Bayes-päättely
Työterveyslaitos
8.-9.2.2018

Ville Hyvönen

1. Johdanto & Bayesin kaava

1. Bayesin kaava
2. Bayes-päättelyn käsitteet
3. Esimerkki Bayes-päättelystä
4. Hyvin lyhyt Bayes-päättelyn historia
5. Motivaatio

Klassinen esimerkki Bayesin kaavasta

- Oletetaan, että meillä on käytössä testi jonkin harvinaisen sairauden toteamiseksi, ja määritellään seuraavat satunnaismuuttujat:

$$B = \begin{cases} 1, & \text{jos henkilöllä on tämä harvinainen sairaus} \\ 0, & \text{jos henkilöllä ei ole tätä sairautta,} \end{cases}$$

$$A = \begin{cases} 1, & \text{jos henkilön testitulos on positiivinen} \\ 0, & \text{jos henkilön testitulos on negatiivinen.} \end{cases}$$

Klassinen esimerkki Bayesin kaavasta

- Testi on hyvin tarkka:

Klassinen esimerkki Bayesin kaavasta

- Testi on hyvin tarkka:
  - testin sensitiivisyys, eli todennäköisyys saada positiviinen testitulos henkilölle, jolla on tutkittava sairaus, on $p(A = 1|B = 1) = 0.95$.

Klassinen esimerkki Bayesin kaavasta

- Testi on hyvin tarkka:
  - testin sensitiivisyys, eli todennäköisyys saada positiviinen testitulos henkilölle, jolla on tutkittava sairaus, on $p(A = 1|B = 1) = 0.95$.
  - testin spesifisyys, eli todennäköisyys saada saada negatiivinen testitulos henkilölle jolla ei ole tutkittavaa sairautta, on $p(A = 0|B = 0) = 0.99$.

Klassinen esimerkki Bayesin kaavasta

- Testi on hyvin tarkka:
  - testin sensitiivisyys, eli todennäköisyys saada positiviinen testitulos henkilölle, jolla on tutkittava sairaus, on $p(A = 1|B = 1) = 0.95$.
  - testin spesifisyys, eli todennäköisyys saada saada negatiivinen testitulos henkilölle jolla ei ole tutkittavaa sairautta, on $p(A = 0|B = 0) = 0.99$.
- Tauti on harvinainen: se esiintyy väestöstä keskimäärin vain yhdellä kymmenestätuhannesta.

Klassinen esimerkki Bayesin kaavasta

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Testi on hyvin tarkka:
  - testin sensitiivisyys, eli todennäköisyys saada positiivinen testitulos henkilölle, jolla on tutkittava sairaus, on $p(A = 1|B = 1) = 0.95$.
  - testin spesifisyys, eli todennäköisyys saada saada negatiivinen testitulos henkilölle jolla ei ole tutkittavaa sairautta, on $p(A = 0|B = 0) = 0.99$.
- Tauti on harvinainen: se esiintyy väestöstä keskimäärin vain yhdellä kymmenestätuhannesta.
  - Kyseessä on seulonta, eli koehenkilö on valittu satunnaisotannalla väestöstä: hänellä ei ole (välttämättä) mitään oireita.

Klassinen esimerkki Bayesin kaavasta

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Testi on hyvin tarkka:
  - testin sensitiivisyys, eli todennäköisyys saada positiviinen testitulos henkilölle, jolla on tutkittava sairaus, on $p(A = 1|B = 1) = 0.95$.
  - testin spesifisyys, eli todennäköisyys saada saada negatiivinen testitulos henkilölle jolla ei ole tutkittavaa sairautta, on $p(A = 0|B = 0) = 0.99$.
- Tauti on harvinainen: se esiintyy väestöstä keskimäärin vain yhdellä kymmenestätuhannesta.
  - Kyseessä on seulonta, eli koehenkilö on valittu satunnaisotannalla väestöstä: hänellä ei ole (välttämättä) mitään oireita.
  - Tällöin todennäköisyys, että koehenkilöllä on tutkittava sairaus on $P(B = 1) = 0.0001$.

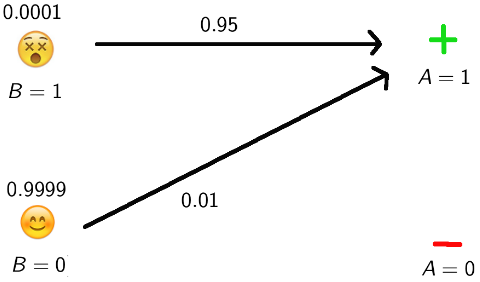
Klassinen esimerkki Bayesin kaavasta

- Mikä on todennäköisyys, että koehenkilöllä on tutkittava sairaus, jos hänen testituloksensa on positiivinen:

$$p(B = 1|A = 1) = ?$$

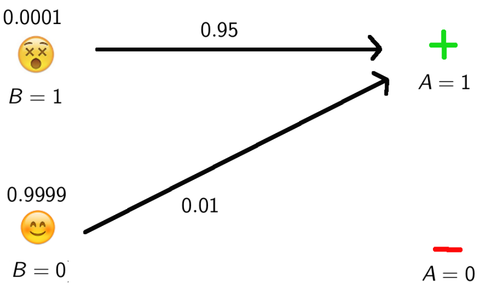
Kokonaistodennäköisyyden kaava

- Positiivinen testitulos voidaan saada kahdella eri tavalla:



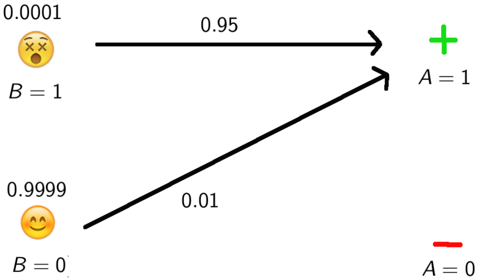
Kokonaistodennäköisyyden kaava

- Positiivinen testitulos voidaan saada kahdella eri tavalla:
  - Oikea positiivinen: koehenkilön, jolla on tutkittava sairaus, testitulos on positiivinen



Kokonaistodennäköisyyden kaava

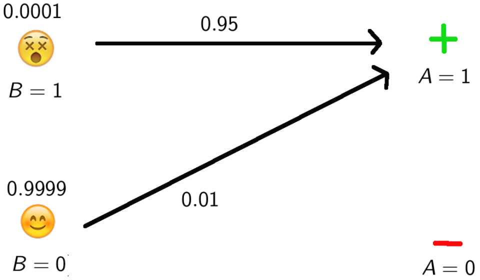
- Positiivinen testitulos voidaan saada kahdella eri tavalla:
  - Oikea positiivinen: koehenkilön, jolla on tutkittava sairaus, testitulos on positiivinen
  - Väärä hälytys: terveen koehenkilön testitulos on positiivinen.



Kokonaistodennäköisyyden kaava

- Siten positiviisen testituloksen todennäköisyys on

$$
\begin{aligned}
P(A=1) &= P(B=0)P(A=1|B=0) + P(B=1)P(A=1|B=1) \\
&= 0.9999 \cdot 0.01 + 0.0001 \cdot 0.95 \\
&= 0.009999 + 0.000095 = 0.010094.
\end{aligned}
$$



Ratkaisu

- Nyt todennäköisyys, että koehenkilöllä on tutkittava sairaus jos hänen testituloksensa on positiivinen saadaan tarkastelemalla kuinka suurella osuudella kaikista positiivisen testituloksen saaneista on tämä sairaus:

$$
\begin{aligned}
p(B=1|A=1) &= \frac{p(A=1,B=1)}{p(A=1)} \\
&= \frac{p(B=1)p(A=1|B=1)}{p(A=1)} \\
&= \frac{0.000095}{0.010094} = 0.0094 \approx 1\%
\end{aligned}
$$

Ratkaisu

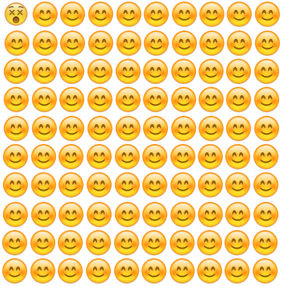
Oikeasti sairaat
(true positives)

😵

Testin tarkkuus =
(precision)

$$\approx \frac{1}{100}$$

Kaikki positiiviset
(true positives + false positives)



Bayesin kaava

- Johdimme juuri Bayesin kaavan! Yleisemmin diskreeteille satunnaismuuttujille:

$$\begin{aligned} p(B = b|A = a) &= \frac{p(B = b)p(A = a|B = b)}{p(A = a)} \\ &= \frac{p(B = b)p(A = a|B = b)}{\sum_{b' \in \mathcal{B}} p(B = b')p(A = a|B = b')} \end{aligned}$$

Bayesin kaava

- Johdimme juuri Bayesin kaavan! Yleisemmin diskreeteille satunnaismuuttujille:

$$
\begin{aligned}
p(B=b|A=a) &= \frac{p(B=b)p(A=a|B=b)}{p(A=a)} \\
&= \frac{p(B=b)p(A=a|B=b)}{\sum_{b' \in \mathcal{B}} p(B=b')p(A=a|B=b')}
\end{aligned}
$$

- Satunnaismuuttujat eivät enää välttämättä ole binäärisiä, joten summataan kaikkien B:n mahdollisten arvojen yli.

Bayesin kaava

- Johdimme juuri Bayesin kaavan! Yleisemmin diskreeteille satunnaismuuttujille:

$$
\begin{aligned}
p(B=b|A=a) &= \frac{p(B=b)p(A=a|B=b)}{p(A=a)} \\
&= \frac{p(B=b)p(A=a|B=b)}{\sum_{b'\in\mathcal{B}} p(B=b')p(A=a|B=b')}
\end{aligned}
$$

- Satunnaismuuttujat eivät enää välttämättä ole binäärisiä, joten summataan kaikkien B:n mahdollisten arvojen yli.
  - Merkitään B:n kaikkien mahdollisten arvojen joukkoa $\mathcal{B}$:llä.

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.
  - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \ldots, Y_n)$ realisaatioina.

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.
  - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \ldots, Y_n)$ realisaatioina.
  - Jatkossa merkitään kuitenkin sekä havaittua aineistoa että satunnaismuuttujia, joiden realisaatioina niitä mallinnetaan, samalla pienellä kirjaimella, ellei erikseen haluta korostaa niiden eroa.

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.
  - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \ldots, Y_n)$ realisaatioina.
  - Jatkossa merkitään kuitenkin sekä havaittua aineistoa että satunnaismuuttujia, joiden realisaatioina niitä mallinnetaan, samalla pienellä kirjaimella, ellei erikseen haluta korostaa niiden eroa.
- Havainnot asuvat **otosavaruudessa** (sample space) $\mathcal{Y}$, voi olla esim.

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.
  - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \ldots, Y_n)$ realisaatioina.
  - Jatkossa merkitään kuitenkin sekä havaittua aineistoa että satunnaismuuttujia, joiden realisaatioina niitä mallinnetaan, samalla pienellä kirjaimella, ellei erikseen haluta korostaa niiden eroa.
- Havainnot asuvat **otosavaruudessa** (sample space) $\mathcal{Y}$, voi olla esim.
  - Kaksialkoinen joukko: $\mathcal{Y} = \{0, 1\}$

Tilastollinen päättely

- Havaitaan **aineisto y** $= (y_1, \dots, y_n)$.
    - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \dots, Y_n)$ realisaatioina.
    - Jatkossa merkitään kuitenkin sekä havaittua aineistoa että satunnaismuuttujia, joiden realisaatioina niitä mallinnetaan, samalla pienellä kirjaimella, ellei erikseen haluta korostaa niiden eroa.
- Havainnot asuvat **otosavaruudessa** (sample space) $\mathcal{Y}$, voi olla esim.
    - Kaksialkoinen joukko: $\mathcal{Y} = \{0, 1\}$
    - Koko reaaliakseli: $\mathcal{Y} = \mathbb{R}$

Tilastollinen päättely

- Havaitaan **aineisto** $\mathbf{y} = (y_1, \ldots, y_n)$.
  - Voidaan mallintaa satunnaismuuttujien $\mathbf{Y} = (Y_1, \ldots, Y_n)$ realisaatioina.
  - Jatkossa merkitään kuitenkin sekä havaittua aineistoa että satunnaismuuttujia, joiden realisaatioina niitä mallinnetaan, samalla pienellä kirjaimella, ellei erikseen haluta korostaa niiden eroa.
- Havainnot asuvat **otosavaruudessa** (sample space) $\mathcal{Y}$, voi olla esim.
  - Kaksialkoinen joukko: $\mathcal{Y} = \{0, 1\}$
  - Koko reaaliakseli: $\mathcal{Y} = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\mathcal{Y} = \mathbb{R}^d$ (Tällöin jokainen yksittäinen havainto on d-komponenttinen vektori: $\mathbf{y}_1 = (y_{i1}, \ldots y_{id})$ kaikille $i = 1, \ldots, n$).

Tilastollinen päättely

- Tavoitteena selvittää aineiston (ajateltun satunnaismuuttujana) **jakauma**.

Tilastollinen päättely

- Tavoitteena selvittää aineiston (ajateltun satunnaismuuttujana) **jakauma**.
- Rajoitumme tässä **parametriseen tilastolliseen päättelyyn**, jossa aineiston jakauma oletetaan tunnetuksi tuntematonta parametrin arvoa vaille.

Tilastollinen päättely

- Tavoitteena selvittää aineiston (ajateltun satunnaismuuttujana) **jakauma**.
- Rajoitumme tässä **parametriseen tilastolliseen päättelyyn**, jossa aineiston jakauma oletetaan tunnetuksi tuntematonta parametrin arvoa vaille.
- Merkitään aineiston **otantajakaumaa** (sampling distribution) (tai tarkemmin otantajakauman tiheys- tai pistetodennäköisyysfunktiota) $p(\mathbf{y}|\boldsymbol{\theta})$

Tilastollinen päättely

- Tavoitteena selvittää aineiston (ajateltun satunnaismuuttujana) **jakauma**.
- Rajoitumme tässä **parametriseen tilastolliseen päättelyyn**, jossa aineiston jakauma oletetaan tunnetuksi tuntematonta parametrin arvoa vaille.
- Merkitään aineiston **otantajakaumaa** (sampling distribution) (tai tarkemmin otantajakauman tiheys- tai pistetodennäköisyysfunktiota) $p(\mathbf{y}|\boldsymbol{\theta})$
  - Esimerkiksi kruunien määrää y n:llä heitolla voidaan mallintaa binomijakaumalla:

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

Tilastollinen päättely

- Tavoitteena selvittää aineiston (ajateltun satunnaismuuttujana) **jakauma**.
- Rajoitumme tässä **parametriseen tilastolliseen päättelyyn**, jossa aineiston jakauma oletetaan tunnetuksi tuntematonta parametrin arvoa vaille.
- Merkitään aineiston **otantajakaumaa** (sampling distribution) (tai tarkemmin otantajakauman tiheys- tai pistetodennäköisyysfunktiota) $p(\mathbf{y}|\boldsymbol{\theta})$
  - Esimerkiksi kruunien määrää y n:llä heitolla voidaan mallintaa binomijakaumalla:

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

  - Tässä tuntematon parametri $\theta \in (0,1)$ on kruunan todennäköisyys.

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$

Mallin parametri(t)

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\Omega = \mathbb{R}^d$

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\Omega = \mathbb{R}^d$
  - Koko reaaliakselin ja positiivisen reaaliakselin karteesinen tulo: $\Omega = \mathbb{R} \times (0, \infty)$

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\Omega = \mathbb{R}^d$
  - Koko reaaliakselin ja positiivisen reaaliakselin karteesinen tulo: $\Omega = \mathbb{R} \times (0, \infty)$
- Parametri voi olla esimerkiksi:
  - Kruunan todennäköisyys $\theta \in (0, 1)$.

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\Omega = \mathbb{R}^d$
  - Koko reaaliakselin ja positiivisen reaaliakselin karteesinen tulo: $\Omega = \mathbb{R} \times (0, \infty)$
- Parametri voi olla esimerkiksi:
  - Kruunan todennäköisyys $\theta \in (0, 1)$.
  - Normaalijakauman odotusarvo ja varianssi $\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

Mallin parametri(t)

- Tavoitteena estimoida **parametrin** $\boldsymbol{\theta} \in \Omega$ arvoja.
  - Voi olla (ja lähes aina onkin) reaaliarvoinen: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$
- Asuu **parametriavaruudessa** (parameter space) Ω, joka voi olla esim:
  - Avoin väli: $\Omega = (0, 1)$
  - Koko reaaliakseli: $\Omega = \mathbb{R}$
  - Koko d-ulotteinen reaaliavaruus: $\Omega = \mathbb{R}^d$
  - Koko reaaliakselin ja positiivisen reaaliakselin karteesinen tulo: $\Omega = \mathbb{R} \times (0, \infty)$
- Parametri voi olla esimerkiksi:
  - Kruunan todennäköisyys $\theta \in (0, 1)$.
  - Normaalijakauman odotusarvo ja varianssi $\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.
  - Lineaarisen regression regressiokertoimet ja varianssi $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d, \sigma^2) \in \mathbb{R}^{d+1} \times (0, \infty)$

Tilastollinen päättely

- Tavoitteena selvittää parametrin θ arvo ehdolla havaittu aineisto $\mathbf{y}$.

Tilastollinen päättely

- Tavoitteena selvittää parametrin θ arvo ehdolla havaittu aineisto $\mathbf{y}$.
- Piste-estimointi: esimerkiksi kruunan todennäköisyyttä voidaan estimoida kruunien havaitulla osuudella $\hat{\theta} = \frac{y}{n}$.

Tilastollinen päättely

- Tavoitteena selvittää parametrin θ arvo ehdolla havaittu aineisto $\mathbf{y}$.
- Piste-estimointi: esimerkiksi kruunan todennäköisyyttä voidaan estimoida kruunien havaitulla osuudella $\hat{\theta} = \frac{y}{n}$.
  - Tämä on binomijakauman parameterin **suurimman uskottavuuden estimaatti**, eli **SU**-estimaatti (maximum likelihood estimate, a.k.a. MLE estimate), joka on siis uskottavuusfunktion suurin arvo, eli uskottavuusfunktion moodi:
$$\hat{\theta}(\mathbf{y}) = \underset{\theta}{\operatorname{argmax}}\ p(\mathbf{y}|\boldsymbol{\theta}).$$

Tilastollinen päättely

- Tilastollisen päättelyn tavoitteena on parametrin todennäköisimpien arvojen estimoinnin lisäksi myös kvantifioida näihin estimaatteihin liittyvää epävarmuutta.

Tilastollinen päättely

- Tilastollisen päättelyn tavoitteena on parametrin todennäköisimpien arvojen estimoinnin lisäksi myös kvantifioida näihin estimaatteihin liittyvää epävarmuutta.
- $\rightarrow$ Kaksi tilastollisen päättelyn teoriaa:

Tilastollinen päättely

- Tilastollisen päättelyn tavoitteena on parametrin todennäköisimpien arvojen estimoinnin lisäksi myös kvantifioida näihin estimaatteihin liittyvää epävarmuutta.
- $\rightarrow$ Kaksi tilastollisen päättelyn teoriaa:
  - Frekventistinen eli klassinen tilastollinen päättely

Tilastollinen päättely

- Tilastollisen päättelyn tavoitteena on parametrin todennäköisimpien arvojen estimoinnin lisäksi myös kvantifioida näihin estimaatteihin liittyvää epävarmuutta.
- $\rightarrow$ Kaksi tilastollisen päättelyn teoriaa:
  - Frekventistinen eli klassinen tilastollinen päättely
  - **Bayes-päättely**

Tilastollinen päättely

- Tilastollisen päättelyn tavoitteena on parametrin todennäköisimpien arvojen estimoinnin lisäksi myös kvantifioida näihin estimaatteihin liittyvää epävarmuutta.
- $\rightarrow$ Kaksi tilastollisen päättelyn teoriaa:
  - Frekventistinen eli klassinen tilastollinen päättely
  - **Bayes-päättely**
- Mallinnetaanko parametria satunnaismuuttujana (Bayes-päättely), vai kiinteänä, mutta tuntemattomana vakiona (frekventistinen teoria)?

Bayes-päättely

- Mallin parametri θ ajatellaan satunnaismuuttujan $\mathbf{\Theta}$ realisaationa.

Bayes-päättely

- Mallin parametri $\boldsymbol{\theta}$ ajatellaan satunnaismuuttujan $\boldsymbol{\Theta}$ realisaationa.
  - Kuten aineiston yhteydessä, myös parametria merkitään jatkossa pienellä (yleensä kreikkalaisella) kirjaimella riippumatta siitä tarkoitetaanko satunnaismuuttujaa, vai jotain sen tiettyä arvoa, ellei sekaannuksen vaaraa ole.

Priorijakauma

- Jos parametri mallinnettaan satunnaismuuttujana, sen pitää noudattaa jotain todennäköisyysjakaumaa.

Priorijakauma

- Jos parametri mallinnettaan satunnaismuuttujana, sen pitää noudattaa jotain todennäköisyysjakaumaa.
- Mallin parametrin jakaumaa $p(\boldsymbol{\theta})$ **ennen** aineiston havaitsemista kutsutaan priorijakaumaksi.

Priorijakauma kolikonheitolle

- Esimerkiksi kruunan todennäköisyyden θ priorijakaumaksi voidaan valita tasajakauma $U(0, 1)$.

Priorijakauma kolikonheitolle

- Esimerkiksi kruunan todennäköisyyden θ priorijakaumaksi voidaan valita tasajakauma $U(0, 1)$.
  - Tämän priorijakauman mukaan kaikki parametrin arvot välillä $(0, 1)$ ovat yhtä todennäköisiä.

Beta-jakauma

- Välin $(0, 1)$ tasajakauma on erikoistapaus **beta-jakaumasta**, nimittäin beta-jakauma $\text{Beta}(1, 1)$.

Beta-jakauma

- Välin $(0, 1)$ tasajakauma on erikoistapaus **beta-jakaumasta**, nimittäin beta-jakauma $\text{Beta}(1, 1)$.
- Beta-jakauma on välillä $(0, 1)$ määritelty todennäköisyysjakauma, jonka tiheysfunktio on

$$p(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1},$$

missä normalisointivakio $B(\alpha, \beta)$ on ns. Eulerin beta-funktio:

$$B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}\, \mathrm{d}\theta.$$

Esimerkkejä beta-jakaumista

- Beta-jakautuneen satunnaismuuttujan $\theta \sim \mathrm{Beta}(\alpha, \beta)$ odotusarvo on $E\theta = \frac{\alpha}{\alpha+\beta}$

Esimerkkejä beta-jakaumista

- Beta-jakautuneen satunnaismuuttujan $\theta \sim \text{Beta}(\alpha, \beta)$ odotusarvo on $E\theta = \frac{\alpha}{\alpha+\beta}$
- Jos ennakkokäsitysten mukaan on todennäköisempää että kruunan todennäköisyys on lähempänä puolta kuin nollaa tai yhtä, voidaan valita prioriksi esimerkiksi Beta$(5, 5)$.

Esimerkkejä beta-jakaumista

- Beta-jakautuneen satunnaismuuttujan $\theta \sim \text{Beta}(\alpha, \beta)$ odotusarvo on $E\theta = \frac{\alpha}{\alpha+\beta}$
- Jos ennakkokäsitysten mukaan on todennäköisempää että kruunan todennäköisyys on lähempänä puolta kuin nollaa tai yhtä, voidaan valita prioriksi esimerkiksi $\text{Beta}(5, 5)$.
  - Odotusarvo on edelleen $E\theta = \frac{5}{5+5} = \frac{1}{2}$, mutta jakauma keskittyneempi arvoihin lähellä puolta.

Esimerkkejä beta-jakaumista

- Beta-jakautuneen satunnaismuuttujan $\theta \sim \mathrm{Beta}(\alpha, \beta)$ odotusarvo on $E\theta = \frac{\alpha}{\alpha+\beta}$
- Jos ennakkokäsitysten mukaan on todennäköisempää että kruunan todennäköisyys on lähempänä puolta kuin nollaa tai yhtä, voidaan valita prioriksi esimerkiksi $\mathrm{Beta}(5, 5)$.
  - Odotusarvo on edelleen $E\theta = \frac{5}{5+5} = \frac{1}{2}$, mutta jakauma keskittyneempi arvoihin lähellä puolta.
  - Priorijakauman parametreja, beta-priorin tapauksessa parametreja α ja β kutsutaan **hyperparametreiksi**.

Posteriorijakauma

- Bayes-päättelyn **keskeinen idea** on arviomme parametrin todennäköisistä arvoista ennen havaintojen tekemistä päivittäinen arvioksi aineiston havaitsemisen **jälkeen**.

Posteriorijakauma

- Bayes-päättelyn **keskeinen idea** on arviomme parametrin todennäköisistä arvoista ennen havaintojen tekemistä päivittäinen arvioksi aineiston havaitsemisen **jälkeen**.
- Parametrin jakaumaa $p(\theta|\mathbf{y})$ ehdolla havaittu aineisto $\mathbf{y}$ kutsutaan parametrin **posteriorijakaumaksi**.

Posteriorijakauma

- Bayes-päättelyn **keskeinen idea** on arviomme parametrin todennäköisistä arvoista ennen havaintojen tekemistä päivittäinen arvioksi aineiston havaitsemisen **jälkeen**.
- Parametrin jakaumaa $p(\theta|\mathbf{y})$ ehdolla havaittu aineisto $\mathbf{y}$ kutsutaan parametrin **posteriorijakaumaksi**.
  - Posteriorijakauma siis kuvaa arviotamme parametrin arvojen todennäköisyyksistä aineiston havaitsemisen jälkeen.

Bayesin kaava

- Priorijakauman päivittäminen posteriorijakaumaksi tapahtuu **Bayesin kaavan** avulla:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

Bayesin kaava

- Priorijakauman päivittäminen posteriorijakaumaksi tapahtuu **Bayesin kaavan** avulla:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

- Periaattellisella tasolla tämä on puolet Bayes-päättelyn keskeisistä kaavoista...

Bayesin kaava

- Priorijakauman päivittäminen posteriorijakaumaksi tapahtuu **Bayesin kaavan** avulla:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

- Periaattellisella tasolla tämä on puolet Bayes-päättelyn keskeisistä kaavoista...
- Ja tätäkin voi vielä hieman yksinkertaistaa: posteriorijakauma on parametrin $\boldsymbol{\theta}$ funktio, joten aineiston reunatodennäköisyys $p(\mathbf{y})$ on vakio (parametrin suhteen).

Bayesin kaava

- Priorijakauman päivittäminen posteriorijakaumaksi tapahtuu **Bayesin kaavan** avulla:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

- Periaattellisella tasolla tämä on puolet Bayes-päättelyn keskeisistä kaavoista...
- Ja tätäkin voi vielä hieman yksinkertaistaa: posteriorijakauma on parametrin $\boldsymbol{\theta}$ funktio, joten aineiston reunatodennäköisyys $p(\mathbf{y})$ on vakio (parametrin suhteen).
  - Usein Bayesin kaavasta käytetään muotoa, jossa posteriori kirjoitetaan verrannollisena otantajakauman ja priorin tuloon:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}).$$

Bayesin kaava

- Priorijakauman päivittäminen posteriorijakaumaksi tapahtuu **Bayesin kaavan** avulla:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

- Periaattellisella tasolla tämä on puolet Bayes-päättelyn keskeisistä kaavoista...
- Ja tätäkin voi vielä hieman yksinkertaistaa: posteriorijakauma on parametrin $\boldsymbol{\theta}$ funktio, joten aineiston reunatodennäköisyys $p(\mathbf{y})$ on vakio (parametrin suhteen).
  - Usein Bayesin kaavasta käytetään muotoa, jossa posteriori kirjoitetaan verrannollisena otantajakauman ja priorin tuloon:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}).$$

  - Verrannollisuussymboli $f(x) \propto g(x)$ tarkoittaa, että on olemassa vakio $c \in \mathbb{R}$ s.e. $f(x) = cg(x)$ kaikille x.

Aineiston reunatodennäköisyys

- Bayesin kaavan nimittäjää, eli aineiston reunatodennäköisyyttä $p(\mathbf{y})$ kutsutaan **reunauskottavuudeksi** (marginal likelihood).

Aineiston reunatodennäköisyys

- Bayesin kaavan nimittäjää, eli aineiston reunatodennäköisyyttä $p(\mathbf{y})$ kutsutaan **reunauskottavuudeksi** (marginal likelihood).
- Jos parametri $\boldsymbol{\theta}$ on diskreetti, reunauskottavuus saadaan laskettua kokonaistodennäköisyyden kaavasta summaamalla kaikkien mahdollisten parametrin arvojen yli (vrt. ensimmäinen esimerkki Bayesin kaavasta):

$$p(\mathbf{y}) = \sum_{\boldsymbol{\theta} \in \Omega} p(\boldsymbol{\theta}) p(\mathbf{y} | \boldsymbol{\theta}).$$

Aineiston reunatodennäköisyys

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Bayesin kaavan nimittäjää, eli aineiston reunatodennäköisyyttä $p(\mathbf{y})$ kutsutaan **reunauskottavuudeksi** (marginal likelihood).
- Jos parametri $\boldsymbol{\theta}$ on diskreetti, reunauskottavuus saadaan laskettua kokonaistodennäköisyyden kaavasta summaamalla kaikkien mahdollisten parametrin arvojen yli (vrt. ensimmäinen esimerkki Bayesin kaavasta):

$$p(\mathbf{y}) = \sum_{\boldsymbol{\theta} \in \Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}).$$

- Jos (ja kun...) parametrilla $\boldsymbol{\theta}$ on jatkuva jakauma, joudutaan integroimaan yli parametriavaruuden:

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}.$$

Pakollinen kolikonheittoesimerkki

- Johdetaan posteriorijakauma mallille, jossa otantajakauma on binomijakauma, ja priorijakauma parametrille θ on betajakauma:

$$
\begin{aligned}
y|\theta &\sim \mathrm{Bin}(n, \theta) \\
\theta &\sim \mathrm{Beta}(\alpha, \beta).
\end{aligned}
$$

Pakollinen kolikonheittoesimerkki

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Johdetaan posteriorijakauma mallille, jossa otantajakauma on binomijakauma, ja priorijakauma parametrille θ on betajakauma:

$$\begin{aligned} y|\theta &\sim \text{Bin}(n, \theta) \\ \theta &\sim \text{Beta}(\alpha, \beta). \end{aligned}$$

- Priorijakauman parametrit voivat olla mitä tahansa vakioita $\alpha, \beta > 0$.

Pakollinen kolikonheittoesimerkki

- Johdetaan posteriorijakauma mallille, jossa otantajakauma on binomijakauma, ja priorijakauma parametrille θ on betajakauma:

$$\begin{aligned} y|\theta &\sim \text{Bin}(n, \theta) \\ \theta &\sim \text{Beta}(\alpha, \beta). \end{aligned}$$

- Priorijakauman parametrit voivat olla mitä tahansa vakioita $\alpha, \beta > 0$.
  - Esimerkiksi jos $\alpha = \beta = 1$, priori on tasajakauma

$$\text{Beta}(1, 1) \overset{d}{=} U(0, 1).$$

Posteriorijakauma kolikonheittoesimerkille

- Sovelletaan verrannollisuusversiota Bayesin kaavasta:

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta)p(y|\theta) \\
&= \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\cdot\binom{n}{y}\theta^{y}(1-\theta)^{n-y} \\
&\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}
\end{aligned}
$$

Posteriorijakauma kolikonheittoesimerkille

- Sovelletaan verrannollisuusversiota Bayesin kaavasta:

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p(y|\theta) \\ &= \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\cdot\binom{n}{y}\theta^{y}(1-\theta)^{n-y} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \end{aligned}$$

- Viimeiselle riville siirryttäessä poistettiin parametrista riippumattomat vakiotermit, ja yhdistettiin θ:n ja $(1-\theta)$:n eksponentit.

Posteriorijakauma kolikonheittoesimerkille

- Sovelletaan verrannollisuusversiota Bayesin kaavasta:

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta)p(y|\theta) \\
&= \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\cdot\binom{n}{y}\theta^{y}(1-\theta)^{n-y} \\
&\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}
\end{aligned}
$$

- Viimeiselle riville siirryttäessä poistettiin parametrista riippumattomat vakiotermit, ja yhdistettiin θ:n ja $(1-\theta)$:n eksponentit.
- Tämä voidaan tunnistaa beta-jakauman $\mathrm{Beta}(y+\alpha,\ n-y+\beta)$ tiheysfunktion ytimeksi (eli normalisoimattomaksi tiheysfunktioksi)

Posteriorijakauma kolikonheittoesimerkille

- Posteriorijakaumaksi saatiin johdettua

$$\theta|y \sim \mathrm{Beta}(y+\alpha,\ n-y+\beta).$$

Posteriorijakauma kolikonheittoesimerkille

- Posteriorijakaumaksi saatiin johdettua

$$\theta | y \sim \mathrm{Beta}(y + \alpha,\ n - y + \beta).$$

- Jos priori on tasajakauma $U(0, 1)$, niin posteriorijakauma on

$$\theta | y \sim \mathrm{Beta}(y + 1,\ n - y + 1).$$

Posteriorijakauman odotusarvo

- Päättelyprosessi tulos on koko posteriorijakauma.

Posteriorijakauman odotusarvo

- Päättelyprosessi tulos on koko posteriorijakauma.
- Monesti posteriorijakauma halutaan kuitenkin tiivistää yhteen piste-estimaattiin.

Posteriorijakauman odotusarvo

- Päättelyprosessi tulos on koko posteriorijakauma.
- Monesti posteriorijakauma halutaan kuitenkin tiivistää yhteen piste-estimaattiin.
- Voidaan käyttää posteriorijakauman moodia, mediaania, tai odotusarvoa.

Posteriorijakauman odotusarvo

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- Päättelyprosessi tulos on koko posteriorijakauma.
- Monesti posteriorijakauma halutaan kuitenkin tiivistää yhteen piste-estimaattiin.
- Voidaan käyttää posteriorijakauman moodia, mediaania, tai odotusarvoa.
- Bayes-päättelyssä piste-estimaattina käytetään yleensä posteriorijakauman odotusarvoa

$$E(\theta | Y = y)$$

Posteriorijakauman odotusarvo

- Päättelyprosessi tulos on koko posteriorijakauma.
- Monesti posteriorijakauma halutaan kuitenkin tiivistää yhteen piste-estimaattiin.
- Voidaan käyttää posteriorijakauman moodia, mediaania, tai odotusarvoa.
- Bayes-päättelyssä piste-estimaattina käytetään yleensä posteriorijakauman odotusarvoa

$$E(\theta|Y = y)$$

  - Voidaan osoittaa, että tämä on keskineliövirheen mielessä optimaalinen piste-estimaatti.

Posteriorijakauman odotusarvo esimerkissä

- Kolikonheittoesimerkissä posteriorijakauman odotusarvo saadaan laskettua helposti beta-jakauman odotusarvona

$$E(\theta|Y=y) = \frac{y+\alpha}{n+\alpha+\beta}.$$

Posteriorijakauman odotusarvo esimerkissä

- Kolikonheittoesimerkissä posteriorijakauman odotusarvo saadaan laskettua helposti beta-jakauman odotusarvona

$$E(\theta|Y = y) = \frac{y + \alpha}{n + \alpha + \beta}.$$

- Käytettäessä tasajakaumaa $U(0, 1)$ priorina, posteriorijakauman odotusarvo on

$$E(\theta|Y = y) = \frac{y + 1}{n + 2}.$$

Posteriorijakauman odotusarvo esimerkissä

- Kolikonheittoesimerkissä posteriorijakauman odotusarvo saadaan laskettua helposti beta-jakauman odotusarvona

$$E(\theta|Y=y) = \frac{y+\alpha}{n+\alpha+\beta}.$$

- Käytettäessä tasajakaumaa $U(0,1)$ priorina, posteriorijakauman odotusarvo on

$$E(\theta|Y=y) = \frac{y+1}{n+2}.$$

- Priorijakauman tulkinta pseudo-havaintoina.
  - Lisätään oikeisiin havaintoihin yhteensä $\alpha+\beta$ pseudo-havaintoa, joista α kruunaa ja β klaavaa.

Posteriorijakauman odotusarvo esimerkissä

- Kolikonheittoesimerkissä posteriorijakauman odotusarvo saadaan laskettua helposti beta-jakauman odotusarvona

$$E(\theta|Y = y) = \frac{y + \alpha}{n + \alpha + \beta}.$$

- Käytettäessä tasajakaumaa $U(0, 1)$ priorina, posteriorijakauman odotusarvo on

$$E(\theta|Y = y) = \frac{y + 1}{n + 2}.$$

- Priorijakauman tulkinta pseudo-havaintoina.
  - Lisätään oikeisiin havaintoihin yhteensä $\alpha + \beta$ pseudo-havaintoa, joista α kruunaa ja β klaavaa.
  - α ja β eivät välttämättä kokonaislukuja, esim. priori Beta$(\frac{1}{2}, \frac{1}{2})$ vastaa puolen kruunan ja puolen klaavan lisäämistä aineistoon.

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$p(\theta|y) = ?$$

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$\cancel{p(\theta|y)} = ?$$

... ei ole hyvin määritelty.

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$\cancel{p(\theta|y)} = ?$$

  ... ei ole hyvin määritelty.

- Mallin parametriin ei voida liittää todennäköisyysväittämiä, vaan se on kiinteä, mutta tuntematon vakio.

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$\cancel{p(\theta|y)} = ?$$

  ... ei ole hyvin määritelty.

- Mallin parametriin ei voida liittää todennäköisyysväittämiä, vaan se on kiinteä, mutta tuntematon vakio.
  - Kiinnitetään parametrin arvo, esim. $H_0 : \theta = 0.5$, ja tarkastellaan otantajakaumaa $p(y|\theta)$.

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$\cancel{p(\theta|y)} = ?$$

  ... ei ole hyvin määritelty.

- Mallin parametriin ei voida liittää todennäköisyysväittämiä, vaan se on kiinteä, mutta tuntematon vakio.
  - Kiinnitetään parametrin arvo, esim. $H_0 : \theta = 0.5$, ja tarkastellaan otantajakaumaa $p(y|\theta)$.
  - Mikä on todennäköisyys havaita vähintään yhtä äärimmäinen aineisto kuin havaittu aineisto kiinnitetyllä parameterin arvolla?

Frekventistinen tilastotiede

- Todennäköisyysjakauma parametrien arvojen yli:

$$\cancel{p(\theta|y) = ?}$$

  ... ei ole hyvin määritelty.
- Mallin parametriin ei voida liittää todennäköisyysväittämiä, vaan se on kiinteä, mutta tuntematon vakio.
  - Kiinnitetään parametrin arvo, esim. $H_0 : \theta = 0.5$, ja tarkastellaan otantajakaumaa $p(y|\theta)$.
  - Mikä on todennäköisyys havaita vähintään yhtä äärimmäinen aineisto kuin havaittu aineisto kiinnitetyllä parameterin arvolla?
  - Esimerkiksi mikä on todennäköisyys, että 10 heitolla saadaan 2 tai alle 2 kruunaa (tai vastaavasti klaavaa), jos kruunan todennäköisyys on 0.5?

- p-arvo: todennäköisyys, että 10 heitolla saadaan 2 tai alle 2 kruunaa (tai vastaavasti klaavaa), jos kruunan todennäköisyys on 0.5:

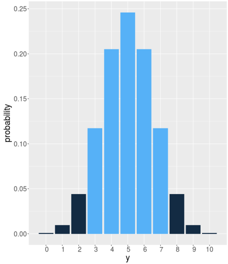
$$p = P(Y \leq 2 \cup Y \geq 8 | \theta = 0.5) \approx 0.109$$

Frekventistinen tilastotiede

- p-arvo: todennäköisyys, että 10 heitolla saadaan 2 tai alle 2 kruunaa (tai vastaavasti klaavaa), jos kruunan todennäköisyys on 0.5:

$$p = P(Y \leq 2 \cup Y \geq 8 | \theta = 0.5) \approx 0.109$$

- Verrataan p-arvoa etukäteen määriteltyyn merkitsevyystasoon (esim. $\alpha = 0.05$): tässä tapauksessa $p = 0.109 > 0.05$, joten nollahypoteesia ei voida hylätä.



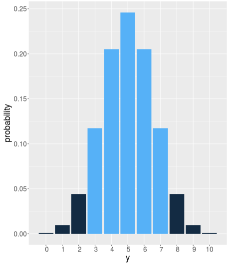
Frekventistinen tilastotiede

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- p-arvo: todennäköisyys, että 10 heitolla saadaan 2 tai alle 2 kruunaa (tai vastaavasti klaavaa), jos kruunan todennäköisyys on 0.5:

$$\cancel{p = P(Y \leq 2 \cup Y \geq 8 \mid \theta = 0.5) \approx 0.109}$$

- Verrataan p-arvoa etukäteen määriteltyyn merkitsevyystasoon (esim. $\alpha = 0.05$): tässä tapauksessa $p = 0.109 > 0.05$, joten nollahypoteesia ei voida hylätä.



Lyhyt katsaus historiaan...

- Esihistoria 1700-1900: Bayes ja Laplace
- Modernin tilastotieteen kehitys 1900-1930
- Frekventismin valtakausi 1930-1960
- Bayes-päättelyn uusi tuleminen 1960-2000
- Nykyaika 2000 -

Thomas Bayes (1701? - 1761)

- Englantilainen presbyteeripappi ja amatöörimatemaatikko.
- Ei julkaissut mitään todennäköisyyteen liittyvää elinaikanaan.
- Bayesin ystävä Richard Price editoi hänen jäämistöstään löytyneen käsikirjoituksen, joka julkaistiin 1763 nimellä *An essay towards solving a problem in the doctrine of chances*.
  - Sisältää erikoistapauksen **Bayesin kaavasta**
  - Fyysinen analogia: parametrin sijainnin parametriavaruudessa kuvaaminen pallon sijaintina biljardipöydällä.

Pierre Simon Laplace (1749-1827)

- Ranskalainen matemaatikko ja fyysikko, "Ranskan Newton"
- Merkittäviä läpimurtoja matemaattisessa fysiikassa, astronomiassa ja erityisesti **todennäköisyyslaskennassa** ja **tilastotieteessä**:
  - todennäköisyysgeneroiva ja karakteristinen funktio
  - pienimmän neliösumman menetelmä
  - keskeinen raja-arvolause
  - **Bayesin kaava** yleisessä muodossa

Pierre Simon Laplace (1749-1827)

- 1745 - 1770 Pariisissa syntyi $n = 493472$ lasta, joista $y = 241945$ oli tyttöjä (ja vastaavasti $n - y = 251527$ poikia).

Pierre Simon Laplace (1749-1827)

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- 1745 - 1770 Pariisissa syntyi $n = 493472$ lasta, joista $y = 241945$ oli tyttöjä (ja vastaavasti $n - y = 251527$ poikia).
- Voidaanko tämän aineiston perusteella päätellä, syntyykö poikia enemmän kuin tyttöjä?

Pierre Simon Laplace (1749-1827)

- 1745 - 1770 Pariisissa syntyi $n = 493472$ lasta, joista $y = 241945$ oli tyttöjä (ja vastaavasti $n - y = 251527$ poikia).
- Voidaanko tämän aineiston perusteella päätellä, syntyykö poikia enemmän kuin tyttöjä?
- Sama malli kuin kolikko-esimerkissä! Laskettava posteriorijakauman perusteella todennäköisyys

$$p(\theta > 0.5|y) = ?$$

Pierre Simon Laplace (1749-1827)

- Posteriorijakaumaksi saadaan

$$\mathrm{Beta}(241945 + 1, 251527 + 1) \stackrel{d}{=} \mathrm{Beta}(241946, 251528)$$

Pierre Simon Laplace (1749-1827)

- Posteriorijakaumaksi saadaan

$$\text{Beta}(241945 + 1, 251527 + 1) \stackrel{d}{=} \text{Beta}(241946, 251528)$$

- Posteriorijakauman odotusarvo on

$$E[\theta|Y = y] = \frac{241946}{241946 + 251527} \approx 0.49$$

Pierre Simon Laplace (1749-1827)

- Posteriorijakaumaksi saadaan

$$\text{Beta}(241945 + 1, 251527 + 1) \stackrel{d}{=} \text{Beta}(241946, 251528)$$

- Posteriorijakauman odotusarvo on

$$E[\theta | Y = y] = \frac{241946}{241946 + 251527} \approx 0.49$$

- Posteriorijakaumasta saadaan laskettua todennäköisyys että tytön todennäköisyys on suurempaa kuin puoli:

$$p(\theta > 0.5) \approx 1.15 \cdot 10^{-42}$$

Pierre Simon Laplace (1749-1827)

- Posteriorijakaumaksi saadaan

$$\text{Beta}(241945 + 1, 251527 + 1) \stackrel{d}{=} \text{Beta}(241946, 251528)$$

- Posteriorijakauman odotusarvo on

$$E[\theta | Y = y] = \frac{241946}{241946 + 251527} \approx 0.49$$

- Posteriorijakaumasta saadaan laskettua todennäköisyys että tytön todennäköisyys on suurempaa kuin puoli:

$$p(\theta > 0.5) \approx 1.15 \cdot 10^{-42}$$

- Laplace oli tämän perusteella "moraalisesti varma", että poikia syntyy enemmän kuin tyttöjä.

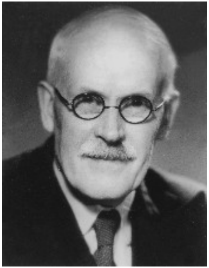
Frekventistinen tilastotiede

- 1900-luvun alussa Karl Pearson, William Gosset (a.k.a. "student"), Jerzy Neyman ja **Ronald Fisher** kehittivät formaalia tilastollisen päättelyn teoriaa.
- Fisher: *Statistical methods for research workers*, 1925: Ensimmäinen soveltajille suunnattu tilastotieteen metodiopas.
- Fisherin teoria tunnetaan nykyään **frekventistisenä** tai **klassisena** tilastotieteenä.

Sir Ronald Fisher

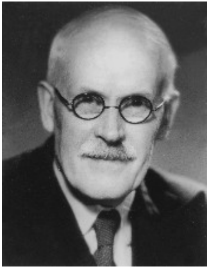
Harold Jeffreys (1891-1989)

- Samaan aikaan brittiläinen matemaatikko ja geofyysikko Harold Jeffreys kehitti 1930-luvulla Bayesiläistä vaihtoehtoa Fisherin p-arvoille ja merkitsevyystesteille: *Bayes-faktorit*



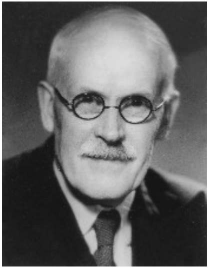
Harold Jeffreys (1891-1989)

- Samaan aikaan brittiläinen matemaatikko ja geofyysikko Harold Jeffreys kehitti 1930-luvulla Bayesiläistä vaihtoehtoa Fisherin p-arvoille ja merkitsevyystesteille: *Bayes-faktorit*
- "Ainoa Bayesiläinen" 30-luvulla



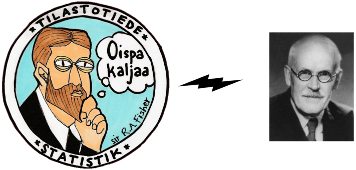
Harold Jeffreys (1891-1989)

- Samaan aikaan brittiläinen matemaatikko ja geofyysikko Harold Jeffreys kehitti 1930-luvulla Bayesiläistä vaihtoehtoa Fisherin *p*-arvoille ja merkitsevyystesteille: *Bayes-faktorit*
- "Ainoa Bayesiläinen" 30-luvulla
- *Theory of probability*, 1939: Pitkän aikaa ainoa teos, jossa selitetään kuinka soveltaa Bayesiläisiä menetelmiä.



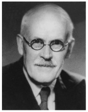
Frekventistinen tilastotiede

- Fisher vastusti hyvin voimakkaasti Bayesiläistä tilastotiedettä



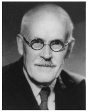
Frekventistinen tilastotiede

- Fisher vastusti hyvin voimakkaasti Bayesiläistä tilastotiedettä
  - *"The theory of inverse probability is founded upon an error, and must be wholly rejected".*



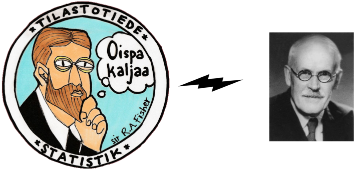
Frekventistinen tilastotiede

- Fisher vastusti hyvin voimakkaasti Bayesiläistä tilastotiedettä
  - *"The theory of inverse probability is founded upon an error, and must be wholly rejected"*.
  - Ihaili Bayesiä ongelman muotoilusta, sen ratkaisemisesta, mutta ennen kaikkea siitä, että tämä ei ikinä julkaissut ratkaisuaan...



Frekventistinen tilastotiede

- Fisher vastusti hyvin voimakkaasti Bayesiläistä tilastotiedettä
  - *"The theory of inverse probability is founded upon an error, and must be wholly rejected".*
  - Ihaili Bayesiä ongelman muotoilusta, sen ratkaisemisesta, mutta ennen kaikkea siitä, että tämä ei ikinä julkaissut ratkaisuaan...
- Kiivasta väittelyä Jeffreys'n kanssa tieteellisissä julkaisuissa ja konferensseissa.



1930 - 1960: Frekventismin valtakausi

- Frekventistisestä tilastotieteestä tuli valtavirtaa, ja Bayesiläisten menetelmien kehittäminen ja käyttäminen lakkasi käytännössä kokonaan useiksi vuosikymmeniksi.

1930 - 1960: Frekventismin valtakausi

- Frekventistisestä tilastotieteestä tuli valtavirtaa, ja Bayesiläisten menetelmien kehittäminen ja käyttäminen lakkasi käytännössä kokonaan useiksi vuosikymmeniksi.
  - Helpompi soveltaa, selkeät säännöt ja ei tarvetta priorijakaumalle.

1930 - 1960: Frekventismin valtakausi

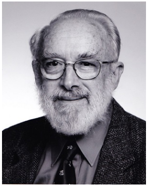
- Frekventistisestä tilastotieteestä tuli valtavirtaa, ja Bayesiläisten menetelmien kehittäminen ja käyttäminen lakkasi käytännössä kokonaan useiksi vuosikymmeniksi.
  - Helpompi soveltaa, selkeät säännöt ja ei tarvetta priorijakaumalle.
  - "Objektiivisempi" lähestymistapa, ei subjektiivisuutta, joka liittyy priorin valintaan.

1930 - 1960: Frekventismin valtakausi

- Frekventistisestä tilastotieteestä tuli valtavirtaa, ja Bayesiläisten menetelmien kehittäminen ja käyttäminen lakkasi käytännössä kokonaan useiksi vuosikymmeniksi.
  - Helpompi soveltaa, selkeät säännöt ja ei tarvetta priorijakaumalle.
  - "Objektiivisempi" lähestymistapa, ei subjektiivisuutta, joka liittyy priorin valintaan.
  - **Laskennallisesti** kevyempää: vain kaikkein yksinkertaisimpien Bayesiläisten mallien sovittaminen mahdollista ennen tietokoneita.

1960 - 2000: Uusi tuleminen

- Teorian kehitys: Dennis Lindley (1923-2013) ja Jimmie Savage (1917-1971) kehittivät aksiomaattista perustaa tilastotieteelle.

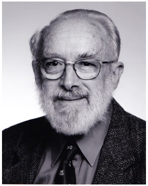


Dennis Lindley

Jimmie Savage

1960 - 2000: Uusi tuleminen

- Teorian kehitys: Dennis Lindley (1923-2013) ja Jimmie Savage (1917-1971) kehittivät aksiomaattista perustaa tilastotieteelle.
  - *Johtaa Bayesiläiseen teoriaan 'lähes vahingossa'*

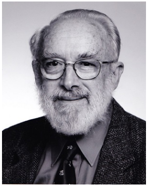


Dennis Lindley

Jimmie Savage

1960 - 2000: Uusi tuleminen

- Teorian kehitys: Dennis Lindley (1923-2013) ja Jimmie Savage (1917-1971) kehittivät aksiomaattista perustaa tilastotieteelle.
  - *Johtaa Bayesiläiseen teoriaan 'lähes vahingossa'*
- Tietokoneiden ja simulaatiomenetelmien, erityisesti Markov chain Monte Carlo (MCMC), kehitys mahdollisti monimutkaisempien mallien sovittamisen.

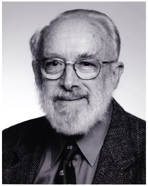


Dennis Lindley

Jimmie Savage

1960 - 2000: Uusi tuleminen

- Teorian kehitys: Dennis Lindley (1923-2013) ja Jimmie Savage (1917-1971) kehittivät aksiomaattista perustaa tilastotieteelle.
  - *Johtaa Bayesiläiseen teoriaan 'lähes vahingossa'*
- Tietokoneiden ja simulaatiomenetelmien, erityisesti Markov chain Monte Carlo (MCMC), kehitys mahdollisti monimutkaisempien mallien sovittamisen.
- Hierarkkisten mallien kehittäminen: hyviä käytännön tuloksia.



Dennis Lindley

Jimmie Savage

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.
  - Pitkät perinteet.

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.
  - Pitkät perinteet.
- Bayesiläiset menetelmät valtaamassa alaa tilastotieteen teoriassa ja monilla sovellusaloilla, kuten biologiassa.

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.
  - Pitkät perinteet.
- Bayesiläiset menetelmät valtaamassa alaa tilastotieteen teoriassa ja monilla sovellusaloilla, kuten biologiassa.
  - Koneoppimisen ja tekoälyn tutkimuksessa käytössä Bayesiläiset menetelmät valtavirtaa (silloin kun ei tehdä pelkkää piste-estimointia).

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.
  - Pitkät perinteet.
- Bayesiläiset menetelmät valtaamassa alaa tilastotieteen teoriassa ja monilla sovellusaloilla, kuten biologiassa.
  - Koneoppimisen ja tekoälyn tutkimuksessa käytössä Bayesiläiset menetelmät valtavirtaa (silloin kun ei tehdä pelkkää piste-estimointia).
- Probabilistinen ohjelmointi (probabilistic programming)

2000-luku: Mitä seuraavaksi?

- Frekventistiset menetelmät edelleen valtavirtaa monilla sovellusaloilla, kuten lääketieteessä.
  - Tarvetta selkeille säännöille: esim. lääkkeiden toimivuuden testauksessa.
  - Pitkät perinteet.
- Bayesiläiset menetelmät valtaamassa alaa tilastotieteen teoriassa ja monilla sovellusaloilla, kuten biologiassa.
  - Koneoppimisen ja tekoälyn tutkimuksessa käytössä Bayesiläiset menetelmät valtavirtaa (silloin kun ei tehdä pelkkää piste-estimointia).
- Probabilistinen ohjelmointi (probabilistic programming)
  - Mahdollistaa Bayes-mallien käyttämisen *lähes* out-of-the-box.

2000-luku

- Nykyään ei enää juurikaan vastakkainasettelua frekventistisen ja Bayes-päättelyn välillä.

2000-luku

- Nykyään ei enää juurikaan vastakkainasettelua frekventistisen ja Bayes-päättelyn välillä.
- Vaikka Bayes-päättelyssä parametria mallinnetaan satunnaismuuttujana, voidaan kuitenkin puhua myös parametrin todellisesta arvosta.

2000-luku

- Nykyään ei enää juurikaan vastakkainasettelua frekventistisen ja Bayes-päättelyn välillä.
- Vaikka Bayes-päättelyssä parametria mallinnetaan satunnaismuuttujana, voidaan kuitenkin puhua myös parametrin todellisesta arvosta.
  - Voidaan puhua estimaattorien asymptoottisista ominaisuuksista, kuten tarkentuvuus (consistency) ja konvergenssinopeus (convergence rate).

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.
  - Bayesiläinen uskottavuusväli: väli, jolle parametrin arvo kuuluu 95% todennäköisyydellä.

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.
  - Bayesiläinen uskottavuusväli: väli, jolle parametrin arvo kuuluu 95% todennäköisyydellä.
- Välineet myös *ennustamiseen*, ei pelkästään selittämiseen.

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.
  - Bayesiläinen uskottavuusväli: väli, jolle parametrin arvo kuuluu 95% todennäköisyydellä.
- Välineet myös *ennustamiseen*, ei pelkästään selittämiseen.
  - Vakuutusala, vedonlyönti, koneoppimis- ja tekoälyjärjestelmät, jne.

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.
  - Bayesiläinen uskottavuusväli: väli, jolle parametrin arvo kuuluu 95% todennäköisyydellä.
- Välineet myös *ennustamiseen*, ei pelkästään selittämiseen.
  - Vakuutusala, vedonlyönti, koneoppimis- ja tekoälyjärjestelmät, jne.
- Kaunis ja yhtenäinen teoria, joka selittää paljon sitä, mitä käytännön sovelluksissa tehdään.

Motivaatio: miksi Bayes?

- Analyysin tulokset helpompia tulkita (oikein!).
  - Voidaan antaa suoraan todennäköisyyksiä parametrin arvoille.
  - Bayesiläinen uskottavuusväli: väli, jolle parametrin arvo kuuluu 95% todennäköisyydellä.
- Välineet myös *ennustamiseen*, ei pelkästään selittämiseen.
  - Vakuutusala, vedonlyönti, koneoppimis- ja tekoälyjärjestelmät, jne.
- Kaunis ja yhtenäinen teoria, joka selittää paljon sitä, mitä käytännön sovelluksissa tehdään.
  - Regularisaatio, smoothing, pseudohavainnot jne. tulkittavissa priorijakaumana.

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!
  - Esim. klassinen lineaarinen regressio: hyvin voimakkaita oletuksia, kuten riippuvuuden lineaarisuus, virhetermien normaalijakautuneisuus ja heteroskedastisuus, jne.

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!
  - Esim. klassinen lineaarinen regressio: hyvin voimakkaita oletuksia, kuten riippuvuuden lineaarisuus, virhetermien normaalijakautuneisuus ja heteroskedastisuus, jne.
- Mallien sovittaminen työlästä: vielä 10 vuotta MCMC-sämpleri uudelle mallille johdettiin ja kirjoitettiin monesti itse

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!
  - Esim. klassinen lineaarinen regressio: hyvin voimakkaita oletuksia, kuten riippuvuuden lineaarisuus, virhetermien normaalijakautuneisuus ja heteroskedastisuus, jne.
- Mallien sovittaminen työlästä: vielä 10 vuotta MCMC-sämpleri uudelle mallille johdettiin ja kirjoitettiin monesti itse
  - Vaatii monen viikon työn tilastotieteilijältä.

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!
  - Esim. klassinen lineaarinen regressio: hyvin voimakkaita oletuksia, kuten riippuvuuden lineaarisuus, virhetermien normaalijakautuneisuus ja heteroskedastisuus, jne.
- Mallien sovittaminen työlästä: vielä 10 vuotta MCMC-sämpleri uudelle mallille johdettiin ja kirjoitettiin monesti itse
  - Vaatii monen viikon työn tilastotieteilijältä.
  - Nykyään kuitenkin yleiskäyttöisiä Hamiltonian Monte Carlo (HMC)-sämplereitä ja niiden helppokäyttöisiä implementaatioita, kuten **Stan**, PyMC3 ja Edward (probabilistic programming tools).

Haasteita

- Eikö priorijakauman valinta tee analyysistä subjektiivista?
  - Myös otantajakauman valinta aina subjektiivinen päätös!
  - Esim. klassinen lineaarinen regressio: hyvin voimakkaita oletuksia, kuten riippuvuuden lineaarisuus, virhetermien normaalijakautuneisuus ja heteroskedastisuus, jne.
- Mallien sovittaminen työlästä: vielä 10 vuotta MCMC-sämpleri uudelle mallille johdettiin ja kirjoitettiin monesti itse
  - Vaatii monen viikon työn tilastotieteilijältä.
  - Nykyään kuitenkin yleiskäyttöisiä Hamiltonian Monte Carlo (HMC)-sämplereitä ja niiden helppokäyttöisiä implementaatioita, kuten **Stan**, PyMC3 ja Edward (probabilistic programming tools).
  - Riittää määritellä pelkästään malli, sovittaminen automaattista.