 4, 5, 6$

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

Todennäköisyyden perusominaisuudet

## Otosavaruus, alkeistapahtumat ja tapahtumat: Esimerkki 2 nopanheitosta 2/3

---

- Otosavaruuden alkiot voidaan esittää seuraavana *taulukkona*:

$(i, j)$	$j = \text{tulos 2. nopan heitosta}$					
$i = \text{tulos 1. nopan heitosta}$	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

## Otosavaruus, alkeistapahtumat ja tapahtumat: Esimerkki 2 nopanheitosta 3/3

---

- Esimerkki tapahtumasta:

$$\begin{aligned} A &= \text{”Kummallakin nopalla saadaan sama silmäluku”} \\ &= \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\} \end{aligned}$$

- Esimerkiksi:

$$(2,2) \in A$$

$$(6,1) \notin A$$

$$A \subset S$$

## Varma tapahtuma ja mahdoton tapahtuma

---

### Varma tapahtuma

- Tapahtuma on **varma**, jos se *esiintyy aina*, kun satunnaisilmiö toistuu.
- *Otosavaruus*  $S$  on varma tapahtuma.

### Mahdoton tapahtuma

- Tapahtuma on **mahdoton**, jos se *ei voi esiintyä koskaan*, kun satunnaisilmiö toistuu.
- *Tyhjä joukko*  $\emptyset$  on mahdoton tapahtuma.



## Varma tapahtuma ja mahdoton tapahtuma: Esimerkit rahan- ja nopanheitosta

---

Esimerkki 1:

- Rahaa heitettäessä tuloksena on aina joko kruuna tai klaava.

- Tapahtuma

$$S = \{\text{Kruuna, Klaava}\}$$

on *varma*.

Esimerkki 2:

- Tavallista noppaa heitettäessä silmäluku 7 ei voi olla tuloksena.

- Tapahtuma  $\{7\}$  on *mahdoton*.

## Todennäköisyyden perusominaisuudet

---

- Olkoon  $S$  otosavaruus, jossa satunnaisilmiötä tarkastellaan.
- **Jokaisen tapahtuman**

$$A \subset S$$

**todennäköisyys**  $\Pr(A)$  on reaaliluku välillä  $[0,1]$ :

$$0 \leq \Pr(A) \leq 1$$

- **Varman tapahtuman  $S$  todennäköisyys on 1:**

$$\Pr(S) = 1$$

- **Mahdottoman tapahtuman  $\emptyset$  todennäköisyys on 0:**

$$\Pr(\emptyset) = 0$$

## Todennäköisyyden perusominaisuudet

# Todennäköisyyksien vertailu

---

- Jos

$$\Pr(A) > \Pr(B)$$

niin sanomme:

”Tapahtuma  $A$  on **todennäköisempi**  
kuin tapahtuma  $B$ ”

*tai*

”Tapahtuma  $B$  on **epätodennäköisempi**  
kuin tapahtuma  $A$ ”

## Todennäköisyyksien vertailu ja todennäköisyyden frekvenssitulkinta

---

- Mitä *todennäköisempi* tapahtuma on, sitä *useammin* tapahtumalla on taipumus esiintyä satunnaisilmiön toistuessa eli sitä *suurempi* on tapahtuman havaittu suhteellinen frekvenssi.
- Mitä *epätodennäköisempi* tapahtuma on, sitä *harvemmin* tapahtumalla on taipumus esiintyä satunnaisilmiön toistuessa eli sitä *pienempi* on tapahtuman havaittu suhteellinen frekvenssi.

## Todennäköisyyden perusominaisuudet

# Lukumääräfunktio

---

- Olkoon

$$n(A)$$

funktio, joka kertoo *joukon A alkioiden lukumäärän*.

- Kutsumme funktiota  $n(\cdot)$  **lukumääräfunktioksi**.
- Jos siis joukon A alkioiden lukumäärä on  $k$ , niin

$$n(A) = k$$

## Todennäköisyyden perusominaisuudet

# Äärelliset otosavaruudet

---

- Olkoon **otosavaruus  $S$  äärellinen joukko** ja olkoon

$$n(S) = n$$

otosavaruuden  $S$  *alkeistapahtumien* eli alkioiden lukumäärä.

- Merkitään alkeistapahtumia seuraavalla tavalla:

$$s_i, i = 1, 2, \dots, n$$

- Tällöin otosavaruus  $S$  voidaan määritellä luettelemalla sen alkiot:

$$S = \{s_1, s_2, \dots, s_n\}$$

## Äärelliset otosavaruudet:

## Alkeistapahtumien todennäköisyydet

---

- Äärellisen otosavaruuden

$$S = \{s_1, s_2, \dots, s_n\}$$

alkeistapahtumien  $s_1, s_2, \dots, s_n$  todennäköisyyksien

$$\Pr(s_i) = p_i, i = 1, 2, \dots, n$$

*on toteuttava ehto*

$$\sum_{i=1}^n p_i = 1$$

## Äärelliset otosavaruudet: Tapahtumien todennäköisyydet

---

- Olkoon  $A$  äärellisen otosavaruuden  $S$  *tapahtuma* eli  $A \subset S$ .
- Tällöin **tapahtuman  $A$  todennäköisyys**  $\Pr(A)$  on

$$\Pr(A) = \sum_{i|s_i \in A} p_i$$

- Summassa lasketaan yhteen kaikki todennäköisyydet

$$p_i = \Pr(s_i)$$

joille  $s_i \in A$ .



## Symmetriset alkeistapahtumat ja niiden todennäköisyydet

---

- Oletetaan, että äärellisen otosavaruuden

$$S = \{s_1, s_2, \dots, s_n\}$$

alkeistapahtumien  $s_1, s_2, \dots, s_n$  todennäköisyydet ovat *yhtä suuria*:

$$\Pr(s_i) = \frac{1}{n}, i = 1, 2, \dots, n$$

- Tällöin sanomme, että alkeistapahtumat  $s_1, s_2, \dots, s_n$  ovat **symmetrisiä**.

## Symmetriset alkeistapahtumat ja klassinen todennäköisyys 1/2

---

- Olemme määritelleet tapahtuman *klassisen todennäköisyyden* tapahtumalle suotuisien tulosvaihtoehtojen suhteellisena frekvenssinä satunnaisilmiön kaikista tulosvaihtoehdoista (ks. kappaletta **Todennäköisyyden määrittelemisen**).
- Olkoot otosavaruuden

$$S = \{s_1, s_2, \dots, s_n\}$$

alkeistapahtumat  $s_1, s_2, \dots, s_n$  *symmetrisiä*:

$$\Pr(s_i) = \frac{1}{n}, i = 1, 2, \dots, n$$

## Symmetriset alkeistapahtumat ja klassinen todennäköisyys 2/2

---

- Olkoon  $A$  otosavaruuden  $S$  tapahtuma, johon liittyvien alkeistapausten lukumäärä on  $k$ :

$$A \subset S$$

$$n(A) = k \leq n = n(S)$$

- Tällöin tapahtuman  $A$  **klassinen todennäköisyys** on

$$\Pr(A) = \frac{k}{n}$$

jossa siis

$$k = n(A)$$

$$n = n(S)$$

## Symmetriaoletus ja uhkapelit

---

- Useimmissa *uhkapeleissä* pelin säännöt edellyttävät, että peliin liittyvät alkeistapahtumat ovat symmetrisiä.
- Tyypillisiä uhkapelien säännöissä esitettyjä symmetria-vaatimuksia ovat seuraavat:
  - (i) Käytettävien pelivälineiden (esim. nopan, rahan tai rulettipyörän) on oltava *fysikaalisesti symmetrisiä*.
  - (ii) Käytettävillä pelivälineillä (esim. arpalipuilla tai korteilla) on oltava *sama todennäköisyys tulla valituiksi tai jaetuiksi*.
- Huomaa, että vaatimus (ii) edellyttää pelivälineiden (esim. arpalippujen tai korttien) huolellista *sekoittamista*.

## Symmetriaoletus: Kommentteja

---

- Otosavaruuden alkeistapahtumien symmetrisyyttä voidaan vain harvoin *perustella* uhkapelien ulkopuolella.
- Oletus alkeistapahtumien symmetrisyydestä on *oletus*, jota voidaan *testata* tilastollisesti, jos ko. satunnaisilmiöstä kerätään *havaintoja*.
- *Klassisen todennäköisyyden määritelmä* edellyttää sitä, että otosavaruus on äärellinen ja sen alkeistapahtumat ovat symmetrisiä.

## Symmetriset alkeistapahtumat: Esimerkki

---

- Satunnaisilmiö:  
Tulos nopanheitosta
- Otosavaruus  $S$ :  
Silmälukujen  $i = 1, 2, 3, 4, 5, 6$  joukko:  
$$S = \{1, 2, 3, 4, 5, 6\}$$
- Oletus nopan **virheettömydestä** voidaan pukea seuraavaan muotoon:  
$$\Pr(i) = \frac{1}{6}, i = 1, 2, 3, 4, 5, 6$$
- Siten oletus noppien virheettömydestä merkitsee oletusta alkeistapahtumien  $i = 1, 2, 3, 4, 5, 6$  *symmetrisyydestä*.