8. Todennäköisyyslaskentaa

Ennen bayesläisen päättelyn käsittelyä kertaamme ehdolliseen todennäköisyyteen liittyvää kalkyyliä. [AS, jakso 1.4; TN I, jaksot 1.7 ja 1.10] [AS ≡ Arjas - Sirén] [TN I ≡ P. Tuominen: Todennäköisyyslaskenta I]

Olkoot A ja B tapahtumia. Oletetaan, että $P(B) > 0$ ja että tiedämme, että B on sattunut (mutta emme tiedä mitään muuta). Miten tällöin pitää määritellä A:n tn? Vastaus: pitää laskea A:n todennäköisyys ehdolla B, eli

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

[TN I, määr. 1.7.1]

Olkoot $P(A) > 0$ ja $P(B) > 0$. Ehdollisen tn:n määritelmästä saadaan (todennäköisyyksien) <u>kertolaskukaava</u> eli <u>ketjusääntö</u>:

$$P(A \cap B) = P(B)\, P(A|B) = P(A)\, P(B|A)$$

Tästä voidaan ratkaista toinen ehdollisista todennäköisyyksistä, $P(B|A)$, jos $P(A|B)$, $P(A)$ ja $P(B)$ tunnetaan:

$$P(B|A) = \frac{P(B)\, P(A|B)}{P(A)}$$

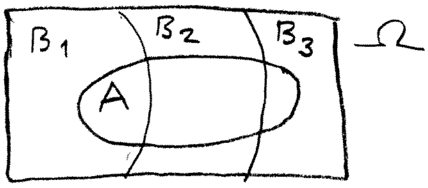
Olkoon $B_1, B_2, \ldots, B_M$ jokin perusjoukon Ω ("varman tapahtuman") ositus ts.

- joukot B_i ovat pistevieraita
- niiden yhdiste $\bigcup_{j=1}^{M} B_j = \Omega$

Oletetaan, että tunnetaan

- tn:t $P(B_k)$, $k = 1, \ldots, M$
- ehdolliset tn:t $P(A \mid B_k)$, $k = 1, \ldots, M$

Havaitaan, että tapahtuma A sattuu. Miten lasketaan tämän tiedon valossa tapahtumien B_k todennäköisyydet $P(B_k \mid A)$?



Ratkaisu:

$$P(B_k \mid A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k)\,P(A \mid B_k)}{P(A)}$$

Miten lasketaan $P(A)$? Ratkaisu: A voidaan osittaa pistevieraisiin paloihin $A \cap B_i$, $i = 1, \ldots, M$, joten (todennäköisyyden additiivisuus:)

$$\begin{aligned} P(A) &= P(A \cap B_1) + \cdots + P(A \cap B_M) \\ &= P(B_1)\,P(A \mid B_1) + \cdots + P(B_M)\,P(A \mid B_M) \end{aligned}$$

Tämä on <u>kokonaistodennäköisyyden kaava.</u>

Nyt saadaan <u>Bayesin kaava</u>:

$$P(B_k \mid A) = \frac{P(B_k)\,P(A \mid B_k)}{\sum_{i=1}^{M} P(B_i)\,P(A \mid B_i)}$$

Laskemista helpottava huomio:
Bayesin kaava voidaan esittää verrannollisuustuloksena

$$P(B_k|A) \propto P(B_k)\, P(A|B_k) =: q_k$$

Tämä tarkoittaa sitä, että

$$P(B_k|A) = c\, q_k = c\, P(B_k) P(A|B_k)$$

jossa $c = 1/P(A)$ on muuttujasta k riippumaton vakio. Tämä vakio voidaan määrittää siitä ehdosta, että

$$1 = \Sigma_k P(B_k|A) = c\, \Sigma q_k$$

$$\Rightarrow \frac{1}{c} = \Sigma q_i = \Sigma P(B_i) P(A|B_i) = P(A)$$

Miksi sitten on ilmeistä, että $\Sigma_k P(B_k|A) = 1$?
Vastaus 1: asian näkee Bayesin kaavasta!
Vastaus 2: ehdollinen tn $B \mapsto P(B|A)$ on todennäköisyysmitta ja (B_k) on Ω:n ositus. Tn:n additiivisuuden nojalla

$$1 = P(\Omega|A) = \Sigma_k P(B_k|A).$$

Bayesläisessä tilastotieteessä

- $P(B_k)$, $k = 1, \ldots, M$ on priorijakauma
  [lat. a priori ≡ ennen (havaintoa)]
- A vastaa havaintoa
- $k \mapsto P(A|B_k)$ on uskottavuusfunktio
- $P(B_k|A)$, $k = 1, \ldots, M$ on posteriorijakauma
  [lat. a posteriori = (havainnon) jälkeen]

Resepti (Prioista ja uskottavuusfunktiosta posterioriin; diskreetti tapaus, äärellinen arvojoukko)

Aineet: Priorijakauma ja uskottavuusfunktio

Ohjeet: 1) Laske luvut q_k kertomalla priori × uskottavuus, eli

$$q_k = P(B_k) \cdot P(A \mid B_k)$$

2) Laske summa $s = \Sigma q_k$.

3) Posteriorijakauma on

$$P(B_k \mid A) = q_k / s.$$

Esimerkkejä: ks. AS esim. 1.1 ja 1.2

Esim 1.1 : $M = 2$

$P(B_1) = 0{,}001$ $\quad$ $P(B_2) = 0{,}999$

$P(A \mid B_1) = 0{,}997$ $\quad$ $P(A \mid B_2) = 0{,}015$

$q_1 = 0{,}001 \times 0{,}997$ $\quad$ $q_2 = 0{,}999 \times 0{,}015$

$s = q_1 + q_2 = 0{,}015982$

$P(B_1 \mid A) = \dfrac{q_1}{s} \approx 0{,}062$ $\quad$ $P(B_2 \mid A) = \dfrac{q_2}{s} \approx 0{,}938$

Suurin osa työstä kuluu siihen, että tilanteen selostuksesta ymmärretään, mikä on priori ja mikä uskottavuusfunktio. Varsinainen lasku on helppo.

[Vrt. AS: jakso 1.3]

9. Bayesläinen päättely: parametri satunnaismuuttujana

Frekventistisessä päättelyssä parametri on tuntematon, mutta kiinteä suure (kiinteä ≡ deterministinen ≡ ei-satunnainen). Tämän takia frekventistisessä tilastotieteessä parametriavaruudessa ei ole määriteltynä todennäköisyysjakaumaa, ja kaikki todennäköisyyttä koskevat lausumat [esim. luottamusvälin luottamustaso] pitää ajatella niin, että aineisto tulkitaan niissä satunnaiseksi. Jotta saataisiin konkreettinen tulkinta, voidaan ajatella, että alkuperäistä koetta toistetaan vastaavissa oloissa hyvin monta kertaa, ja sitten lasketaan kyseisen tapahtuman suhteellinen frekvenssi toistoissa (⇒ frekventistinen päättely).

Bayesläisessä päättelyssä toimitaan toisin:

- parametri ajatellaan satunnaismuuttujaksi
- havaintoja vastaavalle satunnaisvektorille $\underline{X}$ formuloidaan sen ehdollinen todennäköisyysjakauma ehdolla parametri
- kun havainnot $\underline{x}$ on saatu, satunnaisvektorille $\underline{X}$ kiinnitetään arvo $\underline{X} = \underline{x}$
- päättelyn tulos on parametrin posteriorijakauma, joka lasketaan Bayesin kaavalla:

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

Tarkastellaan ensin diskreettiä tapausta, jossa satunnaisen parametrin K arvo kuuluu johonkin äärelliseen joukkoon S_K, ja havaintovektorin

$$\underline{X} = (X_1, X_2, \ldots, X_n)$$

arvo kuuluu johonkin äärelliseen joukkoon $S_{\underline{X}} \subset \mathbb{R}^n$.

Ptnf (pistetodennäköisyysfunktio)
$k \mapsto P(K=k)$ on <u>priorijakauma</u>

Ehdollinen ptnf

$$\underline{x} \mapsto P(\underline{X} = \underline{x} \mid K=k) \qquad \underline{x} \in S_{\underline{X}}, \quad k \in S_K$$

on aineiston <u>otantajakauma</u> ehdolla $K=k$.

Yhdessä nämä kaksi funktiota spesifioivat parametrin ja havaintovektorin $\underline{X}$ yhteisjakauman, sillä (kertolaskukaava!)

$$P(K=k, \underline{X} = \underline{x}) = P(K=k)\, P(\underline{X}=\underline{x} \mid K=k)$$

Nämä kaksi funktiota määrittelevät bayesläisen tilastollisen mallin; malli M^* [AS].

Kun havaintovektorin $\underline{X}$ arvo $\underline{x}$ havaitaan, niin sen jälkeen kaikki tieto parametrin K jakaumasta sisältyy sen <u>posteriorijakaumaan</u>

$$k \mapsto P(K=k \mid \underline{X}=\underline{x})$$

$$\propto \underbrace{P(K=k)}_{\text{priori}}\ \underbrace{P(\underline{X}=\underline{x} \mid K=k)}_{\text{uskottavuusfunktio}}$$

<u>Uskottavuusfunktio</u> on $k \mapsto P(\underline{X}=\underline{x} \mid K=k)$, jossa aineistoa vastaavalle satunnaisvektorille $\underline{X}$ on kiinnitetty sen havaittu arvo $\underline{x}$.

<u>Esimerkki</u>: mustat ja valkoiset pallot kulhossa:
N palloa, joista K (parametri) valkoista
ja N – K mustaa. Poimitaan palloja satunnaisesti
palauttaen.

$$X_i = \begin{cases} 1, & \text{jos } i\text{:s nostettu pallo on valkoinen} \\ 0, & \text{jos } i\text{:s nostettu pallo on musta} \end{cases}$$

$$\begin{aligned} & P(\underline{X} = \underline{x} \mid K = k) \\ &= P(X_1 = x_1, \dots, X_n = x_n \mid K = k) \\ &= \left(\frac{k}{N}\right)^{x_1} \left(1 - \frac{k}{N}\right)^{1-x_1} \cdots \left(\frac{k}{N}\right)^{x_n} \left(1 - \frac{k}{N}\right)^{1-x_n} \\ &= \left(\frac{k}{N}\right)^{T(\underline{x})} \left(1 - \frac{k}{N}\right)^{n - T(\underline{x})} \end{aligned}$$

jossa
$$T(\underline{x}) = \sum_{i=1}^{n} x_i = \text{nostettujen valkoisten lkm}$$
$$n - T(\underline{x}) = \text{nostettujen mustien lkm}$$

$$\frac{k}{N} = P(\text{"nostetaan valkoinen"} \mid K = k)$$
$$1 - \frac{k}{N} = P(\text{"nostetaan musta"} \mid K = k)$$

Huomautus: satunnaismuuttujat $X_1, \dots, X_n$ ovat <u>ehdollisesti riippumattomia</u> ehdolla $K = k$, eli ne ovat riippumattomia niiden <u>ehdollisessa</u> yhteisjakaumassa, sillä

$$P(\underline{X} = \underline{x} \mid K = k) = \prod_{i=1}^{n} P(X_i = x_i \mid K = k)$$

$$P(X_i = x_i \mid K = k) = \left(\frac{k}{N}\right)^{x_i} \left(1 - \frac{k}{N}\right)^{1 - x_i}$$

Ne eivät (yleensä) ole marginaalisesti riippumattomia eli riippumattomia (reuna-) yhteisjakaumassa

$$P(\underline{X} = \underline{x}) = \sum_k P(K = k, \underline{X} = \underline{x}) = \sum_k P(K = k)\, P(\underline{X} = \underline{x} \mid K = k)$$

joka ei yleensä faktoroidu reunajakaumien tuloksi.

Kun havaitaan, että $\underline{X} = \underline{x}$, sen jälkeen lasketaan posteriorijakauma

$$P(K = k \mid \underline{X} = \underline{x}) = \frac{P(K=k)\, P(\underline{X} = \underline{x} \mid K=k)}{\sum_i P(K=i)\, P(\underline{X} = \underline{x} \mid K=i)}$$

[k:n funktiona]

$$\propto P(K=k)\, P(\underline{X} = \underline{x} \mid K=k)$$

"Pallot kulhossa" -esimerkissä riittää tuntea nostettujen valkoisten pallojen lkm $T(\underline{x})$, sillä tämä tunnusluku kertoo havainnoista kaiken tarvittavan (se on ns. tyhjentävä tunnusluku): uskottavuusfunktion $P(\underline{X} = \underline{x} \mid K=k)$ arvot voidaan laskea heti, kun $T(\underline{x})$ tiedetään.

Mitä parametrin satunnaisuus tarkoittaa?

Yksi mahdollisuus: K on oikeasti poimittu jostakin jakaumasta, ja sitten kulho on täytetty.

Toinen mahdollisuus: pidän valkoisten pallojen lukumäärää satunnaismuuttujana, koska en tiedä mikä arvo sillä on. Tässä todennäköisyyden subjektiivisessa tulkinnassa priorijakauma kuvaa kvantitatiivisesti subjektin (esim. minun tai sinun) epävarmuuden parametrin arvosta. Havainnon teon jälkeen tämän epävarmuuden ilmaisee posteriorijakauma, joka saadaan laskettua Bayesin kaavalla.

Bayesläisessä tilastotieteessä käytetään yleensä tätä subjektiivista tulkintaa.

10. Kalkyyliä pistetodennäköisyysfunktioille

Tarkastellaan diskreettejä sm:ia (sm = satunnaismuuttuja), X, Y, K, joista kukin voi saada vain äärellisen määrän eri arvoja; arvojoukot S_X, S_Y, S_K äärellisiä joukkoja. Oletetaan epäoleellisten vaikeuksien välttämiseksi, että

$$P(X=x, Y=y, K=k) > 0 \quad \forall x \in S_X, y \in S_Y, k \in S_K$$

Tällöin kaikki alla käytetyt ehdolliset tn:t ovat hyvin määriteltyjä. Jatkossa jätetään ehdot $x \in S_X$, $y \in S_Y$, $k \in S_K$ kirjoittamatta.

<u>Marginalisointi</u>, eli miten siirrytään yhteisjakaumasta reunajakaumaan:

$$P(X=x, Y=y) = \sum_k P(X=x, Y=y, K=k)$$

(todennäköisyyden additiivisuus!)

$$P(X=x) = \sum_x P(X=x, Y=y)$$

$$= \sum_x \sum_k P(X=x, Y=y, K=k)$$

jne. Yhteenveto:

Yhteis-ptnf:stä (ptnf = pistetodennäköisyysfunktio) saadaan reuna-ptnf summaamalla siitä tarpeeton muuttuja tai tarpeettomat muuttujat pois.

Ehdollinen ptnf ≡ ehdollinen tn. Esim.

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

Kertolaskukaava (eli -sääntö)

$$P(X=x, Y=y) = P(X=x)\, P(Y=y \mid X=x)$$
$$= P(Y=y)\, P(X=x \mid Y=y)$$

(on sama asia kuin ehdollisen ptnf:n määritelmä.)

<u>Riippumattomuus</u>: $X \perp\!\!\!\perp Y$, jos

$$P(X=x, Y=y) = P(X=x)P(Y=y) \quad \forall x, y$$

Jos $X \perp\!\!\!\perp Y$, niin $P(Y=y) = P(Y=y \mid X=x)$ kaikilla x, y, ja $P(X=x) = P(X=x \mid Y=y)$ kaikilla x, y. Tämä seuraa riippumattomuuden määritelmästä sekä kertolaskukaavasta.

Marginalisointi onnistuu samalla tavalla myös ehdolliselle yhteis-ptnf:lle:

$$P(X=x \mid K=k) = \sum_y P(X=x, Y=y \mid K=k)$$

(ehdollinen tn $B \mapsto P(B \mid K=k)$ on additiivinen!)

<u>Ehdollinen riippumattomuus</u>: $X \perp\!\!\!\perp Y \mid K$, eli X ja Y ovat riippumattomia ehdolla K, jos ne ovat riippumattomia ehdollisessa yhteisjakaumassa

$$P(X=x, Y=y \mid K=k) \quad \forall k$$

ts. jos

$$P(X=x, Y=y \mid K=k) = P(X=x \mid K=k)\, P(Y=y \mid K=k)$$

$$\forall x, y, k$$

Vaihtoehtoinen luonnehdinta ehdolliselle riippumattomuudelle (HT):

$$X \perp\!\!\!\perp Y \mid K \quad , \text{ jos}$$

$$P(X = x \mid Y = y, K = k) = P(X = x \mid K = k)$$

tai

$$P(Y = y \mid X = x, K = k) = P(Y = y \mid K = k)$$

$$\forall x, y, k$$

Huomautus: nämä kaikki kaavat pätevät myös, jos X onkin (diskreetti) satunnaisvektori, esim.

$$X = \underline{X} = (X_1, X_2, \ldots, X_n)$$

tai Y on (diskreetti) satunnaisvektori tai K on (diskreetti) satunnaisvektori.

Lisäksi nämä kaavat yleistyvät myös jatkuvien jakaumien tapauksiin, jolloin summien sijasta pitää laskea integraaleja. Palaamme jatkuvaan tapaukseen myöhemmin.

11. Ennustaminen: mikä pallo seuraavaksi?

[AS: jakso 1.5]

Pallot kulhossa: olemme havainneet
$\underline{X} = \underline{x}$ eli $X_1 = x_1,\ X_2 = x_2, \ldots,\ X_n = x_n$

Millä todennäköisyydellä seuraava nosto tuottaa valkoisen pallon, eli $X_{n+1} = 1$?

Tähän kysymykseen on hankala antaa vastausta frekventististen paradigman puitteissa; voitaisiin yrittää ennustaa sijoittamalla parametrin K järkevä frekventistinen estimaatti kaavaan K/N ts. ennuste olisi $\hat{K}/N$. Tällöin täysin sivuutetaan se ongelma, että K:n estimointiin liittyy epävarmuutta.

Bayesläisessä paradigmassa ennustaminen on periaatteessa suoraviivaista: pitää tarkastella ennuste- eli prediktiivistä todennäköisyyttä

$$P(X_{n+1} = 1 \mid \underline{X} = \underline{x})$$

Tämä tapa ottaa huomioon parametrin estimointiin liittyvän epävarmuuden!

Miten ennustetodennäköisyys lasketaan? Yksinkertaisin tapa on hyödyntää ehdollista riippumattomuutta. Kun $K=k$, on

$$P(\underline{X}=\underline{x},\ X_{n+1}=x_{n+1} \mid K=k)$$
$$= \left(\frac{k}{N}\right)^{x_1}\left(1-\frac{k}{N}\right)^{1-x_1}\cdots\left(\frac{k}{N}\right)^{x_n}\left(1-\frac{k}{N}\right)^{1-x_n}\left(\frac{k}{N}\right)^{x_{n+1}}\left(1-\frac{k}{N}\right)^{1-x_{n+1}}$$
$$= P(\underline{X}=\underline{x} \mid K=k)\, P(X_{n+1}=x_{n+1} \mid K=k)$$

Siis $\underline{X} \perp\!\!\!\perp X_{n+1} \mid K$.

Nyt

$$P(X_{n+1}=x_{n+1} \mid \underline{X}=\underline{x})$$
$$= \sum_k P(X_{n+1}=x_{n+1}, K=k \mid \underline{X}=\underline{x}) \quad \text{(marginalisointi)}$$
$$= \sum_k P(K=k \mid \underline{X}=\underline{x})\, P(X_{n+1}=x_{n+1} \mid \underline{X}=\underline{x}, K=k)$$

(kertolaskukaava)

$$= \sum_k P(K=k \mid \underline{X}=\underline{x})\, P(X_{n+1}=x_{n+1} \mid K=k)$$

(sillä $\underline{X} \perp\!\!\!\perp X_{n+1} \mid K$)

Tässä siis summataan posterioritn $P(K=k \mid \underline{X}=\underline{x})$ kertaa uuden havainnon X_{n+1} ehdollinen tn,

$$P(X_{n+1}=x_{n+1} \mid K=k) = \begin{cases} \frac{k}{N}, & \text{jos } x_{n+1}=1 \\ 1-\frac{k}{N}, & \text{jos } x_{n+1}=0 \end{cases}$$

Ennustejakauma on tietenkin ptnf, eli

$$\underbrace{P(X_{n+1}=0 \mid \underline{X}=\underline{x})}_{\geq 0} + \underbrace{P(X_{n+1}=1 \mid \underline{X}=\underline{x})}_{\geq 0} = 1$$

joten riittää laskea toinen tapauksista $x_{n+1}=1$ tai $x_{n+1}=0$

12. Kertolaskukaava jälleen kerran

Edellä sovellettiin kertolaskukaavaa ehdollisille yhteis-ptnf:oille. Sen mukaan ehdollinen yhteis-ptnf voidaan jakaa tekijöihin samalla periaatteella kuin ei-ehdollinen yhteis-ptnf, ts.

$$P(X=x,\ Y=y \mid K=k) = P(X=x \mid K=k)\ P(Y=y \mid X=x, K=k)$$

aina, kun x, y ja k ovat diskreettien satunnaismuuttujien tai -vektorien mahdollisia arvoja ja kun kyseiset ehdolliset tn:t ovat hyvin määriteltyjä. Tämän tuloksen voi perustella tarkistamalla identiteetin

$$P(A \cap B \mid C) = P(A \mid C)\ P(B \mid A \cap C)$$

(kun kyseiset ehdolliset tn:t ovat hyvin määriteltyjä). Tämän jälkeen samastetaan

$$A = \{X = x\},\quad B = \{Y = y\},\quad C = \{K = k\}$$

ja huomataan kaavojen yhtäpitävyys. Tätä varten pitää tietenkin huomata, että pilkun tapahtumien välillä tn-merkinnässä $P(\cdot \mid \cdot)$ tarkoittaa sitä, että kaikki pilkulla erotetut tapahtumat sattuvat, eli pilkku $\equiv$ tapahtumien leikkaus.

Kertolaskusääntö saadaan helposti yleistettyä myös useammalle kuin kahdelle tapahtumalle; nimittäin pätee

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$$
$$P(A_1)\,P(A_2 \mid A_1) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

(mikäli kaikki ehdolliset tn:t ovat hyvin määriteltyjä). Tästä saadaan kertolaskukaava yhteis-ptnf:oille. Analoginen tulos pätee myös ehdollisille tn:ille sekä ehdollisille yhteis-ptnf:oille.

Esimerkki: pallot kulhossa, ilman takaisinpanoa.
Kulhossa on aluksi N palloa, joista K (parametri) on valkoista. Poimimme palloja kulhosta palauttamatta nostettuja palloja kulhoon. Miten lasketaan otantajakauma

$$P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n \mid K=k),$$

kun $n<N$?

Vastaus: käytetään kertolaskukaavaa:

$$P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n \mid K=k)$$
$$= P(X_1=x_1 \mid K=k)$$
$$\times P(X_2=x_2 \mid X_1=x_1, K=k)$$
$$\times \ldots$$
$$\times P(X_n=x_n \mid X_1=x_1, X_2=x_2, \ldots, X_{n-1}=x_{n-1}, K=k)$$

Kukin termeistä on helppo järkeillä pitämällä kirjaa siitä, kuinka monta palloa kulhossa on jäljellä ja kuinka monta niistä on valkoista silloin, kun kyseinen ehto on voimassa.

Esim. $$P(X_1=x_1 \mid K=k) = \begin{cases} \frac{k}{N}, & \text{kun } x_1=1 \\ 1-\frac{k}{N}, & \text{kun } x_1=0 \end{cases}$$

$$P(X_2=1 \mid X_1=0, K=k) = \begin{cases} \frac{k}{N-1}, & \text{kun } 0 \le k \le N-1 \\ 0, & \text{kun } k=N \end{cases}$$

$$P(X_2=1 \mid X_1=1, K=k) = \begin{cases} \frac{k-1}{N-1}, & \text{kun } 1 \le k \le N \\ 0, & \text{kun } k=0 \end{cases}$$

Komplementtitapahtuman $X_2=0$ ehdollinen tn löytyy tietenkin kaavalla

$$P(X_2=0 \mid X_1=x_1, K=k) = 1 - P(X_2=1 \mid X_1=x_1, K=k)$$

Tätä menettelyä on helppo jatkaa eteenpäin. (HT)
Huom. Tässä esimerkissä ei päde s. 39 positiivisuusoletus. Tästä syystä ehdollisen tn:n määritelmää laajennettiin.

Huomaa, että edellä tapahtumat $\{X_1=0, K=N\}$ sekä $\{X_1=1, K=0\}$ ovat mahdottomia (niiden tn on nolla). Käytin konventiota

$$P(B|A) = \begin{cases} \frac{P(A\cap B)}{P(A)}, & \text{kun } P(A)>0 \\ 0, & \text{kun } P(A)=0 \end{cases}$$

Tällöin kertolaskusääntö pysyy voimassa:

1) jos $P(A)>0$, niin
$$P(A)\,P(B|A) = P(A)\,\frac{P(A\cap B)}{P(A)} = P(A\cap B)$$

2) jos $P(A)=0$, niin
$$P(A)\,P(B|A) = 0\cdot 0 = 0,$$
mutta toisaalta $A\cap B \subset A$, joten
$$0 \le P(A\cap B) \le P(A) = 0,$$
joten kertolaskusääntö on voimassa.

Tämä päättely yleistyy ehdolliselle tn:lle sekä niin joukon leikkaukselle, joten otantajakauma (tai uskottavuusfunktio) voidaan tässäkin tapauksessa laskea kertolaskusäännöllä.

Posteriorijakauma saadaan Bayesin kaavalla
$$P(K=k \mid \underline{X}=\underline{x}) = \frac{1}{P(\underline{X}=\underline{x})}\,P(K=k)\,P(\underline{X}=\underline{x} \mid K=k),$$
eikä sen kohdalla jouduta samanlaisiin vaikeuksiin, sillä järkevissä bayeslaisissa malleissa $P(\underline{X}=\underline{x})>0$; $P(K=k \mid \underline{X}=\underline{x})$ häviää niillä k, joilla havainto $\underline{x}$ on mahdoton.

[$P(\underline{X}=\underline{x})$ olisi nolla, jos $P(K=k, \underline{X}=\underline{x})$ olisi identtisesti nolla havaitulle $\underline{x}$. Tällainen havainto $\underline{x}$ olisi mallin mukaan mahdoton, joten malli ei voisi olla järkevä todellisuuden kuvaus.]

13. Ehdolliset jakaumat yleisemmässä tapauksessa

Tähän asti ehdolliset jakaumat on käsitelty suoraan ehdollisen tn:n määritelmän avulla. Nyt siirrymme käsittelemään tilannetta, jossa mukana on jatkuvasti jakautuneita muuttujia.

Hyvä uutinen: kaavat näyttävät samalta kuin ennen, kun summan tilalle kirjoitetaan tarvittaessa integraali.

Huono uutinen: niitä integraaleja harvoin osataan laskea analyyttisesti.

Hyvä uutinen: joissakin tapauksissa integraalit osataan laskea nopean silmäilyn jälkeen (liitto- eli konjugaattijakaumat).

Kun havainto ja parametri ovat diskreettejä satunnaismuuttujia, niin bayesläinen päättely palautuu kertolaskukaavaan,

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y \mid X=x)$$
$$= f_Y(y)\, f_X(x \mid Y=y),$$

jossa ptnf:oita merkitään

$$f_{X,Y}(x,y) = P(X=x, Y=y) \qquad f_X(x) = P(X=x)$$
$$f_Y(y \mid X=x) = P(Y=y \mid X=x) \quad \text{jne.}$$

Sopivasti tulkittuna tämä kaava säilyy voimassa kaikissa jatkossa esiin tulevissa tilanteissa.

Epäolellisten vaikeuksien välttämiseksi teemme positiivisuusoletuksen: tarkastelemme sellaisia yhteisjakaumia, joille

$$f_{X,Y}(x,y) > 0 \iff (f_X(x) > 0 \text{ ja } f_Y(y) > 0)$$

ja vain sellaisia (x,y), joille $f_{X,Y}(x,y) > 0$.

<u>Tapaus: X diskreetti, Y jatkuva</u>

$f_X(x) = P(X = x)$ on X:n reuna-ptnf.

$f_Y(y \mid X = x)$ on Y:n tf (tiheysfunktio) ehdolla $X = x$:

$$P(a \le Y \le b \mid X = x) = \int_a^b f_Y(y \mid X = x)\, dy$$

kaikilla $a \le b$

$f_Y(y)$ on Y:n reuna-tf:

$$P(a \le Y \le b) = \int_a^b f_Y(y)\, dy$$

$f_X(x \mid Y = y)$ on X:n ptnf ehdolla $Y = y$.

Näiden välillä on yhteyksiä:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)\, f_Y(y \mid X = x) \\ &= f_Y(y)\, f_X(x \mid Y = y) \quad \text{kaikilla } x, y \end{aligned}$$

Yhteisjakauman esityksestä $f_{X,Y}(x,y)$ saadaan todennäköisyyksiä summaamalla x:n suhteen ja integroimalla y:n suhteen:

$$P(X \in A, Y \in B) = \sum_{x \in A} \int_B f_{X,Y}(x,y)\, dy$$

kaikilla (järkevillä) joukoilla $A, B \subset \mathbb{R}$

Marginalisointi:

$$f_X(x) = \int f_{X,Y}(x,y)\, dx \quad \left[\begin{array}{l}\text{Määrätty integraali} \\ \text{joukon } S_Y = \{y : f_Y(y) > 0\} \\ \text{yli.}\end{array}\right]$$

$$f_Y(y) = \sum_x f_{X,Y}(x,y)$$

Seuraus: jos jompikumpi yhteisjakauman faktoroinneista tunnetaan, niin niistä toinen pystytään (periaatteessa) laskemaan:

$$f_Y(y \mid X = x) = \frac{f_Y(y)\, f_X(x \mid Y = y)}{f_X(x)}$$

$$f_X(x \mid Y = y) = \frac{f_X(x)\, f_Y(y \mid X = x)}{f_Y(y)}$$

Tapaus: (X,Y):llä jatkuva yhteisjakauma

$f_X(x)$ on X:n reuna-tf

$f_Y(y \mid X=x)$ on Y:n tf ehdolla $X=x$

$f_Y(y)$ on Y:n reuna-tf

$f_X(x \mid Y=y)$ on X:n tf

$f_{X,Y}(x,y)$ on parin (X,Y) yhteis-tf.

Näiden välillä on yhteyksiä:

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y \mid X=x)$$
$$= f_Y(y)\, f_X(x \mid Y=y) \quad \text{kaikilla } x,y.$$

yhteis-tf:stä saadaan todennäköisyyksiä integroimalla kummankin muuttujan suhteen:

$$P(X\in A, Y\in B) = \iint_{A\times B} f_{X,Y}(x,y)\,dx\,dy$$
$$= \int_A \left(\int_B f_{X,Y}(x,y)\,dy \right) dx$$
$$= \int_B \left(\int_A f_{X,Y}(x,y)\,dx \right) dy$$

(tasointegraali voidaan palauttaa iteroiduksi integraaliksi).

Marginalisointi:

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy$$

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx$$

Seuraus: jos jompikumpi yhteis-tf:n faktoroinneista tunnetaan, niin se toinen pystytään (periaatteessa) laskemaan.

Huomautus: jos Y:llä on jatkuva jakauma, niin arvoa $f_X(x \mid Y=y)$ ei voida suoraan ajatella ehdollisena todennäköisyytenä, mutta se voidaan tulkita raja-arvojen kautta.

Jos X diskreetti, Y jatkuva, niin

$$P(X=x \mid y \le Y \le y+h) \xrightarrow[h \to 0+]{} f_X(x \mid Y=y)$$

(ainakin joillakin jatkuvuusoletuksilla).

Jos (X,Y):llä on jatkuva yhteisjakauma, niin kaikilla $a \le b$ pätee

$$P(a \le X \le b \mid y \le Y \le y+h) \xrightarrow[h \to 0+]{} \int_a^b f_X(x \mid Y=y)\,dx$$

(ainakin joillakin jatkuvuusoletuksilla).

Tilastotieteessä havainnot tehdään vain jollakin tarkkuudella, ja tällöin käytännössä approksimoidaan tämän tapaisia suureita vastaavien ehdollisten jakaumien avulla.

Yhteenveto: jokaisessa tapauksessa (kertolaskukaava)

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y \mid X=x) = f_Y(y)\, f_X(x \mid Y=y)$$

Marginalisointi:

$$f_X(x) = \begin{cases} \sum_y f_{X,Y}(x,y), & Y \text{ diskreetti} \\ \int f_{X,Y}(x,y)\,dy, & Y \text{ jatkuva} \end{cases}$$

$$f_Y(y) = \begin{cases} \sum_x f_{X,Y}(x,y), & X \text{ diskreetti} \\ \int f_{X,Y}(x,y)\,dx, & X \text{ jatkuva} \end{cases}$$

$$\left.\begin{matrix} f_X(x) \\ f_X(x \mid Y=y) \end{matrix}\right\} \text{ on } \begin{cases} \text{ptnf}, & \text{jos } X \text{ diskreetti} \\ \text{tf}, & \text{jos } X \text{ jatkuva} \end{cases}$$

$$\left.\begin{matrix} f_Y(y) \\ f_Y(y \mid X=x) \end{matrix}\right\} \text{ on } \begin{cases} \text{ptnf}, & \text{jos } Y \text{ diskreetti} \\ \text{tf}, & \text{jos } Y \text{ jatkuva} \end{cases}$$

Huomautus 1) ptnf on ei-negatiivinen ja sen summa koko avaruuden yli on yksi.
2) tf on ei-negatiivinen ja sen integraali koko avaruuden yli on yksi.

Ehdollinen jakauma verrannollisuusmerkinnän avulla:

Jos Y:llä on jatkuva jakauma, niin

$$f_Y(y \mid X = x) = \frac{1}{f_X(x)} f_Y(y)\, f_X(x \mid Y = y)$$

[y:n funktiona]

$$\propto f_Y(y)\, f_X(x \mid Y = y)$$

Tässä tarvittavan normalisointivakion $f_X(x)$ saa selville normalisoimalla tulon tiheysfunktioksi:

$$f_X(x) = \int f_Y(y)\, f_X(x \mid Y = y)\, dy$$

(Vastaava tarkastelu onnistuu tietenkin myös silloin, kun Y:llä on diskreetti jakauma.)

Kirjallisuudesta löytyy monenlaisia merkintöjä yhteis- reuna- ja ehdollisille jakaumille.
Usein käytetään merkintöjä

$$f_X(x),\ f_Y(y),\ f_{X|Y}(x|y),\ f_{Y|X}(y|x),\ f_{X,Y}(x,y)$$

Bayesläisen päättelyn yhteydessä on tavanomaista jättää alaindeksit kirjoittamatta. Tällöin merkinnät voisivat olla $f(x), f(y), f(x|y), f(y|x), f(x,y)$.
Aloittelijan on näitä merkintöjä vaikea ymmärtää, sillä esim. $f(x|y)$ voi, asiayhteydestä riippuen, tarkoittaa joko

- funktiota $(x,y) \mapsto f_X(x \mid Y = y)$
- funktiota $x \mapsto f_X(x \mid Y = y)$ jollakin kiinteällä y
- funktiota $y \mapsto f_X(x \mid Y = y)$ jollakin kiinteällä x
- arvoa $f_X(x \mid Y = y)$ kiinteillä (x, y).

Tästä huolimatta tällainen geneerinen notaatio ei (taitavasti käytettynä) yleensä johda sekaannukseen.

14. Nasta lasipurkissa: jatkuva parametri

[AS: jakso 1.6]

Nastaa helistetään lasipurkissa ja sen jälkeen rekisteröidään, laskeutuuko se selälleen vai kyljelleen. Koetta toistetaan yhteensä n kertaa.

$$X_i = \begin{cases} 1, & \text{jos nasta tulee selälleen } i\text{:nnellä toistolla} \\ 0, & \text{— " — kyljelleen — " —} \end{cases}$$

Valitaan parametriksi välillä $(0,1)$ oleva reaaliluku θ, joka on tn sille, että nasta päätyy selälleen. Voisimme periaatteessa selvittää θ:n arvon mielivaltaisen tarkasti tekemällä erittäin pitkän koesarjan (mutta jossain vaiheessa jokin menisi rikki: nasta, lasipurkki tai helistäjän käsi).

Emme tunne θ:n arvoa (tällä kertaa kukaan ei oikeasti tunne sitä), joten pidämme sitä satunnaismuuttujana. Huomautus: kreikkalaisten kirjaimien ja parametrien kohdalla käytetään usein samaa symbolia (esim. θ) sekä satunnaismuuttujalle että sen (potentiaaliselle) arvolle. Tehdään [toisin kuin AS, ks. s. 20] kuitenkin selvyyden vuoksi merkintäsopimus:

$\tilde{\theta}$ on parametri ymmärrettynä satunnaismuuttujana

θ on $\tilde{\theta}$:n (potentiaalinen) arvo.

Ennalleko käsitystämme parametrin arvosta kuvaa jokin välillä (0,1) määritelty jatkuva jakauma, jonka tiheysfunktio on

$$f_{\tilde{\theta}}(\theta) =: p(\theta), \qquad 0 < \theta < 1$$

Jos parametrin arvo θ tunnetaan, niin aivan ilmeisesti havaintovektorin otantajakauma on

$$P(\underline{X} = \underline{x} \mid \tilde{\theta} = \theta) = P(\underline{X} = \underline{x} \mid \theta) \quad [\text{merkintä}]$$

$$= \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{T(\underline{x})} (1-\theta)^{n-T(\underline{x})},$$

jossa

$$T(\underline{x}) = \sum_{i=1}^{n} x_i$$

Posteriorijakauma on jatkuva jakauma, jonka tf

$$p(\theta \mid \underline{X} = \underline{x}) := f_{\tilde{\theta}}(\theta \mid \underline{X} = \underline{x})$$

$$= \frac{1}{f_{\underline{X}}(\underline{x})} \, p(\theta) \, P(\underline{X} = \underline{x} \mid \theta)$$

$$\propto p(\theta) \, P(\underline{X} = \underline{x} \mid \theta) \qquad \left[\begin{array}{l} \theta\text{:n funktiona,} \\ 0 < \theta < 1 \end{array} \right]$$

Tässä verrannollisuustuloksessa tarvittava normalisointivakio $f_{\underline{X}}(\underline{x})$ saadaan laskettua integraalina

$$f_{\underline{X}}(\underline{x}) = \int_0^1 p(\theta) \, p(\underline{X} = \underline{x} \mid \theta) \, d\theta$$

(mutta tämä voi olla käytännössä vaikea laskea.)

15. Liittojakauma eli konjugaattijakauma

Joillekin uskottavuusfunktioille on mahdollista löytää parametrinen perhe jakaumia siten, että mikäli priorijakauma valitaan kyseisestä perheestä, niin myös posteriorijakauma kuuluu kyseiseen perheeseen. Nasta lasipurkissa -esimerkeissä päädyttiin ns. binomiuskottavuuteen (binomikokeen uskottavuusfunktioon), ja tälle uskottavuusfunktiolle beta-jakaumat muodostavat tällaisen mukavan jakaumaperheen. Tällöin sanotaan, että beta-jakauma on binomiuskottavuuden <u>liittopriori</u> (conjugate prior) tai että otantajakauma ja priorijakauma ovat toistensa <u>liittojakaumia</u>.

Beta-jakauma on jatkuva jakauma, jonka tiheysfunktio on muotoa

$$f(x) \propto x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

missä $\alpha > 0$ ja $\beta > 0$ ovat jakaumaperheen parametrit. Merkitsemättä jätetty normalisointivakio on

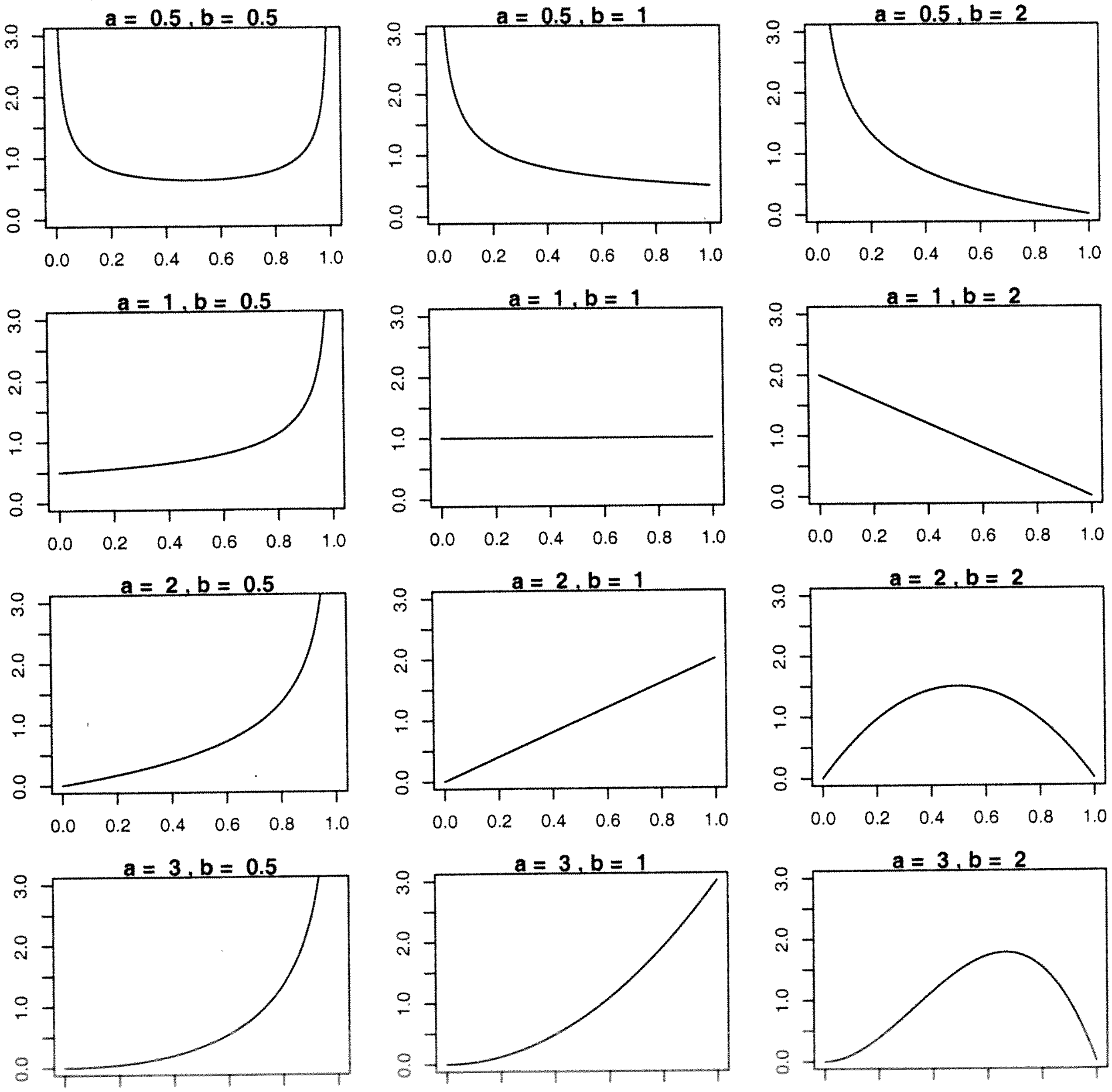
$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

Tämä on ns. (Eulerin) beta-funktio (ja "B" on iso beta-kirjain). Ts. tiheysfunktion täydellinen kaava on (mielivaltaisilla $\alpha,\beta > 0$)

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

Useimmiten riittää tietää tiheysfunktion muoto, $f(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$; toisinaan tarvitaan normalisointivakion lauseketta.

Beta-jakauman tiheysfunktioita eri parametriarvoilla. Tapaus $\alpha = \beta = 1$ on välin $(0,1)$ tasajakauma $Tas(0,1)$. Tiheysfunktiolla on moodi pisteessä $\frac{\alpha-1}{\alpha+\beta-2}$, mikäli $\alpha, \beta > 1$. Jakauman odotusarvo on $\alpha/(\alpha+\beta)$ kaikilla $\alpha, \beta > 0$.



Beta-funktion arvot voidaan laskea, jos osataan laskea (Eulerin) gammafunktion $\Gamma(\cdot)$ arvoja sillä voidaan osoittaa, että

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \qquad \alpha,\beta > 0$$

jossa Γ-funktio voidaan määritellä kaavalla

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx, \quad \text{kun } t > 0$$

Voidaan osoittaa, että $\Gamma(1) = 1$ ja että

$$\Gamma(t+1) = t\,\Gamma(t) \quad \text{kaikilla } t > 0,$$

mistä helposti päätellään, että $\Gamma(n) = (n-1)!$, kun $n \geq 1$ on kokonaisluku.

(Näitä kaavoja ei tarvitse opetella tenttiä varten ulkoa, vaan ne annetaan tehtäväpapereissa, mikäli niitä tarvitaan).

Tarkistetaan, että beta-jakauma on binomiuskottavuuden liittojakauma. Priorin parametrit ovat $\alpha,\beta > 0$.

Tapa 1) Lasketaan normalisointivakio $f_{\underline{X}}(\underline{x}) = P(\underline{X} = \underline{x})$.

$$P(\underline{X} = \underline{x}) = \int_0^1 p(\theta)\, P(\underline{X} = \underline{x} \mid \theta)\, d\theta$$

$$= \int_0^1 \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}\, \theta^{T(\underline{x})}(1-\theta)^{n-T(\underline{x})}\, d\theta$$

[Kirjoitettiin beta-tiheys käyttämällä muuttujana θ:a, sekä binomikokeen uskottavuus]

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 \theta^{\alpha+T(\underline{x})-1}(1-\theta)^{\beta+n-T(\underline{x})-1}\, d\theta$$

$$= \frac{B(\alpha+T(\underline{x}),\, \beta+n-T(\underline{x}))}{B(\alpha,\beta)},$$

joten

$$p(\theta \mid \underline{X} = \underline{x}) = \frac{p(\theta)\, P(\underline{X} = \underline{x} \mid \theta)}{P(\underline{X} = \underline{x})}$$

$= \cdots =$ betajakauman tiheys, jonka parametrit ovat $\alpha + T(\underline{x})$ ja $\beta + n - T(\underline{x})$

Tavassa 1) sattuu laskun varrella helposti virheitä. Tulos on paljon helpompaa johtaa seuraavalla tavalla:

Tapa 2) θ:in funktiona

posteriori $\propto$ priori $\times$ uskottavuus,

joten

$$p(\theta \mid \underline{X} = \underline{x}) \overset{[\theta]}{\propto} p(\theta)\, P(\underline{X} = \underline{x} \mid \theta)$$

$$\propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\, \theta^{T(\underline{x})}(1-\theta)^{n-T(\underline{x})}$$

[Kirjoitettiin $p(\theta)$ (θ:in funktiona!) sekä uskottavuusfunktio, mutta jätettiin kirjoittamatta vakiot, jotka eivät riipu θ:in arvosta]

$$\propto \theta^{\alpha + T(\underline{x}) - 1}(1-\theta)^{\beta + n - T(\underline{x}) - 1}, \quad 0 < \theta < 1$$

Nyt huomataan, että (ainoa!) jatkuva jakauma, jonka tf on (normalisointivakiota vaille) tämän näköinen on beta-jakauma, jonka parametrit ovat $\alpha + T(\underline{x})$ ja $\beta + n - T(\underline{x})$.

Tässä $T(\underline{x})$ on "onnistumisten" $X_i = 1$ lkm, ja $n - T(\underline{x})$ on "epäonnistumisten" $X_i = 0$ lkm missä toistossa, jotka ovat ehdollisesti riippumattomia ehdolla θ [tai pedanttisemmin ilmaistuna: ehdolla $\tilde{\theta} = \theta$].

Resepti (Posterioijakauman johtaminen liitto- eli konjugaattiperheen tapauksessa.)

Ainekset: Priorijakauma ja uskottavuusfunktio

Ohjeet: 1) Kirjoita "priori × uskottavuus" käyttämällä johdonmukaisia merkintöjä: priorijakauman esitys pitää kirjoittaa saman muuttujan (esim. θ) funktiona, jota käytetään parametrille uskottavuusfunktiossa.

2) Kehitä saamaasi lauseketta muuttujan θ funktiona verrannollisuusmerkintöjen avulla. Voit jättää pois tekijöitä, jotka eivät riipu muuttujasta θ.

3) Tunnista lopputuloksesta posteriorijakauman muoto (mistä parametrisesta perheestä se tulee ja mitkä ovat ko. perheen parametrit).

4) Jos et tunnistanut jakaumaa, voi olla että kyseessä ei ole liittojakaumatilanne, joten pitää keksiä jokin muu lähestymistapa.

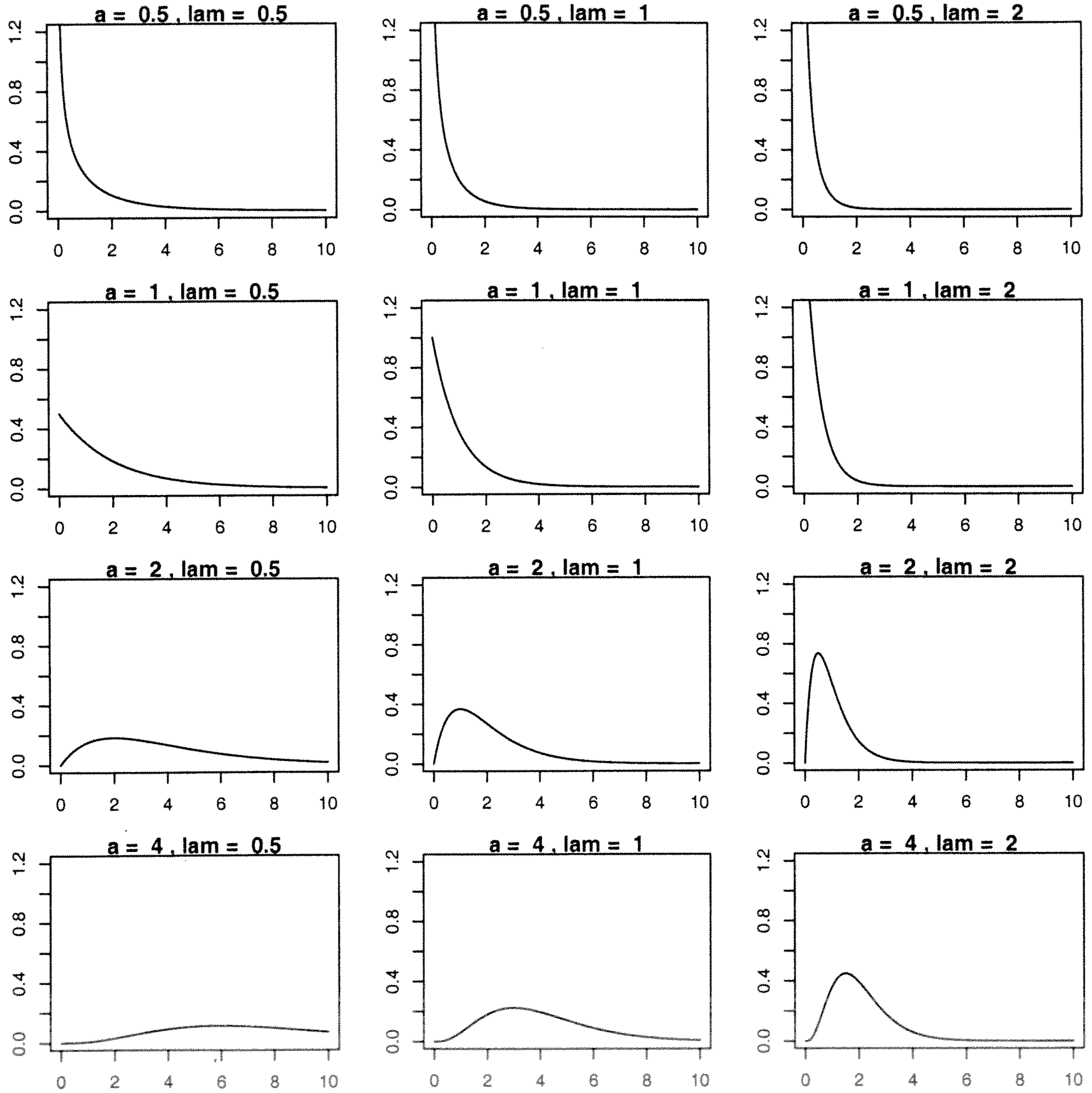
Usein esiintyviä liittopriorieita:

- beetajakauma todennäköisyysparametrille (binomiuskottavuus; geometrisesta jakaumasta tai negatiivisesta binomijakaumasta saatava uskottavuusfunktio)

- gammajakauma positiivisille parametreille (Poissonin jakauma; eksponenttijakauma ym.)

- normaalijakauma (joissakin erikoistapauksissa)

Gammajakauman tiheysfunktioita eri parametriarvoille. Tapaus $\alpha=1$ on eksponenttijakauma. Tiheysfunktiolla on moodi pisteessä $(\alpha-1)/\lambda$, mikäli $\alpha>1$. Jakauman odotusarvo on α/λ ja varianssi on α/λ^2.



Gamma-jakauma on jatkuva jakauma, jonka tiheysfunktio on muotoa

$$f(x) \propto x^{\alpha-1} e^{-\lambda x}, \quad x > 0,$$

missä $\alpha > 0$ ja $\lambda > 0$ ovat jakaumaperheen parametrit. Normalisointivakion kanssa tf on

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{kun } x > 0 \\ 0, & \text{muuten} \end{cases}$$

[Suunnilleen yhtä yleinen parametrointi: α ja $\beta = 1/\lambda$.]

Esimerkki: eksponenttijakautuneet havainnot ja gammapriori. Olkoot $X_1, \ldots, X_n \sim Exp(\theta)$ riippumattomasti, ehdolla θ, ja priori olkoon $Gamma(\alpha, \lambda)$ [α ja λ vakioita]. Tällöin posteriori $p(\theta \mid \underline{x})$ on

$$p(\theta \mid \underline{x}) \propto \underbrace{p(\theta)}_{\text{priorin tf}} \underbrace{f_{\underline{X}}(\underline{x} \mid \theta)}_{\text{uskottavuusfunktio}}$$

$$\propto \underbrace{\theta^{\alpha-1} e^{-\lambda\theta}}_{\substack{\text{muuttujan } \theta \\ \text{funktiona;} \\ \text{jätetään vakio pois}}} \prod_{i=1}^{n} \underbrace{\theta e^{-\theta x_i}}_{\substack{\text{parametrina } \theta; \\ \text{i:s havaittu arvo on } x_i}}$$

$$\propto \theta^{\alpha+n-1} \exp\left(-\theta\left(\lambda + \sum_{i=1}^{n} x_i\right)\right), \quad \theta > 0$$

Tuttu $\Rightarrow$ posteriorijakauma on

$$Gamma\left(\alpha + n, \lambda + \sum_{i=1}^{n} x_i\right)$$