10 Bayes-päättelyn alkeita

- Jos parametrinarvo on kiinnitetty, niin havaintosatunnaisvektorin jakauma mallinnetaan bayesiläisessä päättelyssä täysin samalla tavalla kuin frekventistisessä päättelyssä.
- Frekventistisessä lähestymistavassa parametri on kiinteä (ts. ei-satunnainen) mutta tuntematon.
- Bayesiläisessä lähestymistavassa parametria käsitellään satunnaisena suureena.
- Tämä ehkä vähäiseltä tuntuva ero johtaa suuriin eroihin laskutekniikoissa ja tulosten tulkinnoissa.
- Bayesiläisessä päättelyssä kaikki on toisin kuin frekventistisessä.

10.1 Todennäköisyyslaskentaa

- Oletamme hiljaisesti, että osamäärien nimittäjät ovat esitettävissä kaavoissa erisuuria kuin nolla.
- Olkoot A ja B tapahtumia. Jos tiedämme (täsmälleen sen), että B on sattunut, niin tapahtuman A todennäköisyys lasketaan ehdollisen todennäköisyyden kaavalla

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

- Ehdollisen todennäköisyyden kaavasta saadaan todennäköisyyksien kertolaskukaava

$$P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A). \tag{2}$$

- Tästä nähdään kaava

$$P(A \mid B) = \frac{P(A)\,P(B \mid A)}{P(B)}. \tag{3}$$

Kokonaistodennäköisyyden kaava

Jos tapahtumat $A_1, \dots, A_M$ 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,

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

Bayesin kaava

Kun *kokonaistodennäköisyyden* kaavaa käytetään kaavassa (3) saadaan *Bayesin kaava*

$$P(A_k \mid B) = \frac{P(A_k)\, P(B \mid A_k)}{\sum_{i=1}^{M} P(A_i)\, P(B \mid A_i)} \tag{4}$$

johon bayesiläinen päättely perustuu diskreetin parametrin ja diskreetin havaintovektorin tapauksessa.

Samat kaavat uudestaan

- Käsitellään kahden diskreetin satunnaismuuttujan (tai satunnaisvektorin) $\tilde{\theta}$ ja $\mathbf{Y}$ yhteisjakaumaa, jonka määrää niiden yhteispistetodennäköisyysfunktio

$$f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) = P(\tilde{\theta} = \theta, \mathbf{Y} = \mathbf{y}) = P(\{\tilde{\theta} = \theta\} \cap \{\mathbf{Y} = \mathbf{y}\}). \quad (5)$$

- Satunnaismuuttujan $\tilde{\theta}$ mahdolliset arvot ovat $\theta_1, \theta_2, \ldots$.
- 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.

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

$$f(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y})$$

Reunajakaumat yhteisjakaumasta

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

$$p(\theta) = P(\tilde{\theta} = \theta) = \sum_{\mathbf{y}} P(\tilde{\theta} = \theta, \mathbf{Y} = \mathbf{y}) = \sum_{\mathbf{y}} f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}) \tag{6}$$

$$f(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) = \sum_{\theta} f_{\tilde{\theta},\mathbf{Y}}(\theta, \mathbf{y}). \tag{7}$$

Yhteisjakauma reunajakaumasta ja ehdollisesta jakaumasta

Todennäköisyyksien kertolaskukaavan (2) mukaan yhteispistetodennäköisyysfunktio voidaan jakaa tekijöihin molemmilla seuraavista tavoista

$$f_{\tilde{\theta},\mathbf{Y}}(\theta,\mathbf{y}) = p(\theta)\, f(\mathbf{y} \mid \theta) = f(\mathbf{y})\, p(\theta \mid \mathbf{y}). \tag{8}$$

Bayesin kaava

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

$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{f(\mathbf{y})}. \tag{9}$$

- Bayesiläisessä päättelyssä parametria pidetään satunnaismuuttujana $\tilde{\theta}$ joka saa jonkin arvon parametriavaruudessa Θ.
- Jos aineiston otantajakauma on diskreetti ja parametriavaruus on diskreetti, niin bayesiläinen päättely perustuu suoraan Bayesin kaavaan.

Priorista posterioriin

- 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 $\mathbf{Y} = \mathbf{y}$.

Avainkaava: posteriori $\propto$ priori $\times$ uskottavuus

- Merkintä $\propto$ tarkoittaa verrannollisuutta: posteriori on vakio kertaa priorin ja uskottavuusfunktion tulo, kun lausekkeita ajatellaan muuttujan θ funktiona.
- Ts. kalvon otsikossa väitetään, että

$$p(\theta \mid \mathbf{y}) = C\, p(\theta)\, f(\mathbf{y} \mid \theta), \qquad \text{kaikilla } \theta, \tag{10}$$

  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.

Bayesiläisen ja frekventistisen ajattelun ero

- 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 potentiaalisten aineistojen ominaisuuksien pohtimiseen, joita ei kokeessa havaittu.

Algoritmi: posteriori priorista ja uskottavuudesta

Voimme ajatella, että bayesiläisessä analyysissä diskreetin parametrin tapauksessa lasketaan priorin ja uskottavuuden tulo kaikilla mahdollisilla parametrin arvoilla, jonka jälkeen tulos normalisoidaan pistetodennäköisyysfunktioksi.

1. Laske
$$s = \sum_{\theta} p(\theta)\, f(\mathbf{y} \mid \theta). \tag{11}$$
2. Laske
$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)\, f(\mathbf{y} \mid \theta)}{s}, \qquad \theta \in \Theta. \tag{12}$$

10.2 Pallot kulhossa: diskreetti parametri

- Kulhossa on N palloa: θ valkoista ja $N - \theta$ mustaa.
- Valkoisten pallojen lukumäärä $0 \leq \theta \leq N$ on tuntematon. Palloja nostetaan satunnaisesti ja palauttaen n kertaa.
- Oletetaan nyt, että 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 θ valkoista ja $N - \theta$ mustaa palloa.
- Lopuksi pallojen poiminta tuottaa jonkin jonon $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ onnistumisia (valkoinen pallo: $y_i = 1$) tai epäonnistumisia (musta pallo: $y_i = 0$).

Kohti posteriorijakaumaa

- Vastaavien satunnaismuuttujien $Y_1, \ldots, Y_n$ yhteispistetodennäköisyysfunktio tunnetaan, jos θ tunnetaan, sillä se on

$$f(\mathbf{y} \mid \theta) = \left(\frac{\theta}{N}\right)^{t(\mathbf{y})} \left(1 - \frac{\theta}{N}\right)^{n-t(\mathbf{y})}, \tag{13}$$

  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.

Priorijakauma

- Ennen havaintoja satunnaismuuttujan $\tilde{\theta}$ jakauma eli priorijakauma on tilanteen kuvauksen perusteella diskreetti tasajakauma, jonka pistetodennäköisyysfunktio on

$$p(\theta) = \frac{1}{N+1}, \qquad \theta = 0, 1, \dots, N.$$

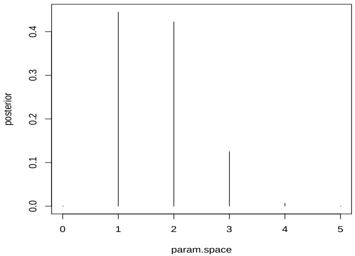
- Esimerkki: Kulhossa on $N = 5$ palloa ja nostoja tehdään $n = 7$ ja tulokset ovat $\mathbf{y} = (1, 0, 0, 0, 1, 0, 0)$, eli onnistumisia on $x = 2$ kappaletta. Lasketaan R:llä priorin ja uskottavuusfunktion tulo, ja normalisoidaan se posteriorijakaumaksi.

Priori ja uskottavuusfunktio

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

## [1] 0.1667 0.1667 0.1667 0.1667 0.1667 0.1667

print(likelihood <- (param.space/N)^x * (1 - param.space/N)^(n -
    x))

## [1] 0.0000000 0.0131072 0.0124416 0.0036864 0.0002048 0.0000000
```

Posteriori

```r
h <- prior * likelihood
posterior <- h/sum(h)
names(posterior) <- as.character(param.space)
print(posterior)

##        0        1        2        3        4        5
## 0.000000 0.445217 0.422609 0.125217 0.006957 0.000000

plot(param.space, posterior, "h")
```



Tulosten tulkintaa

- Tasaisesti priorista seuraa se, että posteriorijakauma on 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.

10.3 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 artikkeleista käyttää tavalla tai toisella bayesiläistä lähestymistapaa.
- Yritän tässä jaksossa valottaa, mikä asia bayesiläisessä päättelyssä aikanaan aiheutti tämän voimakkaan vastustuksen.

Mikä bayesiläisessä lähestymistavassa arvelutti

- Bayesiläistä lähestymistapaa sovelletaan myös tilanteissa, joissa parametrin arvoa ei määrää 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 epävarmuudesta parametrin todellisesta arvosta, kun havainnot on otettu huomioon.

Kiistan ydin

- 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.

Bayesiläinen päättely on subjektiivista

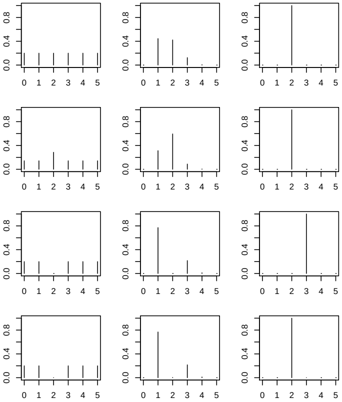
- Posteriorijakauma on tällöin tietenkin subjektiivinen, sillä se riippuu siitä, minkälaista priorijakaumaa kyseinen subjekti pitää hyvänä kuvauksena omasta epävarmuudestaan.
- Priorijakaumalla on voimakas vaikutus posteriorijakaumaan, mikäli otoskoko on pieni.
- Jos otoskoko on suuri, niin tällöin erilaisilla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

Henkilöiden A, B, C ja D posteriorit

Kohta näytetään 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 ennakkokäsitys tukee kahden pallon vaihtoehtoa. B:n priorijakauman mukaan valkoisten pallojen lukumäärä 2 on kaksi kertaa todennäköisempi kuin muut, jotka puolestaan ovat keskenään yhtä todennäköisiä.
- C on dogmaattinen kahden pallon teorian vastustaja. C:n ennakkokäsityksen mukaan arvo 2 on täysin mahdoton, ja kaikki muut mahdollisuudet ovat yhtä todennäköisiä.
- D vastustaa kiivaasti kahden pallon teoriaa. D:n ennakkokäsityksen mukaan arvolla 2 on todennäköisyys $1/1000$, ja kaikki muut arvot ovat keskenään yhtä todennäköisiä.

Tulokset kuvana



Tulosten tulkintaa

- Otoskoolla $n = 7$ posteriorijakaumat riippuvat vahvasti kunkin henkilön priorikäsityksistä, mutta henkilöt C ja D ovat käytännössä yhtä mieltä eri valkoisten pallojen lukumäärän todennäköisyyksistä.
- 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 C:n posterioritodennäköisyys kahdelle valkoiselle pallolle on aina nolla riippumatta lainkaan siitä, mitä havaintoja saadaan.
- Nyt 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 tästä asiasta.

Kommentti

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

10.4 Nasta purkissa: jatkuva parametri

- Tässä tilanteessa binomikokeen onnistumistodennäköisyydellä on jokin arvo avoimelta väliltä $(0, 1)$.
- 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 uusia teknisiä ongelmia.

Jatkuvasti jakautunut parametri

- 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.

Jatkuvassa tapauksessa summaus korvataan integroinnilla

- Suuri osa jakson 10.1 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

- Bayesin kaava saa muodon

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

- Nyt normalisointivakio $f(\mathbf{y})$ 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.
- Posteriorijakauman tiheysfunktion kuva voidaan piirtää käyttämällä verrannollisuustulosta

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

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

10.5 Liittojakauma eli konjugaattijakauma

- Joillekin uskottavuusfunktioille on mahdollista löytää sellainen parametrinen perhe jakaumia, että mikäli priorijakauma valitaan kyseisestä perheestä, niin myös posteriorijakauma kuuluu samaan perheeseen.
- 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.

Priorin päivittäminen posterioriksi

- 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

- Beeta-jakauma on jatkuva jakauma, jonka tiheysfunktio on muotoa

$$g(x) \propto x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{kun } 0 < x < 1, \tag{14}$$

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.

Beeta-jakauman normalisointivakio

- Kaavassa (14) 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

$$g(x) = \begin{cases} \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & \text{kun } 0 < x < 1, \\ 0 & \text{muuten.} \end{cases} \tag{15}$$

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

Beeta-jakauman $\mathrm{Be}(\alpha, \beta)$ ominaisuuksia

- Jakaumaa noudattavan satunnaismuuttujan X arvot ovat välillä $(0, 1)$.
- Sen odotusarvo ja varianssi saadaan laskettua tunnetuilla kaavoilla

$$EX = \frac{\alpha}{\alpha + \beta}, \qquad \mathrm{var}\, X = \frac{\alpha\,\beta}{(\alpha + \beta)^2\,(\alpha + \beta + 1)}. \tag{16}$$

- 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 2: verrannollisuustarkastelu

Muuttujan θ funktiona posteriorijakauman tiheys on verrannollinen priorijakauman tiheyden ja uskottavuusfunktion tuloon, joten välillä $0 < \theta < 1$ pätee verrannollisuus

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

Vertaamalla tulosta kaavaan (14) nähdään, että posteriorijakauma on beeta-jakauma

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

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

10.6 Posteriorijakauman yhteenvetoja

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

Posteriorijakauman keskikohdan luonnehdinta

- Posteriorijakauman keskikohtaa voidaan luonnehtia parametrin posterioriodotusarvolla 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 (16), 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.

Bayesiläinen luottamusväli

- Mikäli posteriorijakauman kvantiilifunktion $q(u)$ arvoja osataan laskea, niin sen jälkeen parametriavaruudesta löytyy helposti väli $[L, U]$, jolle

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

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

$$L = q(\alpha/2), \qquad U = q(1 - \alpha/2).$$

- Liittojakaumien tilanteessa tarvittava kvantiilifunktio löytyy valmiina tilastollisista ohjelmistoista.
- Tälläistä väliä voidaan kutsua tason $1 - \alpha$ todennäköisyysväliksi tai bayesiläiseksi luottamusväliksi.

Bayesiläinen hypoteesintestaus

- Mikäli ollaan kiinnostuneita muotoa

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

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

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

- Jos posteriorijakauman kertymäfunktio osataan laskea ja jos A on väli, niin välin A posterioritodennäköisyys saadaan laskemalla kerymäfunktion arvot välin päätepisteissä sekä laskemalla suuremman arvon ja pienemmän arvon erotus.
- Vastahypoteesin $\tilde{\theta} \notin A$ todennäköisyys havainnon jälkeen saadaan sitten vähentämällä luvusta yksi tarkasteltavan hypoteesin todennäköisyys.

Bayesiläinen hypoteesintestaus liittojakaumien tilanteessa

- Jos käytetään uskottavuusfunktion liittoprioria, niin posteriorijakauman kertymäfunktio löytyy valmiiksi toteutettuna tilastollisista ohjelmistoista.

Kommentteja

- 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.
- Monimutkaisten tilastollisten mallien yhteydessä vaadittavat laskut ovat kuitenkin liian hankalia analyyttisesti suoritettaviksi.

10.7 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.

Monte Carlo -mentelmä bayesiläiseissä laskennassa

- 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.

Mitä kaikkia asioita voidaan selvittää, jos käytössä on otos posteriorijakaumasta

- Posterioriodotusarvoa voidaan arvioida vastaavilla otoskeskiarvoilla,
$$E[\tilde{\theta} \mid \mathbf{y}] \approx \frac{1}{N}\sum_{i=1}^{N} t_i$$
- 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.

Bayesiläisen päättelyn laskentamenetelmistä

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