Luku 10

Uskottavuuspäättelyn perusteet

10.1 Uskottavuuden määritelmä

Tarkastellaan kannatusmittausten tyypillistä asetelmaa. On arvioitava ehdokkaan A kannattajien suhteellinen osuus θ annetulla paikkakunnalla (Katso Esimerkki 8.9). Valitaan äänioikeutetuista satunnaisesti n henkilöä, joilta tiedustellaan, kannattavatko he ehdokasta A. Olkoon X ehdokkaan A kannattajien lukumäärä otoksessa. Populaation koko on suuri verrattuna otoskokoon n, joten otoksessa kannattajien lukumäärän X voidaan olettaa noudattavan likimain binomijakaumaa $\mathrm{Bin}(n, \theta)$, missä θ on todennäköisyys, että satunnaisesti valittu henkilö kannattaa A:ta.

Binomijakaumaa noudattavan satunnaismuuttujan X todennäköisyysfunktio on muotoa

$$f(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n; \quad 0 \le \theta \le 1.$$

Binomijakauman parametrin θ arvojoukko eli parametriavaruus on

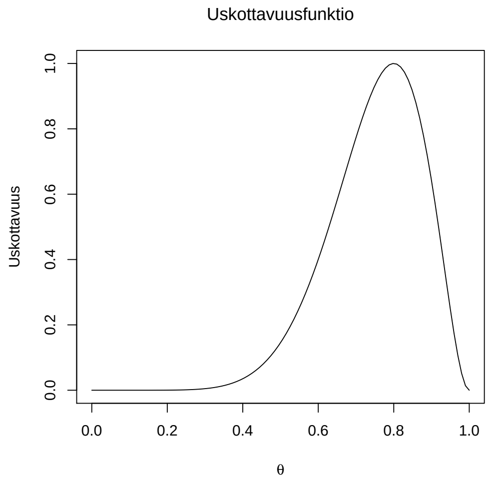
$$\Theta = \{\, \theta \mid 0 \le \theta \le 1 \,\}.$$

Tehtävänämme on arvioida otoksesta saadun havaitun arvon $X = x$ perusteella θ:n arvo eli laskea θ:n estimaatti. Havainnon $X = x$ todennäköisyys on

$$P_\theta(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{10.1.1}$$

Eräs tapa määrittää θ:n estimaatti on tarkastella todennäköisyyttä $P_\theta(X = x)$ parametrin θ funktiona ja etsiä sellainen θ:n arvo, että havainnon x todennäköisyys saavuttaa maksiminsa. Voidaan osoittaa, että havainnon $X = x$ todennäköisyys maksimoituu, kun $\theta = x/n$. Tällaista estimaattia kutsutaan θ:n *suurimman uskottavuuden estimaatiksi* ja sitä merkitään

$$\hat{\theta} = \frac{x}{n}.$$



Kuvio 10.1. Onnistumistodennäköisyyden θ uskottavuus, kun binomijakaumasta $\mathrm{Bin}(10, \theta)$ on saatu havainto $x = 8$.

Jos otoskoko $n = 100$ ja otoksessa oli $X = 10$ A:n kannattajaa, niin θ:n suurimman uskottavuuden estimaatti on $\hat{\theta} = 0.10$. Kun havainto $X = 10$ on saatu ja otoskoko $n = 100$ on kiinnitetty, niin todennäköisyyttä (10.1.1) voidaan tarkastella parametrin θ funktiona:

$$L(\theta) = P_\theta(X = 10) = \binom{100}{10} \theta^{10} (1 - \theta)^{90}. \tag{10.1.2}$$

Funktiota $L(\theta)$ kutsutaan *uskottavuusfunktioksi*. Uskottavuusfunktio sisältää tuntematonta parametria θ koskevaa informaatiota. Informaatio on sikäli epätäydellistä, että parametrin arvo on pääteltävissä vain tietyllä varmuudella. Uskottavuusfunktiosta voidaan kuitenkin arvioida myös tuo varmuuden aste. Tilastollisiin ongelmiin liittyvät tuntemattomat suureet ovat usein *kiinteitä parametreja* (tuntemattomia vakioita) tai *satunnaisparametreja* eli satunnasimuuttujia, joista ei saada havaintoja. Kumpaankin tapauksee liittyy estimointiongelma. Tässä esityksessä tarkastellaan lähes yksinomaan kiinteän parametrin tilannetta.

Määritelmä 10.1 Oletetaan, että diskreetin satunnaismuuttujan X riippuu tuntemattomasta parametrista θ. Silloin uskottavuusfuntion $L(\theta)$ arvo $L(\theta_0)$ on havainnon $X = x$ todennäköisyys, jos $\theta = \theta_0$.

Esimerkissä 8.1 käsiteltiin uskottavuuden (uskottavuusfunktion) käyttöä tilastollisessa päättelyssä. Esitämme nyt tuon ajatuksen uskottavuusperiaatteena.
Uskottavuusperiaate: *Tarkastellaan kahta eri havaintoa, x ja y, jotka on saatu samasta populaatiosta, vaikkakin mahdollisesti eri otantamenetelmillä. Jos näihin havaintoihin perustuvien θ:n uskottavuuksien suhde $L_1(\theta; x)/L_2(\theta; y)$*

ei riipu θ:sta, silloin molemmat havainnot antavat saman θ:a koskevan informaation ja johtavat samoihin θ:a koskeviin johtopäätöksiin.

10.1.1 Diskreetit mallit

Diskreeteissä malleissa havaintoarvon todennäköisyys on hyvin määritelty suure, kuten esimerkiksi binomimallissa (10.1.2).

10.1.2 Jatkuvat mallit

Kun X on jatkuva satunnaismuuttuja, niin tiheysfunktion arvo $f(x;\theta)$ ei ole todennäköisyys $P(X = x)$. Itse asiassa $P(X = x) = 0$ kaikilla x, kun X on jatkuva satunnaismuuttuja. Koska mittaustarkkuus on aina äärellinen, sovelluksissa tapahtuma $\{X = x\}$ tarkoittaa, että x kuuluu johonkin mittaustarkkuuden määrittämään väliin. Tapahtumien $\{X = x\}$ sijasta tarkastellaan siis tapahtumia $\{a < X \leq b\}$, missä $a < b$. Silloin tapahtuman $\{X = x\}$ todennäköisyys on siis muotoa

$$P(a < X \leq b) = \int_a^b f(x;\theta)\,\mathrm{d}x = F(b;\theta) - F(a;\theta).$$

Olkoon $X = x$ havainto jakaumasta $F(x;\theta)$. Jos mittaustarkuus on $\Delta > 0$, niin

$$\begin{aligned} P_\theta(X = x) &= P_\theta(x - \Delta/2 < X \leq x + \Delta/2) \\ &= F(x + \Delta/2;\theta) - F(x - \Delta/2;\theta). \end{aligned} \tag{10.1.3}$$

Jos väli Δ on pieni ja F on kohtuullisen sileä, niin

$$F(x + \Delta/2;\theta) \approx F(x - \Delta/2;\theta).$$

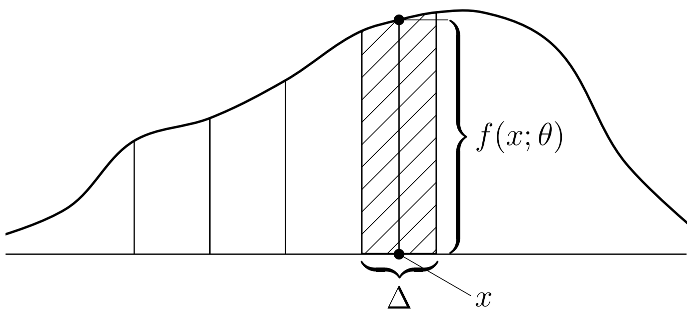
Silloin voidaan käyttää likiarvoa

$$F(x + \Delta/2;\theta) - F(x - \Delta/2;\theta) \approx f(x;\theta)\Delta.$$

Koska mittaustarkkuus Δ ei riipu parametrista θ, niin uskottavuusfunktio on vakio kertaa tiheysfunktion arvo, eli

$$L(\theta) = cf(x;\theta), \tag{10.1.4}$$

missä c on mikä tahansa sopivasti valittu positiivinen vakio.



10.2 Esimerkkejä

Esimerkki 10.1 Tarkastellaan vielä alaluvussa 10.1 esitettyä kannatusmittausta. Ehdokasta A kannatti 100:n äänioikeutetun otoksessa 10, jolloin saatiin uskottavuusfunktio (10.1.2). Jos tiedetään ainoastaan, että A:ta kannatti korkeintaan 10, silloin θ:aa koskeva informaatio voidaan lausua uskottavuusfunktiolla

$$\begin{aligned} L(\theta) &= P_\theta(X \leq 10) \\ &= \sum_{x=0}^{10} \binom{100}{x} \theta^x (1-\theta)^{100-x}. \end{aligned}$$

□

Esimerkki 10.2 Estimoidaan myyrien lukumäärä (N) tietyllä alueella. Tutkimusryhmä pyydystää $N_1 = 25$ myyrää, merkitsee ne ja laskee takaisin luontoon. Myöhemmin pyydystetään $n = 60$ myyrää, joissa on $x = 5$ merkittyä ja $n - x = 55$ merkitsemätöntä. Jos oletetaan, että kaikkien myyrien todennäköisyys joutua pyydykseen on sama, niin N:n uskottavuus voidaan laskea hypergeometrisen todennäköisyysfunktion avulla

$$L(N) = P(X = 5; N) = \frac{\binom{25}{5}\binom{N-25}{55}}{\binom{N}{60}}.$$

□

10.3 Uskottavuuksien yhdistäminen

Tehdään kaksi toisistaan riippumatonta koetta (tai otosta), jotka antavat informaatiota samasta parametrista θ. Havainnon $X = x$ todennäköisyys 1. kokeessa on $f_1(x;\theta)$ ja havainnon $Y = y$ todennäköisyys 2. kokeessa $f_2(y;\theta)$. Teknisesti voimme sanoa, että satunnaismuuttujat X ja Y ovat toisistaan riippumattomat. Meillä on siis kaksi riippumatonta havaintoa x ja y.

Uskottavuuden määritelmän (10.1) mukaan 1. kokeeseen perustuva uskottavuusfunktio on

$$L(\theta; x) = f_1(x;\theta)$$

ja 2. kokeeseen perustuva uskottavuusfunktio

$$L(\theta; y) = f_2(y;\theta).$$

Olkoon satunnaismuuttujien X ja Y yhteisjakauman todennäköisyysfunktio $f(x, y; \theta)$. Yhdistettyyn kokeeseen perustuva uskottavuusfunktio on määritelmän (10.1) mukaan

$$L(\theta; x, y) = f(x, y; \theta).$$

Koska X ja Y ovat riippumattomat, niin $f(x, y; \theta) = f_1(x; \theta) f_2(y; \theta)$. Yhdistettyyn kokeeseen perustuva uskottavuusfunktio voidaan siis kirjoittaa muodossa

$$L(\theta; x, y) = L(\theta; x)\, L(\theta; y). \tag{10.3.1}$$

Ottamalla uskottavuusfunktiosta (10.3.1) puolittain logaritmit saadaan vastaava logaritmoitu uskottavuusfunktio

$$l(\theta; x, y) = l(\theta; x) + l(\theta; y), \tag{10.3.2}$$

missä on merkitty $l(\theta; x) = \log L(\theta; x)$, $l(\theta; y) = \log L(\theta; y)$ ja $l(\theta; x, y) = \log L(\theta; x, y)$.

Kahdesta riippumattomasta kokeesta saatava parametria θ koskeva informaatio voidaan siis yhdistää siten, että kerrotaan yksittäisiin kokeisiin liittyvät uskottavuusfunktiot keskenään ja vastaavasti logaritmoidut uskottavuusfunktiot lasketaan yhteen. On helppo nähdä, että useamman kuin kahden riippumattoman kokeen antamat tulokset voidaan yhdistää vastaavasti kertomalla kokeisiin liittyvät uskottavuusfunktiot ja laskemalla yhteen logaritmoidut uskottavuusfunktiot.

Jos siis $x_1, \ldots, x_n$ on (havaittu) otos jakaumasta, jonka tiheysfunktio (tai todennäköisyysfunktio diskreetin satunnaismuuttujan tapauksessa) on $f(x; \theta)$, niin

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} L(\theta; x_i)$$

ja

$$l(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} l(\theta; x_i).$$

Jos sekaannuksen vaaraa ei ole, voidaan havaintoihin $x_1, \ldots, x_n$ perustuva θ:n uskottavuusfunktio kirjoittaa yksinkertaisesti $L(\theta)$. Huomattakoon, että hajotelma (10.3.1) on sillä tavoin yleinen, että havainnot $\boldsymbol{x} = (x_1, \ldots, x_k)$ ja $\boldsymbol{y} = (y_1, \ldots, y_m)$ voivat olla myös vektoriarvoisia. Jos satunnaismuuttujat $\boldsymbol{X} = (X_1, \ldots, X_k)$ ja $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ ovat riippumattomat, niin

$$f(\boldsymbol{x}, \boldsymbol{y}; \theta) = f_1(\boldsymbol{x}; \theta) f_2(\boldsymbol{y}; \theta)$$

ja hajotelma (10.3.1) pitää paikkansa.

Esimerkki 10.3 Jatketaan esimerkin 10.1 käsittelyä. Oletetaan, että kaksi eri tutkijaa tekevät samaan aikaan toisistaan riippumatta ehdokkaan A kannattusta mittaavan haastattelututkimuksen siten, että 1. tutkija haastatteli 100 ja 2. tutkija 50 äänioikeutettua. Merkitään $n = 100 + 50$. Havaittiin, että 1. tutkijan otoksessa oli 10 ja 2. tutkijan otoksessa 8 ehdokkaan A kannattajaa.

Nyt 1. otokseen perustuva logaritmoitu uskottavuusfunktio on

$$l_1(\theta; 10) = \log \binom{100}{10} + 10 \log(\theta) + 90 \log(1-\theta)$$

ja 2. otokseen perustuva logaritmoitu uskottavuusfunktio on vastaavasti

$$l_2(\theta; 8) = \log \binom{50}{8} + 8 \log(\theta) + 42 \log(1-\theta).$$

Otoksista lasketut suurimman uskottavuuden estimaatit ovat vastaavasti

$$\hat{\theta}_1 = \frac{10}{100} \qquad \text{ja} \qquad \hat{\theta}_2 = \frac{8}{50}.$$

Koska otokset ovat riippumattomat, niin tuloksen (10.3.2) mukaan yhdistettyyn otokseen perustuva logaritmoitu uskottavuusfunktio on

$$\begin{aligned} l(\theta) &= l_1(\theta) + l_2(\theta) \\ &= (10+8)\log(\theta) + (100 + 50 - (10+8))\log(1-\theta) \\ &= 18 \log(\theta) + 132 \log(1-\theta). \end{aligned} \tag{10.3.3}$$

Logaritmoidusta uskottavuusfunktiosta (10.3.3) laskettu θ:n suurimman uskottavuuden estimaatti on

$$\hat{\theta} = \frac{18}{150} = \frac{10}{100} + \frac{8}{50} = \frac{100}{150}\hat{\theta}_1 + \frac{50}{150}\hat{\theta}_2.$$

□

Esimerkki 10.4 Oloon $x_1, \ldots, x_n$ otos normaalijakaumasta $\mathrm{N}(1, \sigma^2)$, missä σ^2 on tuntematon. Yhteen havaintoon x_i perustuva σ^2:n uskottavuusfunktio on

$$L(\sigma^2; x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-1)^2}{2\sigma^2}\right],$$

ja vastaava logaritmoitu uskottavuusfunktio on

$$l(\sigma^2; x_i) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - 1)^2.$$

ja koko otokseen $x_1, \ldots, x_n$ perustuva σ^2:n uskottavuusfunktio on

$$L(\sigma^2; x_1, \ldots, x_n) = \prod_{i=1}^{n} L(\sigma^2; x_i) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-1)^2\right].$$

Tästä saadaan logaritmoitu kokonaisuskottavuus

$$\begin{aligned} \log L(\sigma^2; x_1, \ldots, x_n) &= \sum_{i=1}^{n} l(\sigma^2; x_i) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-1)^2. \end{aligned}$$

□

Esimerkki 10.5 Oloon $x_1, \ldots, x_n$ otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, missä sekä μ että σ^2 ovat tuntemattomia. Esimerkissä 10.4 vain parametri σ^2 oli tuntematon. Tässä tapauksessa uskottavuusfunktio ja logaritmoitu uskottavuusfunktio ovat kahden muuttujan funktiota:

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = \prod_{i=1}^{n} L(\mu, \sigma^2; x_i) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

ja

$$l(\mu, \sigma^2; x_1, \ldots, x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

$\square$

Esimerkki 10.6 Verrataan uuden hoitomenetelmän tehoa käytössä olevan "vanhan" menetelmän tehoon. Uutta hoitoa annettiin n_1:lle potilaalle, joista y_1 parani. Vastaavasti vanhaa hoitoa annettiin n_2:lle potilaalle, joista y_2 parani. Tällaisen hoitokokeen tulokset on esitetty Taulukossa 10.1.

Taulukko 10.1. Hoitokokeiden tulokset parantuneiden lukumäärän mukaan.

Käsittely	Uusi	Vanha
Parantuneiden lkm	y_1	y_2
Epäonnistumisten lkm	$n_1 - y_1$	$n_2 - y_2$
Potilaiden lkm	n_1	n_2

Oletamme, että parantuneiden lukumäärät Y_1 (uusi hoito) ja Y_2 (vanha hoito) noudattavat binomijakaumaa: $Y_i \sim \mathrm{Bin}(n_i, \pi_i)$, $i = 1, 2$. Oletetaan, että lukumäärät Y_1 ja Y_2 ovat toisistaan riippumattomat. Hoitojen tehoa voidaan vertailla onnistumistodennäköisyyksien π_1 (uusi hoito) ja π_2 (vanha hoito) avulla.

Näihin kahteen riippumattomaan kokeeseen perustuvat π_1:n ja π_1:n uskottavuusfunktiot ovat

$$L(\pi_i; y_i) = \binom{n_i}{y_i}\pi_i^{y_i}(1 - \pi_i)^{n_i - y_i}, \ i = 1, 2. \tag{10.3.4}$$

Koska uskottavuusfunktioissa (10.3.4) kerroin $\binom{n_i}{y_i}$ on vakio (parametrin suhteen), se voidaan lausekkeen yksinkertaistamiseksi poistaa kertomalla (10.3.4) vakiolla $1/\binom{n_i}{y_i}$. Kokeiden tulokset voidaan yhdistää yhteen kahden muuttujan uskottavuusfunktioon

$$\begin{aligned} L(\pi_1, \pi_2; y_1, y_2) &= L(\pi_1; y_1)\, L(\pi_2; y_2) \\ &= \pi_1^{y_1}\pi_2^{y_2}(1 - \pi_1)^{n_1 - y_1}(1 - \pi_2)^{n_2 - y_2}. \end{aligned}$$

Vastaava logaritmoitu uskottavuusfunktio on

$$\begin{aligned} l(\pi_1, \pi_2; y_1, y_2) &= l(\pi_1; y_1) + l(\pi_2; y_2) \\ &= \sum_{i=1}^{2} y_i \log \pi_i + \sum_{i=1}^{2} (n_i - y_i) \log(1 - \pi_i). \end{aligned}$$

Kokeella pyritään saamaan tietoa siitä, onko hoitojen tehossa eroa. Lähtökohtana on nollahypoteesi, että hoidoilla ei ole eroa:

$$H_0\colon \pi_1 = \pi_2 = \pi. \tag{10.3.5}$$

Jos nollahypoteesi pitää paikkansa, mallissa on vain yksi tuntematon parametri, koska onnistumistodennäköisyyksien yhteinen arvo on π. Silloin binomimuuttujien yhteenlaskuominaisuuden perusteella tiedämme, että

$$Y_1 + Y_2 \sim \mathrm{Bin}(n, \pi), \tag{10.3.6}$$

missä $n = n_1 + n_2$. Parametrin π logaritmoitu uskottavuusfunktio on

$$l(\pi) = (y_1 + y_2) \log \pi + (n - y_1 - y_2) \log(1 - \pi), \tag{10.3.7}$$

missä $n = y_1 + y_2$. Testauksessa verrataan kahdesta eri binomijakaumasta saatavia onnistumistodennäköisyyksiä yhdistetystä binomimallista (10.3.6) saatavaan onnistumistodennäköisyyteen.

Kahden hoitomenetelmän vertailutekniikka on suoraviivaisesti yleistettävissä usean menetelmän vetailuun, missä vetailtavien menetelmien lukumäärä on $k > 2$. Silloin tulokset ovat Taulukossa 13.1 esitettyä muotoa. Perusmallissa on k tuntematonta parametria ja logaritmoitu uskottavuus-

Taulukko 10.2. Hoitokokeiden tulokset parantuneiden lukumäärän mukaan.

Käsittely	1	2	...	k
Parantuneiden lkm	y_1	y_2	...	y_k
Epäonnistumisten lkm	$n_1 - y_1$	$n_2 - y_2$	...	$n_k - y_k$
Potilaiden lkm	n_1	n_2	...	n_k

funktio on

$$l(\pi_1, \ldots, \pi_k) = \sum_{i=1}^{k} y_i \log \pi_i + \sum_{i=1}^{k} (n_i - y_i) \log(1 - \pi_i).$$

Jos hypoteesi

$$H_0\colon \pi_1 = \cdots = \pi_k = \pi,$$

että hoidoilla ei ole eroa, pitää paikkansa, saadaan logaritmoitu uskottavuusfunktio

$$l(\pi) = y \log \pi + (n - y) \log(1 - \pi), \tag{10.3.8}$$

missä π on hoitojen yhteinen onnistumistodennäköisyys, $y = \sum_{i=1}^{k} y_i$ ja $n = \sum_{i=1}^{k} n_i$. □

Esimerkki 10.7 Tarkastellaan normaalista lineaarista regressiomallia, kuten esimerkissä 8.3. Oletetaan siis, että riippumattomat havainnot $Y_1, Y_2, \ldots, Y_n$ noudattavat normaalijakaumaa

$$Y_i \sim \mathrm{N}(\alpha + \beta x_i, \sigma^2), \quad 1 \le i \le n, \tag{10.3.9}$$

missä $E(Y_i) = \alpha + \beta x_i$, $i = 1, \ldots, n$. Havaittu aineisto muodostuu n:stä arvoparista $(x_1, y_1), \ldots, (x_n, y_n)$. Ennustemuuttujan x arvot $x_1, \ldots, x_n$ ajatellaan tunnetuiksi vakioiksi. Regressiofunktio on siis muotoa

$$E(Y \mid x) = \alpha + \beta x$$

ja kaikilla satunnaismuuttujilla Y_i on sama varianssi σ^2. Malli (10.3.9) voidaan myös lausua muodossa

$$Y_i = \alpha + \beta x_i + V_i, \qquad 1 \le i \le n,$$

missä $V_1, \ldots, V_n$ ovat riippumattomat ja $V_i \sim \mathrm{N}(0, \sigma^2)$, $1 \le i \le n$.

Havaintojen $Y_1, \ldots, Y_n$ yhteisjakauman tiheysfunktio on

$$\begin{aligned} f(y_1, \ldots, y_n \mid \alpha, \beta, \sigma^2) &= \prod_{i=1}^{n} f_i(y_i \mid \alpha, \beta, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right]. \end{aligned}$$

Tästä nähdään, että uskottavuusfunktio on

$$L(\alpha, \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right]$$

ja logaritmoitu uskottavuusfunktio on

$$l(\alpha, \beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2. \tag{10.3.10}$$

□

Esimerkki 10.8 Esimerkissä 8.8 tarkasteltiin lentokoneiden rakennuksessa käytettävien metallin kiinnittimien kestävyyttä logistisen regressiomallin avulla. Selittävänä muuttujana x on kiinnittimeen kohdistuva kuormitus (psi, paunaa/per neliötuuma) ja selitettävänä muuttujana y on särkyvien kiinnittimien lkm annetulla kuormituksella. Oletetaan, että särkyvien kiinnittimien lukumäärät $Y_1, \ldots, Y_{10}$ ovat riippumattomat ja noudattavat binomijakaumaa eli

$$Y_i \sim \mathrm{Bin}[n_i, \pi(x_i)], \; i = 1, \ldots, 10 \tag{10.3.11}$$

ja $Y_i \perp\!\!\!\perp Y_j$. Rikkoontumistodennäköisyys $\pi(x_i)$ riippuu kuormituksesta x_i. Kokeessa käytettiin kymmentä eri kuormitustasoa. Esimerkisi kuormitustasolla $x_1 = 25\,psi$ testattiin $n_1 = 50$ kiinnitintä, joista $y_1 = 10$ rikkoontui. Korkeimmalla kuormitustasolla $x_{10} = 43\,psi$ testattiin $n_{10} = 65$ kiinnitintä, joista rikkoontui $y_{10} = 51$. Testattavia kiinnittimiä oli yhteensä $n = 690$ ja $n = \sum_{i=1}^{10} n_i$.

Koska lukumäärät $Y_1, \ldots, Y_k$ ($k = 10$) ovat riippumattomat, on niiden yhteisjakauman todennäköisyysfunktio on binomioletuksen (10.3.11) nojalla

$$\prod_{i=1}^{k} \binom{n_i}{y_i} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{n_i - y_i}.$$

Havaintoihin $y_1, \ldots, y_k$ perustuva parametrien $\pi(x_1), \ldots, \pi(x_k)$ uskottavuusfunktio voidaan kirjoittaa muodossa

$$L[\pi(x_1), \ldots, \pi(x_k); y_1, \ldots, y_k] = \prod_{i=1}^{k} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{n_i - y_i},$$

kun yhteisjakaumassa esiintyvä vakio $\prod_{i=1}^{k} \binom{n_i}{y_i}$ jätetään pois. Tästä saadaan logaritmoitu uskottavuusfunktio

$$\begin{aligned} l[\pi(x_1), \ldots, \pi(x_k)] &= \sum_{i=1}^{k} y_i \log \pi(x_i) + \sum_{i=1}^{k} (n_i - y_i) \log[1 - \pi(x_i)] \\ &= \sum_{i=1}^{k} y_i \log \frac{\pi(x_i)}{1 - \pi(x_i)} + \sum_{i=1}^{k} n_i \log[1 - \pi(x_i)]. \end{aligned} \tag{10.3.12}$$

Logistisessa regressiomallissa vedon logaritmi

$$\log \frac{\pi(x_i)}{1 - \pi(x_i)} = \alpha + \beta x_i, \quad 1 \le i \le k \tag{10.3.13}$$

oletetaan lineaariseksi. Tästä saadaan rikkoontumistodennäköisyydet

$$\pi(x_i) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}, \quad 1 \le i \le k.$$

Kun todennäköisyydet $\pi(x_i)$, $1 \le i \le k$, lausutaan mallin (10.3.13) mukaisesti parametrien α ja β funktiona ja sijoitetaan lausekkeeseen (10.3.12), saadaan parametrien α ja β uskottavuusfunktio

$$l(\alpha, \beta) = y\alpha + (\sum_{i=1}^{k} x_i y_i)\beta - \sum_{i=1}^{k} n_i \log[1 + \exp(\alpha + \beta x_i)], \tag{10.3.14}$$

missä $y = \sum_{i=1}^{k} y_i$. Funktio $l(\alpha, \beta)$ on nyt α:n ja β:n suhteen epälineaarinen funktio. □

10.4 Yhteys Bayesilaiseen lähestymistapaan

Bayesilaisessa lähestymistavassa tarvitaan parametrin θ priorijakauma $g(\theta)$, jonka avulla voidaan määrittää posteriorijakauma

$$\begin{aligned} f(\theta|x) &= \text{vakio} \times g(\theta) f(x;\theta) \\ &= \text{vakio} \times g(\theta)\, L(\theta) \end{aligned}$$

Bayesilainen menetelmä yhdistää siis informaatiota samalla tavalla kuin uskottavuusmenetelmä: se yhdistää priorin ja uskottavuusfunktion kertomalla ne keskenään.

10.5 Uskottavuussuhde

Oletetaan, että $y = h(x)$ on havainnon x yksi-yksinen eli bijektiivinen muunnos (vrt. Alaluku 5.5.2 s. 173). Silloin on olemassa sellainen funktio $x = g(y)$, että

$$y = h(x) \Leftrightarrow x = g(y).$$

Jos x on jatkuva, niin

$$f_Y(y;\theta) = f_X\big(g(y);\theta\big)|g'(y)|,$$

missä $g'(y) = \frac{\mathrm{d}x}{\mathrm{d}y}$. Silloin muunnettuun havaintoon y perustuva uskottavuus on

$$L(\theta; y) = L(\theta; x)|g'(y)|.$$

On selvää, että x:n ja y:n pitäisi sisältää sama parametria θ koskeva informaatio. Kun parametrin arvoja θ_1 ja θ_2 verrataan uskottavuusfunktion avulla, saadaan

$$\frac{L(\theta_2; y)}{L(\theta_1; y)} = \frac{L(\theta_2; x)}{L(\theta_1; x)}.$$

Vertailtaessa eri parametrin arvoja samassa mallissa, vain uskottavuuksien suhteella on merkitystä, koska vakiot supistuvat pois. Tämä tarkoittaa, että uskottavuusfunktiota tarkasteltaessa voidaan jättää pois sellaiset termit, jotka eivät sisällä parametria θ. Uskottavuusfunktio on siis vakiota vaille yksikäsitteinen. Jos uskottavuusfunktio halutaan tehdä yksikäsitteiseksi, se tavallisesti normeerataan siten, että sen maksimiarvo tulee ykköseksi.

Esimerkki 10.9 Jos x on otos binomijakausta $\mathrm{Bin}(n, \theta)$, niin uskottavuusfunktio on

$$L(\theta) = \theta^x (1-\theta)^{n-x},$$

mistä on jätetty pois kaikki vakiotermit. Tätä muotoa sanotaan usein uskottavuusfunktion ytimeksi. Vastaavasti logaritmoitu uskottavuusfunktio on

$$l(\theta) = \log L(\theta) = x \log(\theta) + (n-x)\log(1-\theta).$$

□

10.6 Uskottavuusfunktion maksimi ja kaarevuus

Suurimman uskottavuuden estimaatti (SUE) antaa tarkastelun kohteena olevan parametrin piste-estimaatin, vaikka uskottavuusmenetelmän käyttö ei rajoitu piste-estimaattien määrittämiseen. Fisher (1922) esitti uskottavuusfunktion suurimman uskottavuuden estimointimenetelmän yhteydessä. Suurimman uskottavuuden estimaatin $\hat{\theta}$ laskemiseksi on siis määritettävä uskottavuusfunktion $L(\theta)$, tai vastaavasti logaritmoidun uskottavuusfunktion $L(\theta)$, maksimikohta parametriavaruudessa Θ.

Uskottavuusfunktio tiivistää havaintojen sisältämän parametreja koskevan informaation. SUE on yksi keino luonnehtia uskottavuusfunktiota. Yleisesti yksi luku ei riitä luonnehtimaan funktiota. Jos logaritmoitu uskottavuusfunktio on kvadraattinen tai jokin kvadraattinen funktio on sen hyvä likiarvo, niin funktion luonnehtimisen tarvitaan vähintään 2 lukua: maksimin sijainti ja funktion kaarevuus maksimissa. Jos logaritmoidulla uskottavuusfunktiolla on maksimikohdan yhpäristössä hyvä kavdraattinen likiarvo, sanomme uskottavuusfunktiota *s*äännölliseksi.

Tavallisesti θ:n SUE $\hat{\theta}$ voidaan määrittää derivointikeinolla. Huomattakoon kuitenkin, ettei tätä keinoa voida tietenkään aina soveltaa. Se ei kuitenkaan tarkoita sitä, etteikö SUE silti voisi olla olemassa. Funktion $l(\theta)$ ensimmäistä derivaattaa kutsutaan *Fisherin pistefunktioksi* tai lyhyesti vain pistefunktioksi ja se määritellään seuraavasti:

$$S(\theta; x) \equiv \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta; x), \tag{10.6.1}$$

missä $l(\theta; x) = \log L(\theta; x)$. Merkitsemme myös lyhyesti $S(\theta) = l'(\theta)$. Parametrin θ SUE $\hat{\theta}$ on *pisteyhtälön*

$$S(\theta) = 0$$

ratkaisu.

Maksimissa logaritmoidun uskottavuusfunktion 2. derivaatta on negatiivinen. Logaritmoidun uskottavuusfunktion kaarevuus pisteessä $\hat{\theta}$ määritellään suureena $I(\hat{\theta})$, missä

$$I(\theta) \equiv -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log L(\theta) = -l''(\theta).$$

Suuri kaarevuus kertoo, että funktiolla on terävä huippu. Tilastollisesti tulkittuna se tarkoittaa suurempaa varmuutta parametrin arvosta. Suuretta

$I(\hat{\theta})$ kutsutaan uskottavuusteoriassa *havaituksi Fisherin informaatioksi* tai lyhyemmin *havaituksi informaatioksi.*

Tavallisesti parametriavaruus Θ on jokin reaalilukuväli ja $l(\theta)$:n 1. ja 2. derivaatta ovat olemassa kaikissa Θ:n sisäpisteissä. Jos nyt $\hat{\theta}$ on Θ:n sisäpiste, niin $l'(\hat{\theta}) = 0$ ja $l''(\hat{\theta}) < 0$. Näiden ehtojen vallitessa on siis

$$S(\hat{\theta}) = 0 \qquad \text{ja} \qquad I(\hat{\theta}) > 0. \tag{10.6.2}$$

Jos edellä mainitut säännöllisyysoletukset eivät pidä paikkaansa, ei $\hat{\theta}$ välttämättä ole uskottavuusyhtälön ratkaisu.

Esimerkki 10.10 Binomijakauman tapauksessa (Esimerkki 8.10)

$$S(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}, \qquad I(\theta) = \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}, \qquad 0 < \theta < 1.$$

Kun $1 \leq x \leq n-1$, niin uskottavuusyhtälöllä $S(\theta) = 0$ on yksikäsitteinen ratkaisu $\theta = x/n$. Koska $I(\theta) > 0$ pisteessä $\theta = x/n$, niin funktiolla $l(\theta)$ [ja funktiolla $L(\theta)$] on maksimi pisteessä $\theta = x/n$. Koska $L(0) = L(1) = 0$, niin $\hat{\theta} = x/n$ on globaali maksimi ja siksi θ:n suurimman uskottavuuden estimaatti. Kun $x = 0$, niin uskottavuusyhtälöllä ei ole ratkaisua, mutta uskottavuusfunktio on silloin

$$L(\theta) = (1-\theta)^n, \qquad 0 \leq \theta \leq 1.$$

Nähdään helposti, että $L(0) = \max_\theta L(\theta)$, joten $\hat{\theta} = 0$. Vastaavasti kun $x = n$, niin $\hat{\theta} = 1$. Näin siis kaava $\hat{\theta} = x/n$ pätee kaikilla havaintoarvoilla, vaikka kaikkia estimaatteja ei saada uskottavuusyhtälön ratkaisuna. □

Esimerkki 10.11 Oletetaan, että puhelinvaihteeseen päivän aikana tulevien "väärien" puheluiden lukumäärä noudattaa Poissonin jakaumaa, jonka odotusarvo on μ. Oletetaan, että jakauma on kaikkina päivinä sama ja eri päivien havainnot ovat toisistaan riippumattomat. Olkoot $x_1, x_2, \ldots, x_n$ eri päivinä havaitut virhepuheluiden lukumäärät. Havainnon x_i todennäköisyys on

$$f(x_i;\mu) = \frac{1}{x_i!}\mu^{x_i}\mathrm{e}^{-\mu}, \qquad x_i = 0, 1, 2, \ldots.$$

Koska eri päivinä havaittujen virhepuheluiden lukumäärät ovat toisistaan riippumattomat, niin otoksen $x_1, x_2, \ldots, x_n$ todennäköisyys on

$$\begin{aligned} f(x_1, x_2, \ldots, x_n; \mu) &= f(x_1;\mu)f(x_2;\mu)\cdots f(x_n;\mu) \\ &= \prod_{i=1}^{n} \frac{1}{x_i!}\mu^{x_i}\mathrm{e}^{-\mu} = \frac{1}{x_1!x_2!\ldots x_n!}\mu^{\sum x_i}\mathrm{e}^{-n\mu}. \end{aligned}$$

Koska uskottavuus ei riipu vakiosta $1/(x_1!x_2!\cdots x_n!)$, se voidaan jättää tarkasteluista pois. Tämä turhista vakioista "puhdistettu"uskottavuusfunktion ydin on

$$L(\mu) = \mu^{\sum x_i}\mathrm{e}^{-n\mu}, \qquad 0 \leq \mu < \infty.$$

Vastaava logaritmoitu uskottavuusfunktio on

$$l(\mu) = \sum x_i \log(\mu) - n\mu.$$

Pistefunktio ja informaatiofunktio ovat vastaavasti

$$S(\mu) = \frac{1}{\mu}\sum x_i - n \qquad \text{ja} \qquad I(\mu) = \frac{\sum x_i}{\mu^2}.$$

Jos $\sum x_i > 0$, niin uskottavuusyhtälöllä $S(\mu) = 0$ on yksikäsitteinen ratkaisu $\mu = \sum x_i/n = \bar{x}$. Koska $I(\mu) > 0$ pisteessä $\mu = \bar{x}$, niin $\bar{x}$ on maksimikohta. Se on myös globaali maksimi, koska $L(0) = 0$ ja $L(\mu) \to 0$, kun $\mu \to \infty$. Jos $\sum x_i = 0$, niin uskottavuusyhtälöllä $S(\mu) = 0$ ei ole ratkaisua, mutta uskottavuusfunktio saavuttaa maksiminsa pisteessä $\mu = 0$. Näin siis $\hat{\mu} = \bar{x}$ kaikilla havaintoarvoilla. Havaintojen x_1, x_2, …, x_n todennäköisyys maksimoituu, kun jakauman tuntematon parametri μ estimoidaan otoskeskiarvolla $\bar{x}$. □

Esimerkki 10.12 Olkoon $x_1, \dots, x_n$ havaittu otos normaalijakaumasta $\mathrm{N}(\theta, \sigma^2)$. Oletamme tässä, että σ^2 on tunnettu vakio. Kun θ:sta riippumattomat termit jätetään pois, saadaan

$$l(\theta) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2,$$

josta seuraa

$$S(\theta) = l'(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \theta).$$

Ratkaisemalla yhtälö $S(\theta) = 0$ saadaan θ:n SU-estimaatiksi $\hat{\theta} = \bar{x}$. Havaituksi informaatioksi saadaan

$$I(\hat{\theta}) = \frac{n}{\sigma^2}.$$

□

Arvioidaan logaritmoitua uskottavuusfunktiota 2. asteen Taylorin polynomilla SUE:n $\hat{\theta}$ ympäristössä. Olkoon $l(\theta)$ parametriavaruudessa Θ määritelty jatkuvan parametrin θ logaritmoitu uskottavuusfunktio. Vastaavasti θ:n pistefunktio on $S(\theta) = l'(\theta)$ ja informaatiofunktio $I(\theta) = -l''(\theta)$. Oletetaan, että $\hat{\theta}$ on olemassa jossain Θ:n sisäpisteessä. Arvioidaan log-uskottavuutta 2. asteen Taylorin polynomilla SUE:n $\hat{\theta}$ ympäristössä. Silloin

$$l(\theta) \approx l(\hat{\theta}) + S(\hat{\theta})(\theta - \hat{\theta}) - \frac{1}{2}I(\hat{\theta})(\theta - \hat{\theta})^2,$$

josta saadaan

$$\log\frac{L(\theta)}{L(\hat{\theta})} \approx -\tfrac{1}{2}I(\hat{\theta})(\theta - \hat{\theta})^2. \qquad (10.6.3)$$

Toisen asteen likiarvon tarkkuutta voidaan tutkia piirtämällä logaritmoitu uskottavuusfunktio ja sen likiarvo samaan kuvioon. Normaalijakauman tapauksessa uskottavuusfunktio on kvadraattinen. Likiarvo mittaa tavallaan uskottavuusfunktion "etäisyyttä" normaalijakauman uskottavuusfunktiosta. Kun likiarvo (10.6.3) derivoidaan puolittain, saadaan

$$S(\theta) \approx -I(\hat{\theta})(\theta - \hat{\theta})$$

tai

$$-\frac{S(\theta)}{I^{1/2}(\hat{\theta})} \approx I^{1/2}(\hat{\theta})(\theta - \hat{\theta})$$

10.7 Uskottavuuden invarianssi

Uskottavuusfunktio ilmaisee kiinteää parametria koskevaa epävarmuutta. Tarkastelemme nyt, kuinka uskottavuusfunktiossa käsitellään parametrin muunnosta. Oletamme aluksi bijektiivisen muunnoksen, mutta periaate pätee yleisemminkin. Esimerksi binomijakaumassa $\mathrm{Bin}(10, \theta)$ on uskottavuussuhde $\theta_1 = 0.8$ vastaan $\theta_2 = 0.3$

$$\frac{L(\theta_1 = 0.8)}{L(\theta_2 = 0.3)} = \frac{\theta_1^8(1-\theta_1)^2}{\theta_2^8(1-\theta_2)^2} = 208.7,$$

kun havainto $x = 8$. Parametrin arvo $\theta = 0.8$ on siis noin 200 kertaa uskottavampi kuin arvo $\theta = 0.3$.

Esimerkki 10.13 Monissa sovelluksissa binomijakauman parametri esitetään logit-asteikolla siten, että

$$\psi \equiv \log\left(\frac{\theta}{1-\theta}\right).$$

Silloin $\psi_1 = \log(0.8/0.2) = 1.39$ ja $\psi_2 = \log(0.3/0.7) = -0.85$, missä $\theta_1 = 0.8$ ja $\theta_2 = 0.3$ kuten edellä. Nyt θ on ψ:n funktiona

$$\theta = \frac{e^{\psi}}{1 + e^{\psi}}.$$

Kun uskottavuusfunktio esitetään ψ:n funktiona ja merkitään sitä $L^*(\psi)$, saadaan uskottavuussuhde

$$\frac{L^*(\psi_1)}{L^*(\psi_2)} = \frac{L(\theta_1)}{L(\theta_2)} = 208.7.$$

Uskottavuussuhde ei muutu parametrimuunnoksessa. Jos θ_i on tapahtuman A_i todennäköisyys $i = 1, 2$, niin $\theta_i/(1-\theta_i$ on tapahtuman A_i *veto*. Tapahtumien A_1 ja A_2 *veto*suhde on

$$\frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}.$$

Logit-asteikko antaa vedon arvojen logaritmit. □

Oletetaan esimerkiksi, että erään elektronisen komponentin elinikä noudattaa eksponenttijakaumaa $\mathrm{Exp}(\theta)$, jolloin sen tiheysfunktio on

$$f(x) = \frac{1}{\theta}\mathrm{e}^{-x/\theta}, \qquad 0 < x < \infty. \tag{10.7.1}$$

Jokaista parametrin θ arvoa vastaa yksi jakauma. Olemme esittäneet eksponenttijakauman tiheysfunktion myös muodossa

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \qquad 0 < x < \infty, \tag{10.7.2}$$

missä $\lambda = 1/\theta$. Tavallisesti parametrisointi valitaan siten, että parametri esittää jotain tärkeää jakauman ominaisuutta tai jakauman matemaattinen esitystapa saadaan yksinkertaiseksi. Parametrisoinnissa (10.7.1) θ on jakauman keskiarvo.

Jos esimerkiksi $\theta = 2$, niin $\lambda = \frac{1}{2}$. Jokaista θ:n valintaa vastaa yksikäsitteinen jakauma parametrisoinnissa (10.7.2) ja kääntäen. Uskottavuusmenetelmällä on se miellyttävä piirre, että menetelmä on invariantti bijektiivisten (yksi-yksisten) parametrimuunnosten suhteen.

10.7.1 Uskottavuus uudessa parametrisoinnissa

Olkoon $\psi \equiv g(\theta)$ parametrimuunnos ja $L^*(\psi)$ parametrin ψ uskottavuusfunktio ja $L(\theta)$ parametrin θ uskottavuusfunktio. Tarkastellaan paria $\{\theta, L(\theta)\}$ parametrin θ uskottavuusfunktion kuvaajana. Silloin parametrin ψ kuvaaja on yksinkertaisesti

$$\begin{aligned}\{\psi, L^*(\psi)\} &= \{g(\theta), L(g(\theta))\}\\ &= \{g(\theta), L(\theta)\},\end{aligned}$$

koska uskottavuuden invarianssiperiaatteen mukaan

$$L^*(\psi) = L(\theta).$$

Tästä seuraa suurimman uskottavuuden estimaatin tärkeä invarianttisuus.

Lause 10.1 *Jos $\hat{\theta}$ on θ:n suurimman uskottavuuden estimaatti ja $\psi = g(\theta)$, niin $\hat{\psi} = g(\hat{\theta})$ on ψ:n suurimman uskottavuuden estimaatti.*

Funktion $g(\theta)$ ei tarvitse olla bijektio. Jos esimerkiksi

$$\psi = g(\theta) = \theta^2,$$

niin esimerkiksi $g(-1) = g(1) = 1$. Jos $L(\theta = 1) = 0.5$ ja $L(\theta = -1) = 0.3$, niin mitä on $L^*(\psi = 1)$? Tässä tapauksessa määritellään

$$\begin{aligned}L^*(\psi = 1) &= \max_{\{\theta, g(\theta)=1\}} L(\theta)\\ &= \max\{0.5, 0.3\}\\ &= 0.5.\end{aligned}$$

Jos siis $\hat{\theta}$ on θ:n suurimman uskottavuuden estimaatti, niin $\hat{\theta}^2$ on θ^2:n suurimman uskottavuuden estimaatti.

Esimerkki 10.14 Jos binomijakaumasta $\mathrm{Bin}(10, \theta)$ saadaan havainto $x = 8$, θ:n SUE on $\hat{\theta} = 0.8$. Silloin parametrin $g(\theta) = \theta/(1-\theta)$ SUE on $\hat{\theta}/(1-\hat{\theta}) = 0.8/0.2 = 4$. □

Huomattakoon, että vastaavaa invarianttisuusominaisuutta ei muilla estimaattoreilla välttämättä ole. Jos esimerkiksi $\tilde{\theta}$ on θ:n harhaton minimivarianssinen estimaattori, ei $g(\tilde{\theta})$ ole yleisesti $g(\theta)$:n harhaton minimivarianssinen estimaattori.

10.8 Pistesuureen jakauma

Olkoon $X_1, X_2, \ldots, X_n$ otos jostain jakaumasta eli X_1, X_2, …, X_n ovat riippumattomat ja noudattavat samaa jakaumaa. Alaluvussa 10.3 tarkasteltiin riippumattomien otosten yhdistämistä. Koska X_1, …, X_n ovat riippumattomat, niin havaintojen $X_1 = x_1$, …, $X_n = x_n$ perusteella määritetty uskottavuusfunktio on muotoa

$$L(\theta; x_1, x_2, \ldots, x_n) = L(\theta; x_1)\, L(\theta; x_2) \cdots L(\theta; x_n).$$

Vastaavasti logaritmoitu uskottavuusfunktio on

$$l(\theta; x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} l(\theta; x_i),$$

missä $l(\theta; x_i) = \log L(\theta; x_i)$, $1 \le i \le n$. Suoraan pistefunktion määritelmästä seuraa, että

$$S(\theta; x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} S(\theta; x_i), \tag{10.8.1}$$

missä $S(\theta; x_i) = l'(\theta; x_i)$.

Kun pistefunktion $S(\theta; x_1, x_2, \ldots, x_n)$ esityksessä (10.8.1) annetaan parametrille jokin arvo $\theta = \theta_0$ ja tarkastellaan havaintoja satunnaismuuttujina, saadaan satunnaismuuttuja

$$S(\theta_0; X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} S(\theta_0; X_i). \tag{10.8.2}$$

Myös satunnaismuuttujaa (10.8.2) kutsutaan pistefunktioksi. Jatkossa pistefunktiota (10.8.2) merkitään usein lyhyesti $S(\theta_0; X_1, X_2, \ldots, X_n) \equiv S(\theta_0)$.

Vastaavasti voidaan myös informaatiofunktiota tarkastella satunnaismuuttujana. Silloin saamme

$$I(\theta_0; X_1, X_2, \ldots, X_n) = -\sum_{j=1}^{n} S'(\theta_0; X_j) = \sum_{j=1}^{n} I(\theta_0, X_j), \tag{10.8.3}$$

missä $S'(\theta_0; X_j)$ on pistefunktion $S(\theta; X_j)$ derivaatan arvo pisteessä $\theta = \theta_0$ ja $I(\theta_0, X_j)$ on yhteen havaintoon X_j perustuva informaatio. Informaation (10.8.3) odotusarvoa

$$\mathcal{I}(\theta_0) = E[I(\theta_0; X_1, X_2, \ldots, X_n)] \tag{10.8.4}$$

kutsutaan *odotetuksi Fisherin informaatioksi* tai lyhyesti *Fisherin informaatioksi*. Osoitamme alaluvussa 10.9.2, että $S(\theta_0)$:n odotusarvo on 0 ja varianssi on odotettu informaatio $\mathcal{I}(\theta_0)$. Keskeisen rajaväittämän mukaan

$$\frac{S(\theta_0)}{\sqrt{\mathcal{I}(\theta_0)}} \approx \mathrm{N}(0, 1),$$

kun n on riittävän suuri.

Esimerkki 10.15 Olkoon $X_1, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\theta, 1)$. Silloin

$$l(\theta) = -\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2$$

ja $l'(\theta) = \sum_{i=1}^{n}(X_i - \theta)$. Kun $\theta = \theta_0$, niin

$$l'(\theta_0) = S(\theta_0; X_1, \ldots, X_n) = \sum_{i=1}^{n}(X_i - \theta_0).$$

Koska $l''(\theta) = -n$, niin $\mathcal{I}(\theta) = I(\theta) = n$. Nyt siis

$$\frac{S(\theta_0; X_1, \ldots, X_n)}{\sqrt{\mathcal{I}(\theta_0)}} = \frac{\sum_{i=1}^{n}(X_i - \theta_0)}{\sqrt{n}} = \frac{\overline{X} - \theta_0}{1/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

□

Esimerkki 10.16 Olkoon $X_1, \ldots, X_n$ otos Poissonin jakaumasta $\mathrm{Poi}(\mu)$. Silloin

$$l(\mu) = \Big(\sum_{i=1}^{n} X_i\Big)\log\mu - n\mu, \quad l'(\mu) = \frac{\sum X_i}{\mu} - n \quad \text{ja} \quad l''(\mu) = -\frac{\sum X_i}{\mu^2},$$

joten $I(\mu) = \sum X_i/\mu^2$. Odotettu informaatio on siis

$$\mathcal{I}(\mu) = E\Big(\frac{\sum X_i}{\mu^2}\Big) = \frac{n\mu}{\mu^2} = \frac{n}{\mu}.$$

Kun oletetaan $\mu = \mu_0$, niin pistefunktio on

$$S(\mu_0; X_1, \ldots, X_n) = \frac{\sum X_i - n\mu_0}{\mu_0} = \frac{n}{\mu_0}(\overline{X} - \mu_0).$$

Tästä seuraa, että

$$\frac{S(\mu_0)}{\sqrt{\mathcal{I}(\mu_0)}} = \frac{(n/\mu_0)(\overline{X} - \mu_0)}{\sqrt{n/\mu_0}} = \frac{\overline{X} - \mu_0}{\sqrt{\mu_0/n}}.$$

Koska $\mathrm{Var}(\overline{X}) = \mu_0/n$, niin keskeisen rajaväittämän mukaan

$$\frac{\overline{X} - \mu_0}{\sqrt{\mu_0/n}} \xrightarrow{d} \mathrm{N}(0,1), \qquad \text{kun } n \to \infty.$$

□

Esimerkki 10.17 Kun $X_1, \ldots, X_n$ on otos eksponenttijakaumasta $\mathrm{Exp}(\theta)$, niin

$$l(\theta) = -n \log \theta - \frac{\sum X_i}{\theta}$$

ja

$$l'(\theta) = -\frac{n}{\theta} + \frac{\sum X_i}{\theta^2} = \frac{1}{\theta^2}\Big(\sum X_i - n\theta\Big) = \frac{n}{\theta^2}(\overline{X} - \theta).$$

Koska $l''(\theta) = n/\theta^2 - \big(2\sum X_i\big)/\theta^3$, niin

$$\mathcal{I}(\theta) = E\bigg(\frac{2\sum X_i}{\theta^3} - \frac{n}{\theta^2}\bigg) = \frac{2n\theta}{\theta^3} - \frac{n}{\theta^2} = \frac{n}{\theta^2}.$$

Tästä seuraa, että kiinnitetyllä θ:n arvolla $\theta = \theta_0$

$$\frac{S(\theta_0)}{\sqrt{\mathcal{I}(\theta_0)}} = \frac{(n/\theta_0^2)(\overline{X} - \theta_0)}{\sqrt{n/\theta_0^2}} = \frac{\overline{X} - \theta_0}{\sqrt{\theta_0^2/n}},$$

joka keskeisen rajaväittämän mukaan lähenee normaalijakaumaa $\mathrm{N}(0,1)$, kun n kasvaa. □

Suurimman uskottavuuden estimaatti on yhtälön $S(\hat{\theta}) = 0$ ratkaisu. Voidaan osoittaa, että $\hat{\theta} \to \theta$ todennäköisyyden mielessä , kun $n \to \infty$. Suurimman uskottavuuden estimaattori $\hat{\theta}$ on siis θ:n tarkentuva estimaattori. Lisäksi voidaan osoittaa, että $I(\hat{\theta})/\mathcal{I}(\theta) \to 1$ todennäköisyyden mielessä, kun $n \to \infty$. Edellisissä esimerkeissä olemme havainneet tuloksen $\mathrm{Var}[S(\theta)] = \mathcal{I}(\theta)$, joka pitää paikkansa yleisesti (tiettyjen ehtojen vallitessa). Siksi

$$\frac{S(\theta)}{\sqrt{I(\hat{\theta})}} \approx \mathrm{N}(0,1), \tag{10.8.5}$$

kun n on riittävän suuri.

Koska $\hat{\theta}$ on lähellä arvoa θ, kun n on suuri, niin likiarvon (10.6.3) mukaan

$$l(\theta) - l(\hat{\theta}) \approx -\tfrac{1}{2}(\hat{\theta} - \theta_0)^2 I(\hat{\theta}).$$

Derivoimalla θ:n suhteen saadaan

$$S(\theta) \approx (\hat{\theta} - \theta) I(\hat{\theta}).$$

Silloin tuloksen (10.8.5) mukaan

$$(\hat{\theta} - \theta)\sqrt{I(\hat{\theta})} \approx \mathrm{N}(0, 1). \tag{10.8.6}$$

Tästä saadaan uskottavuussuureen D likiarvo

$$D = -2[l(\theta) - l(\hat{\theta})] \approx (\hat{\theta} - \theta)^2 \mathcal{I}(\hat{\theta}).$$

Koska (10.8.6):n mukaan $(\hat{\theta} - \theta)\sqrt{I(\hat{\theta})}$ noudattaa likimain normaalijakaumaa $\mathrm{N}(0, 1)$, niin D noudattaa likimain χ^2-jakaumaa vapausastein 1, kun n on suuri.

10.9 Suurimman uskottavuuden menetelmän ominaisuuksia

10.9.1 Odotettu informaatio ja kokeiden suunnittelu

Toistaiseksi on yleensä oletettu, että koe on tehty (havainnot käytettävissä). Olemme tarkastelleet sitä, miten havainnoista saadaan parametria θ koskevaa informaatiota. Tilastollisia menetelmiä voidaan käyttää myös kokeiden suunnitteluvaiheessa päätettäessä siitä, mitä kokeita kannattaa tehdä ja mitä estimaattoreita käyttää. Suuri havaittu Fisherin informaatio $I(\hat{\theta})$ tarkoittaa sitä, että $\hat{\theta}$ on hyvä θ:n estimaatti. Jos siis kokeen tarkoituksena on saada hyvä θ:n estimaatti, kannattaa yrittää valita koe siten, että saadaan mahdollisimman suuri $I(\hat{\theta})$. Emme kuitenkaan voi laskea esimaattia $\hat{\theta}$ ja vastaavasti havaittua Fisherin informaatiota $I(\hat{\theta})$ ennen kuin havainnot ovat käytettävissä eli ennen koetta.

Esimerkki 10.18 Testataan n:n tuotteen kestoaika tarkistamalla, moniko niistä kestää tietyn ennalta asetetun testiajan t_0. Olkoon Y testiajan t_0 kestävien tuotteiden lukumäärä. Oletetaan, että kestoajat X_i noudattavat toisistaan riippumatta eksponenttijakaumaa $\mathrm{Exp}(\theta)$. Miten t_0 olisi valittava, että odotettu informaatio olisi mahdollisimman suuri?

Koska $X \sim \mathrm{Exp}(\theta)$, niin

$$\pi(\theta) = P(X > t_0) = \int_{t_0}^{\infty} \frac{1}{\theta} \mathrm{e}^{-x/\theta}\, \mathrm{d}x = \mathrm{e}^{-t_0/\theta}.$$

Testiajan t_0 kestävien tuotteiden lukumäärä Y noudattaa binomijakaumaa $\mathrm{Bin}(n, \pi)$, jonka logaritmoitu uskottavuusfunktio on

$$l(\theta) = Y \log \pi(\theta) + (n - Y) \log[1 - \pi(\theta)],$$

missä $\pi(\theta) = \mathrm{e}^{-t_0/\theta}$. Derivoimalla $l(\theta)$ kaksi kertaa saadaan

$$I(\theta) = \left[\frac{Y}{\pi^2} + \frac{n-Y}{(1-\pi)^2}\right]\left(\frac{\mathrm{d}\pi}{\mathrm{d}\theta}\right)^2 - \left(\frac{Y}{\pi} - \frac{n-Y}{1-\pi}\right)\frac{\mathrm{d}^2\pi}{\mathrm{d}\theta^2}.$$

Koska $E(Y) = n\pi$, niin

$$\begin{aligned}\mathcal{I}(\theta) = E[I(\theta)] &= \frac{n}{\pi(1-\pi)}\left(\frac{\mathrm{d}\pi}{\mathrm{d}\theta}\right)^2 \\ &= \frac{n\pi t_0^2}{(1-\pi)\theta^4} = \frac{n}{\theta^2}\cdot\frac{\pi}{1-\pi}(\log\pi)^2.\end{aligned}$$

Voidaan osoittaa, että $I(\theta)$ saavuttaa maksiminsa, kun $\pi = 0.203$. Siksi kannattaa yrittää valita t_0 siten, että noin 20 % testattavista tuotteista kestää koko testijakson ajan. □

10.9.2 Pistefunktion ja informaatiofunktion ominaisuuksia

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

$$S(\theta) = S(\theta; x) = \frac{\mathrm{d}}{\mathrm{d}\theta}\, l(\theta; x),$$

missä $l(\theta; x) = \log f(x;\theta)$. Tarkastellaan nyt funktiota satunnaismuuttujana $S = S(\theta; X)$, kuten alaluvussa 10.8.

Esimerkki 10.19 Olkoon $X_1, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\mu, \sigma_0^2)$, missä σ_0^2 on tunnettu. Silloin μ:n logaritmoitu uskottavuusfunktio on

$$l(\mu) = -\frac{1}{2\sigma_0^2}\sum(X_i - \mu)^2$$

ja

$$S = \frac{1}{\sigma_0^2}\sum(X_i - \mu),$$

josta saadaan

$$E(S) = E\left[\frac{1}{\sigma_0^2}\sum(X_i-\mu)\right] = \frac{1}{\sigma_0^2}\sum E(X_i - \mu) = 0$$

ja

$$\begin{aligned}\mathrm{Var}(S) = \mathrm{Var}\left[\frac{1}{\sigma_0^2}\sum(X_i-\mu)\right] &= \frac{1}{\sigma_0^4}\,\mathrm{Var}\left[\sum(X_i-\mu)\right] \\ = \frac{\sum \mathrm{Var}(X_i-\mu)}{\sigma_0^4} &= \frac{n\sigma_0^2}{\sigma_0^4} = \frac{n}{\sigma_0^2}.\end{aligned}$$

Toisaalta $l''(\mu) = -n/\sigma_0^2$, joten $I(\mu) = n/\sigma_0^2$ ei riipu havainnoista ja silloin myös $\mathcal{I}(\mu) = n/\sigma_0^2$. Havaitsemme siis, että $\mathrm{Var}(S) = \mathcal{I}(\mu)$. Seuraavassa näytetään, että vastaava tulos pitää paikkansa kaikille piste- ja informaatiofunktioille, kunhan riittävät säännöllisyysehdot ovat voimassa. □

Voidaan osoittaa yleisesti, että $E(S) = 0$:

$$\begin{aligned} E(S) &= E\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X;\theta)\right] \\ &= \int\left[\frac{1}{f(x;\theta)}\cdot\frac{\mathrm{d}}{\mathrm{d}\theta}f(x;\theta)\right]f(x;\theta)\,\mathrm{d}x \\ &= \frac{\mathrm{d}}{\mathrm{d}\theta}\int f(x;\theta)\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\theta}1 = 0. \end{aligned} \tag{10.9.1}$$

Odotusarvon laskemisessa on oletettu, että derivoinnin ja integroinnin järjestystä voidaan vaihtaa. Koska $E(S) = 0$, niin

$$\mathrm{Var}(S) = E(S^2). \tag{10.9.2}$$

Tätä suuretta kutsutaan *(odotetuksi) Fisherin informaatioksi* ja merkitään

$$\mathrm{Var}(S) = \mathcal{I}(\theta).$$

Toisaalta Fisherin informaatio voidaa lausua muodossa

$$\mathcal{I}(\theta) = -\,E\left[\frac{\mathrm{d}^2\,l(\theta;X)}{\mathrm{d}\theta^2}\right]. \tag{10.9.3}$$

Todetaan ensin, että

$$\begin{aligned} \frac{\mathrm{d}^2\,l}{\mathrm{d}\theta^2} &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{1}{f(x;\theta)}\cdot\frac{\mathrm{d}}{\mathrm{d}\theta}f(x;\theta)\right] \\ &= \frac{1}{f(x;\theta)}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x;\theta) - \frac{1}{[f(x;\theta)]^2}\left[\frac{\mathrm{d}f(x;\theta)}{\mathrm{d}\theta}\right]^2 \\ &= \frac{1}{f(x;\theta)}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x;\theta) - \left[\frac{\mathrm{d}\log f(x;\theta)}{\mathrm{d}\theta}\right]^2. \end{aligned}$$

Koska

$$\begin{aligned} E\left[\frac{1}{f(x;\theta)}\cdot\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x;\theta)\right] &= \int\frac{1}{f(x;\theta)}\cdot\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x;\theta)\cdot f(x;\theta)\,\mathrm{d}x \\ &= \int\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x;\theta)\,\mathrm{d}x = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\int f(x;\theta)\,\mathrm{d}x = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}1 = 0, \end{aligned}$$

niin

$$-\,E\left[\frac{\mathrm{d}^2\,l(\theta;X)}{\mathrm{d}\theta^2}\right] = E\left[\frac{\mathrm{d}\log f(x;\theta)}{\mathrm{d}\theta}\right]^2 = E(S^2).$$

Näin on identiteetti (10.9.3) todistettu. Odotusarvon laskemisessa oletettiin jälleen, että derivoinnin ja integroinnin järjestys voidaan vaihtaa.

Kuten alaluvuissa 10.3 ja 10.8 todettiin, riippumattomista kokeista saatu parametrin θ pistefunktio saadaan laskemalla yhteen riippumattomista kokeista määritetyt pistefunktiot. Informaatiofunktiolla on vastaava ominaisuus. Olkoot X_1 ja X_2 riippumattomat satunnaismuuttujat ja olkoot $S_1(\theta) = S(\theta, X_1)$ ja $S_2(\theta) = S(\theta, X_2)$ vastaavat pistefunktiot. Silloin yhdistetyn kokeen pistefunktio $S(\theta) = S(\theta; X_1, X_2)$ on

$$S(\theta) = S_1(\theta) + S_2(\theta).$$

Ominaisuuden (10.9.1) nojalla $E(S) = 0$ ja

$$\mathrm{Var}(S) = \mathrm{Var}(S_1) + \mathrm{Var}(S_2),$$

sillä S_1 ja S_2 ovat riippumattomat. Koska $\mathrm{Var}(S) = \mathcal{I}(\theta)$, niin odotettu informaatio on additiivinen:

$$\mathcal{I}(\theta) = \mathcal{I}_1(\theta) + \mathcal{I}_2(\theta), \tag{10.9.4}$$

missä $\mathrm{Var}(S_i) = \mathcal{I}_i(\theta)$, $i = 1, 2$.

Jos nyt $X_1, \ldots, X_n$ on otos jakaumasta, jonka tiheysfunktio on $f(x; \theta)$, niin

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

missä $E(S_i) = 0$ ja S_i:t ovat riippumattomat, $1 \le i \le n$. Koska X_i:t noudattavat samaa jakaumaa, niin voidaan merkitä $\mathrm{Var}(S_i) = \mathcal{I}_i(\theta) = i(\theta), i = 1, \ldots, n$ eli jokaisen havainnon Fisherin informaatio on sama. Hajotelmaa (10.9.4) vastaavasti

$$\mathcal{I}(\theta) = \sum_{j=1}^{n} \mathrm{Var}(S_j) = n\, i(\theta).$$

10.9.3 Cramérin ja Raon alaraja

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

Lause 10.2 (Cramér ja Rao) *Olkoon $\hat{\theta}$ parametrin θ harhaton estimaattori. Silloin tiettyjen säännöllisyysehtojen vallitessa*

$$\mathrm{Var}(\hat{\theta}) \ge \frac{1}{\mathcal{I}(\theta)}, \tag{10.9.5}$$

missä $\mathcal{I}(\theta)$ on odotettu informaatio.

Todistus. Koska pistefunktion S odotusarvo on 0, niin

$$\begin{aligned}\mathrm{Cov}(\hat{\theta}, S) &= E[(\hat{\theta} - \theta)S] \\ &= E(\hat{\theta}S) - \theta\, E(S) = E(\hat{\theta}S).\end{aligned}$$

Toisaalta saadaan

$$\begin{aligned}E(\hat{\theta}S) &= \int \hat{\theta}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x;\theta)\right] f(x;\theta)\,\mathrm{d}x \\ &= \int \hat{\theta}\left[\frac{1}{f(x;\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta} f(x;\theta)\right] f(x;\theta)\,\mathrm{d}x \\ &= \int \hat{\theta}\frac{\mathrm{d}}{\mathrm{d}\theta} f(x;\theta)\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\theta}\int \hat{\theta} f(x;\theta)\,\mathrm{d}x \\ &= \frac{\mathrm{d}}{\mathrm{d}\theta}\theta = 1.\end{aligned}$$

Koska

$$[\mathrm{Cov}(\hat{\theta}, S)]^2 \le \mathrm{Var}(\hat{\theta})\,\mathrm{Var}(S),$$

$\mathrm{Cov}(\hat{\theta}, S) = 1$ ja $\mathrm{Var}(S) = \mathcal{I}(\theta)$, niin tästä seuraa Cramérin ja Raon epäyhtälö (10.9.5)

$$\mathrm{Var}(\hat{\theta}) \ge \frac{1}{\mathcal{I}(\theta)}.$$

□

10.9.4 Suurimman uskottavuuden estimaattorin ominaisuuksia

Totesimme alaluvussa 10.8, että $S(\theta_0)$ noudattaa likimain normaalijakaumaa, kun n on suuri:

$$S(\theta_0) \xrightarrow{d} \mathrm{N}[0, \mathcal{I}(\theta_0)], \qquad \text{kun } n \to \infty.$$

Seuraavassa esitetään melko luettelomaisesti suurimman uskottavuuden estimaattorin asymptoottiset ominaisuudet.

Tarkentuva. Suurimman uskottavuuden estimaattori on tarkentuva eräiden yleisten ehtojen vallitessa. Silloin siis $\lim_{n\to\infty} P(|\hat{\theta} - \theta_0| > \varepsilon) = 0$ kaikilla $\varepsilon > 0$, missä $\hat{\theta}$ on parametrin $\theta_0 \in \Theta$ suurimman uskottavuuden estimaattori.

Asymptoottisesti normaalinen. Eräiden yleisten ehtojen vallitessa

$$\hat{\theta} \xrightarrow{d} \mathrm{N}\left(\theta_0, \frac{1}{\mathcal{I}(\theta_0)}\right), \qquad \text{kun } n \to \infty.$$

Kun $\theta = \theta_0$, niin $\hat{\theta}$:n jakauma lähenee normaalijakaumaa n:n kasvaessa.

Esimerkki 10.20 Olkoon $X_1, \dots, X_n$ otos eksponenttijakaumasta $\mathrm{Exp}(1/\beta)$. Silloin $f(x) = \beta \mathrm{e}^{-\beta x}$ ja

$$l(\beta) = n \log \beta - \beta \sum x_i,$$

joten $l''(\beta) = -n/\beta^2$ ja $\mathcal{I}(\beta) = n/\beta^2$. Suurilla n:n arvoilla $\hat{\beta}$ noudattaa likimain normaalijakaumaa:

$$\hat{\beta} \approx \mathrm{N}\left(\beta, \frac{\beta^2}{n}\right).$$

□

Invariantti parametrisoinnin suhteen. Jos $\hat{\theta}$ on θ:n suurimman uskottavuuden estimaattori, niin $g(\hat{\theta})$ on $g(\theta)$:n suurimman uskottavuuden estimaattori, missä g on θ:n monotoninen funktio.

Esimerkki 10.21 Olkoon $X_1, \dots, X_n$ otos eksponenttijakaumasta $\mathrm{Exp}(\theta)$. Parametrin θ suurimman uskottavuuden estimaattori on $\hat{\theta} = \overline{x}$. Lasketaan todennäköisyyden $P(X \geq t_0)$ suurimman uskottavuuden estimaattori. Eksponenttijakauman perusteella $P(X \geq t_0) = \mathrm{e}^{-t_0/\theta} = \pi(\theta)$. Todennäköisyyden π suurimman uskottavuuden estimaattori on

$$\hat{\pi} = \pi(\hat{\theta}) = \mathrm{e}^{-t_0/\overline{x}}.$$

□

Lause 10.3 *Oletetaan, että $\tilde{\theta}$ on θ:n harhaton estimaattori, joka saavuttaa Cramérin ja Raon alarajan. Jos θ:n suurimman uskottavuuden estimaattori $\hat{\theta}$ on yhtälön $\frac{\partial l}{\partial \theta} = 0$ ratkaisu, niin $\tilde{\theta} = \hat{\theta}$.*

Lauseen todistus sivuutetaan.Suurimman uskottavuuden estimaattori voi olla harhainen, mutta se on asymptoottisesti harhaton. Jos $\hat{\theta}$ on harhaton, niin sillä on Cramérin ja Raon lauseessa esitetty minimivarianssi $1/\mathcal{I}(\theta)$. Tällaista estimaattoria sanotaan tehokkaaksi. Jos $\hat{\theta}$ on harhainen, niin $\hat{\theta}$ on asymptoottisesti tehokas eli $\hat{\theta}$:n varianssi lähenee varianssin alarajaa.