Luku 9

Uskottavuuden ominaisuuksia

9.1 Tyhjentävyys

Tyhjentävyys (Fisher 1922) luonnehtii täsmällisesti havaintoihin sisältyvän informaation kvantitatiivisesti. Parametrin θ estimaatti $T(x)$ on tyhjentävä jos se sisältää kaiken havaintoja θ:aa koskevan informaation. Tämä tarkoittaa sitä, että olkoon $U(x)$ mikä tahansa θ:n estimaatori, niin sen jakauma ei riipu θ:sta, jos $T(x)$ on annettu. Jos siis $T(x)$ tunnetaan, $U(x)$ ei anna parametrista θ mitään lisäinformaatiota.

Tyhjentävyyttä koskeva perustulos on se, että uskottavuusfunktio on *minimaalisesti tyhjentävä*. Uskottavuusfunktio tiivistää itseensä kaiken havaintojen parametria θ koskevan informaation. Tätä vähempi informaatio merkitsee informaation menetystä. Tilastollisen kokeen $\mathcal{E}$ malli on kolmikko $\{x, \theta, f_\theta(x)\}$. Määritelmässä x tarkoittaa yleisesti ottaen havaintoja. Todennäköisyysmalli $f_\theta(x)$ luonnehtii sen, miten havainnot on generoitu. Meillä voi olla esimerkiksi otos jostain jakaumasta.

Määritelmä 9.1 Parametrin θ tunnusluku $T(x)$ on tyhjentävä kokeessa $\mathcal{E}$, jos X:n jakauma ei riipu θ:sta, kun T:n arvo $T = t$ on annettu.

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

Esimerkki 9.1 Olkoon $X_1, \ldots, X_n$ otos Poissonin jakaumasta $\mathrm{Poi}(\theta)$. 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 $\mathrm{Poi}(n\theta)$. Ehdollinen todennäköisyys 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. □

Esimerkki 9.2 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

$$P(X_1 = x_1, \ldots, X_n = x_n) = \theta^t (1-\theta)^{n-t}, \tag{9.1.1}$$

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. □

Tyhjentävyys on hyödyllinen käsite tarkasteltaessa havaintoaineiston tiivistämistä. Esimerkiksi havainnot $x_1, \ldots, x_n$ tiivistetään tunnuslukuun $\overline{x}$. Keskiarvon $\overline{x}$ tunteminen ei kuitenkaan riitä, vaan on myös tiedettävä, että havainnot ovat otos Poissonin jakaumasta Poi(θ). Tunnusluvun tyhjentävyyden osoittaminen perusmääritelmän avulla on usein hankalaa. Suoraviivaisempi tapa on käyttää ns. tekijälausetta.

Lause 9.1 *$T(X_1, \ldots, X_n)$ on θ:n tyhjentävä tunnusluku jos ja vain jos tiheysfunktio (tai todennäköisyysfunktio) $f_\theta(x_1, \ldots, x_n)$ voidaan esittää tekijämuodossa*

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

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

Esimerkki 9.3 Olkoon $x_1, \ldots, x_n$ otos normaalijakausta $\mathrm{N}(\mu, \sigma^2)$ ja merkitään $\theta = (\mu, \sigma^2)$. Otoksen tiheysfunktio on

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

Tekijälauseesta 9.1 seuraavat nyt tulokset:

1. Jos σ^2 tunnetaan, niin $\sum x_i$ on μ:n tyhjentävä tunnusluku;
2. Jos μ tunnetaan, niin $\sum (x_i - \mu)^2$ on σ^2:n tyhjentävä tunnusluku;
3. Jos $\theta = (\mu, \sigma^2$ on tuntematon, niin $(\sum x_i, \sum x_i^2)$ on θ:n tyhjentävä tunnuslukupari.

□

9.1.1 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 tai tunnuslukujoukkojen funktio. Toisaalta esimerkiksi järjestystunnuslukujen joukko $(x_(1), \ldots, x_(n))$ ei ole $\overline{x}$:n funktio. Siksi on hyödyllistä tarkastella ns. *m*inimaalisen tyhjentävyyden käsitettä.

Määritelmä 9.2 Tyhjentäviä tunnusluku $T(X_1, \ldots, X_n)$ on minimaalisesti tyhjentävä jos se on minkä tahansa tyhjentävän tunnusluvun funktio.

Tunnusluku on minimaalisesti tyhjentävä, jos havaintoaineistoa ei voida tiivistää enempää. Yleisesti, jos tyhjentävän tunnusluvun dimensio on sama kuin parametriavaruuden dimensio, niin tunnusluku (tai tunnuslukujen joukko) on minimaalisesti tyhjentävä.

Esimerkki 9.4 Jatketaan Esimerkkiä 9.2 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 (9.1) nojalla

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

Yhdistämällä tämä tulokseen (9.1.1) 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 9.5 Oletetaan, että $x_1,\ldots,x_n$ on otos normaalijakaumasta $\mathrm{N}(\mu,1)$. Silloin tekijälauseen 9.1 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

$$\begin{aligned}L(\theta;x_1,\ldots,x_n) &= L(\theta;x_1,\ldots,x_n,t) = f_\theta(x_1,\ldots,x_n,t)\\ &= f_\theta(t)f(x_1,\ldots,x_n|t)\\ &= \text{vakio}\times f_\theta(t)\\ &= \text{vakio}\times L(\theta;t), \qquad (9.1.2)\end{aligned}$$

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

Esimerkissä 9.4 näytettiin, miten todennäköisyysfunktion $f_\theta(x_1, \dots, x_n)$ avulla voidaan osoittaa, että θ:n tyhjentävä tunnusluku on minimaalisesti tyhjentävä. Merkitään nyt lyhyesti $(x_1, \dots, x_n) = \boldsymbol{x}$ ja $(X_1, \dots, X_n) = \boldsymbol{X}$. 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 9.6 Jatketaan Esimerkkiä 9.3. Meillä on otos $x_1, \dots, x_n$ normaalijakaumasta $N(\mu, \sigma^2)$. Kaksiulotteinen tunnusvektori

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

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 θ: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 9.4 voidaan osoittaa, että $t_1, t_2)$ ja L ovat minimaalisesti tyhjentäviä. □

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

$$t(\boldsymbol{x}) = \frac{L_{\boldsymbol{x}}(\cdot)}{L_{\boldsymbol{x}}(\theta_0)}, \tag{9.1.3}$$

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 (9.1.3) on θ:n ja θ_0:n *uskottavuussuhde*. Edellä osoitettiin (identiteetti (9.1.2)), 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 9.2 *Jos t on θ:n tyhjentävä tunnusluku kokeessa $\mathcal{E}$, 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 9.2 seuraa, että mikä tahansa uskottavuusfunktion bijektiivinen kuvaus on minimaalisesti tyhjentävä.

Esimerkki 9.7 Olkoon $x_1, \ldots, x_n$ otos normaalijakumasta $\mathrm{N}(\mu, 1)$. Silloin uskottavuusfunktio

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

$\square$

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 osoitaa, 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.

9.2 Moniparametriset mallit

Useimmat käytännon sovellukset vaativat moniparametristen mallien tarkastelua. Esimerkiksi normaalijakaumassa $\mathrm{N}(\mu, \sigma^2)$ on kaksi parametria, jos sekä odotusarvo että varianssi ovat tuntemattomia. Mallin kompleksisuus määräyryy parametrien, ei havaintojen, lukumäärän mukaan. Tässä alaluvussa tarkastellaan parametrien estimointia uskottavuusmenetelmällä, kun todennäköisyysmallissa on kaksi kaksi tai useampia tuntemattomia parametreja. Kun on havaittu $\boldsymbol{x} = x_1, \ldots, x_n$, niin uskottavuusfunktio on

$$L(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}),$$

missä $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ on havainnon todennäköisyys (diskreetti satunnaismuuttuja) tai tiheysfunktion arvo pisteessä $\boldsymbol{x}$, kun parametrivektori on $\boldsymbol{\theta}$. Koska uskottavuusfunktio on parametrien funktio, se voidaan kertoa parametreista riippumattomalla vakiolla, mikä ei muuta mitään oleellista. Tavallisesti tiheysfunktiossa (tai todennäköisyysfunktiossa) on parametreista riippumaton normeeraustekijä, joka siis voidaan jättää uskottavuusfunktion esityksestä pois (kerrotaan funktio tämän vakion käänteisluvulla).

Esimerkki 9.8 Olkoon $x_1, \ldots, x_n$ nomaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$. Uskottavuusfunktio on

$$L(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2, \tag{9.2.1}$$

missä parametreista riippumattomat vakiot on jätetty pois. Oletetaan, että $n_1, \ldots, n_k$ noudattaa multinomijakaumaa $\mathrm{Mult}(N; \theta_1, \ldots, \theta_k)$, missä $N = \sum_i n_i$ tunnetaan ja $n_1, \ldots, n_k$ ovat tuntemattomia todennäköisyysparametreja. Mallissa on $k-1$ vapaata parametria, koska $\sum_i \theta_i = 1$. Kun parametreista riippumattomat vakiot jätetään pois, saadaan uskottavuusfunktioksi

$$L(\theta_1, \ldots, \theta_k) = \sum_i n_i \log\, \theta_i.$$

Kun $k = 3$, niin $\theta_1 + \theta_2 + \theta_3 = 1$ ja parametriavaruus voidaan esittää kolmiona tasossa. □

9.2.1 Pistevektori ja informaatiomatriisi

Moniparametrisessa mallissa uskottavuusfunktion $L(\boldsymbol{\theta})$ tutkiminen on tietysti huomattavasti haastavampaa kuin yhden parametrin tapauksessa. Olkoon $\boldsymbol{\theta} = \theta_1, \ldots, \theta_p$ ja oletetaan, että $\log L(\boldsymbol{\theta})$ on derivoituva. Silloin pistefunktio on 1. derivaattojen vektori

$$\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta})$$

ja $\boldsymbol{\theta}$:n SUE $\hat{\boldsymbol{\theta}}$ on pisteyhtälön $\boldsymbol{s}(\boldsymbol{\theta})$ ratkaisu. Fisherin informaatio $I(\hat{\boldsymbol{\theta}})$ on toisten derivaattojen matriisi, jonka alkiot ovat

$$I_{ij}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Fisherin informaation idea liittyi logaritmisen uskottavuusfunktion kvadraattiseen likiarvoon. Moniparametrisen uskottavuusfunktion $L(\boldsymbol{\theta})$ likiarvo on muotoa

$$\begin{aligned}\log L(\boldsymbol{\theta}) &\approx \log L(\hat{\boldsymbol{\theta}}) + \boldsymbol{s}(\hat{\boldsymbol{\theta}})(\theta - \hat{\boldsymbol{\theta}}) - \frac{1}{2}(\theta - \hat{\boldsymbol{\theta}})^T I(\hat{\boldsymbol{\theta}})(\theta - \hat{\boldsymbol{\theta}}) \\ &= \log L(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\theta - \hat{\boldsymbol{\theta}})^T I(\hat{\boldsymbol{\theta}})(\theta - \hat{\boldsymbol{\theta}}).\end{aligned}$$

Jos mallissa on esimerkiksi kaksi tuntematonta parametria α ja β, niin parametrien α ja β yhteinen uskottavuusfunktio on

$$L(\alpha, \beta) = f(\boldsymbol{x}; \alpha, \beta), \tag{9.2.2}$$

missä $f(\boldsymbol{x}; \alpha, \beta)$ on havaintojen $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ yhteisjakauman todennäköisyysfunktio tai tiheysfunktio. Logaritmoitua (luonnollinen logaritmi) uskottavuusfunktiota $\log L(\alpha, \beta)$ merkitään $l(\alpha, \beta)$.

Parametriparin (α, β) suurimman uskottavuuden estimaatti $(\hat{\alpha}, \hat{\beta})$ on sellainen arvopari, joka maksimoi funktiot $L(\alpha, \beta)$ ja $l(\alpha, \beta)$. Kahden parametrin tapauksessa pistefunktio on 2×1-vektori

$$\boldsymbol{s}(\alpha, \beta)\begin{pmatrix} s_1(\alpha, \beta) \\ s_2(\alpha, \beta) \end{pmatrix},$$

missä $s_1(\alpha, \beta) = \frac{\partial l}{\partial \alpha}$ ja $s_2(\alpha, \beta) = \frac{\partial l}{\beta}$. Estimaatit $(\hat{\alpha}, \hat{\beta})$ löydetään tavallisesti ratkaisemalla yhtälö $\boldsymbol{s}(\alpha, \beta) = \mathbf{0}$, eli yhtälöpari

$$u_1(\alpha, \beta) = 0 \qquad \text{ja} \qquad u_2(\alpha, \beta) = 0. \tag{9.2.3}$$

Kahden parametrin tapauksessa Fisherin informaatio $I(\hat{\alpha}, \hat{\beta})$ saadaan sijoittamalla estimaatit $\hat{\alpha}$ ja $\hat{\beta}$ toisten derivaattojen matriisiin

$$I(\alpha, \beta) = -\begin{pmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} \end{pmatrix}, \tag{9.2.4}$$

missä alkiot i_{jk} ovat parametrien α:n ja β:n funktiota. Matriisin $I(\alpha, \beta)$ arvo pisteessä $(\hat{\alpha}, \hat{\beta})$ on

$$I(\hat{\alpha}, \hat{\beta}) = \begin{pmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{pmatrix}.$$

Paikallisessa maksimissa $I(\hat{\alpha}, \hat{\beta})$ on positiivisesti definiitti. Silloin $I(\hat{\alpha}, \hat{\beta})$:n alkiot toteuttavat ehdot $\hat{\imath}_{11} > 0$, $\hat{\imath}_{22} > 0$ ja $\hat{\imath}_{11}\hat{\imath}_{22} - \hat{\imath}_{12}^2 > 0$.

Esimerkki 9.9 Mitataan kaksi kappaletta erikseen ja yhdessä, joten saadaan kolme mittaustulosta X_1, X_2 ja X_3. Kappaleiden painot μ_1 ja μ_2 ovat tuntemattomia. Tiedetään kokemuksesta, että mittaukset ovat toisistaan riippumattomat ja noudattavat normaalijakaumaa, jonka varianssi on yksi. Muuttujat X_1, X_2 ja X_3 ovat siis keskenään riippumattomat ja

$$X_1 \sim \mathrm{N}(\mu_1, 1); \qquad X_2 \sim \mathrm{N}(\mu_2, 1); \qquad X_3 \sim \mathrm{N}(\mu_1 + \mu_2, 1).$$

Havaintojen X_1, X_2, X_3 yhteisjakauman tiheysfunktio on

$$f(x_1, x_2, x_3; \mu_1, \mu_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^3 \exp\left\{-\tfrac{1}{2}\left[(x_1-\mu_1)^2+(x_2-\mu_2)^2+(x_3-\mu_1-\mu_2)^2\right]\right\}.$$

Uskottavuusfunktion on sis muotoa

$$L(\mu_1, \mu_2) = \exp\left\{-\tfrac{1}{2}\left[(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 + (x_3 - \mu_1 - \mu_2)^2\right]\right\}$$

ja logaritmoitu uskottavuusfunktio o

$$l(\mu_1, \mu_2) = -\tfrac{1}{2}\left[(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 + (x_3 - \mu_1 - \mu_2)^2\right].$$

Pistefunktion komponentit ovat

$$s_1(\mu_1, \mu_2) = \frac{\partial\, l}{\partial \mu_1} = (x_1 - \mu_1) + (x_3 - \mu_1 - \mu_2);$$

$$s_2(\mu_1, \mu_2) = \frac{\partial\, l}{\partial \mu_2} = (x_2 - \mu_2) + (x_3 - \mu_1 - \mu_2).$$

Toiset derivaatat ovat

$$\frac{\partial^2 l}{\partial \mu_1^2} = -2, \qquad \frac{\partial^2 l}{\partial \mu_2^2} = -2 \qquad \text{ja} \qquad \frac{\partial^2 l}{\partial \mu_1 \partial \mu_2} = -1,$$

joten informaatiomatriisi on

$$I(\mu_1, \mu_2) = -\begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Tässä tapauksessa $I(\mu_1, \mu_2)$ ei riipu parametreista, joten $I(\hat{\mu}_1, \hat{\mu}_2) = I(\mu_1, \mu_2)$. Suurimman uskottavuuden estimaattien määrittämiseksi ratkaistaan yhtälöpari

$$u_1(\mu_1, \mu_2) = 0 \qquad \text{ja} \qquad u_2(\mu_1, \mu_2) = 0.$$

Silloin ensimmäisestä yhtälöstä saadaan

$$\hat{\mu}_1(\mu_2) = \tfrac{1}{2}(x_1 + x_3 - \mu_2),$$

joka on μ_1:n suurimman uskottavuuden estimaatti, jos μ_2 tunnetaan. Kun $\mu_1 = \frac{1}{2}(x_1 + x_3 - \mu_2)$ sijoitetaan toiseen yhtälöön ja ratkaistaan μ_2, saadaan

$$\hat{\mu}_2 = \tfrac{1}{3}(2x_2 + x_3 - x_1) = \tfrac{2}{3}x_2 + \tfrac{1}{3}(x_3 - x_1).$$

Tästä saadaan

$$\hat{\mu}_1 = \tfrac{1}{2}(x_1 + x_3 - \hat{\mu}_2) = \tfrac{1}{3}(2x_1 + x_3 - x_2) = \tfrac{2}{3}x_1 + \tfrac{1}{3}(x_3 - x_2).$$

□

9.3 Profiiliuskottavuus

Vaikka moniparametrisen mallin uskottavuusfunktio on hyvin määritelty, voi tällaisen usean muuttujan funktion käsittely ja sen antaman informaation hahmottaminen olla vaikeaa. Toisaalta saatamme olla kiinnostuneita vain osasta parametreja ja loput ovat ns. *kiusaparametreja*. Tarvittaisiin menetelmä, jolla kiinnostaviin parametreihin sisätyvä informaatio saataisiin esille ja kiusaparametrit voitaisiin eliminoida. Tämä on kuitenkin hankala ongelma eikä sen ratkaisemiseksi ole mitään yhtä yleistä menetelmää.

Jos mallissa on esimerkiksi kaksi parametria (α, β), mutta olemme kiinnostuneita vain parametrista α. Parametri β on silloin kiusaparametri. Parametrien yhteisen uskottavuusfunktion $L(\alpha, \beta)$ avulla voidaan asettaa arvoparit (α, β) uskottavuuden mukaiseen järjestykseen. Haluaisimme kuitenkin mallit vain α:aa koskevan informaation mukaiseen järjestykseen.

Olkoon $\hat{\beta}_\alpha$ parametrin β arvo, joka maksimoi uskottavuusfunktion annetulla α:n arvolla. Määritellään nyt *profiiliuskottavuus* α:n suhteen.

Määritelmä 9.3 Olkoon $L(\alpha, \beta)$ parametrien α ja β yhteinen uskottavuusfunktio. Silloin α:n profiiliuskottavuus on

$$L(\alpha) = \max_\beta L(\alpha, \beta).$$

On syytä korostaa, että jokaista kiinnitettyä α:n arvoa kohti β:n SUE on α:n funktio. Voimme siis kirjoittaa, että

$$L(\alpha) = L(\alpha, \hat{\beta}_\alpha).$$

Profiiliuskottavuutta voidaan käsitellä kuten tavallistakin uskottavuutta. Voimme esimerkiksi normittaa sen sekä laskea siitä uskottavuusvälejä. Yhteydestä tavallisesti käy ilmi, onko kyseessä profiiliuskottavuus vai ”tavallinen”uskottavuus. Jos sekaantumisen vaara on olemassa, merkitään profiiliuskottavuutta $L_p(\alpha)$.

Esimerkki 9.10 Tarkastellaan Esimerkin 9.9 logaritmoitua uskottavuusfunktiota

$$l(\mu_1, \mu_2) = -\tfrac{1}{2}(x_1 - \mu_1)^2 - \tfrac{1}{2}(x_2 - \mu_2)^2 - \tfrac{1}{2}(x_3 - \mu_1 - \mu_2)^2.$$

Kuten Esimerkissä 9.9 todettiin, μ_1:n suurimman uskottavuuden estimaatti μ_2:n funktiona on

$$\hat{\mu}_1(\mu_2) = \tfrac{1}{2}(x_1 + x_3 - \mu_2).$$

Ehdolla, että μ_2 on annettu, voidaan $\hat{\mu}_1(\mu_2)$:n arvo laskea. Normitettu profiiliuskottavuus on

$$L(\mu_2) = \frac{L[\hat{\mu}_1(\mu_2), \mu_2]}{L(\hat{\mu}_1, \hat{\mu}_2)}$$

ja vastaava logaritmoitu profiiliuskottavuuv on

$$l(\mu_2) = l[\hat{\mu}_1(\mu_2), \mu_2] - l(\hat{\mu}_1, \hat{\mu}_2),$$

josta saadaan

$$l(\mu_2) = -\tfrac{3}{4}(\mu_2 - \hat{\mu}_2)^2.$$

Jos $\hat{\mu}_2 = 29.6$, kuten Esimerkissä **??**, niin μ_2:n uskottavuusväli on $27.85 \leq \mu_2 \leq 31.35$. □

9.3.1 Profiiliuskottavuuden kaarevuus

9.4 Havaittu uskottavuus usean parametrin malleissa

9.4.1 Uskottavuuteen perustuvat luottamusalueet

9.4.2 AIC-kriteeriin perustuva kalibrointi