Luku 11

Piste-estimointi

Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka tiheysfunktio on $f(x, \theta)$. Estimoimme jakauman tunnuslukua θ jollakin otoksen tunnusluvulla T. Koska T on otoksen funktio, täydellisempi merkintä olisi $T(X_1, X_2, \ldots, X_n)$. Jos halutaan korostaa, että kyseessä on nimenomaan θ:n estimaattori, merkitään $\hat{\theta}, \tilde{\theta}$ jne. Havaitusta otoksesta $x_1, x_2, \ldots, x_n$ laskettua estimaattorin T arvoa merkitään pienellä kirjaimella, eli $T(x_1, x_2, \ldots, x_n) = t$, ja estimaattorin arvoa kutsutaan estimaatiksi. Sen sijaan $\hat{\theta}$ on merkintä sekä estimaattorille että estimaatille. Olemme edellä jo tarkastelleet useita estimaattoreita. Tavanomaisia odotusarvon μ ja varianssin σ^2 estimaattoreita ovat otoskeskiarvo $\overline{X}$ ja otosvarianssi S^2.

11.1 Piste-estimaattoreiden ominaisuuksia

Luvuissa 8, 9 ja 10 olemme jo tarkastelleet useita estimaattoreiden ominaisuuksia, kuten harhattomuus, minivarianssisuus ja tarkentuvuus. Tässä luvussa kootaan yhteen ja täydennetään edellä esitettyjä tuloksia.

11.1.1 Harhattomuus

Parametrin θ estimaattorin $\hat{\theta}$ hyvyyttä on luontevaa mitata tarkastelemalla poikkeamaa $\hat{\theta} - \theta$. Tässä siis θ on tuntematon kiinteä parametrin arvo ja $\hat{\theta}$ on satunnasimuuttuja. Poikkeamat voivat olla tietysti positiivisia tai negatiivisia, mutta neliöpoikkeama $(\hat{\theta} - \theta)^2 \geq 0$ mittaa poikkeaman suuruutta. Neliöpoikkeaman odotusarvo

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \tag{11.1.1}$$

on keskineliövirhe. Jos estimaattorilla on pieni keskineliövirhe, estimaattori antaa keskimäärin lähellä θ:aa olevia arvoja. Määritelmästä (11.1.1) seuraa suoraan, että

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2,$$

missä $E(\hat{\theta}) - \theta$ on estimaattorin $\hat{\theta}$ harha.

Määritelmä 11.1 Estimaattori $\hat{\theta}$ on parametrin *harhaton estimaattori*, jos $E(\hat{\theta}) = \theta$ kaikilla θ:n arvoilla. Muutoin $\hat{\theta}$ on harhainen ja $[\text{harha}(\hat{\theta})]^2 > 0$ jollakin θ:n arvolla.

Olemme jo aikaisemmin Esimerkissä 9.4 osoittaneet, että otoskeskiarvo $\hat{\mu} = \overline{X}$ ja otosvarianssi $\hat{\sigma}^2 = S^2$ ovat harhattomia estimaattoreita. Sen sijaan esimerkiksi otosvarianssin neliöjuuri $\hat{\sigma} = \sqrt{S^2}$ on σ:n harhainen estimaattori.

Esimerkki 11.1 Olkoon $X_1, \ldots, X_n$ otos Bernoullin jakaumasta $\text{Ber}(\pi)$, $0 < \pi < 1$. Voidaanko π^2 estimoida harhattomasti yhden havainnon perusteella? Oletetaan, että T olisi π^2:n harhatton estimaattori. Silloin

$$\pi^2 = E[T(X)] = \pi T(1) + (1-\pi)T(0) = T(0) + \pi[T(1) - T(0)], \tag{11.1.2}$$

kaikilla $\pi \in (0,1)$. Koska (11.1.2) pitää paikkansa kaikilla $\pi \in (0,1)$, niin yhtälön molemmilla puolilla täytyy π:n potenssien vastaavien kertoimien olla samat. Tämä on mahdotonta, joten π^2:ta ei voida estimoida harhattomasti yhden havainnon perusteella.

Jos $n > 1$, voisimme kokeilla estimaattoria $\bar{X}_n^2$, missä $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ on otoskeskiarvo. Tämä estimaattori ei ole kuitenkaan harhaton. Sen sijaan $\bar{X}_n(n\bar{X}_n - 1)/(n-1)$ on π^2:n harhaton estimaattori, kun $n > 1$. □

Esimerkki 11.2 Olkoon $X_1, \ldots, X_n$ otos Bernoullin jakaumasta $\text{Ber}(\pi)$, $0 < \pi < 1$ kuten Esimerkissä 11.1. Joskus halutaan estimoida $1/\pi$. Luonnollinen estimaattoriehdokas on $1/\bar{X}_n = n/\sum_{i=1}^n X_i$, missä $\sum_{i=1}^n X_i \sim \text{Bin}(n, \pi)$. Koska

$$P(\sum_{i=1}^n X_i = 0) = (1-\pi)^n > 0,$$

$E(1/\bar{X}) = \infty$ kaikilla $n \geq 1$. Sen sijaan estimaattorilla

$$T_n = \frac{n+1}{n\bar{X}+1}$$

on äärellinen odotusarvo, mutta estimaattori on harhainen. □

Harhattomuus on haluttava ominaisuus, mutta ei aina kovin oleellinen.

11.1.2 Tehokkuus

Eräs tapa arvioida estimaattorien hyvyyttä on vertailla niiden keskineliövirheitä. Olkoot $\hat{\theta}_1$ ja $\hat{\theta}_2$ kaksi parametrin θ estimaattoria. Estimaattorin $\hat{\theta}_1$ tehokkuus suhteessa estimaattorin $\hat{\theta}_2$ tehokkuuteet määritellään suhteena

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{MSE}(\hat{\theta}_2)}{\text{MSE}(\hat{\theta}_1)} \tag{11.1.3}$$

ja sitä kutsutaan estimaattoreiden suhteelliseksi tehokkuudeksi. Kun estimaattorit ovat harhattomia, suhteellinen tehokkuus on varianssien suhde.

Jos estimaattori $\hat{\theta}$ on harhaton, voidaan sen varianssille $\mathrm{Var}(\hat{\theta})$, johtaa alaraja, joka on eräänlainen mittatikku estimaattorin hyvyydelle.

Oletetaan, että X noudattaa jakaumaa, jonka tiheysfunktio on $f(x;\theta)$. Fisherin pistefunktio on

$$S(\theta;x) = \frac{\partial\, l(\theta;x)}{\partial\theta}, \tag{11.1.4}$$

missä $l(\theta;x) = \log f(x;\theta)$. Tarkastellaan nyt pistefunktiota S "satunnaisfunktiona" eli

$$S(\theta;X) = \frac{\partial\, l(\theta;X)}{\partial\theta}.$$

Kun havaitaan $X = x$, saadaan funktio (11.1.4). Eri X:n arvoilla saadaan aina eri funktio.

Esimerkki 11.3 Oletetaan, että $X \sim \mathrm{N}(\mu,\sigma^2)$. Silloin μ:n logaritmoitu uskottavuusfunktio on (σ^2 voidaan ajatella vakioksi)

$$l(\mu;X) = -\frac{1}{2\sigma^2}(X-\mu)^2$$

ja

$$S(\mu;X) = \frac{1}{\sigma^2}(X-\mu).$$

Silloin

$$\begin{aligned} E[S(\mu;X)] &= E[(X-\mu)/\sigma^2] = \frac{1}{\sigma^2}\,E(X-\mu) \\ &= \frac{1}{\sigma^2}[E(X)-\mu] = 0. \end{aligned}$$

ja

$$\begin{aligned} \mathrm{Var}[S(\mu;X)] &= \mathrm{Var}\left[\frac{1}{\sigma^2}(X-\mu)\right] \\ &= \frac{1}{(\sigma^2)^2}\,\mathrm{Var}(X-\mu) = \frac{\sigma^2}{(\sigma^2)^2} = \frac{1}{\sigma^2}. \end{aligned}$$

Toisaalta

$$I(\mu;X) = -\frac{\partial^2\, l(\mu;X)}{\partial\mu^2} = -\frac{\partial S(X;\mu)}{\partial\mu} = \frac{1}{\sigma^2},$$

joten tässä tapauksessa havaittu informaatio $I(\mu;X)$ ei riipu havainnoista ja odotettu informaatio on sama kuin havaittu informaatio:

$$E[I(\mu;X)] = \frac{1}{\sigma^2}.$$

Huomaamme, että pistesuureen $S(\mu;X)$ varianssi on odotettu informaatio:

$$\mathrm{Var}[S(\mu;X)] = E[I(\mu;X)].$$

Koska $\hat{\mu} = X$ on μ:n suurimman uskottavuuden estimaattori, niin

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{E[I(\mu; X)]}.$$

□

Esimerkki 11.4 Oletetaan, että $X \sim \mathrm{Bin}(n, \theta)$. Silloin θ:n logaritmoitu uskottavuusfunktio on

$$l(\theta; X) = X \log \theta + (n - X) \log(1 - \theta), \quad 0 < \theta < 1,$$

ja pistefunktio

$$S(\theta; X) = \frac{X}{\theta} - \frac{n - X}{1 - \theta} = \frac{X - n\theta}{\theta(1 - \theta)}.$$

Helposti todetaan, että $E[S(\theta; X)] = 0$ ja

$$\mathrm{Var}[S(\theta; X)] = \frac{n}{\theta(1 - \theta)}.$$

Havaittu informaatio on

$$I(\theta; X) = \frac{X}{\theta^2} + \frac{n - X}{(1 - \theta)^2},$$

josta saadaan odotettu informaatio

$$E[I(\theta; X)] = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1 - \theta)^2} = \frac{n}{\theta(1 - \theta)}.$$

Jälleen pistesuureen $S(\theta; X)$ varianssi, odotettu informaatio ja suurimman uskottavuuden estimaattorin $\hat{\theta} = X/n$ varianssi liittyvät yhteen seuraavasti:

$$\mathrm{Var}[S(\theta; X)] = E[I(\theta; X)]$$

ja

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{E[I(\theta; X)]}.$$

□

Esimerkki 11.5 Kun X noudattaa Poissonin jakaumaa $\mathrm{Poi}(\mu)$, niin

$$l(\mu; X) = X \log \mu - \mu, \quad S(\mu; X) = \frac{X}{\mu} - 1 \quad \text{ja} \quad \frac{\partial^2 \, l(\mu; X)}{\partial \mu^2} = -\frac{X}{\mu^2}.$$

Nyt siis $I(\mu; X) = \frac{X}{\mu^2}$ ja odotettu informaatio on siis

$$E[I(\mu; X)] = E\left(\frac{X}{\mu^2}\right) = \frac{E(X)}{\mu^2} = \frac{1}{\mu}$$

ja välittömästi voidaan myös todeta, että myös

$$\mathrm{Var}[S(\mu; X)] = \frac{1}{\mu}.$$

Parametrin μ suurimman uskottavuuden estimaattori $\hat{\mu} = X$ on ja

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{E[I(\mu; X)]} = \mu.$$

□

Fisherin informaatio on

$$\mathcal{I}(\theta) = E[I(\theta; X)]. \tag{11.1.5}$$

Kun meillä on havainto $X = x$, voidaan laskea havaittu informaatio $I(\theta; x)$ annetulla parametrin arvolla. Suurimman uskottavuuden estimaatti antaa uskottavimman parametrin arvon $\hat{\theta}$ ja $I(\hat{\theta}; x)$ antaa tarkkuuden, millä θ on estimoitu. Informaation määrä riippuu tyypillisesti otoksen koosta, joka on tärkeä näkökohta suunniteltaessa koetta tai aineiston hankintaa. Ennen koetta on hyödyllistä arvioida havaintojen antama informaation määrä. Suunnitteluvaiheessa ei voida laskea havaittua informaatiota, koska havaintoja ei vielä ole. Sen sijaan odotettu informaatio (11.1.5) voidaan laskea. Odotettu informaatio mittaa θ:n estimoinnin keskimääräistä tarkkuutta, joka voidaan saavuttaa toistamalla koetta tai tekemällä toistuvia otoksia.

Olkoon $X_1, \ldots, X_n$ otos jakaumasta, jonka tiheysfunktio (tai todennäköisyysfunktio) on $f(x; \theta)$. Suoraan pistefunktion määritelmästä seuraa, että

$$S(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} S(\theta; X_i),$$

missä $S(\theta; X_i) = \frac{\partial\, l(\theta; X_i)}{\partial \theta}$. Vastaavasti otokseen $X_1, \ldots, X_n$ perustuva informaatiofunktio voidaan esittää summana

$$I(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} I(\theta, X_i),$$

missä $I(\theta; X_i) = -\frac{\partial^2\, l(\theta; X_i)}{\partial \theta^2}$. Koska havainnot X_i, $1 \le i \le n$, noudattavat samaa jakaumaa, jokaisen havainnon antama odotettu informaatio $E[I(\theta, X_i)]$ on sama. Merkitään siksi yhden havainnon odotettua informaatiota

$$i(\theta) = E[I(\theta; X_i)], \tag{11.1.6}$$

missä $I(\theta; X_i) = -\frac{\partial^2\, l(\theta; X_i)}{\partial \theta^2}$. Nyt siis otokseen $X_1, \ldots, X_n$ perustuva odotettu informaatio on

$$\mathcal{I}(\theta) = \sum_{i=1}^{n} E[I(\theta, X_i) = n\, i(\theta). \tag{11.1.7}$$

Esimerkeissä 11.3, 11.4 ja 11.5 totesimme, että $\mathrm{Var}[S(\theta; X)] = i(\theta)$, joten

$$\mathrm{Var}[S(\theta; X_1, \ldots, X_n)] = n\, i(\theta). \tag{11.1.8}$$

Relaatio (11.1.8) voidaan todistaa yleisesti tiettyjen säännöllisyysehtojen vallitessa. Odotettu informaatio voidaan siis laskea joko odotusarvona (11.1.7) tai pistefunktion varianssina (11.1.8). Tavallisesti odotusarvo (11.1.7) on helpompi laskea kuin pistefunktion varianssi.

Alaluvussa 10.9.3 todistetun Cramérin ja Raon lauseen avulla voidaan määrittää harhattoman estimaattorin varianssin alaraja.

Cramérin ja Raon alaraja: Olkoon $\hat{\theta}$ parametrin θ harhaton estimaattori. Silloin tiettyjen säännöllisyysehtojen vallitessa

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{n\, i(\theta)}, \tag{11.1.9}$$

missä $i(\theta)$ on yhden havainnon antama odotettu informaatio.

Nyt määrittelemme *tehokkaan* eli *minimivarianssisen estimaattorin.*

Määritelmä 11.2 Jos $\hat{\theta}$ on parametrin θ harhaton estimaattori ja

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n\, i(\theta)},$$

niin $\hat{\theta}$ on θ:n *harhaton minimivarianssinen estimaattori.*

11.1.3 Tarkentuvuus

Eräs estimaattorille asetettava luonnollinen vaatimus on, että sen tarkkuus kasvaa kun otoskoko kasvaa. Tarkan estimaattorin arvot osuvat suurella todennäköisyydellä lähelle parametrin θ oikeata arvoa. *Tarkentvuvuus* sisältää tämän ajatuksen. Merkitään $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \ldots, X_n)$, missä $(\hat{\theta}_n; n \geq 1)$ on estimaattoreiden jono, missä otoskoko $n = 1, 2, \ldots$ kasvaa.

Määritelmä 11.3 Tunnusluku $\hat{\theta}_n$ on parametrin θ *tarkentuva estimaattori*, jos kaikilla $\varepsilon > 0$

$$P(|\hat{\theta}_n - \theta| \geq \varepsilon) \to 0,$$

kun otoskoko n kasvaa rajatta.

Määritelmä tarkoittaa siis sitä, että estimaattori $\hat{\theta}_n$ suppenee todennäköisyyden mielessä (stokastinen suppeneminen) kohti parametrin arvoa θ eli

$$\hat{\theta}_n \xrightarrow{P} \theta,$$

kun n kasvaa. Luvussa 8 esimerkiksi Lauseessa 9.17 esitettiin stokastista suppenemista koskevia tuloksia. Lauseen 9.17 kohdan *SS1* mukaan

$$\hat{\theta}_n \quad \text{on tarkentuva jos} \quad \mathrm{MSE}(\hat{\theta}_n) \to 0,$$

kun otoskoko n kasvaa rajatta. Tämä tarkentuvuuden ehto on helppo tarkistaa. Jos estimaattori on harhaton, tarkistetaan, meneekö estimaattorin varianssi nollaan otoskoon kasvaessa. Jos estimaattori ei ole harhaton, tarkistetaan lisäksi, meneekö myös estimaattorin harha nollaan otoskoon kasvaessa. Huomattakoon, että tavallisesti merkitsemme θ:n estimaattoria yksinkertaisesti $\hat{\theta}$, kun riippuvuutta n:stä ei tarvitse erityisesti korostaa.

11.2 Estimointimenetelmiä

Kahdessa edellisessä luvussa käsiteltiin suurimman uskottavuuden menetelmää. Lukijalle on tuttu myös pienimmän neliösumman menetelmä ja 9. luvussa mainittiin myös ns. robustitit estimaattorit. Seuraavassa alaluvussa käsitellään lyhyesti momenttimenetelmää.

11.2.1 Momenttimenetelmä

Momenttimenetelmän ideana on se, että asetetaan populaation momentit ja vastaavat otoksen momentit yhtä suuriksi. Olkoon $X_1, \ldots, X_n$ otos jakaumasta, jolla on olemassa riittävä määrä momentteja. Momentit riippuvat estimoitavista tuntemattomista parametreista, joita merkitään $\theta_1, \ldots, \theta_k$. Estimoitavia parametreja on siis k kappaletta. Merkitään otokosesta laskettua r. momenttia m_r ja se on

$$m_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

Vastaava populaation r. momentti on

$$\alpha_r = E(X^r),$$

missä X:n tiheysfunktio on $f(x; \theta_1, \ldots, \theta_k)$.

Merkitään momenttiestimaattoreita $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$. Kun asetetaan otosmomentit ja populaation momentit yhtä suuriksi, saadaan yhtälöryhmä

$$\begin{aligned} \alpha_1(\theta_1, \ldots, \theta_k) &= m_1 \\ \alpha_2(\theta_1, \ldots, \theta_k) &= m_2 \\ &\vdots \\ \alpha_k(\theta_1, \ldots, \theta_k) &= m_k, \end{aligned}$$

joka ratkaistaan parametrien $\theta_1, \ldots, \theta_k$ suhteen. Menetelmä on yksinkertainen ja se antaa tarkentuvia estimaattoreita.

Esimerkki 11.6 Olkoon $X_1, \ldots, X_n$ otos normaalijakumasta $\mathrm{N}(\mu, \sigma^2)$. Määritetään parametrien μ ja σ^2 momenttiestimaattorit. Koska otosmomentit ovat

$$m_1 = \bar{X} \quad \text{ja} \quad m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

ja vastaavat populaation momentit

$$\alpha_1 = \mu \quad \text{ja} \quad \alpha_2 = \sigma^2 + \mu^2,$$

saadaan yhtälöryhmä

$$\begin{aligned} \mu &= \bar{X} \\ \sigma^2 + \mu^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2. \end{aligned}$$

Tästä ratkaisemalla saadaan momenttiestimaattorit

$$\begin{aligned} \tilde{\mu} &= \bar{X} \\ \tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2. \end{aligned}$$

□

11.2.2 Bayesin menetelmä

Bayesilaisessa lähestymistavassa parametria θ käsitellään satunnaismuuttujana. Tuntematonta parametria θ koskeva epävarmuus esritetään priorijakauman $\pi(\theta)$ avulla. Kun havainto on saatu, parametrin jakauma päivitetään Bayesin kaavan avulla. Näin saadaan ns. posteriorijakauma

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{g(x)}, \tag{11.2.1}$$

missä $g(x)$ on X:n reunajakauman tiheysfunktio (tai todennäköisyysfunktio). Reunajakauma lasketaan kaavalla

$$g(x) = \begin{cases} \sum_\theta f(x|\theta)\pi(\theta), & \text{diskreetissä tapauksessa;} \\ \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\,\theta, & \text{jatkuvassa tapauksessa.} \end{cases}$$

Kun minimoidaan odotettu kvadraattinen tappio (riski)

$$E(b-\theta)^2 = \int (b-\theta)^2\pi(\theta|x)d\,\theta$$

b:n suhteen, saadaan posteriorijakauman odotusarvo (keskiarvo)

$$b = \int \theta\pi(\theta|x)d\,\theta.$$

Bayesin estimaatiksi $\hat{\theta}_b$ valitaan posteriorijakauman odotusarvo. Jos valitaan toinen tappiofunktio, päädytään tavallisesti eri estimaattoriin. Bayesilainen menetelmä yhdistää siis informaatiota kertomalla priorin ja uskottavuusfunktion keskenään.

11.2.3 Suurimman uskottavuuden estimaattorin (SUE) ominaisuuksia

Edellä olemme käsitelleet useita suurimman uskottavuuden estimaattorin ominaisuuksia. Esitimme myös harhattaoman estimaattorin varianssille alarajan Fisherin informaation avulla (Cramérin ja Raon lause). Seuraavassa esitämme vielä suurimman uskottavuuden estimaattorin tärkeimmät ominaisuudet luettelonomaisesti.

1. Suurimman uskottavuuden estimaattorit eivät ole välttämättä harhattomia. Jos esimerkiksi $X_1, \ldots, X_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, niin σ^2:n SUE $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ on harhainen. Vaikka jotkut SUE:t ovat harhaisia, kaikki SUE:t ovat *tarkentuvia*, joten kaikki SUE:t ovat *asymptoottisesti harhattomia*.

2. Jos $\hat{\theta}$ on θ:n SUE ja g jokin funktio, niin silloin $g(\hat{\theta})$ on $g(\theta)$:n SUE. Tämä on SUE:n *invarinttisuusominaisuus* (Katso alaluku 10.7 ja Lause 10.1). Jos esimerkiksi $\bar{X}$ on θ:n SUE, niin $\bar{X}^2$ on θ^2:n SUE.

3. Oletetaan, että $f(x;\theta)$ toteuttaa tietyt säännöllisyysoletukset ja θ:lla on harhaton estimaattori $\tilde{\theta}$, joka saavuttaa Cramerin ja Raon rajan ($\tilde{\theta}$ on tehokas). Oletetaan, että θ:n SUE $\hat{\theta}$ on pisteyhtälön $S(\theta) = 0$ ratkaisu. Silloin $\tilde{\theta} = \hat{\theta}$. On kuitenkin huomattava, että kaikki suurimman uskottavuuden estimaattorit eivät ole tehokkaita. Kuitenkin, jos tehokas estimaattori on olemassa, se on SUE.

4. Kun $f(x;\theta)$ toteuttaa tietyt säännöllisyysoletukset, niin otokseen $X_1, \ldots, X_n$ perustuva θ:n suurimman uskottavuuden estimaattori $\hat{\theta}_n$ noudattaa asymptoottisesti normaalijakaumaa, jonka odotusarvo on θ ja varianssi on $1/\mathcal{I}(\theta)$. Jos siis $n \to \infty$, niin

   $$\hat{\theta}_n \sim \mathrm{N}(\theta, \frac{1}{\mathcal{I}(\theta)}). \tag{11.2.2}$$

   Huomattakoon, että $\mathcal{I}(\theta) = n\, i(\theta)$, missä $i(\theta)$ on yhden havainnon antama informaatio. Tulokseen (11.2.2) perustuvat monet suurten otosten testit ja luottamusvälit. Suurimman uskottavuuden estimaattorin $\hat{\theta}_n$ asymptoottinen varianssi siis saavuttaa Cramérin ja Raon alarajan. SUE $\hat{\theta}_n$ on siis *asymptoottisesti tehokas estimaattori.* Riittävän suurilla otoksilla normaalijakauma on hyvä $\hat{\theta}_n$:n jakauman likiarvo.

Voidaan osoittaa, että seuraavat tulokset pitävät paikkansa:

$$\begin{aligned}
\sqrt{\mathcal{I}(\theta)}(\hat{\theta}_n - \theta) &\xrightarrow{d} \mathrm{N}(0,1)\\
\sqrt{I(\theta)}(\hat{\theta}_n - \theta) &\xrightarrow{d} \mathrm{N}(0,1)\\
\sqrt{\mathcal{I}(\hat{\theta}_n)}(\hat{\theta}_n - \theta) &\xrightarrow{d} \mathrm{N}(0,1)\\
\sqrt{I(\hat{\theta}_n)}(\hat{\theta}_n - \theta) &\xrightarrow{d} \mathrm{N}(0,1).
\end{aligned}$$

Kaksi viimeistä tulosta ovat käytännössä tärkeimmät ja ne tarkoittavat sitä, että suurilla n:n arvoilla

$$\hat{\theta}_n \sim \mathrm{N}(\theta, \frac{1}{\mathcal{I}(\hat{\theta}_n)}) \tag{11.2.3}$$

$$\hat{\theta}_n \sim \mathrm{N}(\theta, \frac{1}{I(\hat{\theta}_n)}). \tag{11.2.4}$$

On suositeltavaa käyttää $\hat{\theta}_n$:n varianssin estimaattina lauseketta $1/I(\hat{\theta}_n)$. Tavallisesti kuitenkin käsittelemämme jakaumat kuuluvat eksponentiaaliseen perheeseen ja silloin $\mathcal{I}(\hat{\theta}_n) = I(\hat{\theta}_n)$

11.3 Delta-menetelmä

Määritelmä 11.4 Funktion $g(x)$ r. asteen *Taylorin polynomi* pisteessä a on

$$T_r(x) = g(a) + g'(a)(x-a) + \frac{g''(a)}{2!}(x-a)^2 + \cdots + \frac{g^{(r)}(a)}{r!}(x-a)^r, \tag{11.3.1}$$

missä $g^{(r)}(x) = \frac{\mathrm{d}^r}{\mathrm{d}x^r} g(x)$ on funktion $g(x)$ r. derivaatta.

Taylorin lauseen mukaan

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0, \tag{11.3.2}$$

jos $g^{(r)}(a)$ on olemassa. Funktio $g(x)$ voidaan lausua pisteen $x = a$ ympäristössä muodossa

$$g(x) = T_r(x) + R_{r+1}(x),$$

missä $R_{r+1}(x) = g(x) - T_r(x)$ on jäännöstermi, joka siis toteuttaa ehdon 11.3.2.

Oletetaan, että X on satunnaismuuttuja, jonka odotusarvo on $E(X) = \mu \neq 0$. Jos estimoidaan funktiota $g(\mu)$, niin sen Taylorin polynomiin perustuva 1. kertaluvun likiarvo pisteessä μ on

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu). \tag{11.3.3}$$

Jos käytetään $g(\mu)$:n estimaattorina funktiota $g(X)$, niin

$$E[g(X)] \approx g(\mu)$$

ja

$$\mathrm{Var}[g(X)] = [g'(\mu)]^2 \,\mathrm{Var}(X).$$

Esimerkki 11.7 Tarkastellaan odotusarvon $E(X) = \mu \neq 0$ funktion $g(\mu) = 1/\mu$ estimointia. Olkoon estimaattorina $1/X$. Silloin edellisen mukaan

$$E\left(\frac{1}{X}\right) \approx \frac{1}{\mu}$$

ja

$$\operatorname{Var}\left(\frac{1}{X}\right) \approx \left(\frac{1}{\mu}\right)^4 \operatorname{Var}(X).$$

$\square$

Lause 11.1 (Delta-menetelmä) *Olkoon $\{X_n\}$ sellainen satunnaismuuttujien jono, että $\sqrt{n}(X_n - \theta)$ lähenee jakaumamielessä normaalijakaumaa $\mathrm{N}(0, \sigma^2)$. Oletetaan, että annetulla funktiolla g on määrätyllä arvolla θ derivaatta $g'(\theta) \neq 0$. Silloin*

$$\sqrt{n}\,[g(X_n) - g(\theta)] \to \mathrm{N}\big(0, \sigma^2[g'(\theta)]^2\big)$$

jakaumamielessä.

Esimerkki 11.8 Olkoon $X_1, \ldots, X_n$ otos jakaumasta $\mathrm{Ber}(p)$. Onnistumisen todennäköisyyden p estimaattori on tavallisesti $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Onnistumisen mahdollisuus (odds) $p/(1-p)$ on vedonlyönnissä ja biostatiikassa tavanomainen parametri. Voimme käyttää $p/(1-p)$:n estimaattorina $\hat{p}$:n funktiota $\hat{p}/(1-\hat{p})$. Mitä voimme sanoa tämän estimaattorin ominaisuuksista? Nyt estimoidaan siis funktiota $g(p) = p/(1-p)$. Koska $g'(p) = 1/(1-p)^2$, niin lausekkeen 11.3.3 mukaan

$$\begin{aligned}\operatorname{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) &\approx [g'(p)]^2 \operatorname{Var}(\hat{p}) \\ &= \left[\frac{1}{(1-p)^2}\right]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}.\end{aligned}$$

$\square$

Lauseessa 11.1 oletettiin, että $\sqrt{n}(X_n - \theta) \xrightarrow{d} Z \sim \mathrm{N}(0, \sigma^2)$. Usein varianssi riippuu θ:sta eli $\sigma^2 = \sigma^2(\theta)$. Silloin lauseen 11.1 mukaan

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} \mathrm{N}(0, \sigma^2(\theta)[g'(\theta)]^2).$$

Jos meillä on esimerkiksi otos Poissonin jakaumasta $\mathrm{Poi}(\lambda)$, niin

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} \mathrm{N}(0, \lambda).$$

Esimerkki 11.9 Edellä mainitun Poissonin jakauman tapauksessa $\theta = \lambda$ ja $\sigma(\theta) = \sqrt{\lambda}$. Määritetään muunnos $g(\lambda)$ siten, että $g(\bar{X}_n)$:n varianssi on vakio. Olkoon c vakio ja valitaan g siten, että

$$\sigma^2(\theta)[g'(\theta)]^2 = c^2 \text{ eli } g'(\lambda) = \frac{c}{\sigma(\lambda)}.$$

$\square$

Silloin siis $g(\lambda) = 2c\sqrt{\lambda}$ ja valitsemalla $c = 1/2$ saadaan $g(\lambda) = \sqrt{\lambda}$ ja

$$\frac{\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda})}{1/2} \xrightarrow{d} \mathrm{N}(0, 1),$$

missä $\sigma^2(\lambda)[g'(\lambda)]^2 = c^2 = \frac{1}{4}$. Nyt siis $\sqrt{\bar{X}_n}$ noudattaa asymptoottisesti normaalijakaumaa $\mathrm{N}(\sqrt{\lambda}, \frac{1}{4n})$.

11.4 Tyhjentävyys

11.4.1 Perusidea

Eräs tilastollisen päättelyn peruskysymyksiä on, voidaanko havainnot korvata jollakin tunnusluvulla menettämättä oleellista informaatiota. Tyhjentävyyden käsite (Fisher 1922) luonnehtii havaintoihin sisältyvän parametria θ koskevan informaation kvantitatiivisesti. Parametrin θ estimaatti $T(x)$ on tyhjentävä jos se sisältää kaiken parametria θ koskevan informaation. Tyhjentävyyttä koskeva perustulos on se, että uskottavuusfunktio tiivistää itseensä kaiken havaintojen antaman parametria θ koskevan informaation.

Määritelmä 11.5 Oletetaan, että $X_1, \ldots, X_n$ on otos jakaumasta, jonka tiheysfunktio (tai todennäköisyysfunktio) on $f(x;\theta)$. Olkoon $t(x_1, \ldots, x_n)$ havaintojen $x_1, \ldots, x_n$ funktio (joka ei riipu θ:sta) ja $t(X_1, \ldots, X_n)$ on vastaava satunnaismuuttuja. Funktio $T = t(X_1, \ldots, X_n)$ on parametrin θ tyhjentävä tunnusluku, jos satunnaismuuttujien $X_1, \ldots, X_n$ ehdollinen jakauma ehdolla $T = t$ ei riipu θ:sta.

Intuitiivisesti määritelmä tarkoittaa sitä, että X:n arvot voidaan generoida X:n ehdollisesta jakaumasta $f(x|T = t;\theta)$, koska ehdollinen jakauma ei riipu θ:sta. Ei ole mielekästä esimerkiksi väittää, että otoskeskiarvo on populaation keskiarvon tyhjentävä tunnusluku. Usein näin on asia, mutta väitteen todenperäisyys riippuu populaation todennäköisyysmallista $f(x;\theta)$.

Esimerkki 11.10 Olkoon $X_1, \ldots, X_n$ otos Poissonin jakaumasta Poi(θ). Silloin $T(X_1, \ldots, X_n) = \sum X_i$ on θ:n tyhjentävä tunnusluku. Näytetään, että otoksen $X_1, \ldots, X_n$ jakauma ei riipu θ:sta, kun $T = t$ on annettu. Huomattakoon, että riippumattomien Poissonin jakaumaa noudattavien satunnaismuuttujien summa T noudattaa Poissonin jakaumaa Poi($n\theta$). Havaintojen ehdollinen todennäköisyys ehdolla $T = t$ on

$$
\begin{aligned}
P(X_1 = x_1, \ldots, X_n = x_n | T = t) &= \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P(T = t)} \\
&= \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = t)} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i} / \prod x_i!}{e^{-n\theta}(n\theta)^t / t!} \\
&= \frac{t!}{\prod x_i!} \prod \left(\frac{1}{n}\right)^{x_i},
\end{aligned}
$$

kaikille $x_1, \ldots, x_n$, joille $\sum x_i = t$, ja muutoin todennäköisyys on nolla. Ehdollinen todennäköisyys ei siis riipu θ:sta, joten T on θ:n tyhjentävä tunnusluku. Huomattakoon, että myös jokainen tunnusluvun $t = \sum x_i$ bijektiivinen kuvaus on θ:n tyhjentävä tunnusluku. $\square$

Esimerkki 11.11 Koneella tehdään n tuotetta. Tuote on hyväksyttävä todennäköisyydellä θ ja viallinen todennäköisyydellä $1-\theta$. Oletetaan, että eri tuotteiden laadun välillä ei ole riippuvuutta. Olkoon $X_i = 1$, jos tuote on hyväksyttävä ja $X_i = 0$ muutoin. Tehdään siis n riippumatonta Bernoullin koetta ja

$$P(X_1 = x_1, \ldots, X_n = x_n) = \theta^t (1-\theta)^{n-t}, \tag{11.4.1}$$

missä x_i saa arvon 0 tai 1 ja $t = \sum x_i$. Otoksen $X_1, \ldots, X_n$ ehdollinen jakauma ehdolla $T = \sum X_i = t$ on

$$\begin{aligned} f(x_1, \ldots, x_n | t) &= \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}, \end{aligned}$$

joka ei riipu θ:sta. T on siis tyhjentävä tunnusluku. □

11.4.2 Tekijälause

Tyhjentävyys on hyödyllinen käsite tarkasteltaessa havaintoaineiston tiivistämistä. Tiivistetään esimerkiksi havainnot $x_1, \ldots, x_n$ tunnuslukuun $\overline{x}$. Keskiarvon $\overline{x}$ tunteminen ei kuitenkaan riitä tyhjentävyyden toteamiseksi, vaan on myös tiedettävä, mistä jakaumasta havainnot ovat peräisin. Tunnusluvun tyhjentävyyden osoittaminen perusmääritelmän avulla on usein hankalaa. Useimmiten helpompi tapa on käyttää ns. tekijälausetta.

Lause 11.2 *Tunnusluku $t(x_1, \ldots, x_n)$ on parametrin θ tyhjentävä tunnusluku jos ja vain jos tiheysfunktio (tai todennäköisyysfunktio) $f(x_1, \ldots, x_n; \theta)$ voidaan esittää tekijämuodossa*

$$f(x_1, \ldots, x_n; \theta) = g(t(x_1, \ldots, x_n), \theta) h(x_1, \ldots, x_n),$$

missä $h(x_1, \ldots, x_n)$ ei riipu θ:sta.

Esimerkki 11.12 Olkoon $x_1, \ldots, x_n$ otos normaalijakausta $\mathrm{N}(\mu, \sigma^2)$ ja merkitään $\theta = (\mu, \sigma^2)$. Otoksen tiheysfunktio on

$$\begin{aligned} f(x_1, \ldots, x_n; \mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu \sum x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2} \right\}. \end{aligned}$$

Tekijälauseesta 11.2 seuraavat nyt tulokset:

1. Jos σ^2 tunnetaan, niin $\sum x_i$ on μ:n tyhjentävä tunnusluku (μ tuntematon);

2. Jos μ tunnetaan, niin $\sum(x_i - \mu)^2$ on σ^2:n tyhjentävä tunnusluku (σ^2 tuntematon);

3. Jos $\theta = (\mu, \sigma^2)$ on tuntematon, niin $(\sum x_i, \sum x_i^2)$ on θ:n tyhjentävä tunnuslukupari.

□

11.4.3 Minimaalinen tyhjentävyys

Jos $x_1, \ldots, x_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, 1)$, silloin esimerkiksi $\sum x_i$ ja $\overline{x}$ ovat μ:n tyhjentäviä tunnuslukuja. Myös lukupari $(\sum_{i=1}^m x_i, \sum_{i=m+1}^n x_i)$ ja järjestystunnuslukujen joukko $(x_{(1)}, \ldots, x_{(n)})$ ovat μ:n suhteen tyhjentäviä. Huomaa, että $\overline{x}$ on muiden tyhjentävien tunnuslukujen funktio. Toisaalta esimerkiksi järjestystunnuslukujen joukko $(x_{(1)}, \ldots, x_{(n)})$ ei ole $\overline{x}$:n funktio. Siksi on hyödyllistä tarkastella ns. *minimaalisen tyhjentävyyden* käsitettä. Intuitiivisesti luonnehtien minimaalisesti tyhjentävä tunnusluku tiivistää informaatiota mahdollisimman paljon siten, että mitään oleellista parametria koskevaa informaatiota ei menetetä.

Määritelmä 11.6 Tyhjentäviä tunnusluku $T = t(X_1, \ldots, X_n)$ on minimaalisesti tyhjentävä jos se on tyhjentävä ja jokaista toista tyhjentävää tunnuslukua V kohti on olemassa sellainen funktio g, että $T = g(V)$.

Tunnusluku on minimaalisesti tyhjentävä, jos havaintoaineistoa ei voida tiivistää enempää. Jos tyhjentävän tunnusluvun dimensio on sama kuin parametriavaruuden dimensio, niin tyhjentävä tunnusluku on minimaalisesti tyhjentävä.

Esimerkki 11.13 Jatketaan Esimerkkiä 11.11 ja osoitetaan, että $t = \sum x_i$ on θ:n tyhjentävä tunnusluku, kun otos $x_1, \ldots, x_n$ on Bernoullin jakaumasta $\mathrm{Ber}(\theta)$. Olkoon $s = s(x_1, \ldots, x_n)$ mikä tahansa toinen θ:n tyhjentävä tunnusluku. Silloin Tekijälauseen (11.2) nojalla

$$f(x_1, \ldots, x_n; \theta) = g(s, \theta)h(x_1, \ldots, x_n).$$

Yhdistämällä tämä tulokseen (11.4.1) huomaamme, että

$$\theta^t(1-\theta)^{n-t} = g(s, \theta)h(x_1, \ldots, x_n)$$

kaikilla θ:n arvoilla. Nyt siis millä tahansa annetuilla θ:n arvoilla θ_1 ja θ_2 kummallakin puolella muodostetut suhteet ovat yhtä suuret:

$$(\theta_1/\theta_2)^t[(1-\theta_1)/(1-\theta_2)]^{n-t} = \frac{g(s, \theta_1)}{g(s, \theta_2)}.$$

Jos erityisesti valitaan $\theta_1 = 2/3$ ja $\theta_2 = 1/3$ ja otetaan puolittain logaritmit, saadaan

$$t = \frac{1}{2\log 2} \log \frac{2^n g(s, 2/3)}{g(s, 1/3)}.$$

Tästä nähdään, että t on s:n funktio, joten t on minimaalisesti tyhjentävä. □

Esimerkki 11.14 Oletetaan, että $x_1, \ldots, x_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, 1)$. Silloin tekijälauseen 11.2 nojalla $\overline{x}$ on μ:n tyhjentävä tunnusluku. Havaintojen $x_1, \ldots, x_n$ perusteella saadaan uskottavuusfunktio

$$L(\mu; x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum(x_i - \mu)^2\right\}.$$

Tästä saadaan

$$\begin{aligned}\log L(\theta; x_1, \ldots, x_n) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum(x_i - \overline{x})^2 - \frac{n}{2}(\overline{x} - \mu)^2 \\ &= \text{vakio} - \frac{n}{2}(\overline{x} - \mu)^2,\end{aligned}$$

joten uskottavuusfunktio riippuu havainnoista vain otoskeskiarvon $\overline{x}$ kautta.

Oletetaan nyt, että vain otoskeskiarvo olisi kirjattu. Koska $\overline{X} \sim \mathrm{N}(1, 1/n)$, niin havaintoon $\overline{x}$ perustuva uskottavuusfunktio on

$$L(\mu; \overline{x}) = \frac{1}{\sqrt{2\pi/n}} \exp\left\{-\frac{n}{2}(\overline{x} - \mu)^2\right\},$$

tai vastaavasti

$$\log L(\mu; \overline{x}) = \text{vakio} - \frac{n}{2}(\overline{x} - \mu)^2.$$

Näin siis koko aineistoon $x_1, \ldots, x_n$ perustuva uskottavuus on sama kuin otoskeskiarvoon $\overline{x}$ perustuva uskottavuus. □

Edellisessä esimerkissa esitettyä tulosta vastaava tulos koskee kaikkia tyhjentäviä tunnuslukuja. Jos t on jokin havaintojen $x_1, \ldots, x_n$ funktio, niin havaintoihin $x_1, \ldots, x_n$ perustuva uskottavuus on sama kuin havaintoihin $x_1, \ldots, x_n$ ja t perustuva uskottavuus. Jos t on tyhjentävä, niin

$$\begin{aligned}L(\theta; x_1, \ldots, x_n) &= L(\theta; x_1, \ldots, x_n, t) = f(x_1, \ldots, x_n, t; \theta) \\ &= f(t; \theta) f(x_1, \ldots, x_n | t) \\ &= \text{vakio} \times f(t; \theta) \\ &= \text{vakio} \times L(\theta; t), \end{aligned} \tag{11.4.2}$$

mikä tarkoittaa, että $L(\theta; x_1, \ldots, x_n)$ voidaan laskea pelkästään t:n avulla.

Esimerkissä 11.13 näytettiin, miten todennäköisyysfunktion $f_\theta(x_1, \ldots, x_n)$ avulla voidaan osoittaa, että θ:n tyhjentävä tunnusluku on minimaalisesti tyhjentävä. Merkitään nyt lyhyesti $(x_1, \ldots, x_n) = \boldsymbol{x}$ ja $(X_1, \ldots, X_n) = \boldsymbol{X}$. Jokaista annettua havaintovektoria $\boldsymbol{x}$ (otosta) vastaa uskottavuusfunktio L siten, että

$$L_{\boldsymbol{x}}(\theta) = f(\boldsymbol{x}; \theta), \ \theta \in \Theta.$$

$L_{\boldsymbol{x}}$ on siis kuvaus otosavaruudelta S_X ($\boldsymbol{X}$:n arvojoukko) uskottavuusfuntioden joukkoon

$$\mathcal{L} = \{L_{\boldsymbol{x}} : \theta \to f(\boldsymbol{x};\theta),\ \boldsymbol{x} \in S_X\}.$$

Tämä on tunnusluku, jonka arvot ovat funktioita. Jos $\boldsymbol{X} = \boldsymbol{x}$ niin tunnusluku L saa arvon $L_{\boldsymbol{x}}$. Jos X on diskreetti, niin annetulla parametrin θ arvolla $L_{\boldsymbol{x}}(\theta)$ on todennäköisyys, että havaitaan $\boldsymbol{x}$. Jatkuvassa tapauksessa $L_{\boldsymbol{x}}(\theta)$ on verrannollinen todennäköisyyteen, että havainto sattuu pieneen $\boldsymbol{x}$:n sisältävään suorakaiteeseen. Kun ajattelemme θ:n funktiota $L_{\boldsymbol{x}}(\theta)$, se antaa kaikkien θ:n arvojen uskottavuuden, kun on havaittu $\boldsymbol{x}$.

Esimerkki 11.15 Jatketaan Esimerkkiä 11.12. Meillä on otos $x_1, \ldots, x_n$ normaalijakaumasta $N(\mu, \sigma^2)$. Kaksiulotteinen tunnusvektori

$$\boldsymbol{t} = (t_1, t_2) = (\sum x_i, \sum x_i^2)$$

on parametrivektorin (μ, σ^2) tyhjentävä tunnusvektori. Kun merkitään $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$, saadaan

$$L_{\boldsymbol{x}}(\theta) = (2\pi\theta_2)^{-n/2} \exp\left\{-\frac{n\theta_1^2}{2\theta_2}\right\} \exp\left\{-\frac{1}{2\theta_2}(t_2 - 2\theta_1 t_1)\right\}.$$

Nyt $\boldsymbol{\theta}$:n funktiona $L_{\boldsymbol{x}}(\cdot)$ määrittää tunnusvektorin (t_1, t_2), koska esimerkiksi

$$t_2 = 2\log L_{\boldsymbol{x}}(0,1) - n\log 2\pi$$

ja vastaavasti t_1 on määriteltävissä $L_{\boldsymbol{x}}(0,1)$:n ja $L_{\boldsymbol{x}}(1,1)$:n avulla. Näin uskottavuusfunktio L on yhtäpitävä tunnusvektorin (t_1, t_2) kanssa, joten L on tyhjentävä. Päättelemällä samalla tavalla kuin Esimerkissä 11.13 voidaan osoittaa, että (t_1, t_2) ja L ovat minimaalisesti tyhjentäviä. □

Uskottavuusfunktioon $L_{\boldsymbol{x}}(\cdot)$ läheisesti liittyvä funktio saadaan määrittelemällä tunnusluku

$$t(\boldsymbol{x}) = \frac{L_{\boldsymbol{x}}(\cdot)}{L_{\boldsymbol{x}}(\theta_0)}, \tag{11.4.3}$$

missä θ_0 on sellainen θ:n arvo, että $L_{\boldsymbol{x}}(\theta_0) > 0$. Tekijälauseen avulla voidaan näyttää, että $t(\boldsymbol{x})$ on tyhjentävä, kun lauseessa merkitään

$$g(t,\theta) = \frac{L_{\boldsymbol{x}}(\theta)}{L_{\boldsymbol{x}}(\theta_0)}$$

ja $h(\boldsymbol{x}) = L_{\boldsymbol{x}}(\theta_0)$. Suure (11.4.3) on θ:n ja θ_0:n *uskottavuussuhde*. Edellä osoitettiin (identiteetti (11.4.2)), että uskottavuusfunktio voidaan määrittää pelkästään tunusluvun t avulla, jos t on tyhjentävä. Olipa t siis mikä tahansa tyhjentävä tunnusluku, $L_{\boldsymbol{x}}(\cdot)$ voidaan lausua t:n funktiona. Uskottavuusfunktio on siis minimaalisesti tyhjentävä.

Lause 11.3 *Jos t on θ:n tyhjentävä tunnusluku, niin θ:n koko aineistoon $x_1, \ldots, x_n$ perustuva uskottavuus on sama kuin pelkästään tunnuslukuun t perustuva uskottavuus. Siksi uskottavuusfunktio on minimaalisesti tyhjentävä.*

On syytä muistaa, että kaksi uskottavuusfunktiota ovat samat, jos ne ovat verrannolliset. Lauseesta 11.3 seuraa, että mikä tahansa uskottavuusfunktion bijektiivinen kuvaus on minimaalisesti tyhjentävä.

Esimerkki 11.16 Olkoon $x_1, \ldots, x_n$ otos normaalijakumasta $\mathrm{N}(\mu, 1)$. Silloin uskottavuusfunktio

$$L(\mu; x_1, \ldots, x_n) = \text{vakio} \times \exp[-\frac{n}{2}(\overline{x} - \mu)^2]$$

on otoskeskiarvon $\overline{x}$ bijektiivinen kuvaus. Jos meillä on kaksi eri otosta $x_1, \ldots, x_n$ ja $y_1, \ldots, y_n$, niin

$$L(\mu; x_1, \ldots, x_n) = \text{vakio} \times L(\mu; y_1, \ldots, y_n) \text{ jos ja vain jos } \overline{x} = \overline{y}.$$

Tämä osoittaa, että $\overline{x}$ on minimaalisesti tyhjentävä. Vastaavanlaisella päättelyllä voidaan osoittaa, että $(\overline{x}, s^2)$ on parametrivektorin (μ, σ^2) minimaalisesti tyhjentävä tunnuslukupari, kun ostos on normaalijakumasta $\mathrm{N}(\mu, \sigma^2)$ ja s^2 on otosvarianssi. □

Monotoninen uskottavuussuhde

Minimaalinen tyhjentävyys voidaan luonnehtia uskottavuussuhteen

$$t(\boldsymbol{x}) = \frac{L_{\boldsymbol{x}}(\cdot)}{L_{\boldsymbol{x}}(\theta_0)}$$

avulla. Koska $t(\boldsymbol{x})$ on uskottavuusfunktion bijektiivinen kuvaus, se on minimaalisesti tyhjentävä. Millä tahansa θ:n arvojen $\theta_0 \neq \theta_1$ sellaisilla valinnoilla, että $L_{\boldsymbol{x}}(\theta_0) > 0$, uskottavuussuhde $L_{\boldsymbol{x}}(\theta_1)/L_{\boldsymbol{x}}(\theta_0)$ on $t(\boldsymbol{x})$:n bijektiivinen kuvaus. Jos $t(\boldsymbol{x})$ on skalaari, ominaisuutta kutsutaam *monotoniseksi uskottavuussuhteeksi*. Kun θ_1 lähestyy θ_0:aa, niin

$$\begin{aligned}\frac{L_{\boldsymbol{x}}(\theta_1)}{L_{\boldsymbol{x}}(\theta_0)} \approx & \frac{L_{\boldsymbol{x}}(\theta_0) + L'_{\boldsymbol{x}}(\theta_0)(\theta_1 - \theta_0)}{L_{\boldsymbol{x}}(\theta_0)} \\ = & 1 + \frac{\partial \log L_{\boldsymbol{x}}(\theta_0)}{\partial \theta_0}(\theta_1 - \theta_0).\end{aligned}$$

Tämä osoittaa, että $t(\boldsymbol{x})$ on minimaalisesti tyhjentävä, jos pistesuure $\frac{\partial \log L_{\boldsymbol{x}}(\theta_0)}{\partial \theta_0}$ on $t(\boldsymbol{x})$:n bijektiivinen funktio.

11.5 Eksponentiaalinen perhe

Eksponentiaalinen jakaumaperheeseen kuuluvat useimmat tavallisimmat jakaumat kuten normaali-, binomi- ja gammajakaumat sekä Poissonin jakauma. Eksponentiaalisen perheeseen jakauman tiheysfunktio (tai todennäköisyysfunktio) on muotoa

$$f(x;\theta) = \exp[t(x)\eta(\theta) - A(\theta) + c(x)], \tag{11.5.1}$$

kun $x \in S$, missä arvoalue S ei riipu parametrista θ. Funktiot $\eta(\cdot), t(\cdot), A(\cdot)$ ja $c(\cdot)$ ovat annettuja. Parametria $\eta(\cdot)$ kutsutaan luonnolliseksi parametriksi ja tunnuslukua $t(\cdot)$ luonnolliseksi tunnusluvuksi.

Yleisen p-parametriseen eksponentiaaliseen perheeseen kuuluvan jakauman logaritmoitu tiheysfunktio (tai todennäköisyysfunktio) on muotoa

$$\log f(x;\boldsymbol{\theta}) = \sum_{i=1}^{p} \eta_i(\boldsymbol{\theta})t_i(x) - A(\boldsymbol{\theta}) + c(x), \tag{11.5.2}$$

missä $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ ja $A(\boldsymbol{\theta}), c(x), \eta_i(\boldsymbol{\theta})$ ja $T_i(x)$ ovat tunnettuja funktioita. Lisäksi muuttujan x arvoalue ei riipu tuntemattomasta parametrista $\boldsymbol{\theta}$. Parametreja η_i kutsutaan luonnollisiksi parametreiksi ja tunnuslukuja $T_i(x)$ luonnollisiksi tunnusluvuiksi. Oletetaan lisäksi, että luonnolliset parametrit η_i eivät ole keskenään lineaarisesti riippuvia eivätkä luonnolliset tunnusluvut $T_i(x)$ keskenään. Voidaan osoittaa, että luonnolliset tunnusluvut $T_i(x)$ ovat minimaalisesti tyhjentäviä.

Eksponentiaalinen perhe on täysiasteinen, jos luonnollisten parametrien avaruus sisältää avoimen joukon. Esimerkiksi 2-ulotteinen neliö 2-ulotteisessa (2D) Euklidisessa avaruudessa sisältää avoimen joukon, mutta esimerkiksi suora 2D-avaruudessa ei sisällä avointa joukkoa. Tyypillisesti eksponentiaalinen perhe on täysiasteinen, jos luonnollisten parametrien lukumäärä on sama kuin tuntemattomien parametrien lukumäärä.

Jos luonnollisten parametrien välillä on epälineaarisia riippuvuuksia, niin luonnollisten tyhjentävien parametrien lukumäärä on suurempi kuin vapaitten parametrien lukumäärä ja silloin eksponentiaalista perhettä kutsutaan kaareutuvaksi. Monet teoreetiset tulokset pitävät paikkansa vain täysiasteisissa malleissa. Eksponentiaalinen perhe sisältää sekä jatkuvien että diskreettien satunnaismuuttujien jakaumia. Kuten jo edellä mainittiin, normaali-, binomi- ja gammajakauma kuuluvat eksponentiaaliseen perheeseen, kun taas Cauchyn jakauma ja t-jakauma eivät kuulu.

Esimerkki 11.17 Tarkastellaan normaalijakaumaa $\mathrm{N}(\boldsymbol{\theta})$, missä $\boldsymbol{\theta} = (\mu, \sigma^2)$ ja molemmat parametrit ovat tuntemattomia. Silloin jakauman tiheysfunktion logaritmi on

$$\log f(x;\mu,\sigma^2) = \frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2).$$

Tämä normaalijakauma on täysiasteinen eksponentiaalisen perheen malli, luonnolliset parametrit ovat $\eta_1 = \frac{\mu}{\sigma^2}$ ja $\eta_2 = -\frac{1}{2\sigma^2}$ sekä luonnolliset tunnusluvut $t_1(x) = x$ ja $t_2(x) = x^2$.

Oletetaan nyt, että satunnaismuuttujan $X \sim \mathrm{N}(\boldsymbol{\theta})$ variaatiokerroin on tunnettu vakio $a = \sigma/\mu$, missä $\mu > 0$. Silloin $\mathrm{N}(\mu, a^2\mu^2)$ kuuluu kaareutuvaan eksponentiaaliseen perheeseen. Vaikka mallissa on vain yksi tunematon parametri, luonnolliset tunnusluvut $t_1(x) = x$ ja $t_2(x) = x^2$ ovat edelleen minimaalisesti tyhjentäviä. □

Esimerkki 11.18 Poissonin jakauman $\mathrm{Poi}(\mu)$ tapauksessa todennäköisyysfunktion logaritmi on

$$\log f(x;\mu) = x\log\mu - \mu - \log x!.$$

Siksi $\mathrm{Poi}(\mu)$ on täysiasteinen yhden parametrin eksponentiaalisen perheen malli.

Jos X noudattaa sellaista typistettyä Poissonin jakaumaa, että arvoa $X = 0$ ei havaita, niin arvojen $x = 1, 2, \ldots$ todennäköisyydet ovat

$$f_\mu(x) = \frac{e^{-\mu}\mu^x/x!}{1 - e^{-\mu}},$$

joten

$$\log f(x;\mu) = x\log\mu - \mu - \log(1 - e^{-\mu}) - \log x!.$$

Tämä on myös yhden parametrin eksponentiaalisen perheen malli, jolla on sama luonnollinen parametri ja tunnusluku, mutta funktio $A(\mu)$ on muuttuu. □

Jos $X_1, \ldots, X_n$ on otos eksponentiaalisen perheen jakaumasta, niin otoksen yhteisjakauma kuuluu eksponentiaaliseen perheeseen. Olkoon $x_1, \ldots, x_n$ otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, niin yhteisjakauman tiheysfunktion logaritmi on

$$\log f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{\mu\sum_i x_i}{\sigma^2} - \frac{\sum_i x_i^2}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2).$$

Tässä jakaumalla on samat luonnolliset parametrit $\eta_1 = \frac{\mu}{\sigma^2}$ ja $\eta_2 = -\frac{1}{2\sigma^2}$ kuin normaalijakaumalla Esimerkissä 11.17. Yhteisjakauman luonnolliset tunnusluvut ovat $t_1(x_1, \ldots, x_n) = \sum_i x_i$ ja $t_2(x_1, \ldots, x_n) = \sum_i x_i^2$.

Tarkastellaan nyt lähemmin funktion $A(\eta)$ roolia eksponentiaalisen perheen jakauman tiheysfunktion/todennäköisyysfunktion lausekkeessa. Oletetaan, että X on satunnaismuuttuja, jonka tiheysfunktio on $\exp[c(x)]$ ja momenttifunktio

$$M(\eta) = E(e^{\eta X}).$$

Olkoon $A(\eta) \equiv \log M(\eta)$, jota kutsutaan kumulantteja generoivaksi funktioksi. Silloin

$$\int e^{\eta x + c(x)} d\,x = e^{A(\eta)}$$

tai

$$\int e^{\eta x - A(\eta) + c(x)} d\, x = 1$$

kaikilla η. Näin siis funktio

$$f(x;\eta) = e^{\eta x - A(\eta) + c(x)} \tag{11.5.3}$$

määrittelee eksponentiaaliseen perheeseen kuuluvan tiheysfunktion (tai todennäköisyysfunktion). Tätä menetelmää kutsutaan satunnasimuuttujan eksponenttikäännöksi. Alkuperäinen satunnasimuuttuja X liittyy arvoon $\eta = 0$.

Jokaiselle muotoa (11.5.3) olevalle funktiolle pätee, että

$$\mu = E(X;\eta) = A'(\eta)$$

ja

$$\mathrm{Var}(X;\eta) = A''(\eta) = \frac{\partial}{\partial \eta} E(X;\eta) = v(\mu).$$

Siksi $A(\eta)$ määrittelee tietyn yhteyden odotusarvon ja varianssin välille.

Eksponentiaalinen hajontamalli

Oletetaan nyt, että parametrien θ ja ϕ logaritmoitu uskottavuusfunktio on muotoa

$$\log L(\theta,\phi) = \frac{x\theta - A(\theta)}{\phi} + c(x,\phi), \tag{11.5.4}$$

missä $A(\theta)$ ja $c(x,\phi)$ ovat annettuja funktioita. Tässä esitysmuodossa parametria θ kutsutaan kanoniseksi parametriksi ja ϕ:tä hajontaparametriksi. Valitsemalla funktiot $A(\theta)$ ja $c(x,\phi)$ eri tavoin saadaan määriteltyä erilaisia todennäköisyysmalleja. Jos X on diskreetti, niin silloin täytyy olla

$$\sum_x \exp\left\{\frac{x\theta - A(\theta)}{\phi} + c(x,\phi)\right\} = 1,$$

mikä määrittelee $A(\theta)$:n ja $c(x,\phi)$:n välille tietyn yhteyden.

Hajontaparametri sallii varianssin vaihdella vapaasti odotusarvosta riippumatta:

$$\mu = E_\theta(X) = A'(\theta)$$

ja

$$\begin{aligned}\mathrm{Var}(X;\theta) =& \phi\, A''(\theta)\\ =& \phi\, \frac{\partial}{\partial \theta} E(X;\theta) = \phi\, v(\mu).\end{aligned}$$

Tämä on eksponentiaalisen hajontamallin suurin etu jäykempään malliin (11.5.2) verrattuna.

Käytännössä funktio $A(\theta)$ annetaan mallissa (11.5.4) eksplisiittisesti, kun taas funktio $c(x, \phi)$ jätetään tavallisesti tarkemmin määrittelemättä. Tämä ei ole ongelma niin kauan kuin on kyse vain θ:n estimoinnista, koska pistefunktio ei riipu $c(x, \phi)$:sta. Sen sijaan ϕ:n estimoimiseksi tarvitaan koko uskottavuusfunktio.

Esimerkki 11.19 Normaalijakauma $\mathrm{N}(\mu, \sigma^2)$ logaritmoitu uskottavuusfunktio voidaan kirjoittaa muodossa

$$\log L(\mu, \sigma^2) = \frac{x\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\log\sigma^2 - \frac{x^2}{2\sigma^2}.$$

Normaalijakauma on siis eksponentiaalisen perheen malli, jonka luonnollinen parametri on μ, hajontaparametri $\phi = \sigma^2, A(\mu) = \mu^2/2$ ja $c(x, \phi) = -\frac{1}{2}\log\phi - \frac{1}{2}x^2/\phi$. Tässä siis $c(x, \phi)$:n lauseke tunnetaan. □

Esimerkki 11.20 Poissonin jakauman $\mathrm{Poi}(\mu)$ logaritmoitu uskottavuusfunktio on muotoa

$$\log L(\mu) = x\log\mu - \mu - \log x!.$$

Tässä mallissa luonnollinen parametri on $\theta = \log\mu$, hajontaparametri $\phi = 1$ ja $A(\theta) = \mu = e^\theta$.

Pitämällä $A(\theta) = e^\theta$ vakiona ja vaihtelemalla hajonataparametria ϕ voidaan muodostaa Poissonin malleja, joissa on joko yli- tai alihajontaa. Samalla odotusarvolla $E(X) = \mu$ saadaan

$$\mathrm{Var}(X) = \phi A''(\theta) = \phi\mu.$$

Funktion $c(x, \phi)$ täytyy toteuttaa identiteetti

$$\sum_{x=0}^{\infty} \exp\left[\frac{x\theta - e^\theta}{\phi} + c(x, \phi)\right] = 1,$$

kaikilla θ ja ϕ. □

Esimerkki 11.21 Tarkastettiin 20 uuden auton otoksessa värivikojen lukumäärä käyttäen tiettyä kokeellista menetelmää. Tuloksiksi saatiin seuraavat lukumäärät:

`0 10 1 1 1 2 1 4 11 0 5 2 5 2 0 2 0 1 3 0`

Otoskeskiarvo on $\overline{x} = 2.55$ ja otosvarianssi $s^2 = 9.85$, mikä osoittaa ylihajontaa. Jos käytetään Poissonin tyyppistä mallia, jossa $A(\theta) = e^\theta$, niin

$$\mathrm{Var}_\theta(X) = \phi\, E_\theta(X).$$

Momenttimenetelmällä hajontaparametrin ϕ estimaatti on $\hat{\phi} = 9.84/2.55 = 3.86$. Aivan ilmeisesti hajontaparametrin arvo on suurempi kuin 1. □

Minimaalinen tyhjentävyys ja eksponentiaalinen perhe

Minimaalisen tyhjentävyyden ja eksponentiaalisen perheen välillä on tärkeä teoreettinen yhteys. Lievien oletusten vallitessa tyhjentävä (minimaalisesti tyhjentävä) tunnusluku (tunnusvektori) on olemassa jos ja vain jos $f_\theta(x)$ kuuluu täysiasteiseen eksponentiaaliseen perheeseen. Eksponentiaalisen perheen rakenteesta taas seuraa, että uskottavuusfunktiolla on yksikäsitteinen maksimi ja SUE on tyhjentävä. Jos siis tyhjentävä estimaatti (tunnusluku) on olemassa, niin se on SUE.