Luku 8

Bayes-päättelyn alkeita

Bayesiläisessä päättelyssä havaintosatunnaisvektorin jakauma mallinnetaan täysin samalla tavalla kuin frekventistisessä lähestymistavassa silloin, kun parametrin arvo on kiinnitetty. Frekventistisessä lähestymistavassa parametri on kiinteä (ts. ei-satunnainen) mutta tuntematon. Bayesiläisessä lähestymistavassa parametria käsitellään satunnaisena.

Tämä ehkä vähäiseltä tuntuva ero johtaa suuriin eroihin laskutekniikoissa ja tulosten tulkinnoissa. *Bayesiläisessä päättelyssä kaikki on toisin* kuin frekventistisessä.

8.1 Todennäköisyyslaskentaa

Tämän jakson kaavoissa oletetaan hiljaisesti, että osamäärien nimittäjät ovat erisuuria kuin nolla.

Olkoot A ja B tapahtumia. Jos tiedämme (täsmälleen sen), että B on sattunut, niin tapahtuman A todennäköisyys lasketaan *ehdollisen todennäköisyyden* kaavalla

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{8.1}$$

Ehdollisen todennäköisyyden kaavasta saadaan todennäköisyyksien *kertolaskukaava*

$$P(A \cap B) = P(B)\, P(A \mid B) = P(A)\, P(B \mid A). \tag{8.2}$$

Tästä nähdään kaava

$$P(A \mid B) = \frac{P(A)\, P(B \mid A)}{P(B)}. \tag{8.3}$$

Jos tapahtumat $A_1, \ldots, A_M$ ovat jokin perusjoukon ositus ts.

- joukot A_i ovat erillisiä: jos $i \neq j$, niin $A_i \cap A_j = \emptyset$.
- niiden yhdiste on koko perusjoukko,

niin $P(B)$ voidaan laskea (todennäköisyyden additiivisuuden perusteella) seuraavasti,

$$\begin{aligned} P(B) &= P(\cup_{i=1}^{M}(B \cap A_i)) = \sum_{i=1}^{M} P(B \cap A_i) \\ &= \sum_{i=1}^{M} P(A_i)\, P(B \mid A_i) \end{aligned}$$

Kun tätä ns. *kokonaistodennäköisyyden* kaavaa käytetään kaavassa (8.3) saadaan *Bayesin kaava*

$$P(A_k \mid B) = \frac{P(A_k)\, P(B \mid A_k)}{\sum_{i=1}^{M} P(A_i)\, P(B \mid A_i)} \tag{8.4}$$

johon bayesiläinen päättely perustuu diskreetin parametrin ja diskreetin havaintovektorin tapauksessa.

Kirjoitetaan edelliset kaavat vielä siinä tapauksessa, jossa käsitellään kahden diskreetin satunnaismuuttujan (tai satunnaisvektorin) $\tilde{\theta}$ ja $\mathbf{Y}$ yhteisjakaumaa, jonka määrää niiden yhteispistetodennäköisyysfunktio

$$f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) = P(\tilde{\theta} = \theta, \mathbf{Y} = \mathbf{y}) = P(\{\tilde{\theta} = \theta\} \cap \{\mathbf{Y} = \mathbf{y}\}). \tag{8.5}$$

Satunnaismuuttujan (tai satunnaisvektorin) $\tilde{\theta}$ mahdolliset arvot ovat $\theta_1, \theta_2, \ldots$ ja satunnaismuuttujan (tai satunnaisvektorin) $\mathbf{Y}$ mahdolliset arvot ovat $\mathbf{y}_1, \mathbf{y}_2, \ldots$. Yhteispistetodennäköisyysfunktion arvoja ja muiden pistetodennäköisyysfunktioiden arvoja lasketaan jatkossa sellaisissa pisteissä $(\theta, \mathbf{y})$, joiden θ-koordinaatti on jokin $\tilde{\theta}$ mahdollisista arvoista ja $\mathbf{y}$-koordinaatti on jokin $\mathbf{Y}$:n mahdollisista arvoista.

Käytetään seuraavia merkintöjä.

- $p(\theta)$ on satunnaismuuttujan $\tilde{\theta}$ reunajakauman ptnf, eli
$$p(\theta) = P(\tilde{\theta} = \theta)$$
- $f(\mathbf{y} \mid \theta)$ on satunnaisvektorin $\mathbf{Y}$ pntf, kun $\tilde{\theta} = \theta$, eli
$$f(\mathbf{y} \mid \theta) = P(\mathbf{Y} = \mathbf{y} \mid \tilde{\theta} = \theta).$$
- $p(\theta \mid \mathbf{y})$ on satunnaismuuttujan $\tilde{\theta}$ ptnf, kun $\mathbf{Y} = \mathbf{y}$, eli
$$p(\theta \mid \mathbf{y}) = P(\tilde{\theta} = \theta \mid \mathbf{Y} = \mathbf{y})$$
- $f(\mathbf{y})$ on satunnaisvektorin $\mathbf{Y}$ reunajakauman ptnf, eli
$$f(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y})$$

Satunnaismuuttujien $\tilde{\theta}$ ja $\mathbf{Y}$ reunapistetodennäköisyysfunktiot saadaan niiden yhteispistetodennäköisyysfunkitosta kokonaistodennäköisyyden kaavalla, nimittäin

$$p(\theta) = \sum_{\mathbf{y}} P(\tilde{\theta} = \theta, \mathbf{Y} = \mathbf{y}) = \sum_{\mathbf{y}} f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) \tag{8.6}$$

$$f(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) = \sum_{\theta} f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}). \tag{8.7}$$

Todennäköisyyksien kertolaskukaavan (8.2) mukaan yhteispistetodennäköisyysfunktio voidaan jakaa tekijöihin molemmilla seuraavista tavoista

$$f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) = p(\theta)\, f(\mathbf{y} \mid \theta) = f(\mathbf{y})\, p(\theta \mid \mathbf{y}). \tag{8.8}$$

Bayesin kaava tapahtuman $\tilde{\theta} = \theta$ todennäköisyydelle ehdolla $\mathbf{Y} = \mathbf{y}$ saadaan ratkaistua edellisestä identiteetistä, nimittäin

$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{f(\mathbf{y})}. \tag{8.9}$$

Bayesiläinen päättely perustuu Bayesin kaavaan. Nyt parametria pidetään satunnaismuuttujana $\tilde{\theta}$ joka saa jonkin arvon parametriavaruudessa Θ. Jos parametri on diskreetti, niin bayesiläinen päättely perustuu suoraan edellä kirjoitettuihin kaavoihin.

Ennen (lat. *a priori*) havaintojen tekoa parametrilla on ns. *priorijakauma* (engl. *prior distribution*) eli jakauma, jonka ptnf on $p(\theta)$. Seuraavaksi tehdään havainto $\mathbf{Y} = \mathbf{y}$. Bayesiläinen päättely tarkoittaa sitä, että havaintojen jälkeen (lat. *a posteriori*) priorikäsitys päivitetään siirtymällä havaintoja vastaavaan ehdolliseen jakaumaan. Ennakkokäsitys päivitetään Bayesin kaavalla havaintojen jälkeiseksi käsitykseksi käyttämällä hyväksi priorijakaumaa ja havaintoja vastaavaa uskottavuusfunktiota $f(\mathbf{y} \mid \theta)$. Tulos on parametrin *posteriorijakauma* (engl. *posterior distribution*), eli parametrin ehdollinen jakauma $p(\theta \mid \mathbf{y})$ ehdolla $\mathbf{Y} = \mathbf{y}$.

Bayesin kaava kannattaa pitää mielessä muodossa

$$\text{posteriori} \propto \text{priori} \times \text{uskottavuus} \tag{8.10}$$

Tässä merkintä $\propto$ tarkoittaa verrannollisuutta, ts. edellä väitetään, että posteriori on vakio kertaa priorin ja uskottavuusfunktion tulo. Tässä posterioria $p(\theta \mid \mathbf{y})$ ajatellaan muuttujan θ funktiona kuten myös priorin ja uskottavuusfunktion tuloa. Ts. edellä väitetään, että

$$p(\theta \mid \mathbf{y}) = C\, p(\theta)\, f(\mathbf{y} \mid \theta), \qquad \text{kaikilla } \theta, \tag{8.11}$$

ja tämän on totta, sillä

$$C = \frac{1}{f(\mathbf{y})} = \frac{1}{\sum_{\theta=0}^{N} p(\theta)\, f(\mathbf{y} \mid \theta)}$$

Verrannollisuusvakio C toki riippuu havainnoista $\mathbf{y}$, mutta muuttujan θ funktiona ajateltuna se on vakio.

Bayesiläisessä analyysissä havaintoja vastaavalle satunnaisvektorille kiinnitetään sen havaittu arvo, ja sitten pohditaan eri parametrinarvojen todennäköisyyksiä. Frekventistisessä analyysissä parametri on kiinteä, ja todennäköisyyslaskentaa käytetään aineistoa vastaavan satunnaisvektorin jakauman ja siitä johdettujen tunnuslukujen jakaumien johtamiseen erilaisilla hypoteettisilla parametrinarvoilla. Useimmat frekventistisen tilastotieteen käsitteet perustuvat sellaisten aineistojen ominaisuuksien pohtimiseen, joita ei kokeessa havaittu.

Voimme ajatella, että bayesiläisessä analyysissä diskreetin parametrin tapauksessa lasketaan priorin ja uskottavuuden tulo kaikilla mahdollisilla parametrin arvoilla, jonka jälkeen tulos normalisoidaan pistetodennäköisyysfunktioksi jakamalla laskettujen arvojen summalla. Tämän algoritmin vaiheet ovat

1. Laske
$$s = \sum_{\theta} p(\theta)\, f(\mathbf{y} \mid \theta). \tag{8.12}$$
2. Laske
$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{s}, \qquad \theta \in \Theta. \tag{8.13}$$

Näemme sovelluksia seuraavissa jaksossa, joissa palaamme käsittelemään tilastollista päättelyä luvun 2 esimerkeissä.

8.2 Pallot kulhossa: diskreetti parametri

Kulhossa on N palloa ja niistä θ kpl on mustia ja $N - \theta$ valkoisia. Lukumäärä $0 \leq \theta \leq N$ on tuntematon. Palloja nostetaan satunnaisesti ja palauttaen n kertaa.

Nyt luontoäiti on laittanut pallot kulhoon sillä tavalla, että hän ensin arpoi valkoisten pallojen lukumääräksi θ yhden luvuista $0, 1, \ldots, N$ siten, että kaikki vaihtoehdot ovat yhtä todennäköisiä. Sen jälkeen hän laittoi kulhoon vastaavan lukumäärän valkoisia ja mustia palloja. Lopuksi pallojen poiminta tuottaa jonkin jonon $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ onnistumisia (valkoisen pallon nosto koodataan arvolla $y_i = 1$) tai epäonnistumisia (mustan pallon nosto koodataan arvolla $y_i = 0$). Vastaavien satunnaismuuttujien $Y_1, \ldots, Y_n$ yhteispistetodennäköisyysfunktio tunnetaan, jos θ tunnetaan, sillä se on
$$f(\mathbf{y} \mid \theta) = \left(\frac{\theta}{N}\right)^{t(\mathbf{y})} \left(1 - \frac{\theta}{N}\right)^{n-t(\mathbf{y})}, \tag{8.14}$$
jossa $t(\mathbf{y}) = \sum_i y_i$ on onnistumisten lukumäärä.

Kun nyt on havaittu tietty jono $\mathbf{y}$ niin sitten kysytään, millä todennäköisyydellä kulhossa on θ kappaletta valkoisia palloja. Kun tähän kysymykseen vastataan kaikilla mahdollisilla parametrin θ arvoilla $0, 1, \ldots, N$, saadaan tuloksena posteriorijakauma.

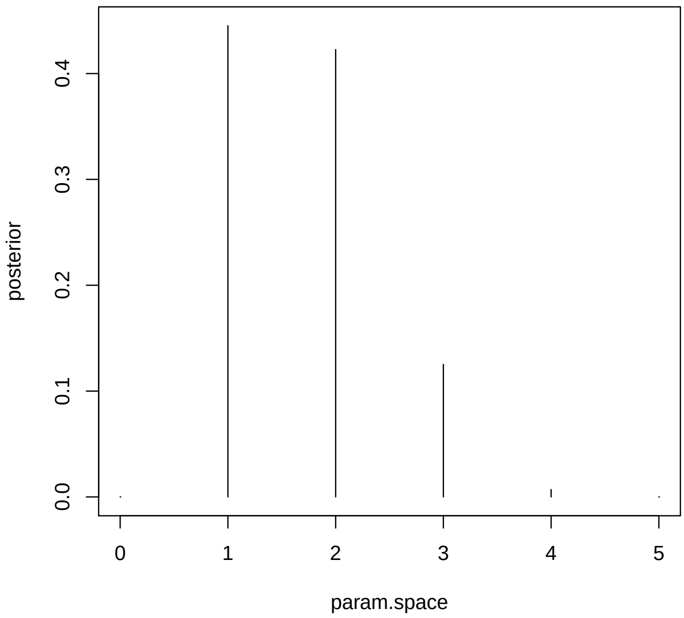
Ennen havaintoja satunnaismuuttujan $\tilde{\theta}$ jakauma eli priorijakauma on tilanteen kuvauksen perusteella diskreetti tasajakauma, jonka pistetodennäköisyysfunktio on
$$p(\theta) = \frac{1}{N+1}, \qquad \theta = 0, 1, \ldots, N.$$
Posteriorijakauma lasketaan seuraavassa esimerkissä soveltamalla kaavoja (8.12)–(8.13).

Esimerkki 8.1. Kulhossa on $N = 5$ palloa ja nostoja tehdään $n = 7$ ja tulokset ovat $\mathbf{y} = (1, 0, 0, 0, 1, 0, 0)$, eli onnistumisia on 2 kappaletta. Lasketaan R:llä priorin ja uskottavuusfunktion tulo, ja normalisoidaan se posteriorijakaumaksi.

```
> N <- 5
> n <- 7
> k <- 2
> param.space <- 0:N
> print(prior <- rep(1, N + 1) / (N+1))

[1] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
```

```
> print(likelihood <- (param.space / N)^k * (1 - param.space / N)^(n-k))
[1] 0.0000000 0.0131072 0.0124416 0.0036864 0.0002048 0.0000000
> h <- prior * likelihood
> print(posterior <- h / sum(h))
[1] 0.000000000 0.445217391 0.422608696 0.125217391 0.006956522 0.000000000
> names(posterior) <- as.character(param.space)
> print(posterior)
          0           1           2           3           4           5
0.000000000 0.445217391 0.422608696 0.125217391 0.006956522 0.000000000
> plot(param.space, posterior, 'h')
```



Tarkkaavainen lukija huomasi, että tasaisesti priorista seurasi se, että posteriorijakauma oli yhtä kuin uskottavuusfunktio normalisoituna todennäköisyysjakaumaksi. Havaintojen jälkeen parametrin todennäköisin arvo on 1, mutta arvo 2 on lähes yhtä todennäköinen. Huomaa, että nyt on paikallaan puhua parametrin todennäköisyydestä (ennen ja jälkeen havaintojen teon); enää ei tarvitse puhua esim. uskottavuudesta. △

8.3 Priorin ja posteriorin tulkitseminen epävarmuuden kuvauksina

Edellisen jakson laskuissa ei ole mitään kiistanalaista, vaan ne seuraavat suoraan todennäköisyyslaskennan sääntöjen avulla tilanteen kuvauksesta. Kuitenkin bayesiläistä päättelyä pidettiin tilastotieteilijöiden piirissä yleisesti ainakin 1920–1960 luvuilla vanhentuneena ja suorastaan tuomittavana tapana lähestyä tilastollisen päättelyn ongelmia. Nykypäivänä tilastotieteen ammattilehdissä suurin osa artikkeleistä käyttää tavalla tai toisella bayesiläistä lähestymistapaa. Yritän tässä jaksossa valottaa, mikä asia bayesiläisessä päättelyssä aikanaan aiheutti tämän voimakkaan vastustuksen.

Ongelmaksi koettiin se, että bayesiläistä lähestymistapaa sovellettiin tilanteissa, joissa parametrin arvoa ei määrännyt mikään satunnaismekanismi. Tällöin priorijakauma ei kuvaa todellista arpomista, vaan se pitää ymmärtää kvantitatiivisena kuvauksena soveltajan epävarmuudesta parametrin todellisesta arvosta ennen havaintojen tekemistä. Posteriorijakauman tulkinnaksi tulee puolestaan se, että se on kvantitatiivinen kuvaus menetelmän soveltajan epävarmuudesta parametrin todellisesta arvosta, kun havainnot on otettu huomioon. Kiistan ydin oli siinä, että frekventistisen tilastotieteen perustajat eivät hyväksyneet sitä ajatusta, että todennäköisyysjakauma saataisiin tulkita subjektiivisena epävarmuuden kuvauksena. Heidän mielestään todennäköisyyden käsite oli objektiivinen, ja sitä saadaan käyttää ainoastaan sellaisissa tilanteessa, jossa jotakin koetta (jossakin mielessä) toistetaan useita kertoja.

Nykyaikana bayesiläisessä päättelyssä lähes aina käytetään todennäköisyyden subjektiivista tulkintaa epävarmuuden kvantitatiivisena kuvauksena silloin, kun puhutaan parametrin todennäköisyysjakaumasta. Tätä ajatusta ei nykyään enää koeta ongelmallisena.

Posteriorijakauma on tällöin tietenkin subjektiivinen, sillä se riippuu siitä, minkälaista priorijakaumaa kyseinen subjekti pitää hyvänä kuvauksena omasta epävarmuudestaan. Priorijakaumalla on voimakas vaikutus posteriorijakaumaan, mikäli otoskoko on pieni. Jos otoskoko on suuri, niin tällöin erilaisilla järkevillä prioreilla saavutetaan lähes samanlainen posteriorijakauma. Otoskoon kasvaessa järkevien soveltajien subjektiiviset posteriorijakaumat alkavat siis muistuttaa yhä enenenevässä määrin toisiaan.

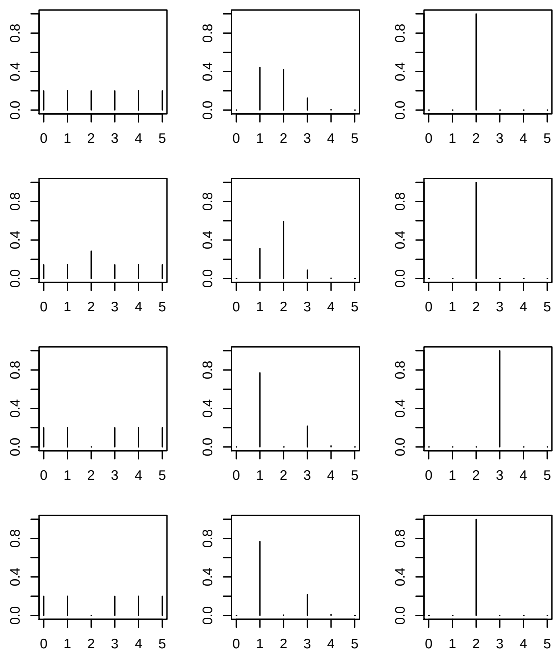
Lasketaan esimerkin vuoksi neljän eri henkilön A, B, C ja D posteriorijakaumat pallot kulhossa -tilanteessa: ensin seitsemän noston (2 onnistumista) ja sitten 300 noston (133 onnistumista) jälkeen.

- A:n priorijakauma on tasajakauma arvoilla $0, 1, \ldots, 5$.
- B:n priorijakauman mukaan valkoisten pallojen lukumäärä 2 on kaksi kertaa todennäköisempi kuin muut, jotka puolestaan ovat keskenään yhtä todennäköisiä.
- C:n ennakkokäsityksen mukaan arvo 2 on täysin mahdoton, ja kaikki muut mahdollisuudet ovat yhtä todennäköisiä.
- D:n ennakkokäsityksen mukaan arvolla 2 on todennäköisyys 1/1000, ja kaikki muut arvot ovat keskenään yhtä todennäköisiä.

Eri henkilöiden priorijakaumat ja posteriorijakaumat on esitetty kuvassa 8.3

Otoskoolla $n = 7$ nähdään, että posteriorijakaumat riippuvat vahvasti kunkin henkilön priorikäsityksistä, mutta henkilöt C ja D ovat käytännössä yhtä mieltä eri valkoisten pallojen lukumäärän todennäköisyyksistä. Sen sijaan otoskoolla $n = 300$ kaikkien muiden henkilöiden paitsi C:n posteriorikäsitykset ovat käytännössä yhtenevät. C sulkee omalla priorin valinnallaan kokonaan pois sen mahdollisuuden, että valkoisia palloja voisi kulhossa olla kaksi. Tämän takia

Kuva 8.1 Vaakariveillä on henkilöiden A, B, C ja D priorijakaumat, posteriorijakaumat $n = 7$ toiston ja $n = 300$ toiston jälkeen



C:n posterioritodennäköisyys kahdelle valkoiselle pallolle on aina nolla riippumatta lainkaan siitä, mitä havaintoja saadaan. Tällä kertaa aineisto todistaa otoskoolla $n = 300$ vahvasti sen puolesta, että valkoisia palloja olisi kulhossa todellisuudessa kaksi kappaletta, ja henkilöt A, B ja D ovat käytännössä täysin vakuuttuneita siitä, että valkoisia palloja on kulhossa kaksi.

Sellaista priorin valintaa (kuten C:n priori), joka sulkee osan parametriavaruudesta kokonaan pois tarkastelusta voidaan pitää bayesiläisen päättelyn pelisääntöjen vastaisena.

8.4 Nasta purkissa: jatkuva parametri

Nyt binomikokeen onnistumistodennäköisyydellä on jokin arvo avoimella välillä $(0, 1)$. Jos tämä arvo θ tunnetaan, niin havaintosatunnaisvektorin pistetodennäköisyysfunktio on

$$f(\mathbf{y} \mid \theta) = \theta^{t(\mathbf{y})} (1 - \theta)^{n-t(\mathbf{y})},$$

kun $\mathbf{y} = (y_1, \ldots, y_n)$ on jono nollia tai ykkösiä, ja $t(\mathbf{y}) = \sum_i y_i$ on onnistumisten lukumäärä n toistossa. Nyt parametriavaruus on jatkuva, ja tämä tuo mukanaan tiettyjä teknisiä ongelmia.

Parametri $\tilde{\theta}$ on nyt jatkuvasti jakautunut satunnaismuuttuja, jonka arvot kuuluvat parametriavaruuteen $\Theta = (0, 1)$. Ennen havaintojen tekoa soveltajan pitää onnistua kuvaamaan oma epävarmuutensa parametrin arvoista priorijakaumalla, jonka tiheysfunktiota merkitsemme

$$p(\theta), \qquad \theta \in \Theta.$$

Havaintojen jälkeen siirrytään tarkastelemaan parametrin ehdollista tiheysfunktiota eli sen posteriorijakaumaa.

Suuri osa jakson 8.1 kaavoista pitää sellaisenaan paikkansa myös tässä uudessa tilanteessa, mutta summaus pitää korvata integroinnilla. Yhteisjakauma voidaan esittää funktion

$$f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) = p(\theta)\, f(\mathbf{y} \mid \theta)$$

avulla. Se on muuttujan θ suhteen tiheysfunktio ja muuttujan $\mathbf{y}$ suhteen pistetodennäköisyysfunktio, ts. todennäköisyyksiä lasketaan integroimalla muuttujan θ suhteen ja summaamalla muuttujan $\mathbf{y}$ suhteen.

Bayesin kaava saa tutun muodon

$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{f(\mathbf{y})},$$

mutta nyt normalisointivakio $f(\mathbf{y})$ eli havaintosatunnaisvektorin reunajakauman pistetodennäköisyysfunktion arvo pitää laskea integroimalla,

$$f(\mathbf{y}) = \int_0^1 p(\theta)\, f(\mathbf{y} \mid \theta)\, \mathrm{d}\theta.$$

Tätä integraalia ei yleisesti ottaen osata laskea analyyttisesti. Sitä voi toki yrittää approksimoida jollakin numeerisella menetelmällä.

Posteriorijakauman tiheysfunktion kuva voidaan piirtää suoraan käyttämällä verrannollisuustulosta

$$p(\theta \mid \mathbf{y}) \propto p(\theta)\, f(\mathbf{y} \mid \theta),$$

mikäli tyydytään siihen, että y-akselin skaala jää selvittämättä.

8.5 Liittojakauma eli konjugaattijakauma

Joillekin uskottavuusfunktioille on mahdollista löytää sellainen parametrinen perhe jakaumia, joilla on seuraava miellyttävä ominaisuus. Mikäli priorijakauma valitaan kyseisestä perheestä, niin myös posteriorijakauma kuuluu samaan perheeseen. Näemme kohta, että binomikokeen uskottavuusfunktiolle beeta-jakaumat muodostavat tällaisen perheen. Tämä voidaan ilmaista sanomalla, että beeta-jakauma on binomiuskottavuuden *liittopriori* (engl. *conjugate prior*) tai että havaintosatunnaisvektorin otantajakauma ja priorijakauma ovat toistensa liittojakaumia.

Mikäli tutkijan ennakkokäsitys parametrinarvosta voidaan esittää jollakin liittoperheen jakaumalla, niin tällöin posteriorijakauma saadaan laskettua johtamalla päivityskaavat, joilla priorijakauman parametrit päivitetään posteriorijakauman parametreiksi. Kutsutaan näitä liittojakaumaperheen parametreja selvyyden vuoksi hyperparametreiksi, jotta ne saadaan erotettua tilastollisen mallin parametrista θ.

Beeta-jakauma on jatkuva jakauma, jonka tiheysfunktio on muotoa

$$g(x) \propto x^{\alpha-1} (1-x)^{\beta-1}, \quad \text{kun } 0 < x < 1, \tag{8.15}$$

missä $\alpha > 0$ ja $\beta > 0$ ovat jakaumaperheen parametrit. Tämä lauseke määrittelee yksikäsiteisesti tietyn todennäköisyysjakauman, jonka tiheysfunktio saadaan selville jakamalla lauseke sen välin $(0, 1)$ yli lasketulla integraalilla, sillä tiheysfunktion integraalin koko satunnaismuuttujan arvoalueen yli täytyy olla yksi. Kaavassa (8.15) merkitsemättä jätetty normalisointivakio saadaan ns. Eulerin beetafunktion

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

avulla (jossa B on iso beeta-kirjain). Beeta-jakauman tiheysfunktion täydellinen kaava on

$$g(x) = \begin{cases} \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & \text{kun } 0 < x < 1, \\ 0 & \text{muuten.} \end{cases} \tag{8.16}$$

Erityisesti valinnoilla $\alpha = 1$ ja $\beta = 1$ saadaan välin $(0, 1)$ tasajakauma.

Beeta-jakauman $\text{Beta}(\alpha, \beta)$ ominaisuudet tunnetaan. Sitä noudattavan satunnaismuuttujan X arvot ovat välillä $(0, 1)$, ja esim. sen odotusarvo ja varianssi saadaan laskettua tunnetuilla kaavoilla

$$EX = \frac{\alpha}{\alpha + \beta}, \qquad \operatorname{var} X = \frac{\alpha \, \beta}{(\alpha + \beta)^2 \, (\alpha + \beta + 1)}. \tag{8.17}$$

Jakauman moodi (eli tiheysfunktion maksimipiste) on

$$\frac{\alpha - 1}{\alpha + \beta - 2},$$

mikäli $\alpha > 1$ ja $\beta > 1$ (muilla parametrinarvoilla jakaumalla ei ole hyvin määriteltyä moodia).

Tarkistetaan nyt, että beeta-jakauma on binomiuskottavuuden liittopriori. Priorin hyperparametrit ovat $\alpha, \beta > 0$ ja onnistumisia havaitaan k kappaletta. Tarkistuksen voi tehdä kahdella erilaisella tavalla.

Tapa 1: normalisointivakion laskeminen integroimalla

Integroidaan tulo priori kertaa uskottavuus. Huomaa, että priorijakauman tiheysfunktiossa argumenttina pitää käyttää integrointimuuttujaa, joka on θ.

$$
\begin{aligned}
f(\mathbf{y}) &= \int_0^1 p(\theta)\, f(\mathbf{y} \mid \theta)\, \mathrm{d}\theta \\
&= \int_0^1 \frac{1}{B(\alpha,\beta)}\, \theta^{\alpha-1}\, (1-\theta)^{\beta-1}\, \theta^k\, \theta^{n-k}\, \mathrm{d}\theta \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 \theta^{\alpha+k-1}\, (1-\theta)^{\beta+n-k-1}\, \mathrm{d}\theta = \frac{B(\alpha+k, \beta+n-k)}{B(\alpha,\beta)}.
\end{aligned}
$$

Integraali osattiin laskea, koska huomattiin käyttää Eulerin beeta-funktion määritelmää.

Bayesin kaavan mukaan

$$
\begin{aligned}
p(\theta \mid \mathbf{y}) &= \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{f(\mathbf{y})} \\
&= \frac{\frac{1}{B(\alpha,\beta)}\, \theta^{\alpha-1}\, (1-\theta)^{\beta-1}\, \theta^k\, \theta^{n-k}}{B(\alpha+k, \beta+n-k)/B(\alpha,\beta)} \\
&= \frac{1}{B(\alpha+k, \beta+n-k)}\, \theta^{\alpha+k-1}\, (1-\theta)^{\beta+n-k-1}, \qquad 0 < \theta < 1
\end{aligned}
$$

Tästä nähdään, että posteriorijakauma on beeta-jakauma

$$\mathrm{Beta}(\alpha+k, \beta+n-k).$$

Tapa 2: verrannollisuustarkastelu

Edellisessä tavassa tulee matkan varrella helposti virheitä. Tulos on paljon helpompi johtaa seuraavalla tekniikalla. Muuttujan θ funktiona posteriorijakauman tiheys on verrannollinen priorijakauman tiheyden ja uskottavuusfunktion tuloon, joten välillä $0 < \theta < 1$ pätee verrannollisuus

$$
\begin{aligned}
p(\theta \mid \mathbf{y}) &\propto p(\theta)\, f(\mathbf{y} \mid \theta) \\
&= \theta^{\alpha-1}\, (1-\theta)^{\beta-1}\, \theta^k\, \theta^{n-k} \\
&= \theta^{\alpha+k-1}\, (1-\theta)^{\beta+n-k-1}.
\end{aligned}
$$

Vertaamalla tulosta kaavaan (8.15) nähdään, että posteriorijakauma on beeta-jakauma

$$\mathrm{Beta}(\alpha+k, \beta+n-k),$$

sillä ainoa todennäköisyysjakauma, jonka kantaja on $(0, 1)$ ja jonka tiheysfunktio on verrannollinen johdettuun funktioon on tämä beeta-jakauma.

8.6 Posteriorijakauman yhteenvetoja

Kun posteriorijakauma on selvitetty, niin lopuksi voidaan yrittää laskea siitä tiettyjä yhteenvetoja. Jos posteriorijakauma on ennestään tuttu yksinkertainen jakauma, niin tämä vaihe voidaan sivuuttaa.

Posteriorijakauman keskikohtaa voidaan luonnehtia parametrin posteriorio-dotusarvolla eli binomikokeen tapauksessa luvulla

$$E[\tilde{\theta} \mid \mathbf{y}] = \int_0^1 \theta\, p(\theta \mid \mathbf{y})\, \mathrm{d}\theta.$$

Beeta-jakaumalle tulos voidaan lukea suoraan beeta-jakauman odotusarvon kaavasta (8.17), kun siihen sijoitetaan posteriorijakauman hyperparametrit, jotka ovat binomikokeessa

$$\alpha_1 = \alpha + k, \qquad \beta_1 = \beta + n - k.$$

Vastaavasti voidaan laskea posteriorimoodi eli posteriorijakauman moodi. Posterioriodotusarvoa ja posteriorimoodia voidaan ajatella bayesiläisinä piste-estimaatteina. Posteriorijakauman keskittyneisyyttä voidaan kuvailla esim. laskemalla parametrin posteriorivarianssi eli posteriorijakauman varianssi.

Beeta-jakauman kvantiilifunktio $q(u)$ osataan laskea numeerisesti, ja tällä perusteella parametriavaruudesta löytyy helposti väli $[L, U]$, jolle

$$P(L \le \tilde{\theta} \le U \mid \mathbf{Y} = \mathbf{y}) = 1 - \alpha$$

millä tahansa annetulla $0 < \alpha < 1$. Eräs tällainen väli saadaan jakamalla virhetodennäköisyys α tasan alemman ja ylemmän jakauman hännän kesken valitsemalla

$$L = q(\alpha/2), \qquad U = q(\alpha/2).$$

Havainnon jälkeen parametri kuuluu tälle välille todennäköisyydellä $1 - \alpha$. Tälläistä väliä voidaan kutsua tason $1 - \alpha$ todennäköisyysväliksi tai bayesiläiseksi luottamusväliksi.

Mikäli ollaan kiinnostuneita muotoa

$$\tilde{\theta} \in A$$

olevasta hypoteesista, niin sen todennäköisyys havainnon jälkeen voidaan (periaatteessa) laskea integroimalla posteriorijakauman tiheysfunktiota

$$P(\tilde{\theta} \in A \mid \mathbf{Y} = \mathbf{y}) = \int_A p(\theta \mid \mathbf{y})\, \mathrm{d}\theta.$$

Beeta-jakauman kertymäfunktio osataan laskea numeerisesti, ja tämän takia yllä merkitty integraali osataan käytännössä laskea välittömästi numeerisesti, mikäli A on väli. Vastahypoteesin $\tilde{\theta} \notin A$ todennäköisyys havainnon jälkeen saadaan sitten vähentämällä luvusta yksi tarkasteltavan hypoteesin todennäköisyys.

Koska parametria käsitellään satunnaismuuttujana, niin päästään puhumaan parametrin todennäköisyyksistä tai parametrin todennäköisyysjakaumasta tai kyseisen jakauman ominaisuuksista havainnon tekemisen jälkeen. Johdot ovat periaatteessa täysin suoraviivaisia ja niissä tarvitaan ainoastaan todennäköisyyslaskentaa, eikä mitään sen ulkopuolisia periaatteita. Toisinaan vaadittavat laskut ovat kuitenkin liian hankalia analyyttisesti suoritettaviksi.

8.7 Bayesiläisen päättelyn laskentamenetelmiä

Posteriorijakauma saadaan selvitettyä kaavojen avulla seuraavissa tilanteissa.

1. Jos parametri on diskreetti ja sillä on äärellinen määrä mahdollisia arvoja.
2. Jos käytetään priorijakaumaa, joka kuuluu uskottavuusfunktion liittoperheeseen.

Muissa tapauksissa ei saada johdettua käyttökelpoisia analyyttisiä kaavoja.

Nykyään bayesiläinen analyysi tehdään tietokoneen avulla. Laskentamenetelmät perustuvat tyypillisesti satunnaisuuden simulointiin tietokoneen avulla eli ns. Monte Carlo -menetelmiin.

Perustana on se ajatus, että koska posteriorijakauma on todennäköisyysjakauma, niin sitä voidaan yrittää simuloida tietokoneella. On kehitetty menetelmiä, joissa lasketaan suuri määrä arvoja $t_1, \ldots, t_N$, joita voidaan pitää sellaisten satunnaismuuttujien $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ havaittuina arvoina, joista kukin on jakautunut posteriorijakauman mukaisesti.

Tämän jälkeen otosta $t_1, \ldots, t_N$ käsitellään data-analyysin keinoin.

- Posterioriodotusarvoa voidaan arvioida vastaavilla otoskeskiarvoilla,
$$E[\tilde{\theta} \mid \mathbf{y}] \approx \frac{1}{N} \sum_{i=1} t_i$$
- Parametrille voidaan muodostaa bayesiläisiä luottamusvälejä otoksesta laskettujen kvantiilipisteiden avulla.
- Posteriorijakauman pistetodennäköisyysfunktiota voidaan arvioida vastaavilla suhteellisilla frekvensseillä (diskreetin parametrin tapauksessa).
- Posteriorijakauman tiheysfunktiota voidaan arvioida esim. histogrammilla (jatkuvan parametrin tapauksessa).
- Mielivaltaisen parametrin funktion $g(\tilde{\theta})$ posteriorijakauman ominaisuuksia (esim. posterioriodotusarvoa tai posteriorijakaumaa) voidaan selvittää otoksesta $g(t_1), \ldots, g(t_N)$ aivan samoilla menetelmillä.

Tällaiset menetelmät ovat suhteellisen uusia: intensiivinen kehitysvaihe alkoi 1980-luvun lopulla. Niiden ansiosta nykyään on mahdollista analysoida lähes mielivaltaisen monimutkaisia bayesiläisiä tilastollisia malleja.