Resepti (Posteriorijakauman 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.

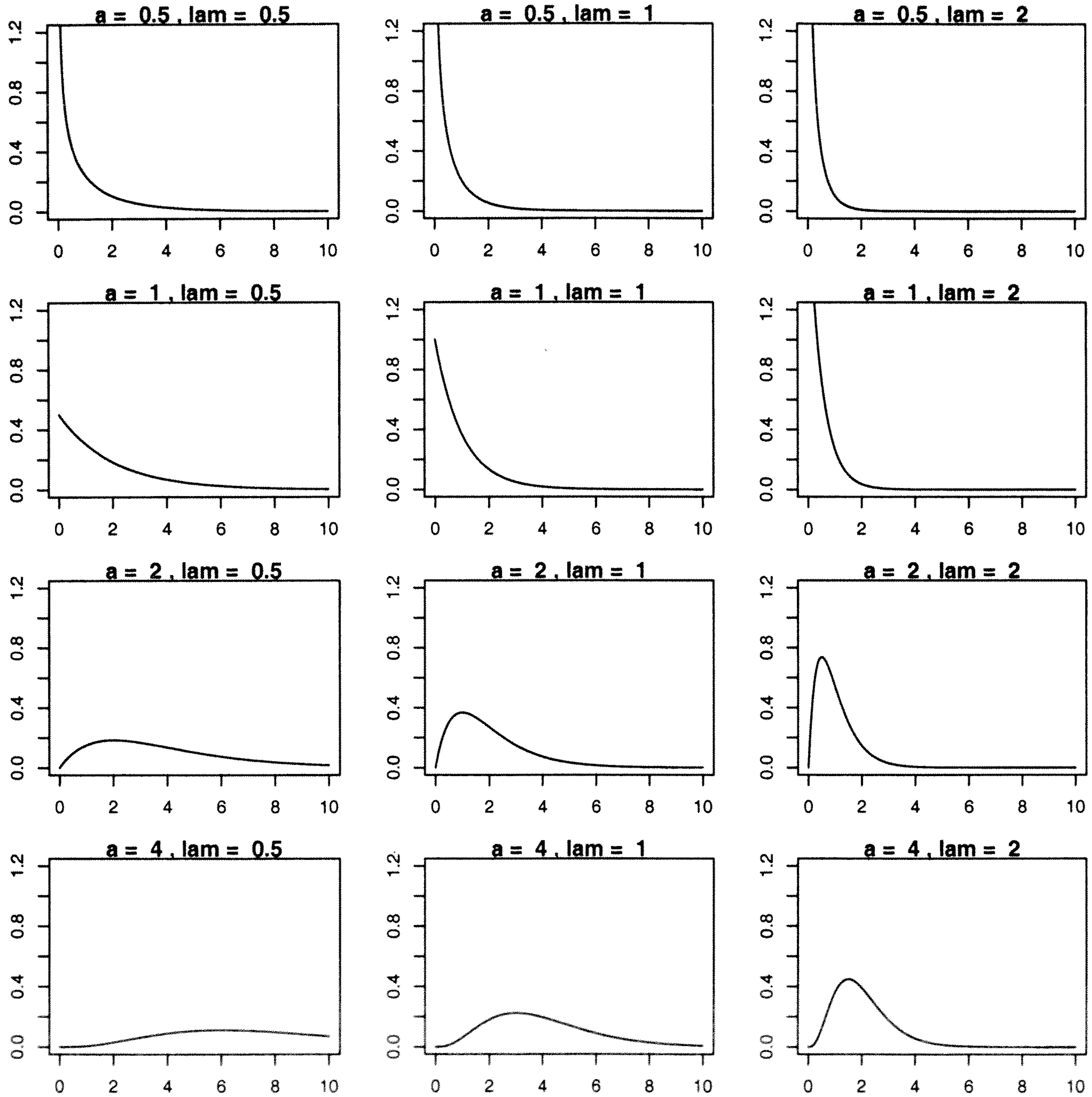
Usein esiintyviä liittoprioreita:

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

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

- normaalijakauma (joissakin erikoistapauksissa)

Gammajakauman tiheysfunktioita eri parametriarvoilla. 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

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

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

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

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

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

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

Tuttu $\Rightarrow$ posteriorijakauma on

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

[Vrt. AS, jakso 2.7]

16. Posteriorijakauma tunnistettu – mitä seuraavaksi?

Posteriorijakauma on täydellinen kuvaus epävarmuudelle havainnon jälkeen. Jos se kuuluu johonkin yleisesti tunnettuun parametriseen perheeseen, riittää kertoa jakaumaperheen nimi sekä posteriorin parametrit ko. perheessä (näitä voidaan kutsua hyperparametreiksi jotta ne eivät menisi sekaisin tilastollisen mallin parametri(e)n kanssa).

Usein kuitenkin posteriori tahdotaan kuvailla joidenkin tunnuslukujen avulla. Jos g on parametriavaruudessa määritelty funktio, niin $g(\theta)$:n <u>posterioriodotusarvo</u> on

$$E[g(\theta) \mid \underline{X} = \underline{x}] \quad \left[\overset{\text{pedanttisemmin}}{=} E[g(\tilde{\theta}) \mid \underline{X} = \underline{x}] \right]$$

$$= \int g(\theta)\, p(\theta \mid \underline{x})\, d\theta$$

(integraalin tilalla on summa, jos parametri on diskreetti; integrointijoukko ≡ parametriavaruus; tässä $p(\theta \mid \underline{x})$ on posteriorijakauman tf). Erityisesti parametrin <u>posterioriodotusarvo</u> on

$$E[\theta \mid \underline{X} = \underline{x}] \quad [= E[\tilde{\theta} \mid \underline{X} = \underline{x}]] = \int \theta \, p(\theta \mid \underline{x})\, d\theta$$

Valinnalla $g(\theta) = \theta^2$ saadaan posteriorijakauman toinen momentti $E[\theta^2 \mid \underline{X} = \underline{x}]$, ja näistä saadaan posteriorijakauman varianssi eli <u>posteriorivarianssi</u>

$$\mathrm{Var}(\theta \mid \underline{X} = \underline{x}) \quad [= \mathrm{Var}(\tilde{\theta} \mid \underline{X} = \underline{x})]$$

$$= E[\theta^2 \mid \underline{X} = \underline{x}] - \{E[\theta \mid \underline{X} = \underline{x}]\}^2$$

Posteriorijakauman sijaintia kuvaileviatunnuslukuja:

- posteriorodotusarvo (engl. posterior mean, posterior expectation) lienee yleisin
- posteriorimoodia (eli posteriorijakauman moodia) voidaan käyttää silloin, kun posteriorijakauman tiheys on yksihuippuinen. Englanniksi käytetään nimityksiä posterior mode tai maximum a posteriori estimate, MAP estimate.

Normaalijakaumalle odotusarvo = moodi. Muille jakaumille näin ei välttämättä ole:

Beta(α,β): odotusarvo $= \frac{\alpha}{\alpha+\beta}$; moodi $= \frac{\alpha-1}{\alpha+\beta-2}$ (kun $\alpha,\beta>1$)

Gamma(α,λ): odotusarvo $= \frac{\alpha}{\lambda}$; moodi $= \frac{\alpha-1}{\lambda}$ (kun $\alpha>1$)

Posteriorijakauman hajontaa kuvaillaan useimmiten joko posteriorivarianssilla tai tämän luvun neliöjuurella eli posteriorikeskihajonnalla.

Bayesläisessä päättelyssä voidaan laskea myös muotoa

$$P(\theta \in A \mid \underline{X} = \underline{x}) \left[= P(\tilde{\theta} \in A \mid \underline{X} = \underline{x})\right]$$
$$= \int_A p(\theta \mid \underline{x})\, d\theta$$

olevia todennäköisyyksiä (joita frekventistisen paradigman puitteissa ei saa edes ajatella!). Voidaan suoraan kysyä (posteriori-)todennäköisyyttä sille, että $\theta > \theta_1$ tai $\theta_1 < \theta < \theta_2$. Voidaan myös etsiä <u>posteriorivälejä</u> eli <u>bayesläisiä luottamusvälejä</u> ts. etsiä luvut θ_1 ja θ_2 siten, että

$$P(\theta_1 \le \tilde{\theta} \le \theta_2 \mid \underline{X} = \underline{x}) = 1-\alpha,$$

jossa α on annettu luku, esim. 0.05.
[engl. <u>credible interval</u>, posterior (probability) interval, Bayesian confidence interval, ...]

Tällaisen välin päätepisteet voidaan valita esim. niin, että posteriorijakauman molempiin häntiin jää saman verran todennäköisyysmassaa, eli päätepisteet θ_1 ja θ_2 voidaan määrätä ehdosta

$$P(\theta < \theta_1 \mid \underline{X} = \underline{x}) = \frac{\alpha}{2} \quad \text{ja} \quad P(\theta > \theta_2 \mid \underline{X} = \underline{x}) = \frac{\alpha}{2}.$$

Nämä pisteet ovat hyvin määriteltyjä, mikäli posteriorijakauma on jatkuva jakauma (mikä on päättelyssä tyypillisin tilanne). Esim. beta-, normaali- tai gammajakaumalle nämä pisteet voidaan laskea sopivalla tietokoneohjelmalla. [Esim. R-tilasto-ohjelmistossa funktioilla qbeta(), qnorm(), qgamma().]

Varoitus: termi "luottamusväli" (ilman lisämääreitä) on frekventistisen tilastotieteen käsite. Usea tilastotieteen soveltaja ajattelee, että luottamustasolla $(1-\alpha)$ laskettu luottamusväli sisältää parametrin arvon todennäköisyydellä $(1-\alpha)$, mikä on <u>täysin väärä tulkinta</u>; aineistosta laskettu luottamusväli joko sisältää tai ei sisällä parametrin arvoa, mutta jos luottamusvälejä laskettaisiin uusille, samasta tilanteesta kerätyille aineistoille, niin suhteellinen osuus niille toistoille, joilla luottamusväli sisältää parametrin olisi noin $(1-\alpha)$ [tämä on oikea tulkinta].

Bayesläisen lähestymistavan väliestimaatin oikea tulkinta on sitä vastoin juuri se, että väli sisältää parametrin arvon todennäköisyydellä $(1-\alpha)$, kun todennäköisyyden ymmärretään tarkoittavan ehdollista todennäköisyyttä sen jälkeen, kun aineisto on havaittu.

[AS, jakso 2.7 s. 38-39]

17. Odotusarvon estimointi Bayes-menetelmällä – normaalijakautunut otos ja varianssi tunnettu

Käsitellään vielä yksi esimerkki liittojakaumiin perustuvasta päättelystä. Oletamme, että havaintoja vastaavat satunnaismuuttujat ovat ehdollisesti riippumattomia ehdolla parametri ja että
$X_i \sim N(\theta, \sigma^2)$, $i=1,\dots,n$ (ehdolla θ).
Priorijakaumaksi otamme normaalijakauman $N(\mu_0, \sigma_0^2)$. Posteriorijakauman tiheysfunktio on

$$p(\theta \mid \underline{x}) \propto p(\theta) \, f_{\underline{X}}(\underline{x} \mid \theta)$$
$$\left[= \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{1}{2}\frac{(\theta-\mu_0)^2}{\sigma_0^2}\right) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right)\right]$$
$$\propto \exp\left(-\frac{1}{2} q(\theta)\right)$$

jossa eksponenttifunktion sisällä oleva θ:n kvadraattinen funktio on

$$q(\theta) = \frac{(\theta-\mu_0)^2}{\sigma_0^2} + \sum_{i=1}^{n} \frac{(\theta - x_i)^2}{\sigma^2}$$
$$= \theta^2 \left[\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right]$$
$$- 2\theta \left[\frac{\mu_0}{\sigma_0^2} + \frac{1}{\sigma^2} \Sigma x_i \right] + c_1,$$

jossa c_1 on θ:n funktiona ajateltuna vakio (se tietenkin riippuu aineistosta sekä vakioista $\mu_0, \sigma_0^2, \sigma^2$). Huomaa, että tästä additiivisesta vakiosta tulee multiplikatiivinen eksponenttifunktion soveltamisen jälkeen, ts.

$$\exp\left(-\frac{1}{2} q(\theta)\right) = \exp\left(-\frac{1}{2} (g(\theta) + c_1)\right)$$
$$= e^{-\frac{1}{2}c_1} \exp\left(-\frac{1}{2} g(\theta)\right) \propto \exp\left(-\frac{1}{2} g(\theta)\right)$$

Toisaalta, normaalijakauman $N(\mu_1, \sigma_1^2)$ tf on, θ:in funktiona ilmaistuna

$$\frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{1}{2}\frac{(\theta-\mu_1)^2}{\sigma_1^2}\right)$$

$$\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_1^2}\theta^2 - 2\frac{\mu_1}{\sigma_1^2}\theta\right]\right)$$

(jossa jätettiin kirjoittamatta θ:sta riippumaton vakiotermi μ_1^2/σ_1^2, eksponenttifunktion sisällä).

Kun tätä tulosta verrataan posteriorijakauman esitykseen polynomin $q(\theta)$ avulla, huomataan että posteriorijakauma on normaalijakauma $N(\mu_1, \sigma_1^2)$, jossa

$$\begin{cases} \dfrac{1}{\sigma_1^2} = \dfrac{1}{\sigma_0^2} + \dfrac{n}{\sigma^2} \\ \dfrac{\mu_1}{\sigma_1^2} = \dfrac{\mu_0}{\sigma_0^2} + \dfrac{1}{\sigma^2}\sum_{i=1}^{n} x_i \end{cases}$$

Tästä lasketaan helposti ensin $\frac{1}{\sigma_1^2}$, sitten σ_1^2 ja lopulta

$$\mu_1 = \sigma_1^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{1}{\sigma^2}\sum x_i\right).$$

Huomautus: Bayesläisen tilastotieteen johdoissa kannattaa usein ajatella, että normaalijakauman parametrit ovat odotusarvo μ sekä ns. tarkkuus, joka määritellään varianssin käänteislukuna, $\psi = \frac{1}{\sigma^2}$. Päivityskaava priorijakauman ja uskottavuusfunktion tarkkuusparametreille on yksinkertainen:

$$\psi_1 = \psi_0 + n\psi \qquad \left[\psi_1 = \frac{1}{\sigma_1^2},\ \psi_0 = \frac{1}{\sigma_0^2},\ \psi = \frac{1}{\sigma^2}\right]$$

Varianssiparametreille saadaan hankalampi kaava

$$\sigma_1^2 = \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{n}{\sigma^2}}$$

[Toisin kuin varianssille, tarkkuusparametrille ei ole yhtä yleisesti sovittua merkintää.]

18. Epäaito priori [AS, s. 40–41]

Jos edellisen jakson tilanteessa annetaan priorijakauman tarkkuuden $\psi_0 = \frac{1}{\sigma_0^2}$ lähestyä nollaa (eli varianssin σ_0^2 lähestyä ääretöntä), niin rajalla posteriorijakaumaksi saadaan

$$N(\bar{x}, \tfrac{1}{n}\sigma^2) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Tämä on siis rajalla saatava θ:in posteriorijakauma, kun havainnot on annettu; [frekventistisessä tilastotieteessä keskiarvon otantajakauma
$\bar{X} \sim N(\theta, \frac{1}{n}\sigma^2)$, kun $X_1, \ldots, X_n \perp\!\!\!\perp \sim N(\theta, \sigma^2)$.]

Samaan tulokseen päädyttäisiin ottamalla (formaalisti) priorijakauman lausekkeeksi $h(\theta) \equiv 1$ ja soveltamalla Bayesin kaavaa (ts. kertolaskukaavaa). Nyt kuitenkin

$$\int_{-\infty}^{\infty} h(\theta)\, d\theta = \infty$$

eikä funktio $h(\theta)$ tämän takia määrää mitään reaaliakselin todennäköisyysjakaumaa.

Varsinkin aikaisemmin suositeltiin tällaisten epäaitojen priorien (engl. improper prior) käyttöä. Niihin liittyy se ongelma, että epäaidon priorin avulla laskettu posteriori saattaa joskus olla epäaito (ts. myös sen integraali saattaa olla ääretön), jolloin se ei määrää mitään todennäköisyysjakaumaa.

Mm. tästä syystä epäaitojen priorien käyttö on vähenemässä, eikä niiden käyttöä voi suositella ainakaan vasta-alkajille.

19. Lisää ennustamisesta

Havainnot: $\underline{X} = (X_1, \dots, X_n)$ (satunnaismuuttujina)
$\underline{x} = (x_1, \dots, x_n)$ (havaitut arvot)

Tulevaisuudessa tehtävä uusi havainto: X_{n+1} (satunnaismuuttuja). Mitä voimme sanoa X_{n+1}:n jakaumasta havaintojen perusteella?

Bayesläisessä tilastotieteessä kysymykselle on yksikäsitteinen ratkaisu: X_{n+1}:n jakauma havaintojen jälkeen on sen ennustejakauma,

$$f_{X_{n+1}}(x_{n+1} \mid \underline{X} = \underline{x}) = \int f_{\tilde{\theta}, X_{n+1}}(\theta, x_{n+1} \mid \underline{X} = \underline{x})\, d\theta$$

$$= \int \underbrace{f_{\tilde{\theta}}(\theta \mid \underline{X} = \underline{x})}_{\text{posteriori} \equiv p(\theta \mid \underline{x})} f_{X_{n+1}}(x_{n+1} \mid \tilde{\theta} = \theta, \underline{X} = \underline{x})\, d\theta$$

Huomaa, että parametrin arvoon liittyvä epävarmuus otetaan huomioon integroimalla θ pois parametrin ja uuden havainnon yhteisjakaumasta ehdolla havainnot [näin siirrytään yhteisjakaumasta reunajakaumaan!].

Usein oletetaan, että $\underline{X} \perp\!\!\!\perp X_{n+1} \mid \theta$ eli että uusi havainto ja jo tehdyt havainnot ovat ehdollisesti riippumattomia. Tällöin

$$f_{X_{n+1}}(x_{n+1} \mid \tilde{\theta} = \theta, \underline{X} = \underline{x}) = f_{X_{n+1}}(x_{n+1} \mid \tilde{\theta} = \theta),$$

ja viimeinen jakauma on tunnettu tehtävän spesifikaation perusteella.

Jos $X_{n+1} \perp\!\!\!\perp \underline{X} \mid \theta$, niin on siis

$$f_{X_{n+1}}(x_{n+1} \mid \underline{X} = \underline{x}) = \int p(\theta \mid \underline{x})\, f_{X_{n+1}}(x_{n+1} \mid \theta)\, d\theta.$$

Tämän integraalin voi usein laskea liittojakaumien tilanteessa. Ennustejakauman momentteja voi usein laskea helposti seuraavalla periaatteella. Ennustejakauman odotusarvo eli ehdollinen odotusarvo $E[X_{n+1} \mid \underline{X} = \underline{x}]$ voidaan laskea seuraavalla kaavalla (kun X_{n+1}:llä on diskreetti jakauma ja $X_{n+1} \perp\!\!\!\perp \underline{X} \mid \theta$):

$$E[X_{n+1} \mid \underline{X} = \underline{x}] = \sum_{x_{n+1}} x_{n+1}\, f_{X_{n+1}}(x_{n+1} \mid \underline{X} = \underline{x})$$

$$= \sum_{x_{n+1}} x_{n+1} \int p(\theta \mid \underline{x})\, f_{X_{n+1}}(x_{n+1} \mid \theta)\, d\theta$$

$$= \int p(\theta \mid \underline{x}) \underbrace{\sum_{x_{n+1}} x_{n+1}\, f_{X_{n+1}}(x_{n+1} \mid \theta)}_{E[X_{n+1} \mid \theta]}\, d\theta$$

(tunnetaan käytännössä)

(Tässä pitää jotenkin pystyä perustelemaan summauksen ja integroinnin järjestyksen vaihto, mutta emme jää nyt pohtimaan tätä asiaa.)

Korkeampia momentteja saadaan laskettua samalla tavalla, esim.

$$E[X_{n+1}^2 \mid \underline{X} = \underline{x}] = \int p(\theta \mid \underline{x}) \underbrace{E[X_{n+1}^2 \mid \theta]}_{\text{tyypillisesti tunnetaan}}\, d\theta$$

jolloin ennustejakauman varianssi saadaan selville (mikäli integraalit saadaan laskettua), sillä

$$\mathrm{Var}(X_{n+1} \mid \underline{X} = \underline{x}) = E[X_{n+1}^2 \mid \underline{X} = \underline{x}] - \{E[X_{n+1} \mid \underline{X} = \underline{x}]\}^2$$

20. Bayes-päättelyn laskentaa

Liittoperheillä bayesläinen päättely onnistuu tiettyyn rajaan asti, mutta jos jotakin mallissa muutetaan, niin tyypillisesti liitto-ominaisuus katoaa.

Nykyään Bayes-päättelyn laskentamenetelmät perustuvat tyypillisesti satunnaisuuden simulointiin tietokoneella eli ns. Monte Carlo -menetelmien käyttöön. Ne perustuvat sille ajatukselle, että koska posteriorijakauma on todennäköisyysjakauma, niin sitä voidaan (yrittää) simuloida, eli tietokoneella voidaan laskea arvoja $t_1, t_2, \ldots, t_N$ (N voi olla hyvin suuri), joita voidaan käytännössä pitää sellaisten satunnaismuuttujien $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_N$ havaittuina arvoina, joista kukin on jakautunut posteriorijakauman mukaisesti. Tällaisia simulointimenetelmiä käsitellään esim. Laskennallisen tilastotieteen kursseilla. Tämän jälkeen otosta $t_1, \ldots, t_N$ käsitellään data-analyysin keinoilla: esim. posterioriodotusarvoja voidaan arvioida kaavoilla

$$E[\tilde{\theta} \mid \underline{X} = \underline{x}] \approx \frac{1}{N} \sum_{i=1}^{N} t_i$$

$$E[g(\tilde{\theta}) \mid \underline{X} = \underline{x}] \approx \frac{1}{N} \sum_{i=1}^{N} g(t_i),$$

posteriorijakauman tiheysfunktiota voidaan arvioida histogrammin tai kehittyneempien tiheysfunktion estimointimenetelmien avulla jne. Tällaiset menetelmät ovat suhteellisen uusia, ja niiden ansiosta voidaan käsitellä lähes mielivaltaisen monimutkaisia bayesläisiä tilastollisia malleja.