MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

5A Bayesläinen tilastollinen päättely

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen

Bayesläinen päättely jatkuvilla malleilla

Kolikonheiton bayesläinen malli

Esimerkki: Kolikko

Tasaiseksi väitettyä kolikkoa heitettäessä ($0 =$ klaava, $1 =$ kruuna) saatiin tulossarja

$$\vec{x} = (0, 0, 0, 0, 0, 0, 0, 0, 1, 0)$$

Kruunan odotettua esiintyvyyttä kuvaavan parametrin p suurimman uskottavuuden estimaatti $\hat{p}(x) = 10\%$ poikkeaa vahvasti tasosta 50%. Miten tähän pitäisi suhtautua?

- Tulee hankkia lisää dataa (entä jos ei saatavilla?)
- Tulee jättää kokeen tulos huomiotta (ja uskoa sokeasti siihen, että kolikko on tasainen?)

Voidaanko yo. kokeen tulos sovittaa aiempaan tietämykseen kruunan odotetusta esiintyvyydestä?

- Tällöin pitää kvantifioida termi *tietämys*

Tietämyksen mallintaminen

Yksilön (ihminen tai kone) tietämystä tuntemattoman parametrin arvosta mallinnetaan tulkitsemalla parametri satunnaismuuttujaksi Θ.

$\mathbb{P}(a \leq \Theta \leq b) \ = \ 95\%$ tarkoittaa, että yksilö uskoo parametrin arvon sijaitsevan välillä $[a, b]$ todennäköisyydellä 95%.

- Frekventistit rajoittuvat analysoimaan vain objektiivisia todennäköisyyksiä (toistettavissa olevat tilastokokeet)
  - Protonin massa on välillä $a \pm 10^{-12}$ tn:llä 99.999999%
  - Tupakointi lisää riskiä sairastua keuhkosyöpään
- Todennäköisyyden käyttäminen tietämyksen kvantifiointiin tekee todennäköisyydestä subjektiivisen käsitteen
  - Huomenna sataa tn:llä 10%
  - Aarnio tuomittiin rikoksesta todennäköisin syin

Relativismi: Todennäköisyys on suhteellista

Thomas Bayes 1701–1761

Sisältö

Esimerkki: Tuntematon kolikko

Laatikossa on kolme tasaista kolikkoa (kruunan tn $\theta = 0.5$) sekä yksi lievästi vino ($\theta = 0.6$) ja yksi vahvasti vino ($\theta = 0.9$).

Satunnaisesti valittua kolikkoa heitettäessä havaitaan klaava.

Millä tn heitetty kolikko oli tasainen?

Kolikon tyypin Θ jakauma ennen datan havaitsemista on

θ	0.5	0.6	0.9
$\mathbb{P}(\Theta = \theta)$	0.6	0.2	0.2

Osituskaavasta

$$\mathbb{P}(\text{klaava}) \;=\; 0.6 \times 0.5 \;+\; 0.2 \times 0.4 \;+\; 0.2 \times 0.1 \;=\; 0.4$$

Bayesin kaavasta

$$\mathbb{P}(\Theta = 0.5 \,|\, \text{klaava}) \;=\; \frac{\mathbb{P}(\Theta = 0.5)\mathbb{P}(\text{klaava} \,|\, \Theta = 0.5)}{\mathbb{P}(\text{klaava})} \;=\; \frac{0.6 \times 0.5}{0.4} \;=\; 0.75$$

Kolikon tyypin jakauma datan havaitsemisen jälkeen

θ	0.5	0.6	0.9
$\mathbb{P}(\Theta = \theta \,\|\, \text{klaava})$	0.75	0.20	0.05

Tietämyksen päivityskaava: Diskreetti malli

Tietämys Θ:n arvosta ennen datan x_1 havaitsemista:

- Priorijakauma $p_0(\theta) \;=\; \mathbb{P}(\Theta = \theta)$

Tietämys Θ:n arvosta datan x_1 havaitsemisen jälkeen:

- Posteriorijakauma $p_1(\theta \,|\, x_1) \;=\; \mathbb{P}(\Theta = \theta \,|\, X_1 = x_1)$

Datalähteen stokastinen malli:

- Uskottavuusfunktio $f(x_1 \,|\, \theta) \;=\; \mathbb{P}(X_1 = x_1 \,|\, \Theta = \theta)$

Fakta

Posteriorijakauma $p_1(\theta \,|\, x_1)$ saadaan priorijakaumasta $p_0(\theta)$ painottamalla ja normittamalla sitä uskottavuudella $f(x_1 \,|\, \theta)$:

$$p_1(\theta \,|\, x_1) \;=\; \frac{p_0(\theta)\, f(x_1 \,|\, \theta)}{\sum_{\theta'} p_0(\theta')\, f(x_1 \,|\, \theta')}.$$

Tietämyksen päivityskaava: Todistus

Osituskaavasta

$$
\begin{aligned}
\mathbb{P}(X_1 = x_1) &= \sum_{\theta'} \mathbb{P}(\Theta = \theta')\, \mathbb{P}(X_1 = x_1 \,|\, \Theta = \theta') \\
&= \sum_{\theta'} p_0(\theta')\, f(x_1 \,|\, \theta),
\end{aligned}
$$

joten Bayesin kaavasta

$$
\begin{aligned}
p_1(\theta \,|\, x_1) &= \mathbb{P}(\Theta = \theta \,|\, X_1 = x_1) \\
&= \frac{\mathbb{P}(\Theta = \theta)\mathbb{P}(X_1 = x_1 \,|\, \Theta = \theta)}{\mathbb{P}(X_1 = x_1)} \\
&= \frac{p_0(\theta) f(x_1 \,|\, \theta)}{\mathbb{P}(X_1 = x_1)} \\
&= \frac{p_0(\theta) f(x_1 \,|\, \theta)}{\sum_{\theta'} p_0(\theta')\, f(x_1 \,|\, \theta)}.
\end{aligned}
$$

Esimerkki: Tuntematon kolikko

Tuntematon parametri: Heitetyn kolikon tyyppi Θ

Priorijakauma $p_0(\theta) = \mathbb{P}(\Theta = \theta)$

Data $x_1 = 0$ (havaittiin klaava)

Uskottavuus $f(x_1 \mid \theta) = \mathbb{P}(X_1 = x_1 \mid \Theta = \theta)$

θ	Priori $p_0(\theta)$	Uskottavuus $f(0 \mid \theta)$	Normittamaton posteriori	Posteriori $p_1(\theta \mid 0)$
0.5	0.6	0.5	0.30	0.75
0.6	0.2	0.4	0.08	0.20
0.9	0.2	0.1	0.02	0.05

Esimerkki: Tuntematon kolikko

Kolikon tyypin priorijakauma:

θ	0.5	0.6	0.9
$p_0(\theta)$	0.6	0.2	0.2

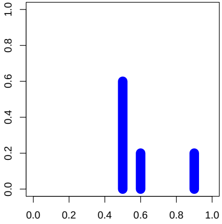
Uskottavuusfunktio datapisteelle $x_1 = 0$ (klaava):

θ	0.5	0.6	0.9
$f(0 \mid \theta)$	0.5	0.4	0.1

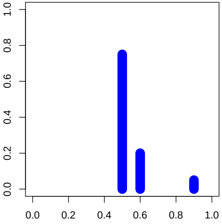
Kolikon tyypin posteriorijakauma:

θ	0.5	0.6	0.9
$p_1(\theta \mid 0)$	0.75	0.20	0.05

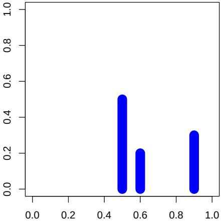
Kolikon tyypin priori- ja posteriorijakaumat



$p_0(\theta)$



$p_1(\theta \mid 0)$



$p_1(\theta \mid 1)$

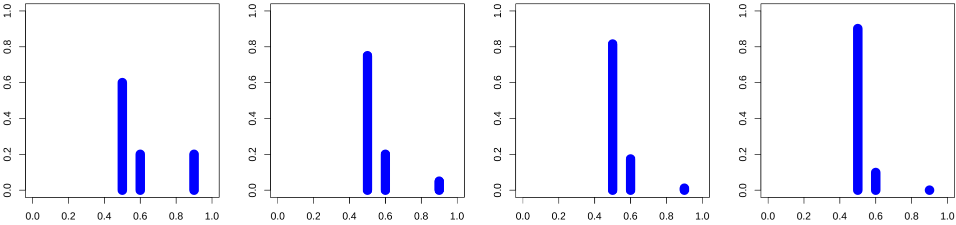
Tuntematon kolikko: Monta havaintoa

Laatikossa on kolme tasaista kolikkoa (kruunan tn $\theta = 0.5$) sekä yksi lievästi vino ($\theta = 0.6$) ja yksi vahvasti vino ($\theta = 0.9$).
Satunnaisesti valittua kolikkoa heitettäessä havaitaan 2 klaavaa.
Millä tn heitetty kolikko oli tasainen?
Uskottavuusfunktio datajoukolle $(x_1, x_2) = (0, 0)$,

θ	0.5	0.6	0.9
$f(0,0 \mid \theta)$	0.25	0.16	0.01

θ	Priori $p_0(\theta)$	Uskottavuus $f(0,0 \mid \theta)$	Normittamaton posteriori	Posteriori $p_1(\theta \mid 0,0)$
0.5	0.6	0.25	0.150	0.815
0.6	0.2	0.16	0.032	0.174
0.9	0.2	0.01	0.002	0.001

Kolikon tyypin priori- ja posteriorijakauma



$p_0(\theta)$ $\quad$ $p_1(\theta \mid 0)$ $\quad$ $p_1(\theta \mid 00)$ $\quad$ $p_1(\theta \mid 00000)$

Sadan klaavan havaitsemisen jälkeen posteriorijakauman massa keskittyy tähtitieteellisen pientä poikkeamaa vaille arvoon 0.5.

Tämä tuntuu paradoksaaliselta, sillä tn saada 100 klaavaa peräkkäin tasaisella kolikolla on 2^{-100}.

Paradoksi selittyy priorin valinnalla: Ylläoleva priori $p_0(\theta)$ kuvastaa absoluuttista 100% varmuutta siitä, että kolikko ei puolla klaavan suuntaan: $\mathbb{P}(\Theta < 0.5) = 0$.

Tietämyksen 2-vaiheinen päivityskaava

Tietämys Θ:n arvosta ennen datan havaitsemista:

- Priorijakauma $p_0(\theta) \;=\; \mathbb{P}(\Theta = \theta)$

Tietämys Θ:n arvosta datan havaitsemisen jälkeen:

- Posteriorijakauma $p_1(\theta \,|\, x_1) \;=\; \mathbb{P}(\Theta = \theta \,|\, X_1 = x_1)$.
- Posteriorijakauma $p_2(\theta \,|\, x_1, x_2) \;=\; \mathbb{P}(\Theta = \theta \,|\, X_1 = x_1, X_2 = x_2)$.

Datalähteen stokastinen malli:

- Uskottavuusfunktio $f_1(x_1 \,|\, \theta) \;=\; \mathbb{P}(X_1 = x_1 \,|\, \Theta = \theta)$
- Uskottavuusfunktio $f_2(x_2 \,|\, \theta, x_1) \;=\; \mathbb{P}(X_2 = x_2 \,|\, \Theta = \theta, X_1 = x_1)$

Fakta

Posteriorijakauma $p_2(\theta \,|\, x_2)$ saadaan posteriorijakaumasta $p_1(\theta \,|\, x_1)$ painottamalla ja normittamalla sitä uskottavuudella $f_2(x_2 \,|\, \theta, x_1)$:

$$p_2(\theta \,|\, x_1, x_2) \;=\; \frac{p_1(\theta \,|\, x_1)\, f_2(x_2 \,|\, \theta, x_1)}{\sum_{\theta'} p_1(\theta' \,|\, x_1)\, f_2(x_2 \,|\, \theta', x_1)}.$$

Tietämyksen 2-vaiheinen päivityskaava: Todistus

Kun $D_1 = \{X_1 = x_1\}$ ja $D_2 = \{X_2 = x_2\}$, saadaan tulokaavasta

$$\begin{aligned}\mathbb{P}(\Theta = \theta, D_1, D_2) &= \mathbb{P}(D_1)\mathbb{P}(\Theta = \theta \,|\, D_1)\mathbb{P}(D_2 \,|\, \Theta = \theta, D_1) \\ &= \mathbb{P}(D_1)\, p_1(\theta \,|\, x_1)\, f_2(x_2 \,|\, \theta, x_1),\end{aligned}$$

osituskaavasta (ehdoton)

$$\begin{aligned}\mathbb{P}(D_1, D_2) &= \sum_{\theta'} \mathbb{P}(\Theta = \theta', D_1, D_2) \\ &= \sum_{\theta'} \mathbb{P}(D_1)\, p_1(\theta' \,|\, x_1)\, f_2(x_2 \,|\, \theta', x_1),\end{aligned}$$

ja nämä yhdistämällä

$$\begin{aligned}p_2(\theta \,|\, x_1, x_2) &= \frac{\mathbb{P}(\Theta = \theta, X_1 = x_1, X_2 = x_2)}{\mathbb{P}(X_1 = x_1, X_2 = x_2)} \\ &= \frac{\mathbb{P}(D_1)\, p_1(\theta \,|\, x_1)\, f_2(x_2 \,|\, \theta, x_1)}{\sum_{\theta'} \mathbb{P}(D_1)\, p_1(\theta' \,|\, x_1)\, f_2(x_2 \,|\, \theta', x_1)} \\ &= \frac{p_1(\theta \,|\, x_1)\, f_2(x_2 \,|\, \theta, x_1)}{\sum_{\theta'} p_1(\theta' \,|\, x_1)\, f_2(x_2 \,|\, \theta', x_1)}.\end{aligned}$$

Tietämyksen päivitys: Yhteenveto

- Priorijakauma $p_0(\theta)$ mallintaa yksilön tietämystä tuntemattoman parametrin arvosta Θ
- Uskottavuusfunktio $f(x \mid \theta)$ vastaa datalähteen stokastista mallia
- Posteriorijakauma mallintaa yksilön tietämystä, johon on yhdistetty priorijakauma sekä havaittu data
- Posteriorijakauma $p_1(\theta \mid x)$ lasketaan priorijakaumasta ja havaitusta datasta x painottamalla prioritodennäköisyyksiä uskottavuusfunktion arvoilla ja sen jälkeen normittamalla
- Yksilön tietämyksen mallintamista subjektiivisilla todennäköisyyksillä kutsutaan bayeslaiseksi lähestymistavaksi

Sisältö

Esim. Kohinainen kanava

- Pisteestä A lähetetään (tuntematon) signaali θ
- Pisteessä B vastaantotetun signaalin arvo on normaalijakautunut odotusarvona θ ja keskihajontana $\sigma = 2$.

Kun sama signaali lähetettiin 3 kertaa peräkkäin, vastaanotettiin arvot $\vec{x} = (3, 8, 7)$.

- Lähetetyn signaalin arvon SU-estimaatti on vastaanotettujen arvojen keskiarvo $m(\vec{x}) = (3 + 8 + 7)/3 = 6$

Pisteessä B arvellaan ennalta, että lähetetyn signaalin Θ arvo on normaalijakautunut odotusarvona $\mu_0 = 5$ ja keskihajontana $\sigma_0 = 1$.

Bayesläinen normaalimalli

Tuntemattoman parametrin priorijakauma: $\Theta \sim \mathrm{Nor}(\mu_0, \sigma_0)$,

$$p_0(\theta) \;=\; (2\pi\sigma_0^2)^{-1/2} e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}}$$

Datan uskottavuusfunktio: $(X_i \,|\, \theta) \sim \mathrm{Nor}(\theta, \sigma)$,

$$\begin{aligned} f(x_i \,|\, \theta) &= (2\pi\sigma^2)^{-1/2} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} \\ f(x_1, \ldots, x_n \,|\, \theta) &= f(x_1 \,|\, \theta) \cdots f(x_n \,|\, \theta) \end{aligned}$$

Esim. Kohinainen kanava:

- Lähetetyn signaalin priori: $\Theta \sim \mathrm{Nor}(\mu_0, \sigma_0)$, $\mu_0 = 5$, $\sigma_0 = 1$
- Vastaanotettu signaali: $(X_i \,|\, \theta) \sim \mathrm{Nor}(\theta, \sigma)$, $\sigma = 2$

Bayesläisen normaalimallin posteriorijakauma

Priorijakauma $\Theta \sim \mathrm{Nor}(\mu_0, \sigma_0)$
Uskottavuus: $(X_i \,|\, \theta) \sim \mathrm{Nor}(\theta, \sigma)$

Fakta

Bayeslaisen normaalimallin posteriorijakauma havaitun datajoukon $\vec{x} = (x_1, \ldots, x_n)$ *suhteen on normaalijakauma* $\mathrm{Nor}(\mu_1, \sigma_1)$*, missä*

$$\mu_1 = \frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2} m(\vec{x})}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \qquad \sigma_1 = \frac{1}{\sqrt{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}},$$

ja $m(\vec{x}) = \frac{1}{n}\sum_{i=1}^n x_i$ *on havaitun datajoukon keskiarvo.*

Esim. Kohinainen kanava

Pisteessä B vastaantotetun signaalin arvo on normaalijakautunut odotusarvona θ ja keskihajontana $\sigma = 2$.
Pisteessä B arvellaan ennalta, että lähetetyn signaalin Θ arvo on normaalijakautunut odotusarvona $\mu_0 = 5$ ja keskihajontana $\sigma_0 = 1$.

Lähetetyn signaalin posteriorijakauma vastaanotettujen arvojen $\vec{x} = (3, 8, 7)$ suhteen on $\mathrm{Nor}(\mu_1, \sigma_1)$, missä

$$\mu_1 = \frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2} m(\vec{x})}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} = \frac{\frac{1}{1^2} \times 5 + \frac{3}{2^2} \times 6}{\frac{1}{1^2} + \frac{3}{2^2}} \approx 5.43$$

ja

$$\sigma_1 = \frac{1}{\sqrt{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}} = \frac{1}{\sqrt{\frac{1}{1^2} + \frac{3}{2^2}}} \approx 0.756$$

Kohinainen kanava: Piste- ja väliestimaatit

Lähetetyn signaalin posteriorijakauma datan $\vec{x} = (3, 8, 7)$ suhteen on $\mathrm{Nor}(\mu_1, \sigma_1)$, missä $\mu_1 = 5.43$ ja $\sigma_1 = 0.756$.
Lähetetyn signaalinarvon bayesläisiä piste-estimaatteja:

- Posteriorijakauman odotusarvo: $\mu_1 = 5.43$
- Suurimman posterioritodennäköisyyden estimaatti: $\mu_1 = 5.43$

Määritä väli, joka sisältää lähetetyn signaalin todellisen arvon 90% todennäköisyydellä. Ratkaistaan c yhtälöstä

$$0.90 \;=\; \mathbb{P}(\Theta = \mu_1 \pm c) \;=\; \mathbb{P}\left(\frac{\Theta - \mu_1}{\sigma_1} = 0 \pm c/\sigma_1\right) \;=\; \mathbb{P}(|Z| \leq c/\sigma_1)$$

Taulukoista: $\mathbb{P}(|Z| \leq 1.64) = 0.90$, joten $c = 1.64 \times 0.756 = 1.24$.
Väli $5.43 \pm 1.24 = [4.19, 6.67]$ siis peittää lähetetyn signaalin arvon 90% todennäköisyydellä. (Piirrä posteriorijakauman kuva ja vertaa priorijakaumaan)

Sisältö

Tuntematon kolikko

Tuntematonta kolikkoa heitettäessä (0=klaava, 1=kruuna) on havaittu data $x = (0,0,0,0,0,0,1,0,1,0)$. Kolikosta ei ole mitään taustatietoja. Määritä parametrin Θ (kruunan tn) posteriorijakauma.

Valitaan prioriksi jatkuvan välin $[0,1]$ tasajakauma tiheysfunktiona

$$p_0(\theta) = \begin{cases} 1, & \theta \in [0,1], \\ 0, & \text{muuten.} \end{cases}$$

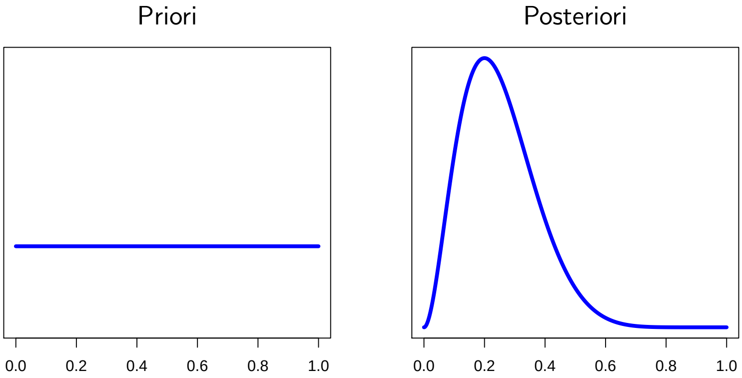
Uskottavuusfunktio $f(x \mid \theta) = \theta^2 (1-\theta)^8$
Posteriorijakauman tiheysfunktio

$$p_1(\theta \mid x) = c\, p_0(\theta) f(x \mid \theta) = \begin{cases} c\,\theta^2 (1-\theta)^8, & \theta \in [0,1], \\ 0, & \text{muuten,} \end{cases}$$

missä normitusvakio $c = \left(\int_0^1 t^2 (1-t)^8 \, dt\right)^{-1}$

Tuntematon kolikko

Data: $\vec{x} = (0, 0, 0, 0, 0, 0, 1, 0, 1, 0)$



$p_0(\theta)\, d\theta = 1\, d\theta$

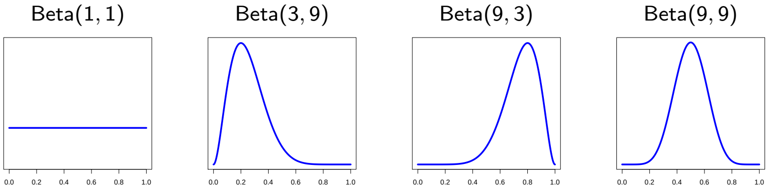
$p_1(\theta|\vec{x}) d\theta = c\, \theta^2 (1-\theta)^8\, d\theta$

Beta-jakauma

Beta(a, b)-jakauman parametreina $a > 0$ ja $b > 0$ tiheysfunktio on

$$f(\theta) \;=\; \begin{cases} c\,\theta^{a-1}(1-\theta)^{b-1}, & \text{kun } \theta \in [0,1], \\ 0, & \text{muuten}, \end{cases}$$

normitusvakiona $c \;=\; \frac{(a+b-1)!}{(a-1)!(b-1)!}$.



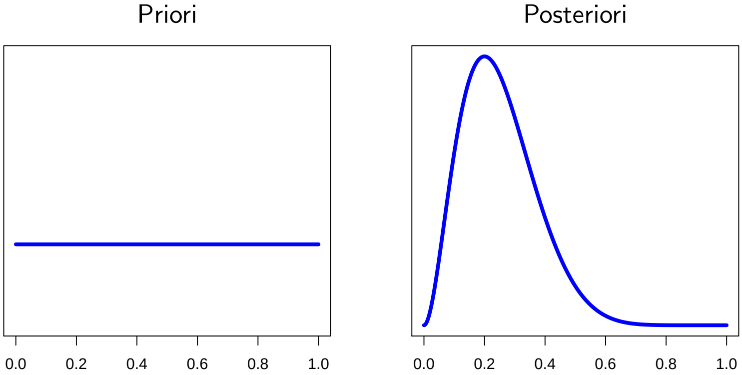
- Arvojoukko $= [0, 1]$
- Odotusarvo $\mu = \frac{a}{a+b}$ ja keskihajonta $\sigma = \sqrt{\frac{\mu(1-\mu)}{a+b+1}}$

```
dbeta(theta,a,b); pbeta(theta,a,b)
```

Tuntematon kolikko

Data: $\vec{x} = (0, 0, 0, 0, 0, 0, 1, 0, 1, 0)$

Priori: Tasajakauma Beta$(1, 1)$
Posteriori: Beta$(3, 9)$



$p_0(\theta)\, d\theta = 1\, d\theta$

$p_1(\theta|x) d\theta = c\, \theta^2 (1-\theta)^8\, d\theta$

Tuntematon kolikko: Kruunien lukumäärä

Kolikkoa n kertaa heitettäessä havaittiin x kruunaa. Kolikosta ei ole taustatietoja. Määritä parametrin Θ (kruunan tn) posteriorijakauma.

Priorijakauman tiheysfunktio: $p_0(\theta) = 1$, $\theta \in [0, 1]$
Uskottavuusfunktio datapisteelle x saadaan $\mathrm{Bin}(n, \theta)$-jakaumasta

$$f(x \mid \theta) \;=\; \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Posterioritiheys

$$p_1(\theta \mid x) \;=\; \frac{p_0(\theta) f(x \mid \theta)}{\int p_0(t) f(x \mid \theta') \, d\theta'} \;=\; c\, \theta^x (1-\theta)^y$$

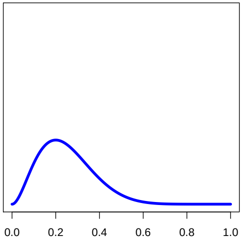
on $\mathrm{Beta}(x+1, y+1)$, missä $y = n - x$ on klaavojen lkm.

Huom

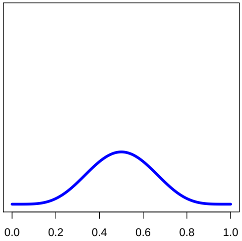
- Kun $n = 10$ ja $k = 2$, saadaan sama posteriori $\mathrm{Beta}(3, 9)$, mitä yksityiskohtaiselle datalle $\vec{x} = (0, 0, 0, 0, 0, 0, 1, 0, 1, 0)$.
- Normitusvakion c arvo määräytyy ehdosta $\int_0^1 p_1(\theta \mid x) d\theta = 1$. Beta-jakauman taulukoista $\implies c = \frac{(x+y+1)!}{x!y!}$

Tuntematon kolikko: Kruunien lukumäärä

$n = 10$

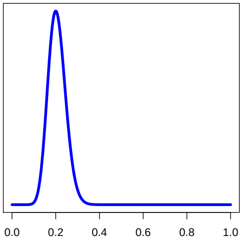


Beta$(3, 9)$: $x = 2$, $y = 8$

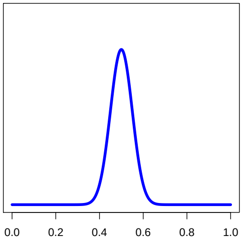


Beta$(6, 6)$: $x = 5$, $y = 5$

$n = 100$



Beta$(21, 81)$: $x = 20$, $y = 80$



Beta$(51, 51)$: $x = 50$, $y = 50$

Loppuviikolla vertaillaan bayesläisiä väliestimaatteja frekventistisiin luottamusväleihin.