Luku 7

Johdanto tilastolliseen päättelyyn

7.1 Tilastollisen ongelman luonne

Tilastollisen mallintaminen ja päättely käsittelee havaintojen vaihtelua ja epävarmuutta. Ei ole itsestään selvää, että tuollaisesta aiheesta voidaan esittää mitään täsmällistä tai tieteellistä. Tilastotieteessäkin on omaksuttu erilaisia lähestymistapoja epävarmuuden käsittelyyn. Kaksi pääkoulukuntaa ovat Bayesilainen ja frekventistinen koulukunta. Seuraava esitys perustuu pääoson uskottavuuden (uskottavuusfunktio) käsitteeseen, joka on kummankin edellä mainitun koulukunnan keskeinen peruskäsite .

Aspiriiniaineisto

Taulukon 7.1 tulokset ovat peräisin tutkimuksesta, jossa selvitettiin, ehkäiseekö aspiriini aivohalvauksia ja sydäninfarkteja (infarctus myocardii acutus) (Steering Committee of the Physicians' Health Study Research Group 1989). Tutkimuksessa satunnaistettiin 22071 tervettä henkilöä aspiriiniryhmään ja lumeryhmään. Aspiriiniryhmään kuuluvat saivat päivittäin pienen annoksen aspiriinia ja lumeryhmään kuuluvat vastaavasti lumetabletin. Henkilöiden terveydentilaa seurattiin keskimäärin 5 vuotta. Asetelma on satunnaistettu kliininen koe, jossa tutkittiin aspiriinin käytön vaikutusta sydäninfarktikuolleisuuteen. Tutkimukseen osallistuneet eivät tienneet, kumpaan ryhmään he kuuluivat.

Pääkysymys on siis tämä: Onko aspiriinista hyötyä sydäninfarktin ehkäisyssä? Aspiriiniryhmässä on vähemmän sydäninfarkteja kuin lumeryhmässä, 139 vastaan 239. Mitä tämä todistaa, mitä luvuista voidaan päätellä? Onko todistusaineisto kyllin vahva, jotta voimme vastata kysymykseen? Sivuvaikutuksia, jos niitä mitataan aivohalvausten määrällä, oli enemmän aspiriiniryhmässä. Kuitenkaan lukujen 119 vastaan 98 osoittama ero ei tunnu vakuuttavalta. Tällaisiin kysymyksiin vastaaminen edellyttää *tilastollista* eli *stokastista mallia*, joka kuvaa havintojen stokastista käyttäytymistä.

Taulukko 7.1 Aspiriinin käytön mahdollinen aivohalvauksia ja sydäninfarkteja ehkäisevä vaikutus.

Ryhmä	Sydänkohtaus	Aivohalvaus	Yhteensä
Aspiriini	139	119	11037
Lume	239	98	11034
Yhteensä	378	217	22071

Suhteellinen riski

$$\frac{139/11037}{239/11034} = 0.58$$

on eräs vakiintunut vertailla kahta suhteellista osuutta. Aspiriinin hyöty suhteellisena riskinä on 0.58. Jos suhteellinen riski olisi 1, se tarkoittaisi, että aspiriinilla ei ole vaikutusta. Ykköstä selvästi pienempi arvo osoittaa, että aspiriinista on hyötyä. Onko siis 0.58 tarpeeksi paljon ykköstä pienempi? Tässä esimerkissä voidaan olettaa binomimalli, jossa sydäninfarktiin sairastuneiden lukumäärä aspiriiniryhmässä noudattaa binomijakaumaa $\mathrm{Bin}(\theta_1, n_1)$ ja lumeryhmässä binomijakaumaa $\mathrm{Bin}(\theta_2, n_2)$, missä $n_1 = 11037, n_2 = 11034$ ja todennäköisyydet θ_1 sekä θ_2 ovat tuntemattomia parametreja. Tuntematon suhteellinen riski on $\theta_1/\theta_2 \equiv \theta$.

Olemme siis laskeneet suhteellisen riskin estimaatin $\hat{\theta} = 0.58$. Siihen ei ole liitetty mitään arvon luotettavuudesta kertovaa mittaa, joten se ei yksin pysty vastaamaan alkuperäiseen kysymykseen. Antaako koe niin paljon informaatiota, että voimme estimaatin $\hat{\theta}$ perusteella väittää θ:n olevan paljon pienempi kuin 1. Ajatellaan, että olisi tehty 10 kertaa laajempi koe ja olisi havaittu 1390 vastaan 2390 sydänkohtausta. Silloin jälleen olisi $\hat{\theta} = 0.58$, mutta intuitiivisesti tämän kokeen tulos tuntuu vakuuttavammalta. Estimaattiin pitää liittää jokin sen täsmällisyyttä kuvaava mitta, jonka avulla voimme arvioida estimaatin luotettavuutta. Tässä on päädytty tilastollisen päättelyn perusongelmaan: Miten havaintojen avulla voidaan tehdä tuntemattomia parametreja koskevia päteviä päätelmiä?

Jos esimerkiksi todennäköisyys sairastua johonkin tautiin muuttuu jossain väestössä vaikkapa todennäköisyydestä 0.0001 todennäköisyyteen 0.0010, on muutos suhteellisesti ottaen erittäin suuri. Jonkin tavallisen tapahtuman kohdalla yhtä suuri todennäköisyyden muutos, esimerkiksi todennäköisyydestä 0.2001 todennäköisyyteen 0.2010 ei ole merkittävä. Samoilla todennäköisyyksien muutoksilla lähellä ääripäitä 0 ja 1 on usein suurempi merkitys kuin vaihtelualueen keskivaiheilla. Todennäköisyyksien suhteen avulla voidaan tarkastella suhteellista muutosta.

Taulukossa 7.2 suhteellinen riski on

$$\theta = \theta_1/\theta_2,$$

joka voi saada minkä tahansa ei-negatiivisen arvon. Todennäköisyyksien arvoilla $\theta_1 = 0.0010$ ja $\theta_2 = 0.0001$ suhteellinen riski on $\theta = 0.0010/0.0001 = 10$ ja arvoilla $\theta_1 = 0.2010$ ja $\theta_2 = 0.2001$ suhteellinen riski on $\theta = 0.2010/0.2001 = 1.004$.

Taulukko 7.2 Sairastumistodennäköisyys Aspiriiniyhmässä on θ_1 ja Lumeryhmässä θ_2. Indikaattori $Y = 1$, kun henkilöllä on infarkti ja muutoin $Y = 0$, joten $P(Y = 1|\text{Aspiriiniryhmä}) = \theta_1$ ja $P(Y = 1|\text{Lumeryhmä}) = \theta_2$.

	Y		
	1	0	Yhteensä
Aspiriini	θ_1	$1 - \theta_1$	1.0
Lume	θ_2	$1 - \theta_2$	1.0

7.2 Tilastolliset mallit

Tilastotiede käsittelee sitä, mitä voimme oppia havainnoista. Tilastollisessa mallintamisessa havaintoja tarkastellaan satunnaiskokeen tuloksena, satunnaisilmiönä. Ajatelkaamme, että havainnot ovat "mustan laatikon"tuottamia. Syötemuuttujien vektori $\boldsymbol{x} = (x_1, \dots, x_p)$ (selittävät muuttujat, riippumattomat muuttuja) menee sisään laatikkoon ja vastemuuttujat (riippuvat muuttujat, selitettävät muuttujat) $\boldsymbol{y} = (y_1, \dots, y_m)$ tulevat ulos:

$$\boldsymbol{x} \longrightarrow \boxed{\text{luonto}} \longrightarrow \boldsymbol{y}$$

Laatikon sisällä luonto liittää jonkin funktion avulla selittävät muuttujat ja vasteet yhteen. Havaintojen analysoinnin tavoitteet voidaan karkeati jakaa kahteen ryhmään:

Ennustaminen. Mallilla halutaan ennustaa, mitä vatemuuttujan arvoja saadaa tulevaisuuden syötteillä.

Tieto riippuvuuksista. Halutaan saada selvyys siitä, miten luonto on liittänyt yhteen syötteet ja vastemuuttujat.

Tilastolliset tulokset ja näkemykset voidaan täsmällisimmin ilmaista matematiikan keinoin, mutta yhteys havaintoihin ja tieteelliseen päättelyyn on ominaista tilastotieteelliselle ajattelutavalle. Monet tilastolliset tulokset ovat syntyneet ja syntyvät vastauksena melko konkreettisiin kysymyksiin, joihin ei aina ole olemassa yleistä eleganttia ratkaisua. Tällaisten ongelmien valtava määrä tekee vaikeaksi kehittää kaiken kattavaa teoriaa, vaikka toisaalta yhteisiä yleisiä periaatteita voidaan esittää. Tällaiset periaatteet voidaan kiteyttää tilastollisen mallin käsitteeseen.

Havaintojen parametrinen mallintaminen

Tässä lahestymistavassa oletetaan mustan laatikon sisälle jokin havaintoja gereroiva todennäköisyysmalli ja funktion f yleinen matemaattinen muoto oletetaan tunnetuksi. Havaintojen malli on esimerkiksi muotoa

$$\text{vaste} = f(\text{selittäjät, satunnaisvirhe, parametrit}). \tag{7.2.1}$$

Parametrit estimoidaan havainnoista ja sen jälkeen mallia voidaan käyttää riippuvuuksien tarkasteluun tai ennustamiseen.

Musta laatikko voisi näyttää esimerkiksi seuraavalta:

$$\boldsymbol{x} \longrightarrow \boxed{\text{lineaarinen regressio}} \longrightarrow y$$

Tässä siis ajatellaan, että lineaarinen regressiomalli kuvaa riittävän hyvin vasteen y riippuvuutta selittäjistä $\boldsymbol{x}$. Silloin siis f on 1. asteen polynomi (parametrien suhteen). Mallin pätevyys pyritään vahvistamaan havaintojen avulla. Silloin tehdään esimerkiksi yhteensopivuustestejä ja tarkastellaan havaintoihin sovitetun mallin residuaaleja.

Esimerkki 7.1 Count Rumford Baierilainen oli ensimmäisiä, joka teki lämpöfysiikan kokeita. Vuonna 1798 hän teki kokeen, jossa kanuunan putki kuumennettiin $130°F$ lämpötilaan ($°F = \,°C \times 1.8 + 32$). Sitten putken annettiin jäähtyä ja lämpötila mitattiin tietyin väliajoin. Ulkolämpötila kokeen aikana oli $60°F$. Newtonin jäähtymislaki sanoo, että $d\,f/d\,t = -\theta(f - t_0)$, missä t_0 on ulkolämpötila. Silloin putken lämpötilan hetkellä t pitäisi olla

$$f(t,\theta) = 60 + 70\,e^{-\theta t}.$$

Kun mittauksia tehdään käytännössä, havainnot eivät aivan täsmällisesti toteuta lakia. Poikkeamat tulkitaan mittausvirheiksi. Tilastollinen malli, jossa mittausvirheet otetaan huomioon, on muotoa

$$Y = f(t,\theta) + \epsilon, \tag{7.2.2}$$

missä $f(t,\theta) = 60 + 70\,e^{-\theta t}$. Mittausvirhe ϵ on satunnaismuuttuja, jonka odotusarvo $E(\epsilon)$ oletetaan nollaksi.

Mittausvirheet noudattavat yleensä erittäin hyvin normaalijakaumaa, siksi tilastollista päättelyä ajatellen oletetaankin tavallisesti, että ϵ nouddattaa normaalijakaumaa. Näillä tilastollisilla oletuksilla malli (7.2.2) voidaan luonnehtia seuraavasti:

$$Y \sim \mathrm{N}(60 + 70\,e^{-\theta t}, \sigma^2).$$

Varsinainen fysikaalisesti kiinnostava estimoitava parametri on θ ja σ^2 koejärjestelyyn liittyvä virhevarianssi. $\square$

Algoritminen mallintaminen

Algoritminen lähestymistapa on saavuttanut suosiota ja sovellusmahdollisuuksia tietokoneiden laskentakapasiteetin kasvun myötä. Tässä ajattelutavassa mustan laatikon sisältö on monimutkainen ja tuntematon. Pyritään löytämään funktio $f(\boldsymbol{x})$ algoritmisesti– algoritmi laskee $\boldsymbol{x}$:n perusteella ennusteita $\boldsymbol{y}$:lle. Algoritmi siis pyritään muokkaamaan sellaiseksi, että se antaa hyviä ennusteita. Musta laatikko näyttäisi tältä:

$$\boldsymbol{x} \longrightarrow \boxed{\text{tuntematon}} \longrightarrow y$$

Esimerkiksi neuroverkot kuuluvat tähän kategoriaan. Mallin pätevyyttä arvioidaan ennustevirheen avulla.

Tavallisesti havaintojen vaihtelu jaetaan systemaattiseen osaan ja satunnaisosaan on additiivisesti:

$$\begin{aligned} \text{havainnot} &= f(\text{selittävät muuttujat, parametrit}) + \text{satunnaisosa} \\ &= \text{systemaattinen osa} + \text{satunnaisosa}. \end{aligned}$$

Esitys (7.2.1) on itse asiassa varsin yleinen, joka sallii monimutkaisetkin vaikutusmekanismit. Yksinkertaistavia oletuksia kuitenkin tarvitaan, jotta mallit pystytään ymmätämään ja analysoimaan. Havaintojen oletetaan olevan peräisin jostain jakaumaperheestä, tavallisimmin ns. parametrisesta jakaumaperheestä. Systemaattinen osa on esimerkiksi havaintojen Y_1, Y_2, …, Y_n odotusarvoja $E(Y_i)$, $1 \leq i \leq n$, koskeva oletus, joka lausutaan vaikkapa regressiofunktiona. Tavallisesti odotusarvo riippuu joistain selittävistä muuttujista (eli kovariaatista). Tilastollisen mallin voidaan sanoa olevan havaintojen yhteisjakaumaa ja systemaattista osaa koskevien oletusten joukko.

Esimerkki 7.2 Oletetaan, että $Y_1, Y_2, \ldots, Y_n$ on otos normaalijakaumasta $\mathrm{N}(\mu_i, \sigma^2)$, missä $E(Y_i) = \mu_i$. Oletetaan lisäksi, että $E(Y_i)$ riippuu lineaarisesti selittävästä muuttujasta x, joten

$$\mu_i = \alpha + \beta x_i, \qquad 1 \leq i \leq n. \tag{7.2.3}$$

Malli voidaan kirjoittaa myös muodossa

$$Y_i = \mu_i + \varepsilon_i,$$

missä $\varepsilon_i = Y_i - E(Y_i)$. Virhetermi ε_i noudattaa normaalijakaumaa $\mathrm{N}(0, \sigma^2)$. □

Kokeellisessa tilanteessa selittäjä x on koevakio. Tutkija voi päättää, mitkä arvot hän x:lle valitsee. Esimerkiksi törmäystestissä valitaan törmäysnopeudet $x_1, \ldots, x_n$. Näillä selittäjän arvoilla mitataan vastemuuttujan (tai vastemuuttujien) arvot. Regressioanalyysia käytetään kuitenkin myös ei-kokeellisessa tilanteessa, jossa tutkija ei voi kontrolloida x:n arvoja. Silloin x on satunnaismuuttuja, jonka arvo havainnoidaan usein samanaikaisesti vastemuuttujan kanssa. On huomattava, että regressiomallissa (7.2.3) tarkastellaan ehdollista odotusarvoa

$$E(Y|x) = \mu(x),$$

missä Y:n ehdollisen odotusarvon oletetaan olevan x:n lineaarinen funktio $\mu(x) = \alpha + \beta x$. Suoran kertoimet α ja β ovat tuntemattomia parametreja, jotka estimoidaan havainnoista.

Tilastotieteen oppikirjoissa lähdetään tavallisesti liikkeelle melko teknisesti. Sanotaan, että havainnot $Y_1, \ldots, Y_n$ ovat otos jostain tuntenmattomasta jakaumasta F, missä F on siis jakauman kertymäfunktio. Tavallisesti jakaumasta tehdään joitain oletuksia. Tilanne voi olla esimerkiksi sellainen, että jakauma voidaan oletettaa symmetriseksi. Tällä kurssilla käytetään useimmiten parametrista lähestymistapaa, jolloin jakauman ajatellaan kuuluvan johonkin parametriseen jakaumaperheeseen

$$\mathcal{F} = \{\, F(x;\theta),\ \theta \in \Theta \,\}$$

missä $F(x;\theta)$ on kertymäfunktio jokaisella kiinnitetyllä θ:n arvolla.

Käsittelemissämme päättelyongelmissa operoimme tavallisesti tiheysfunktioiden avulla, joten jakaumaperhe on silloin suoraviivaisempaa luonnehtia tiheysfunktioiden joukkona

$$\mathcal{F} = \{\, f(x;\theta),\ \theta \in \Theta \,\}.$$

Suure θ on siis parametri ja sen arvojoukko Θ on parametriavaruus. Valitsemalla yksi parametrin θ arvo saadaan täysin määrätty jakauma. Edellä olemme nähneet, että θ voi olla selittäjien funktio. Kun parametrin θ arvo valitaan havaintojen perusteella, saadaan θ:n *piste-estimaatti*. Parametrin (parametrien) arvo määrittämistä havaintojen perusteella sanotaan *piste-estimoinniksi.*

Esimerkki 7.3 Tarkastellaan auto-onnettomuuksien vakavuusastetta, kun selittäjänä on kuljettajan ikä. Usein väitetään, että nuoret kuljettajat aiheuttavat keskimääräistä enemmän vakavia onnettomuuksia.

Taulukko 7.1. Vakavien onnettomuuksien lukumäärä alueella A tammikuussa vuonna 2000.

Yli 21-vuotiaat		Alle 21-vuotiaat	
Kuolemaan johtaneet	Muut	Kuolemaan johtaneet	Muut
Y_1	Y_2	Y_3	Y_4
11	62	4	7

Oletetaan, että onnettomuuksien lukumäärä kuukaudessa noudattaa Poissonin jakaumaa Poi(λ). Tarkastellaan neljää onnettomuustyyppiä, jotka on määritelty kuljettajan iän ja onnettomuuden vakavuusasteen mukaan. Eri tyyppisten onnettomuuksien lukumäärien $Y_i, 1 \leq i \leq 4$, oletetaan noudattavan toisistaan riippumatta Poissonin jakaumaa $\mathrm{Poi}(\lambda_i)$. Oheisessa taulukossa on annettu eräs aineisto. Silloin esimerkiksi kuolemaan johtaneiden onnettomuuksien lukumäärä Y_3 alle 21-vuotiaiden ryhmässä noudattaa Poissonin jakaumaa $\mathrm{Poi}(\lambda_3)$. Parametrit λ_1, λ_2, λ_3 ja λ_4 ovat satunnaismuuttujien Y_1, Y_2, Y_3 ja Y_4 odotusarvoja. Odotusarvo λ_i kertoo onnettomuusasteen i. kategoriassa. Vastaavasti esimerkiksi yli 21-vuotiaiden onnettomuusaste on $\lambda_1 + \lambda_2$ ja alle 21-vuotiaiden $\lambda_3 + \lambda_4$. Merkitään $\theta_1 = \lambda_1 + \lambda_2$ ja $\theta_2 = \lambda_3 + \lambda_4$. Näin todennäköisyys, että yli 21-vuotias aiheuttaa kohtalokkaan onnettomuuden, on

$$\pi_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

ja alle 21-vuotiaan todennäköisyys aiheuttaa kohtalokas onnettomuus on

$$\pi_2 = \frac{\lambda_3}{\lambda_3 + \lambda_4}.$$

Nelikko $(\theta_1, \pi_1, \theta_2, \pi_2)$ muodostaa uuden parametrisoinnin, joka saattaa olla tulkinnallisesti selkeämpi ja mielenkiintoisempi kuin alkuperäinen. □

7.3 Estimoinnista

Tarkastelemme nyt satunnaismuuttujia, joiden todennäköisyysfunktion (tai tiheysfunktion) funktionaalinen muoto tunnetaan, mutta jakauma riippuu jostain tuntemattomasta parametrista θ. Oletetaan, että parametrin θ mahdolliset arvot kuuluvat johonkin annettuun joukkoon Θ, jota kutsutaan *parametriavaruudeksi*. Tiedetään esimerkiksi, että jonkin tuotteen elinaika X noudattaa eksponenttijakaumaa

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

missä $\theta \in \Theta = \{\, \theta \mid 0 < \theta < \infty \,\}$. Parametriavaruus Θ on siis positiivisten reaalilukujen joukko. Haluamme valita funktioperheestä

$$\mathcal{F} = \{\, f(x;\theta),\ \theta \in \Theta \,\}$$

yhden tiheysfunktion, joka esittää parhaiten tuotteen elinaikaa. Valitaan siis yksi parametrin θ arvo eli parametrin θ *piste-estimaatti*, joka määrittää jakauman.

Parametrin arvo arvioidaan eli estimoidaan havaintojen perusteella. Teemme jakaumasta havainnon $X = x$ ja estimoimme parametrin θ arvon havainnon x perusteella. Parametrin θ estimoimiseen käytettävää otosfunktiota $T(X)$ kutsutaan parametrin θ *estimaattoriksi* ja estimaattorin $T(X)$ arvoa $t = T(x)$ kutsutaan parametrin θ *estimaatiksi*. Estimaattori pyritään valitsemaan siten, että se antaa hyviä arvioita parametrista θ.

Esimerkki 7.4 Estimoidaan ehdokkaan A kannattajien suhteellinen osuus θ eräässä suuressa kaupungissa. Valitaan kaupungin äänioikeutetuista satunnaisesti n henkilöä, joilta tiedustellaan, kannattavatko he ehdokasta A. Olkoon X ehdokkaan A kannattajien lukumäärä otoksessa. Koska populaation koko on suuri verrattuna otoskokoon n, voidaan olettaa, että $X \sim \mathrm{Bin}(n, \theta)$, 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 parametriavaruus on $\Theta = \{\, \theta \mid 0 \le \theta \le 1 \,\}$. Tehtävänämme on määrittää θ:n estimaattori $T(X)$ siten, että havaitun arvon $X = x$ perusteella saadaan hyvä θ:n piste-estimaatti $T(x)$. Havainnon $X = x$ todennäköisyys on

$$P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{7.3.1}$$

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

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

□

7.4 Suurimman uskottavuuden menetelmä

Kun todennäköisyyttä (7.3.1) tarkastellaan parametrin θ funktiona, huomaamme, että tekijä $\binom{n}{x}$ ei riipu parametrista θ. Siksi määrittelemmekin *uskottavuusfunktion* seuraavasti:

$$L(\theta) = c \cdot f(x; \theta), \tag{7.4.1}$$

missä $f(x; \theta)$ on todennäköisyysfunktio ja $f(x; \theta)$ on siis havainnon $X = x$ todennäköisyys. Vakio c ei riipu parametrista θ, vaikkakin se voi riippua havainnosta x. Vakio c pyritään valitsemaan siten, että $L(\theta)$:lle saadaan yksinkertainen lauseke. Uskottavuusfunktioon perustuvat päätelmät eivät riipu vakion c valinnasta.

Tavallisesti uskottavuusfunktio tulee olemaan useiden tekijöiden tulo ja mm. siitä syystä on osoittautunut käteväksi työskennellä uskottavuusfunktion logaritmin avulla. *Logaritmoitu uskottavuusfunktio* $l(\theta)$ on uskottavuusfunktion luonnollinen logaritmi eli

$$l(\theta) = \log L(\theta). \tag{7.4.2}$$

Esityksestä (7.4.1) seuraa, että

$$l(\theta) = \log c + \log f(x;\theta),$$

missä vakio c ei riipu θ:sta. Jatkossakin kaikki logaritmit ovat luonnollisia logaritmeja, ellei toisin mainita.

Suurimman uskottavuuden estimaatti (SUE) $\hat{\theta}$ on se parametrin θ arvo, joka maksimoi havainnon x todennäköisyyden $f(x;\theta)$. Sama arvo $\hat{\theta}$ maksimoi myös funktiot $L(\theta)$ ja $l(\theta)$. Suurimman uskottavuuden estimaatti $\hat{\theta}$ on siis uskottavuusfunktion ja logaritmoidun uskottavuusfunktion maksimikohta. Tavallisesti tarkastellaan logaritmoitua uskottavuusfunktiota, koska se on usein matemaattisesti yksinkertaisempi kuin uskottavuusfunktio. Logaritmoidulla uskottavuusfunktiolla on myös teoreettisesti merkittävä tilastollinen tulkinta.

Esimerkki 7.5 Tarkastellaan edelleen Esimerkkiä 7.4, jossa havaintojen todennäköisyysfunktio on $f(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$. Kun uskottavuusfunktiossa (7.4.1) valitaan vakion arvoksi $c = 1/\binom{n}{x}$, saadaan esitysmuoto

$$L(\theta) = \theta^x(1-\theta)^{n-x}, \qquad 0 \le \theta \le 1.$$

Tässä uskottavuusfunktion esityksessä ei ole ”turhia” vakiotekijöitä. Tätä uskottavuusfunktion esitysmuotoa kutsutaan myös *uskottavuusfunktion ytimeksi*. Yleensä esitämme uskottavuusfunktion tässä *ydinmuodossa*. Logaritmoitu uskottavuusfunktio on

$$l(\theta) = x\log\theta + (n-x)\log(1-\theta), \qquad 0 < \theta < 1.$$

Parametrin θ suurimman uskottavuuden estimaatti on siis se θ:n arvo, joka maksimoi funktion $l(\theta)$. Huomattakoon, että $l(\theta)$ ei ole määritelty välin $[0,1]$ päätepisteissä, mutta $L(\theta)$ on. $\square$

Luku 8

Uskottavuuspäättelyn perusteet

8.1 Uskottavuuden määritelmä

Tarkastellaan kannatusmittausten tyypillistä asetelmaa. On arvioitava ehdokkaan A kannattajien suhteellinen osuus θ annetulla paikkakunnalla (Katso Esimerkki 7.4). 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 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

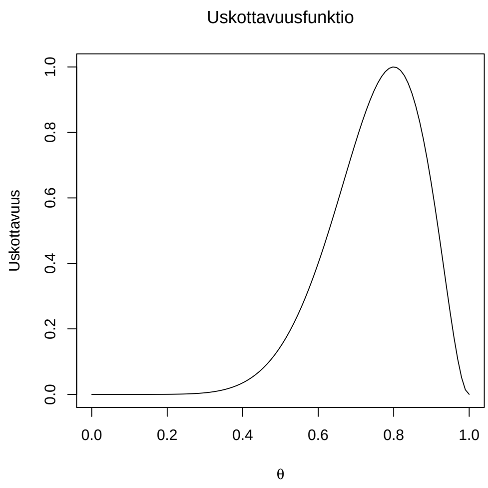
$$\Theta = \{\, \theta \mid 0 \le \theta \le 1 \,\}.$$

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

$$P_\theta(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{8.1.1}$$

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 8.1. 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 (8.1.1) voidaan tarkastella parametrin θ funktiona:

$$L(\theta) = P_\theta(X = 10) = \binom{100}{10} \theta^{10}(1-\theta)^{90}. \tag{8.1.2}$$

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 8.1 Jos tilastollinen malli on parametrisoitu kiinteällä tuntemattomalla parametrilla θ, niin uskottavuusfuntio $L(\theta)$ on havainnon x todennäköisyys parametrin θ funktiona.

8.1.1 Diskreetit mallit

Diskreeteissä malleissa havaintoarvon todennäköisyys on hyvin määritelty suure, kuten esimerkiksi binomimallissa (8.1.2).

8.1.2 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

$$\begin{aligned} P_\theta(X = x) &= P_\theta(x - \Delta/2 < X \leq x + \Delta/2) \\ &= F(x + \Delta/2;\theta) - F(x - \Delta/2;\theta). \end{aligned} \tag{8.1.3}$$

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

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

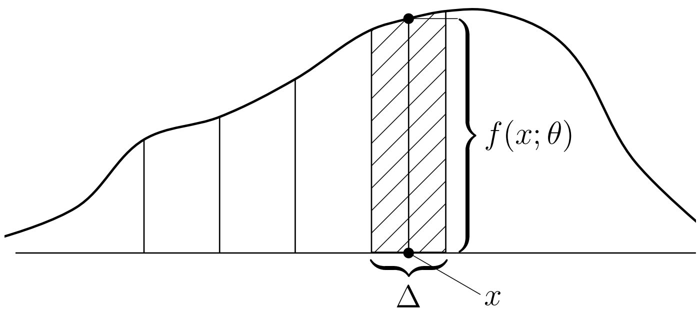
Silloin voidaan käyttää likiarvoa

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

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

$$L(\theta) = cf(x;\theta), \tag{8.1.4}$$

missä c on mikä tahansa sopivasti valittu positiivinen vakio.



8.2 Esimerkkejä

Esimerkki 8.1 Tarkastellaan vielä alaluvussa 8.1 esitettyä kannatusmittausta. Ehdokasta A kannatti 100:n äänioikeutetun otoksessa 10, jolloin saatiin uskottavuusfunktio (8.1.2). Jos tiedetään ainoastaan, että A:ta kannatti korkeintaan 10, silloin θ:aa koskeva informaatio voidaan lausua uskottavuusfunktiolla

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

□

Esimerkki 8.2 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

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

□

8.3 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 (8.1) mukaan 1. kokeeseen perustuva uskottavuusfunktio on

$$L_1(\theta) = f_1(x; \theta)$$

ja 2. kokeeseen perustuva uskottavuusfunktio

$$L_2(\theta) = f_2(y; \theta).$$

Olkoon satunnaismuuttujien X ja Y yhteisjakauman todennäköisyysfunktio $f(x, y; \theta)$. Yhdistettyyn kokeeseen perustuva uskottavuusfunktio on määritelmän (8.1) mukaan

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

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

$$L(\theta) = L_1(\theta)\, L_2(\theta). \tag{8.3.1}$$

Ottamalla uskottavuusfunktiosta (8.3.1) puolittain logaritmit saadaan vastaava logaritmoitu uskottavuusfunktio

$$l(\theta) = l_1(\theta) + l_2(\theta), \tag{8.3.2}$$

missä on merkitty $l(\theta) = \log L(\theta)$.

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 (todennäköisyysfunktio) on $f(x;\theta)$, niin

$$L(\theta) = \prod_{i=1}^{n} L_i(\theta)$$

ja

$$l(\theta) = \sum_{i=1}^{n} l_i(\theta).$$

Jos halutaan korostaa, että uskottavuusfunktio perustuu havaintoihin $x_1, \ldots, x_n$, niin merkitään $L(\theta; x_1, \ldots, x_n)$. Huomattakoon, että hajotelma (8.3.1) 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 havainnot $\boldsymbol{x}$ ja $\boldsymbol{y}$ ovat riippumattomat, niin

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

ja hajotelma (8.3.1) pitää paikkansa.

Esimerkki 8.3 Jatketaan esimerkin 8.1 käsittelyä. Oletetaan, että kaksi eri tutkijaa tekevät samaan aikaan toisistaan riippumatta ehdokkaan A kannatusta 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

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

ja 2. otokseen perustuva logaritmoitu uskottavuusfunktio on vastaavasti

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

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 (8.3.2) mukaan yhdistettyyn otokseen perustuva logaritmoitu uskottavuusfunktio on

$$\begin{aligned} l(\theta) &= l_1(\theta) + l_2(\theta) \\ &= (10+8)\log(\theta) + (100+50-(10+8))\log(1-\theta) \\ &= 18\log(\theta) + 132\log(1-\theta). \end{aligned} \tag{8.3.3}$$

Logaritmoidusta uskottavuusfunktiosta (8.3.3) 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 8.4 Oloon $x_1, \ldots, x_n$ otos normaalijakausta $\mathrm{N}(\theta, \sigma^2)$, missä σ^2 on tuntematon. Yhden havainnon x_i vaikutus uskottavuuteen on

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

ja logaritmoitu kokonaisuskottavuus on

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^{n} l_i(\theta) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2. \end{aligned}$$

□

8.4 Yhteys Bayesilaiseen lähestymistapaan

Bayesilaisessa lähestymistavassa tarvitaan parametrin θ priori $f(\theta)$, jonka avulla voidaan määrittää posteriori

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

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

8.5 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 8.5 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

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

□

8.6 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 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 $l(\theta)$ 1. derivaattaa kutsutaan *Fisherin pistefunktioksi* tai lyhyesti vain pistefunktioksi ja se määritellään seuraavasti:

$$S(\theta; x) \equiv \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta; x), \tag{8.6.1}$$

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

$$S(\theta) = 0$$

ratkaisu.

Maksimissa logaritmoidun uskottavuusfunktion 2. derivaatta on negatiivinen. Siksi 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

$$S(\hat{\theta}) = 0 \qquad \text{ja} \qquad I(\hat{\theta}) > 0. \tag{8.6.2}$$

Jos edellä mainitut säännöllisyysoletukset eivät pidä paikkaansa, ei $\hat{\theta}$:a välttämättä saada uskottavuusyhtälön ratkaisuna.

Esimerkki 8.6 Binomijakauman tapauksessa (Esimerkki 7.5)

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

Kun $1 \le x \le n-1$, 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

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

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 8.7 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

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

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 $x_1, x_2, \ldots, x_n$ todennäköisyys maksimoituu, kun jakauman tuntematon parametri μ estimoidaan otoskeskiarvolla $\bar{x}$. □

Esimerkki 8.8 Olkoon $x_1, \ldots, x_n$ 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

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

□

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

$$\log\frac{L(\theta)}{L(\hat{\theta})} \approx -\tfrac{1}{2}I(\hat{\theta})(\theta - \hat{\theta})^2. \tag{8.6.3}$$

Toisen asteen likiarvon tarkkuutta voidaan tutkia piittämällä log-uskottavuusfunktio ja likiarvo samaan kuvioon. Esimerkiksi normaalijakauman tapauksessa likiarvo on tarkka. Kun likiarvo (8.6.3) derivoidaan puolittain, saadaan

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

tai

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

8.7 Uskottavuusalueet

Miten esitetään uskottavuusfunktioon sisältyvä parametria koskeva informaatio? Eräs suoraviivainen tapa on piirtää uskottavuusfunktion kuvaaja. Se tapa voi kuitenkin olla varsin epäkäytännöllinen ja usean parametrin tapauksessa joskus jopa suorastaan mahdoton toteuttaa.

8.7.1 Suora uskottavuuspäättely

Parametrin θ uskottavuusalueeseen kuuluvat ne θ:n arvot, joilla on tarpeeksi suuri uskottavuus:

$$\text{ua}(\theta; c) = \{\, \theta \mid \frac{L(\theta)}{L(\hat{\theta})} > c \,\}. \tag{8.7.1}$$

Tavallisesti uskottavuusalue $\text{ua}(\theta, c)$ on reaalilukuväli, jolloin uskottavuusaluetta kutsutaan uskottavuusväliksi (uv). Huomaa, että $\frac{L(\theta)}{L(\hat{\theta})}$ on normeerattu uskottavuus. Kynnysarvon c arvon määrittäminen on tietysti käytännön sovelluksissa tärkeä tehtävä. Parametrin arvojen uskottavuutta verrataan suurimman uskottavuuden estimaatin uskottavuuteen. Se on havainnollista ilmoittaa prosenttilukuna. Koska $0 < \frac{L(\theta)}{L(\hat{\theta})} \leq 1$, niin voimme kutsua väliä $\text{ua}(\theta, c)$ myös $100c$ %:n uskottavuus. Tällä välillä uskottavuus on enenemmän kuin $100c$ %:a maksimiuskottavuudesta $L(\hat{\theta})$.

Uskottavuusalue on usein kätevämpää määrittää normeeratun uskottavuusfunktion logaritmin avulla. Uskottavuusalue (8.7.1) voidaan määritellä joukkona

$$\text{ua}(\theta; c) = \{\, \theta \mid l(\theta) - l(\hat{\theta}) > \log\, c \,\}.$$

Esimerkiksi 50 %:n uskottavuusväli on epäyhtälön

$$l(\theta) - l(\hat{\theta}) > \log(0.5) = -0.69$$

toteuttavien θ:n arvojen joukko.

Parametria θ koskeva informaatio voidaan kuvata havainnollisesti $\frac{L(\theta)}{L(\hat{\theta})}$:n tai $l(\theta) - l(\hat{\theta})$:n kuvaajan avulla. Kun funktio $l(\theta)$ on riittävän säännöllinen, kuten useimmissa esimerkeissämme, saadaan parametrista riittävän tarkka yleiskuva esittämällä suurimman uskottavuuden estimaatti $\hat{\theta}$ ja pari kolme uskottavuusväliä.

Esimerkki 8.9 Tuotantolijalla testattiin 100:n tuotteen satunnaisotos. Virheellisten tuotteiden lukumäärä X otoksessa noudattaa binomijakaumaa $\text{Bin}(100, \pi)$. Otoksessa on kaksi viallista tuotetta ($X = 2$). Logaritmoitu uskottavuusfunktio on silloin

$$l(\theta) = 2\log(\theta) + 98\log(1 - \theta)$$

ja $\hat{\theta} = 0.02$. Logaritmoidun uskottavuusfunktion maksimi

$$l(\hat{\theta}) = \log(0.02) + 98\log(0.98) = -9.80.$$

Logaritmoitu normeerattu uskottavuusfunktio on siis

$$l(\theta) - l(\hat{\theta}) = 2\log(\theta) + 98\log(1 - \theta) + 9.80.$$

Esimerkiksi 50 % uskottavuusväli on

$$\{\, \theta \mid l(\theta) - l(\hat{\theta}) - \log(0.5) \geq 0 \,\} = \{\, \theta \mid 2\log(\theta) + 98\log(1-\theta) + 9.80 + 0.69 \geq 0 \,\}.$$

$\square$

8.7.2 Todennäköisyyteen perustuva päättely

Uskottavuusväliin kynnysarvon c määrittäminen on ongelma, koska c ei viittaa suoraan mihinkään havaittavaan. Mitä tarkoittaa vaikkapa 5%:n uskottavuus? Se riippuu esimerkiksi myös otosavaruuden laajuudesta. Sen sijaan 5%:n (0.05)todennäköisyys voidaan tulkita suhteelliseksi frekvenssiksi toistokokeessa.

Frekventistinen parametria θ koskeva päättely perustuu θ:n estimaattorin $\hat{\theta}$ jakaumaan. Esimerkissä 8.8 tarkasteltiin normaalijakauman $/N(\theta, \sigma^2)$ tuntemattoman parametrin θ estimointia, kun σ^2 oletettiin tunnetuksi. Esimerkistä 8.8 seuraa, että

$$\log \frac{L(\theta)}{L(\hat{\theta})} = l(\theta) - l(\hat{\theta}) = -\frac{n}{2\sigma^2}(\bar{X} - \theta)^2$$

Tiedämme, että satunnaisotoksen $X_1, \ldots, X_n$ funktiona otoskeskeskiarvo

$$\bar{X} = \sum_{i=1}^{n} X_i$$

noudattaa normaalijakaumaa $\mathrm{N}(\theta, \frac{\sigma^2}{n})$. Siksi

$$\frac{n}{\sigma^2}(\bar{X} - \theta)^2 \sim \text{Khi2}(1)$$

noudattaa Khiin nelöjakaumaa vapausastein 1. Suuretta

$$W \equiv 2 \log \frac{L(\theta)}{L(\hat{\theta})} \sim \text{Khi2}(1) \tag{8.7.2}$$

Wilksin uskottavuussuhteeksi. Normaalijakauman tapauksessa W:n jakauma on täsmälleen Khi2 ja varsin yleisin oletuksin W:n jakauma on likimain Khi2.

Wilksin suureen 8.7.2 avulla saadaan nyt uskottavuusvälille todennäköisyystulkinta, sillä

$$\begin{aligned} P(\log \frac{L(\theta)}{L(\hat{\theta})} > c) &= P(2 \log \frac{L(\hat{\theta})}{L(\theta)} < -2 \log c) \\ &= P(\chi_1^2 < -2 \log c) \end{aligned}$$

Kun nyt valitsemme Khin neliöjakauman Khi2(1) $100(1-\alpha)$%:n fraktiilin $\chi^2_{1,1-\alpha}$ ja sen avulla kynnysarvon

$$c = e^{-\frac{1}{2}\chi^2_{1,1-\alpha}}, \tag{8.7.3}$$

saadaan

$$P(\log \frac{L(\theta)}{L(\hat{\theta})} > c) = P(\chi_1^2 < \chi^2_{1,1-\alpha}) = 1 - \alpha.$$

Kun valitaan kynnys (8.7.3), niin uskottavuusväli

$$\{\theta : \frac{L(\theta)}{L(\hat{\theta})} > c\}$$

on θ:n $100(1-\alpha)$%:n luottamusväli.

8.7.3 Uskottavuusvälit vs. luottamusvälit

Jos uskottavuusvälillä on teoreettisesti perusteltu luottamustaso, kutsumme väliä luottamusväliksi. Muutoin kutsumme sitä uskottavuusväliksi.

Esimerkki 8.10 Henkilö poimii satunnaisesti kokonaisluvun θ ja pyytää sinua arvaamaan luvun annetun aineiston avulla. Hän heittää lanttia kahdesti (sinä et näe tuloksia) ja raportoi heittojen tulokset seuraavasti: Jos tulee kruuna, hän ilmoittaa tuloksen $\theta+1$, muutoin hän hän ilmoittaa $\theta-1$. Havainnot X_1, X_2 ovat siis otos sellaisesta jakaumasta, että $P(X_i=\theta+1)=P(X_i=\theta-1)=0.5,\ i=1,2$. Esimerkiksi hän saattaa rapotoida havainnot $X_1=5$ ja $X_2=5$.

Seuraavalla arvauksella on 75%:n todennäköisyys olla oikea:

$$C(x_1,x_2)=\begin{cases}\frac{1}{2}(x_1+x_2), & \text{jos } x_1\neq x_2\\ x_1-1, & \text{jos } x_1=x_2.\end{cases}$$

Tavanomaisen luottamusajattelun mukaan edellä esitetyllä arvauksella on 75%:n luottamustaso. Jos $x_1\neq x_2$, meillä on kuitenkin 100%:n luottamus, että arvaus on oikea. Muutoin meillä on vain 50%:n varmuus. Jos havaitaan $x_1\neq x_2$, on järjetöntä väittää, että meillä on ainoastaan 75%:n luottamus arvaukseen $\frac{1}{2}(x_1+x_2)$. Puhtaasti uskottavuuteen perustuvan päättelyn logiikka ei johda tässä tapauksessa ristiriitaisuuksiin. Siinä raportoidaan kummankinhavainnon $\{x_1,x_2\}$ tapauksessa, mikä on parametria θ koskeva epävarmuus. Se ei kuitenkaan kerro mitään siitä, kuinka usein arvaus osuu oikeaan pitkässä sarjassa. □

Esimerkki 8.11 Valitaan satunnaisotos $X_1,\ldots,X_n$ tasajakaumasta $\mathrm{Tas}(0,\theta)$. Silloin

$$P_\theta(X_1\le x,\ldots,X_n\le x)=\prod_{i=1}^{n}[P_\theta(X_i\le x)]=F^n(x;\theta),$$

missä

$$F(x;\theta)=\begin{cases}\frac{1}{\theta}, & \text{kun } 0\le x\le\theta\\ 0, & \text{muualla.}\end{cases}$$

Silloin θ:n uskottavuusfunktio (turhat vakiot jätetty pois) on

$$L(\theta;x_1,\ldots,x_n)=\begin{cases}\frac{1}{\theta^n}, & \text{kun } 0\le x_i\le\theta, i=1,\ldots,n.\\ 0, & \text{muualla.}\end{cases}$$

Uskottavuusfunktio voidaan kirjoittaa lyhyesti muodossa

$$L(\theta;x_1,\ldots,x_n)=\frac{1}{\theta^n}I_{[0\le x(n)\le\theta]},$$

missä $x(n)=\max(x_1,\ldots,x_n)$ ja $I_{[0\le x_{(n)}\le\theta]}$ on joukon $[0\le x_{(n)}\le\theta]$ indikaattorifunktio. Funktion määrittelyalue riippuu siis parametrista θ. Funktio

$L(\theta; x_1, \ldots, x_n) = \frac{1}{\theta^n}$ on θ:n vähenevä funktio, kun $\theta \geq x_{(n)}$. Funktio saavuttaa siis maksiminsa, kun $\theta = x_{(n)}$, joten θ:n SUE on

$$\hat{\theta} = x_{(n)}.$$

$\square$

8.7.4 Uskottavuussuhteeseen perustuva testaus

Uskottavuussuhdetta voidaan käyttää suoraan testaukseen. Esimerkiksi nollahypoteesin $H_0 : \theta = \theta_0$ uskottavuus on θ_0:n normeerattu uskottavuus

$$\frac{L(\theta_0)}{L(\hat{\theta})}.$$

H_0 hylätään, jos sen uskottavuus on "liian pieni". Mutta kuinka pieni on liian pieni?

Säännöllisissä parametrisissa malleissa voidaan laskea ns. testimenettelyyn liittyvä P-arvo Wilksin suureen (8.7.2) avulla. Jos havaitaan $L(\theta_0)/\,L(\hat{\theta}) = c$, niin (8.7.2):n perusteella P-arvo on

$$P(\chi_1^2 > -2 \log c).$$

Normaalijakuman tapauksessa P-arvo on tarkka, mutta muutoin vain likimäärin oikea.

8.7.5 SUE:n hajonta ja Waldin testisuure

Säännöllisissä malleissa log-uskottavuusfunktion kavdraattinen likiarvo on hyvä ja voimme kirjoittaa

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

Silloin uskottavuusväli $\{\theta : \; L(\theta)/\,L(\hat{\theta}) > c\}$ on likimain

$$\hat{\theta} \pm \sqrt{-2 \log c} \times I(\hat{\theta})^{-1/2}.$$

Normaalijakauman keskiarvon luottamustaso Esimerkissä 8.8 saadaan määrittämällä todennäköisyys

$$P\chi_1^2 < -2 \log c.$$

Esimerksi

$$\hat{\theta} \pm 1.96 I(\hat{\theta})^{-1/2}$$

on θ:n täsmällinen 95%:n luottamusväli. Jos otos ei ole normaalijakaumasta, väli on ainoastaan likimain 95%:n luottamusväli.

Samoin kuin normaalijakauman keskiarvomallissa, on $I(\hat{\theta})^{-1/2}$ SUE:n $\hat{\theta}$ hajonta. Sitä käytettään testattessa hypoteesia $H_0 : \theta = \theta_0$ *Waldin testisuureen*

$$z = \frac{\hat{\theta} - \theta_0}{\mathrm{std}(\hat{\theta})}$$

avulla tai laskettaessa *Waldin luottamusvälejä*. Esimerkiksi Waldin 95%:n luottamusväli θ:lle on

$$\hat{\theta} \pm 1.96 \mathrm{std}(\hat{\theta}).$$

Jos havainnot ovat normaalijakausta, noudattaa z H_0:n vallitessa täsmällisesti normaalijakaumaa ja likimain ei-normaalisessa mallissa. Suuri $|z|$:n arvo liittyy H_0:n vähäiseen uskottavuuteen. Kun esimerkiksi $|z| > 2$, niin uskottavuus on pienempi kuin 0.15 ja p-arvo alle 0.05.

Jos uskottavuusfunktion muoto poikkeaa paljon kvadraattisesta, niin Waldin luottamusväli on uskottavuuden kannalta huono, koska välillä on arvoja, joiden uskottavuus on pienempi kuin joillain välin ulkopuolisilla pisteillä.

Waldin välejä kutsutaan SUE-perustaisiksi väleiksi. *Uskottavuus-perustaisiksi* luottamusväleiksi kutsutaan väleistä $\{\theta : \frac{L(\theta)}{L(\hat{\theta})} > c\}$ johdettuja luottamusvälejä. Waldin välit ovat aina symmetrisiä, mutta uskottavuus-perustaiset välit voivat olla epäsymmetrisiä.

8.8 Uskottavuuden invarianssi

Uskottavuusfunktio esittää 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 8.12 Monissa sovelluksissa binomijakauman parametri esitetään logit-asteikolla siten, että

$$\psi \equiv \log\{\frac{\theta}{1-\theta}\}.$$

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 *vetosuhde* 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

$$f(x) = \frac{1}{\theta}\mathrm{e}^{-x/\theta}, \qquad 0 < x < \infty. \tag{8.8.1}$$

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

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \qquad 0 < x < \infty, \tag{8.8.2}$$

missä $\lambda = 1/\theta$. Tavallisesti parametrisointi valitaan siten, että parametri esittää jotain tärkeää jakauman ominaisuutta tai siten, että jakauman matemaattinen esitystapa saadaan yksinkertaiseksi. Parametrisoinnissa (8.8.1) θ on jakauman keskiarvo.

Jos esimerkiksi $\theta = 2$, niin $\lambda = \frac{1}{2}$. Jokaista θ:n valintaa vastaa yksikäsitteinen jakauma parametrisoinnissa (8.8.2) ja kääntäen. Uskottavuusmenetelmällä on se miellyttävä piirre, että menetelmä on invariantti bijektiivisten (yksi-yksisten) parametrimuunnosten suhteen.

8.8.1 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

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

Täastä seuraa se tärkeä suurimman uskottavuuden estimaatin inavarianttisuus.

Lause 8.1 *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 8.13 Jos binomijakaumasta $\text{Bin}(10, \theta)$ 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.

8.8.2 Kvadraattisen likiarvon parantaminen

Käytännössä saatetaan hakea sellaista parametrimuunnosta $\psi = g(\theta)$, että ψ:n uskottavuusfunktio olisi säännöllisempi kuin θ:n. Jos ψ:n uskottavuusfunktio on likimain kvadraattinen, niin θ:n luottamusväli voidaan määrittää ψ:n luottamusvälin avulla. Silloin $g(\theta)$ Waldin luottamusväli on

$$g(\hat{\theta}) \pm 1.96 \text{std}(g(\hat{\theta})),$$

josta saadaan θ:n luottamusväli. Voidaan osoitaa, että

$$\text{std}(g(\hat{\theta})) = \text{std}(\theta) |\frac{\partial g}{\partial \hat{\theta}}|.$$

Esimerkki 8.14 Oletetaan, että binomijakaumasta $\text{Bin}(10, \theta)$ saadaan havainto $x = 8$. Silloin θ:n SUE on $\hat{\theta} = 0.8$. Tarkastellaan logaritmoitua vetoa

$$\psi = g(\theta) = \log \frac{\theta}{1-\theta}.$$

Nyt ψ:n SUE on

$$\hat{\psi} = \log \frac{0.8}{0.2} = 1.39.$$

Lasketaan ensin

$$\frac{\partial g}{\partial \theta} = \frac{1}{\theta} + \frac{1}{1-\theta},$$

josta

$$\begin{aligned}\mathrm{std}(\psi) &= \mathrm{std}(\theta)\left(\frac{1}{\hat{\theta}} + \frac{1}{1-\hat{\theta}}\right)\\ &= \left(\frac{1}{x} + \frac{1}{n-x}\right)\\ &= 0.79.\end{aligned}$$

Waldin 95%:n luottamusväli ψ:lle on

$$1.39 \pm 1.96 \times 0.79,$$

joten $-0.16 < \psi < 2.94$. Kun tämä muunnetaan takaisin alkuperäiseen parametrisointiin, saadaan

$$0.46 < \theta < 0.95.$$

□

8.8.3 Uskottavuuteen perustuva luottamusväli on parempi kuin Waldin luottamusväli

Ei ole välttämättä helppoa tietää, millä muunnoksella uskottavuusfunktio saadaan säännöllisemmäksi. Uskottavuuteen perustuvat välit välttävät tämän vaikeuden. Uskottavuuteen perustuva θ:n 95%:n luottamusväli on

$$\left\{\theta : 2\log\frac{L(\hat{\theta})}{L(\theta)} \leq 3.84\right\}$$

ja Waldin väli on

$$\hat{\theta} \pm 1.96\mathrm{std}(\hat{\theta}).$$

Vaikka molemmat välit perustuvat normaalijakauman likiarvoon, on uskottavuuteen perustuva väli parempi.

Waldin väli on täsmällinen vain silloin, jos

$$\frac{\hat{\theta}-\theta}{\mathrm{std}(\hat{\theta})} \sim \mathrm{N}(0,1).$$

Toisaalta uskottavuuteen perustuva väli on täsmällinen, jos on olemassa sellainen bijektiivinen muunnos $g(\cdot)$, että

$$\frac{g(\hat{\theta})-g(\theta)}{\mathrm{std}(g(\hat{\theta}))} \sim \mathrm{N}(0,1). \tag{8.8.3}$$

Tämä perustuu uskottavuussuhteen invarianttisuuteen. Huomattakoon, että meidän ei tarvitse tuntea muunnosfunktiota g. Jos $A < g(\theta) < Y$ on $g(\theta)$:n 95%:n luottamusväli ja $g(\hat{\theta})$ noudattaa (8.8.3):n perusteella normaalijakaumaa, väli (A, Y) on uskottavuusväli, jossa uskottavuusfunktion kynnysarvo on 0.15. Silloin uskottavuussuhteen invarianttisuuden nojalla θ:n uskottavuusväli samalla kynnysarvolla on

$$g^{-1}(A) < \theta) < g^{-1}(Y),$$

jonka luottamustaso on 95%.