

MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

5A Bayesläinen tilastollinen päättely

Jukka Kohonen

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2019–2020
Periodi IV

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Esimerkki: Kolikko

Tasaiseksi väitettyä kolikkoa heitettiin 5 kertaa. Saatiin tulossarja (0 = klaava, 1 = kruuna)

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

Aineisto on pieni. Keskeinen raja-arvolause ei sovellu.

Vaikka aineisto on pieni, vaikuttaa $\hat{p} = 0$ kertovan kolikosta ainakin jotain. Mitä aineiston perusteella pitäisi päätellä ja tehdä?

- Hankitaan lisää dataa? (entä jos ei saatavilla?)
- Jätetään kokeen tulos huomiotta (ja uskotaan edelleen, että kolikko on tasainen?)

Onhan totta, että tasaisesta kolikosta *voidaan* saada tämä tulossarja (todennäköisyydellä ...)

- Yritetään sovittaa kokeen tulos aiempaan tietämykseen kruunan odotetusta esiintyvyydestä?

Tällöin pitää kvantifioida termi *tietämys* ja selvittää säännöt, *miten* näitä sovitetaan toisiinsa.

Tietämyksen mallintaminen

- Kiinnostuksen kohteena on jokin tuntematon **suure**, esimerkiksi datalähteen jakauman parametri θ .
- Jos suureen arvosta on *jotakin* käsitystä, tämä kuvataan **todennäköisyysjakaumalla**.
- Suuretta kuvataan tällöin *satunnaismuuttujalla* Θ .
- “Satunnaisuus” on tässä ymmärrettävä laajemmin *epävarmuutena*.
- Esim. $\mathbb{P}(a \leq \Theta \leq b) = 95\%$ tarkoittaa, että käytettävissä olevan informaation nojalla suureen arvon katsotaan olevan välillä $[a, b]$ todennäköisyydellä 95%.

Informaation lähteenä esim.

- symmetriaoletukset tai -havainnot (kolikon rakenne)
- prosessin tuntemus (kolikonheiton mekaniikka)
- havainnot ja kokemus (heittotilastot)

Frekventistinen ja bayesläinen tilastotiede

Frekventistinen tilastotiede liittyy todennäköisyydet ilmiöihin, jotka syntyvät jonkinlaisesta satunnaisesta *prosessista* (yleensä toistettavasti).

Esim. kolikonheittoprosessin *tulokset* X_1, X_2, X_3, X_4, X_5 ja niistä lasketut *tunnusluvut* kuten $(X_1 + X_2 + X_3 + X_4 + X_5)/5$.

Bayesläinen tilastotiede laajentaa todennäköisyyden käsitteen myös muihin suureisiin, joiden *arvo on tuntematon*, vaikka ne eivät olisi perinteisesti ajatellen “satunnaisia” tai toistettavia.

Esim. kolikonheittoprosessia kuvaava *parametri* p .

(Nimi “bayesläinen” tulee siitä, että tässä käytetään paljon Bayesin kaavaa.)

Frekventistinen ja Bayesläinen tulkinta — Otantatutkimus

Puoluetta A kannattaa osuus p populaatiosta.

- **(F)** p on tuntematon, mutta silti eräs kiinteä luku eikä siihen liity satunnaisuutta. Emme liitä p :n arvoihin todennäköisyyttä.
- **(B)** p on tuntematon suure, käsittelemme sitä satunnaismuuttujana ja liitämme sen arvoihin todennäköisyyksiä.

Valitaan satunnaiset 1000 ihmistä. Heistä k kpl kannattaa A:ta.

- **(F&B)** k on satunnaismuuttuja, jonka jakauma riippuu p :stä.
(Bayes-tulkinnassa *ehdollinen jakauma*)

Molemmissa käytetään matemaattisesti täysin samaa todennäköisyysslaskentaa. Ero on siinä, mihin muuttujiin sitä sovelletaan.

Frekventistinen ja Bayesläinen tulkinta — Kolikonheitto

Eräs kolikko tuottaa kruunia todennäköisyydellä p .

- **(F)** p on tuntematon, mutta silti eräs kiinteä luku eikä siihen liity satunnaisuutta. Emme liitä p :n arvoihin todennäköisyyttä.
- **(B)** p on tuntematon suure, käsittelemme sitä satunnaismuuttujana ja liitämme sen arvoihin todennäköisyyksiä.

Heitetään kolikkoa 5 kertaa ja saadaan k kruunaa.

- **(F&B)** k on satunnaismuuttuja, jonka jakauma riippuu p :stä.
(Bayes-tulkinnassa *ehdollinen jakauma*)

Molemmissa käytetään matemaattisesti täysin samaa todennäköisyyyslaskentaa. Ero on siinä, mihin muuttujiin sitä sovelletaan.

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Esimerkki: Tuntematon kolikko

Laatikossa on 10 erilaista kolikkoa: kuusi reilua (kruunan tn $\theta = 0.5$), kaksi vinoa ($\theta = 0.25$ ja 0.75) ja kaksi täysin huijauskolikkoa ($\theta = 0$ ja 1). Laatikosta valitaan satunnainen kolikko. Sen tyyppin Θ jakauma on:

θ	0	0.25	0.5	0.75	1
$\mathbb{P}(\Theta = \theta)$	0.1	0.1	0.6	0.1	0.1

Esim. alaspäin vino kolikko ($\theta = 0.25$) tulee valituksi tn:llä 0.1.

Valittua kolikkoa heitettäessä havaitaan klaava.

Mikä on tn, että kyseinen kolikko on alaspäin vino?

Osituskaavasta

$$\mathbb{P}(\text{klaava}) = (0.1 \cdot 1) + (0.1 \cdot 0.75) + (0.6 \cdot 0.5) + (0.1 \cdot 0.25) + (0 \cdot 1) = 0.5$$

Bayesin kaavasta

$$\mathbb{P}(\Theta = 0.25 \mid \text{klaava}) = \frac{\mathbb{P}(\Theta = 0.25) \cdot \mathbb{P}(\text{klaava} \mid \Theta = 0.25)}{\mathbb{P}(\text{klaava})} = \frac{0.1 \cdot 0.75}{0.5} = 0.15$$

Tuntematon kolikko jatkuu

Heitimme kolikkoa kerran ja saimme klaavan.

Voimme nyt laskea kullekin kolikkotyypille todennäköisyyden, että kädessämme oleva kolikko on kyseistä tyyppiä. (Samat laskut kuin edellä, jokaiselle θ :n arvolle erikseen.)

θ	0	0.25	0.5	0.75	1
$\mathbb{P}(\Theta = \theta \mid \text{klaava})$	0.20	0.15	0.60	0.05	0.00

Tämä on kolikkotyypin *posteriorijakauma* yhden havainnon jälkeen.

Miten yhden klaavan havaitseminen vaikutti todennäköisyyksiin?

- Klaavoja helposti tuottavien tyyppien tn *kasvoi*.
- Klaavoja harvoin tuottavien tyyppien tn *pieneni*.
- Tyypin $\Theta = 1$ tn pieneni peräti nolnaan. Tämä on järkevää: havaitsimme klaavan, kolikko ei siis tuota pelkkiä kruunia.

Tyypilliset lyhennysmerkinnät

Pitkä merkintä	Lyhennys	Selitys
$f_X(x)$	$f(x)$	X :n tiheys
$f_\Theta(\theta)$	$f(\theta)$	Θ :n tiheys (“priori”)
$f_{X \Theta}(x \theta)$	$f(x \theta)$	X :n tiheys, jos $\Theta = \theta$ (“uskottavuus”)
$f_{\Theta X}(\theta x)$	$f(\theta x)$	Θ :n tiheys, jos $X = x$ (“posteriori”)

Pitkässä merkinnässä alaindeksi selventää, mistä satunnaismuuttujista puhutaan.

Lyhennysmerkinnässä alaindeksi jätetään pois, joten f voi viitata useisiin eri funktioihin. Mihin funktioon, pitää yleensä ymmärtää argumentista sulkujen sisällä.

Tarvittaessa selvennetään alaindeksillä. (Tarkoittaako $f(5|3)$ funktiota $f_{X|\Theta}$ vai $f_{\Theta|X}$? Nämä ovat aivan eri funktiot.)

(Luentomonisteessa käytetään f :ää datan tiheysfunktioille ja p :tä parametrin tiheysfunktioille.)

Tietämyksen päivityskaava: Diskreetti malli

Tietämys Θ :n arvosta ennen datan havaitsemista:

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

Tietämys Θ :n arvosta datan $X = x$ havaitsemisen jälkeen:

- Posteriorijakauma $f(\theta | x) = \mathbb{P}(\Theta = \theta | X = x)$

Datalähteen stokastinen malli:

- Uskottavuusfunktio $f(x | \theta) = \mathbb{P}(X = x | \Theta = \theta)$

Fakta

Posteriorijakauma saadaan laskemalla pisteittäin (= kullekin mahdolliselle arvolle θ) *priorin* ja uskottavuuden tulo, ja lopuksi *normittamalla* niin että summa on 1:

$$f(\theta | x) = \frac{f(\theta) f(x | \theta)}{\sum_{\theta'} f(\theta') f(x | \theta')}.$$

Tietämyksen päivityskaava: Todistus

Osituskaavasta

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

joten Bayesin kaavasta

$$\begin{aligned}f(\theta | x) &= \mathbb{P}(\Theta = \theta | X = x) \\ &= \frac{\mathbb{P}(\Theta = \theta) \mathbb{P}(X = x | \Theta = \theta)}{\mathbb{P}(X = x)} \\ &= \frac{f(\theta) f(x | \theta)}{\mathbb{P}(X = x)} \\ &= \frac{f(\theta) f(x | \theta)}{\sum_{\theta'} f(\theta') f(x | \theta')}.\end{aligned}$$

Esimerkki: Tuntematon kolikko

Tuntematon parametri: Heitetyn kolikon tyyppi Θ

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

Data $x = 0$ (havaittiin klaava)

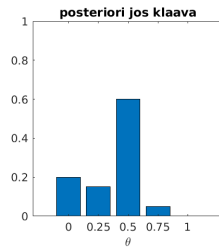
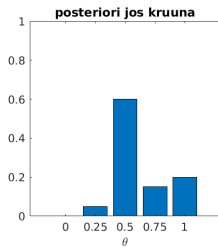
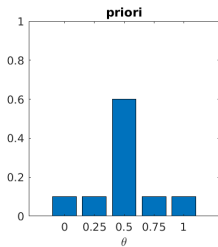
Uskottavuus $f(x | \theta) = \mathbb{P}(X = x | \Theta = \theta)$

θ	Priori $f(\theta)$	Uskottavuus $f(0 \theta)$	Tulo	Posteriori $f(\theta 0)$
0	0.1	1.00	0.100	0.20
0.25	0.1	0.75	0.075	0.15
0.5	0.6	0.50	0.300	0.60
0.75	0.1	0.25	0.025	0.05
1	0.1	0.00	0.000	0.00
summa	1.0		0.500	1.00

“Tulo” = priorin ja uskottavuuden tulo eli ns. *normittamaton* posteriorijakauma. Kun se vielä normitetaan (niin että summautuu ykköseksi), saadaan posteriorijakauma.

Huom. Uskottavuusfunktion arvot eivät keskenään muodosta mitään todennäköisyysjakaumaa.

Kolikon tyypin priori- ja posteriorijakaumat



Tuntematon kolikko: Monta havaintoa

Laatikossa on 10 kolikkoa kuten edellä. Valitaan yksi kolikko satunnaisesti, heitetään sitä kaksi kertaa. Havaitaan 2 klaavaa.

Millä tn heitetty kolikko oli tasainen?

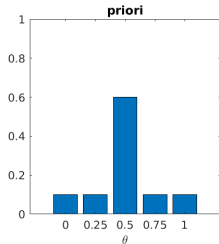
Uskottavuusfunktio datajoukolle $(x_1, x_2) = (0, 0)$ on:

θ	0	0.25	0.5	0.75	1
$f(0, 0 \theta)$	1.0000	0.5625	0.2500	0.0625	0.0000

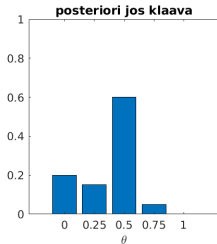
θ	Priori $f(\theta)$	Uskottavuus $f(0, 0 \theta)$	Tulo	Posteriori $f(\theta 0, 0)$
0	0.1	1.0000	0.1000	0.32
0.25	0.1	0.5625	0.0563	0.18
0.5	0.6	0.2500	0.1500	0.48
0.75	0.1	0.0625	0.0063	0.02
1	0.1	0.0000	0.0000	0.00
summa	1.0		0.3125	1.00

“Tulo” = priorin ja uskottavuuden tulo eli ns. *normittamaton* posteriorijakauma. Kun se vielä normitetaan (niin että summautuu ykköseksi), saadaan posteriorijakauma.

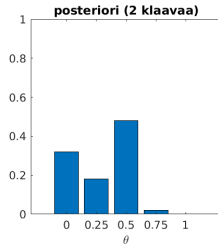
Kolikon priori ja erilaisia posterioreja



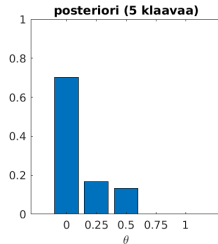
$$f(\theta)$$



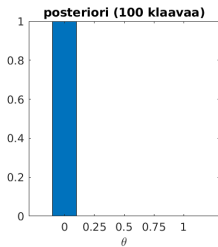
$$f(\theta | 0)$$



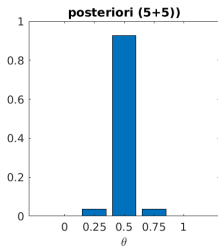
$$f(\theta | 00)$$



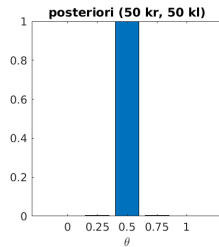
$$f(\theta | 00000)$$



$$f(\theta | 0000\dots)$$



$$f(\theta | 0000011111)$$



$$f(\theta | 00\dots11\dots)$$

Vaihtoehto: Tietämyksen päivittäminen 2 vaiheessa

Tietämys Θ :n arvosta ennen datan havaitsemista:

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

Tietämys Θ :n arvosta datan havaitsemisen jälkeen:

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

Datalähteen stokastinen malli:

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

Fakta

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

$$f(\theta | x_1, x_2) = \frac{f(\theta | x_1) f(x_2 | \theta, x_1)}{\sum_{\theta'} f(\theta' | x_1) f(x_2 | \theta', x_1)}.$$

Tietämyksen 2-vaiheinen päivityskaava: Todistus

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

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

osituskaavasta (ehdoton)

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

ja nämä yhdistämällä

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

Tietämyksen päivitys: Yhteenveto

- Priorijakauma $f(\theta)$ mallintaa tietämystä tuntemattoman parametrin arvosta Θ ennen havaintoja
- Uskottavuusfunktio $f(x | \theta)$ vastaa datalähteen stokastista mallia
- Posteriorijakauma mallintaa tietämystä, johon on yhdistetty priorijakauma sekä havaittu data
- Posteriorijakauma $f(\theta | x)$ lasketaan kertomalla pisteittäin (= kullekin θ erikseen) priori ja uskottavuus, ja lopuksi normittamalla aidoksi tn-jakaumaksi (niin että summa=1)
- Tätä kutsutaan Bayes-päätteeksi (Bayes-inferenssiksi).

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Tuntematon kolikko

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

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

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

Uskottavuusfunktio $f(\vec{x} | \theta) = \theta^2(1 - \theta)^8$

Posteriorijakauman tiheysfunktio

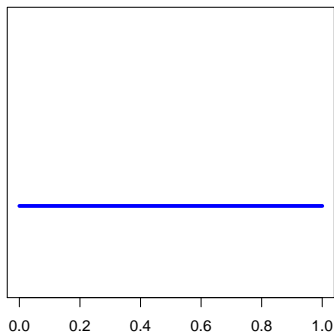
$$f(\theta | \vec{x}) = c f(\theta) f(\vec{x} | \theta) = \begin{cases} c \theta^2(1 - \theta)^8, & \theta \in [0, 1], \\ 0, & \text{muuten,} \end{cases}$$

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

Tuntematon kolikko

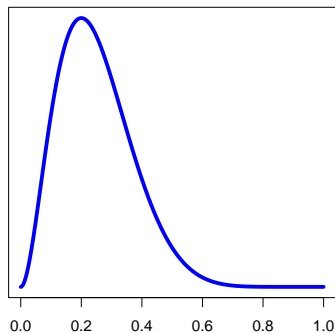
Data: $\vec{x} = (0, 0, 0, 0, 0, 0, 1, 0, 1, 0)$

Priori



$$f(\theta) d\theta = 1 d\theta$$

Posteriori



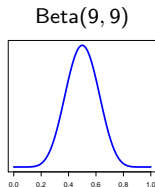
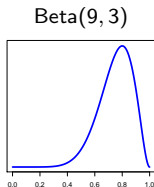
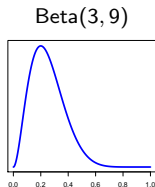
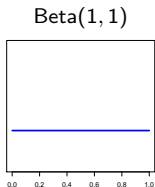
$$f(\theta|\vec{x})d\theta = c \theta^2(1 - \theta)^8 d\theta$$

Beetajakauma

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)`

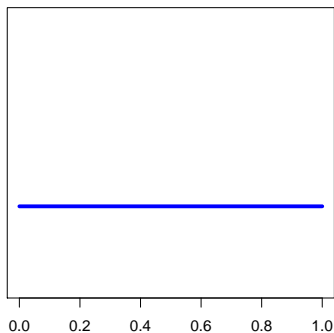
Tuntematon kolikko

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

Priori: Tasajakauma Beta(1, 1)

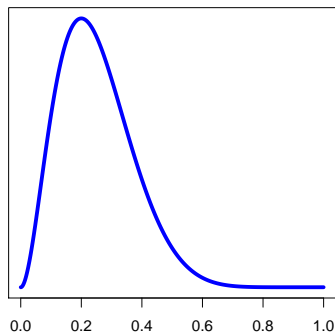
Posteriori: Beta(3, 9)

Priori



$$f(\theta) d\theta = 1 d\theta$$

Posteriori



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

Vaihtoehto: Datana lukumäärä eikä yksittäiset heitot

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

Priorijakauman tiheysfunktio: $f(\theta) = 1, \theta \in [0, 1]$

Uskottavuusfunktio datapisteelle x saadaan $\text{Bin}(n, \theta)$ -jakaumasta

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

Posterioritiheys

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\int f(t)f(x|\theta') d\theta'} = c \theta^x (1-\theta)^y$$

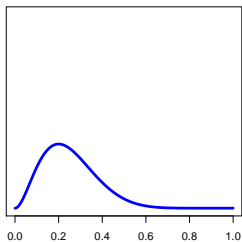
on $\text{Beta}(x+1, y+1)$, missä $y = n - x$ on klaavojen lkm.

Huom

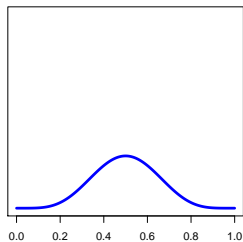
- Kun $n = 10$ ja $k = 2$, saadaan sama posteriori $\text{Beta}(3, 9)$, mitä yksityiskohtaiselle datalle $\vec{x} = (0, 0, 0, 0, 0, 0, 1, 0, 1, 0)$.
- Normitusvakion c arvo määräytyy ehdosta $\int_0^1 f(\theta|x) d\theta = 1$.
Beta-jakauman taulukoista $\implies c = \frac{(x+y+1)!}{x!y!}$

Tuntematon kolikko: Kruunien lukumäärä

$n = 10$

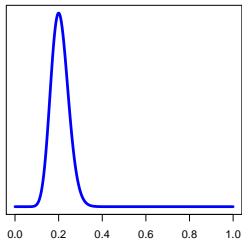


Beta(3, 9): $x = 2$, $y = 8$

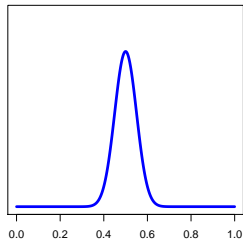


Beta(6, 6): $x = 5$, $y = 5$

$n = 100$



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



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

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Posteriorin laskenta käytännössä

Jokaiselle θ -parametrin mahdolliselle arvolle,

$$\text{posteriori} = \frac{\text{piori} \cdot \text{uskottavuus}}{\text{normalisointivakio}}$$

eli

$$f(\theta | x) = \frac{f(\theta) \cdot f(x | \theta)}{\dots}$$

missä \dots on vakio (ei θ :sta riippuva) jolla saadaan aikaan $\sum_{\theta} f(\theta | x) = 1$.

Usein helpointa: Lasketaan ensin *normalisoimaton posteriori*

$$f(\theta) \cdot f(x | \theta).$$

- Se ei ole tn-jakauma, mutta \dots
- siitä jo näkee *suhteelliset* erot,
- maksimikohdan voi etsiä yhtä hyvin ilman normalisointia,
- ja sen voi sitten normalisoida.

Posteriorin normalisointi

Normalisoimaton posteriori

$$f(\theta) \cdot f(x | \theta)$$

on *melkein* tn-jakauma Θ :lle, paitsi että sen summa

$$c = \sum_{\theta} \left(f(\theta) \cdot f(x | \theta) \right)$$

voi olla $\neq 1$. Jaetaan kaikki arvot c :llä \Rightarrow saadaan aito tn-jakauma, nimittäin Θ :n posteriorijakauma.

Jos Θ on jatkuva, summan tilalla on integraali.

Kolikko laatikosta (diskreetti parametri)

Parametri Θ kuvaa millainen kolikko on poimittu

Priori $f(\theta) = \mathbb{P}(\Theta = \theta)$

Data $x = 0$ (yksi klaava)

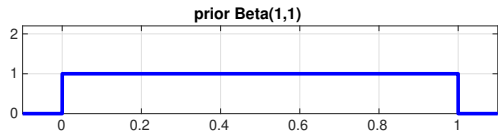
Uskottavuus $f(0 | \theta) = \mathbb{P}(X = 0 | \Theta = \theta) = 1 - \theta$

θ	Priori $f(\theta)$	Uskottavuus $f(0 \theta)$	Tulo	Posteriori $f(\theta 0)$
0	0.1	1.00	0.100	0.20
0.25	0.1	0.75	0.075	0.15
0.5	0.6	0.50	0.300	0.60
0.75	0.1	0.25	0.025	0.05
1	0.1	0.00	0.000	0.00
Σ	1.0		0.500	1.00

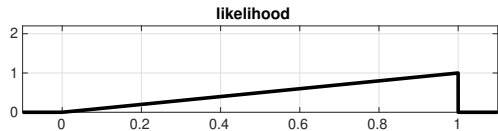
Jo normalisoimattomasta posteriorista (“tulo”) nähdään mitkä parametrin arvot ovat toisia todennäköisempiä. other).

Aito posteriori saadaan jakamalla kukin tulon arvo niiden summalla (tässä 0.5).

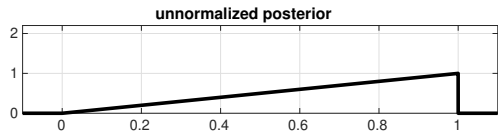
Kolikko, jatkuva parametri — tasapriori, eka tulos 1



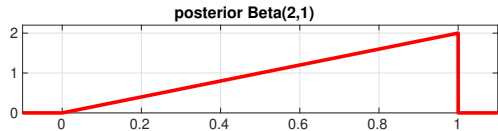
1



θ

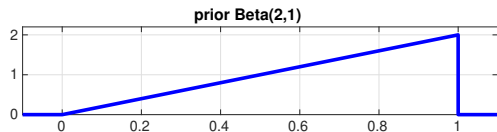


θ

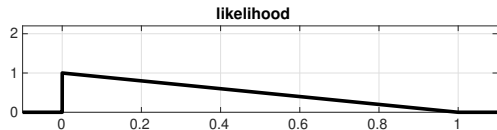


2θ

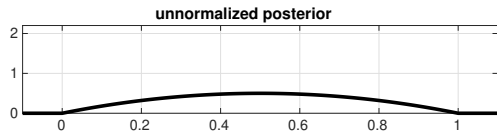
Seuraava päivitys, kun toinen tulos 0



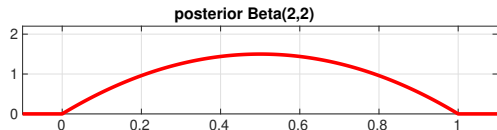
$$2\theta$$



$$(1 - \theta)$$

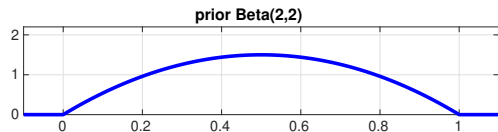


$$2\theta(1 - \theta)$$

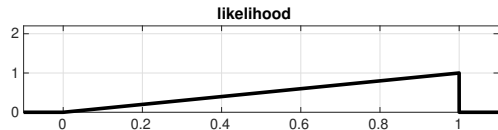


$$6\theta(1 - \theta)$$

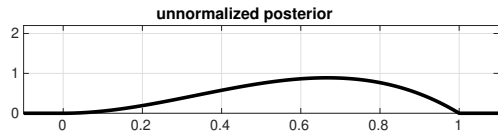
Seuraava päivitys, kun kolmas tulos 1



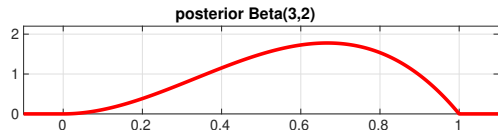
$$6\theta(1 - \theta)$$



$$\theta$$



$$6\theta^2(1 - \theta)$$



$$12\theta^2(1 - \theta)$$

Entäs priorii?

Tasapriori $\text{Beta}(1,1)$ sanoi, että (ennen havaintoja) pidimme kaikkia parametriarvoja välillä $[0, 1]$ yhtä todennäköisinä. Tämä on eräessä mielessä “tuntematon kolikko”.

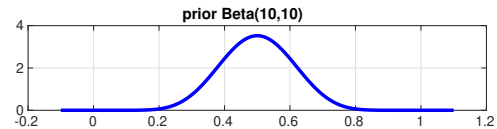
Mutta reaali maailmassa vastaantulevista kolikoista ehkä pidämme jo *a priori* todennäköisenä, että ne ovat (suunnilleen) reiluja.

Entä jos otamme prioriksi esim. $\text{Beta}(10,10)$, joka on symmetrinen, mutta keskittyy enemmän keskikohtaan?

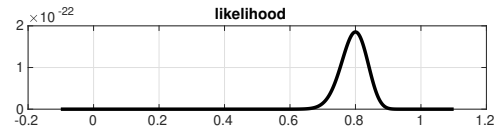
(Matemaattisesti tähän päädytään, jos kolikon priorii oli alun perin tasainen, mutta olemme sitten nähneet 9 kruunaa ja 9 klaavaa.)

Kokeillaan. Aloitetaan symmetrisellä “luultavasti lähellä reilua” uskomuksella, ja sitten havaitaan 80 kruunaa ja 20 klaavaa.

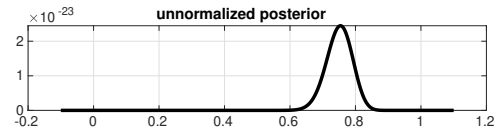
Prior "lähellä reilua", epäsymmetrinen (80+20)



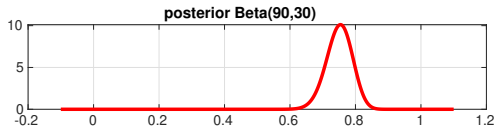
$$\propto \theta^9(1 - \theta)^9$$



$$\propto \theta^{80}(1 - \theta)^{20}$$



$$\propto \theta^{89}(1 - \theta)^{29}$$



$$\propto \theta^{89}(1 - \theta)^{29}$$

Sisältö

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Esim. Kohinainen kanava

- Pisteestä A lähetetään (tuntematon) signaali θ
- Pisteessä B vastaanotetun 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: $\Theta \sim \text{Nor}(\mu_0, \sigma_0)$,

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

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

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

Esim. Kohinainen kanava:

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

Bayesläisen normaalimallin posteriorijakauma

Priorijakauma $\Theta \sim \text{Nor}(\mu_0, \sigma_0)$

Uskottavuus: $(X_i | \theta) \sim \text{Nor}(\theta, \sigma)$

Fakta

Bayesläisen normaalimallin posteriorijakauma havaitun datajoukon $\vec{x} = (x_1, \dots, x_n)$ suhteen on normaalijakauma $\text{Nor}(\mu_1, \sigma_1)$, missä

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

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

Esim. Kohinainen kanava

Pisteessä B vastaanotetun 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 $\text{Nor}(\mu_1, \sigma_1)$, 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 $\text{Nor}(\mu_1, \sigma_1)$, missä $\mu_1 = 5.43$ ja $\sigma_1 = 0.756$.

Lähetetyn signaalin arvon 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

Tietämyksen kvantifiointi ja subjektiivinen todennäköisyys

Tietämyksen päivittäminen – Diskreetti data, diskreetti parametri

Diskreetti data, jatkuva parametri

Posteriorin laskenta käytännössä

Jatkuva data, jatkuva parametri

Bayes-inferenssin etuja

Joitakin Bayes-inferenssin etuja

Kun parametria Θ käsitellään satunnaismuuttujana, niin saadaan:

- **Matemaattinen yhtenäisyys.** “Parameterit” and “havainnot” ovat kaikki “suureita”, joihin sovelletaan samoja tn-laskennan laskusääntöjä.
- **Paljon sovelluksia.** Tällä menetelmällä voidaan laskea posterioreja
 - vähästä datasta, jopa $n = 1$, jolloin esim. normaaliapproksimaatiota ei voi ollenkaan käyttää
 - paljosta datasta
 - ei-normaalista datasta (esim. eksponenttijakautuneet havainnot)
 - kun parametrissa on käytettävissä hyödyllistä ennakkotietoa
 - monimutkaisilla malleilla, joilla voi kuvata monimutkaisia reaali maailman ilmiöitä
- **Aito tn-jakauma.** Inferenssin tuloksena on Θ :lle tn-jakauma. Se kertoo Θ :sta paljon enemmän kuin yksittäinen piste- tai väliestimaatti. Jakaumaa voi tarkastella visuaalisesti ja siitä voi laskea mitä tahansa tunnuslukuja (esim. odotusarvo, moodi, minkä tahansa välin todennäköisyys).

Seuraavalla luennolla tutkitaan tarkemmin, mitä posteriorijakaumalla voi tehdä, mm. lasketaan väliestimaatteja.