Pekka Nieminen Pentti Saikkonen

Tilastollisen päättelyn kurssi

Helsingin yliopisto
Matematiikan ja tilastotieteen laitos
2013

1. laitos 2004
2. laitos 2005
3. laitos 2006, korjattu 2007, 2009 ja 2013

http://www.iki.fi/pjniemin/paattely.pdf

Saatteeksi

Nämä luentomuistiinpanot syntyivät tilastollisen päättelyn kurssilla, jota luennoin lukuvuonna 2003–2004 suorittaessani siviilipalvelusta Helsingin yliopiston tilastotieteen laitoksessa (vuodesta 2004 matematiikan ja tilastotieteen laitos). Korjauksia ja joitakin muutoksia sekä lisäyksiä olen tehnyt lukuvuosina 2004–2005 ja 2005–2006. Suuri osa asiasisällöstä ja paikoitellen myös esitystapa pohjautuvat professori Pentti Saikkosen keväällä 2003 pitämiin luentoihin. Joitakin vaikutteita on otettu myös professori Anders Ekholmin luentomonisteesta *Johdatus uskottavuuspäättelyyn* (Helsingin yliopisto, tilastotieteen laitos).

Tarkoituksena kurssilla on tutustua eräisiin klassisen tilastollisen päättelyn keskeisiin käsitteisiin ja ajatuskulkuihin, joiden oikea ymmärtäminen on välttämätöntä tilastollisia menetelmiä sovellettaessa ja kehitettäessä. Näkökulma on uskottavuuspohjainen: uskottavuusfunktiolla ja sen johdannaisilla on tärkeä rooli sekä käytännön laskuissa että esiteltävässä teoriassa. Toisaalta juuri lainkaan ei kiinnitetä huomiota tilastollisten mallien valintaan ja rakentamiseen liittyviin kysymyksiin, vaikka ne ovatkin tilastollisessa tutkimustyössä ensiarvoisen tärkeitä.

Muutamat todistukset on merkitty asteriskilla (*). Ne ovat täydentävää ainesta, eikä niitä ole vaadittu kokeessa osattaviksi.

Huomautuksia virheistä ja parannusehdotuksia voi lähettää sähköpostitse osoitteeseen `pjniemin@iki.fi`.

18. 3. 2006 Pekka Nieminen

Sisältö

1 Johdanto

1.1 Tilastollisen päättelyn lähtökohta

1.1.1 Aineisto ja tilastollinen malli. Tilastollisessa tutkimuksessa on analysoitavana *aineisto* $\boldsymbol{y} = (y_1, \ldots, y_n)$, joka koostuu reaaliarvoisista *havainnoista* y_i. Näihin havaintoihin oletetaan liittyvän epävarmuutta tai satunnaisvaihtelua. Havainnot voivat olla peräisin satunnaiskokeesta tai satunnaisotannalla poimitusta perusjoukon osasta, jolloin satunnaisvaihtelu on seurausta aineiston hankintatavasta: kokeen tai otannan toistaminen saisi aikaan toisenlaiset havainnot. Tai voidaan ajatella, että satunnaisvaihtelu on ikään kuin sisäänrakennettuna tutkittavaan ilmiöön: ”luonto” tai ”sattuma” on useista mahdollisista tulosvaihtoehdoista valinnut juuri sen, jota saatu aineisto edustaa. Lisäksi empiirisessä tutkimuksessa mittausvirheet aina aiheuttavat epävarmuutta havaintoihin. Kaiken kaikkiaan havaintoja $y_1, \ldots, y_n$ voidaan pitää eräiden satunnaismuuttujien $Y_1, \ldots, Y_n$ toteutuneina arvoina eli realisaatioina.

Tilastollisen päättelyn yleisenä tavoitteena on tehdä aineiston $\boldsymbol{y}$ perusteella päätelmiä siitä todennäköisyysjakaumasta, jota satunnaisvektori $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ noudattaa, eli siitä satunnaisilmiöstä, joka on tutkimuksen kohteena ja josta saatu aineisto on peräisin. Tunnusomaista on myös esittää arvioita näiden päätelmien luotettavuudesta. Mahdollisten todennäköisyysjakaumien joukko on normaalisti jollain tavalla rajattu, ja niitä kuvataan joko yhteispistetodennäköisyysfunktion (yptf) tai yhteistiheysfunktion (ytf) $f_{\boldsymbol{Y}}$ avulla. Tätä kutsutaan *tilastolliseksi malliksi*. Kyseeseen tulevien mallien valinta perustuu tutkittavaa ilmiötä kuvaavaan taustatietoon ja -teoriaan, esimerkiksi fysikaalisiin lainalaisuuksiin, joita ilmiön tiedetään noudattavan.

Useimmissa tällä kurssilla käsiteltävissä esimerkeissä havainnot $y_1, \ldots, y_n$ ovat jonkin tietyn numeerisen muuttujan arvot n eri havaintoyksikön osalta, jolloin tavallisesti voidaan olettaa, että vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomat. Tällöin on tietysti

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{Y_1}(y_1) \cdots f_{Y_n}(y_n).$$

Käytännön sovelluksissa aineisto kuitenkin monesti sisältää useita samaan havaintoyksikköön liittyviä muuttujien arvoja tai se voi olla myös *aikasarja* eli saman suureen toteutuneet arvot peräkkäisinä ajanhetkinä. Näissä tapauksissa riippumattomuus harvoin toteutuu ja riippuvuusmekanismin selvittäminen on oleellinen osa aineiston analysointia ja tilastollisen mallin spesifiointia.

Huomattakoon, että havaintojen määrä n on tällä kurssilla kiinteä luku, joka annetaan tutkimukseen käytettävää koeasetelmaa valittaessa ja tilastollista mallia määriteltäessä. Sitä ei siis pidetä aineistosta riippuvana suureena.

1.1.2 Parametrinen päättely. Useimmissa tilastollisen päättelyn sovellustilanteissa otetaan lähtökohdaksi, että tilastollinen malli on tunnettu yhtä tai useampaa reaalista parametria vaille. Vastaavalla yptf/ytf:lla on siis tunnettu funktionaalinen lauseke $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, joka riippuu *parametrista* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. (Tapauksessa $d = 1$ merkitään tietysti $\boldsymbol{\theta} = \theta$.) Mallia kutsutaan tällöin *parametriseksi malliksi*. Mahdolliset parametriarvot muodostavat joukon $\Omega \subset \mathbb{R}^d$, jota kutsutaan *parametriavaruudeksi* ja jonka määrittely on osa mallin spesifiointia. Parametrin $\boldsymbol{\theta}$ arvo on tuntematon, ja tavoitteena on tehdä siitä päätelmiä aineiston $\boldsymbol{y}$ perusteella. Tällöin tilastollista päättelyä sanotaan *parametriseksi päättelyksi*, ja siihen rajoitutaan tällä kurssilla.

1.2 Esimerkkejä parametrisista malleista

Seuraavat esimerkit osoittavat, että edellä kuvattu parametristen tilastollisten mallien asetelma kattaa suuren määrän erilaisia malleja erilaisissa sovellustilanteissa.

1.2.1 Suhteellinen osuus: toistokoemalli. Tehdas on valmistanut suuren määrän hehkulamppuja, ja halutaan tutkia, kuinka suuri osa niistä on rikkinäisiä. Poimitaan kokoa n oleva otos lamppuja ja määritellään

$$y_i = \begin{cases} 0, & \text{jos } i\text{:s lamppu on ehjä,} \\ 1, & \text{jos } i\text{:s lamppu on rikki,} \end{cases}$$

kun $i = 1, \ldots, n$. Aineisto on $\boldsymbol{y} = (y_1, \ldots, y_n)$. Vastaavat satunnaismuuttujat Y_i noudattavat Bernoulli-jakaumaa $B(\theta)$, jossa θ on rikkinäisten lamppujen osuus koko tuotannossa. Siis $P\{Y_i = 1\} = \theta$ ja $P\{Y_i = 0\} = 1 - \theta$, eli kunkin Y_i:n pistetodennäköisyysfunktio (ptf) on

$$f_{Y_i}(y_i;\theta) = \theta^{y_i}(1-\theta)^{1-y_i}, \qquad y_i = 0, 1.$$

Koska kyseessä on otos suuresta perusjoukosta, voidaan olettaa, että $Y_1, \ldots, Y_n \perp\!\!\!\perp$. Siten tilastolliseksi malliksi saadaan yptf

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i;\theta) = \theta^k(1-\theta)^{n-k},$$

jossa $k = k(\boldsymbol{y}) = y_1 + \cdots + y_n$ on rikkinäisten lamppujen lukumäärä otoksessa. Kyseessä on parametrinen malli, jonka parametri θ on yksiulotteinen ja jonka parametriavaruus on väli $(0, 1)$ tai jopa $[0, 1]$. Tutkijan mielenkiinto kohdistuu tuntematonta parametria θ koskevien päätelmien tekoon.

1.2.2 Kestoiät. Halutaan selvittää, mikä on erään toisen tehtaan valmistamien tietyntyyppisten sähkölaitteiden keskimääräinen kestoikä. Poimitaan jälleen kokoa n oleva otos ja mitataan kunkin otosyksikön kestoikä; olkoot ne $y_1, \ldots, y_n > 0$. Todennäköisyyslaskennassa on opittu, että erilaisia elinaikoja voidaan usein mallittaa melko hyvin eksponenttijakaumalla.† Oletetaan, että eksponenttijakauma soveltuu myös nyt tarkasteltavien kestoikien kuvaamiseen. Havaintoja vastaavat satunnaismuuttujat

† Perusteluna tälle oli eksponenttijakauman "muistamattomuusominaisuus": jos $Y \sim Exp(\lambda)$, niin $P\{Y \geq t + h \,|\, Y \geq t\} = P\{Y \geq h\}$.

ovat siis $Y_1, \ldots, Y_n \sim Exp(\lambda)$ $\perp\!\!\!\perp$, jossa $\lambda > 0$ on tuntematon parametri. Tilastollisen mallin spesifioi nyt ytf

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda y_i} = \lambda^n e^{-\lambda(y_1 + \cdots + y_n)} = \lambda^n e^{-\lambda n \overline{y}},$$

jossa $\overline{y} = (y_1 + \cdots + y_n)/n$ on otoksen laitteiden kestoikien keskiarvo.

Koska eksponenttijakaumalle pätee $E(Y_i) = 1/\lambda$ ja mielenkiinnon kohteena on nimenomaan kestoiän odotusarvo, on ehkä luontevampaa käyttää parametria $\mu = 1/\lambda$, jolloin malliksi saadaan

$$f^*_{\boldsymbol{Y}}(\boldsymbol{y}; \mu) = f_{\boldsymbol{Y}}(\boldsymbol{y}; 1/\mu) = \frac{1}{\mu^n} e^{-n\overline{y}/\mu}.$$

Mallit $f_{\boldsymbol{Y}}$ ja $f^*_{\boldsymbol{Y}}$ ovat tilastollisesti täysin ekvivalentit, koska ne kuvaavat samaa todennäköisyysjakaumien joukkoa, mutta jälkimmäinen voi olla usein tulkinnallisesti mukavampi. Tämä on esimerkki *uudelleenparametroinnista*.

1.2.3 Normaalihavainnot. Yksi perustavimmista tilastollisista malleista on riippumaton otos normaalijakaumasta $N(\mu, \sigma^2)$. Siis $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Mallin parametri on kaksiulotteinen vektori (μ, σ^2), ja parametriavaruus on ylempi puolitaso $\mathbb{R} \times (0, \infty)$. Malliin liittyvä ytf on

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - \mu)^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \Big\}.$$

1.2.4 Lineaarinen regressiomalli. Tutkija haluaa selvittää erään metallisen työkalun lujuuden riippuvuutta siitä lämpötilasta, jossa työkalu on valmistettu. Hän tuottaa n työkalua eri lämpötiloissa $x_1, \ldots, x_n$ ja mittaa niiden lujuudet $y_1, \ldots, y_n$ (sopivissa yksiköissä). Hän otaksuu, että keskimäärin lujuus riippuu likipitäen lineaarisesti lämpötilasta, joten vastaavista satunnaismuuttujista $Y_1, \ldots, Y_n$ on järkevää olettaa, että eräillä kertoimilla α, β pätee

$$E(Y_i) = \alpha + \beta x_i, \tag{1.1}$$

kun $i = 1, \ldots, n$. Lisäksi lujuudessa on eri syistä esiintyvää satunnaisvaihtelua, jonka tutkija olettaa normaaliseksi ja varianssiltaan vakioksi (lämpötilasta riippumattomaksi). Siis

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2). \tag{1.2}$$

Jos vielä oletetaan eri työkalut toisistaan riippumattomiksi eli $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$, saadaan tilastollisen mallin lausekkeeksi

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \Big\}.$$

Kyseessä on yhden selittäjän *lineaarinen regressiomalli*. Tässä on syytä panna merkille, että havainnot $Y_1, \ldots, Y_n$ eivät ole samoin jakautuneita.

Mallin rakentamisessa on hyödyllistä erottaa kaksi vaihetta. Ensinnäkin siihen liittyi *rakenneoletus* (1.1), joka kuvaa *vastemuuttujan* Y_i vaihtelun systemaattisen osan eli sen riippuvuuden *selittävästä muuttujasta* x_i. Toiseksi kaava (1.2) toi mukanaan *jakaumaoletuksen*, joka kuvaa vaihtelun satunnaisen osan. Mallin parametri on kolmiulotteinen vektori $(\alpha, \beta, \sigma^2)$, ja parametriavaruus määräytyy ehdoista $\alpha, \beta \in \mathbb{R}$ ja

$\sigma^2 > 0$. Tutkijan pääasiallinen mielenkiinto luultavasti kohdistuu komponentteja α ja β koskevien päätelmien tekoon, kun taas tuntematon varianssi σ^2 on eräänlaisen *kiusaparametrin* asemassa.

Mallin määrittelystä käy ilmi, että selittävän muuttujan arvoihin x_i ei liitetty satunnaisvaihtelua vaan niitä pidettiin kiinteinä annettuina lukuina; itse asiassa niitä ei edes luettu aineiston osaksi ainakaan siinä teknisessä mielessä kuin "aineisto" on luvun alussa määritelty. Tarkasteltavassa esimerkkitilanteessa tämä tuntuu loogiselta siksi, että valmistuslämpötila lienee tutkijan valittavissa ennen kokeen suorittamista. Monissa sovellustilanteissa näin ei kuitenkaan ole. Esimerkki: Tutkitaan suomalaisten avioparien kohdalla sitä, miten vaimon pituus on selitettävissä miehen pituuden avulla. Tätä varten poimitaan satunnaisotos aviopareista ja mitataan pituudet. Tällöin miesten pituuksiin x_i liittyy satunnaisvaihtelu aivan samalla tavalla kuin vaimojenkin pituuksiin y_i (suoritetun satunnaisotannan johdosta), joten periaatteessa mallittajan tulisi tutkia eräiden satunnaismuuttujaparien (X_i, Y_i) yhteisjakaumaa. Näin ei kuitenkaan yleensä menetellä vaan tässäkin tapauksessa tilastollinen analyysi suoritetaan pitäen lukuja x_i kiinteinä eli ikään kuin ehdollistamalla tapahtumien $X_i = x_i$ suhteen. Tämän taustalla on tilastollisen päättelyn ns. ehdollistamisperiaate.

1.2.5 Autoregressiivinen aikasarja. Olkoot $U_1, \ldots, U_n \sim N(0, \sigma^2)$ $\perp\!\!\!\perp$, ja määritellään satunnaismuuttujat $Y_1, \ldots, Y_n$ palautuskaavalla

$$Y_i = \beta Y_{i-1} + U_i, \qquad i = 1, \ldots, n, \tag{1.3}$$

kun oletetaan, että $Y_0 = y_0$ on tunnettu vakio. Satunnaisvektorin $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ ytf on tällöin (ks. teht. 1.3)

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta y_{i-1})^2\Big\}.$$

Prosessia (Y_i), jolle (1.3) on voimassa, kutsutaan *ensimmäisen kertaluvun autoregressiiviseksi aikasarjaksi*. Sitä ja sen yleistyksiä käytetään esimerkiksi ekonometriassa erilaisten ajassa etenevien taloudellisten ilmiöiden mallittamiseen. Mallin parametri on (β, σ^2), ja siitä on tapana olettaa, että $|\beta| < 1$ ja tietysti $\sigma^2 > 0$. Yhtälön (1.3) valossa on ilmeistä, että $Y_1, \ldots, Y_n$ eivät ole riippumattomia mikäli $\beta \neq 0$.

1.3 Parametrisen päättelyn tavoitteet

Tällä kurssilla käsitellään lähinnä seuraavia kolmea kysymystä:

Piste-estimointi. Aineiston perusteella on määritettävä sellainen parametriavaruuden piste, joka on jossain mielessä hyvä tai jopa paras arvio eli *estimaatti* tuntemattomalle parametrille $\boldsymbol{\theta}$. Yhtäältä tarkastellaan menetelmiä, joilla tällaisia estimaatteja generoidaan (esim. suurimman uskottavuuden menetelmä), ja toisaalta kriteerejä, joilla niiden hyvyyttä mitataan.

Luottamusvälit ja -joukot. On rajattava parametriavaruuden Ω osajoukko, joka suurella varmuudella sisältää tuntemattoman parametrin $\boldsymbol{\theta}$ oikean arvon. Yksiulotteisen parametrin tapauksessa nämä joukot ovat yleensä välejä, joten niiden muodostamista kutsutaan myös *väliestimoinniksi*.

Hypoteesien testaaminen. On selvitettävä, onko aineisto sopusoinnussa annetun nollahypoteesin $\boldsymbol{\theta} \in \Omega_0$ kanssa vai tukeeko se ennemminkin vastahypoteesia $\boldsymbol{\theta} \in \Omega_1$, jossa tavallisesti $\Omega = \Omega_0 \cup \Omega_1$ ja $\Omega_0 \cap \Omega_1 = \emptyset$.

Näiden lisäksi ainakin seuraavat kaksi tärkeää tehtävää kuuluvat tilastollisen päättelyn piiriin, mutta niitä ei juuri käsitellä tällä kurssilla:

Ennustaminen. Oletetaan, että mallista $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ on havaittu aineisto $\boldsymbol{y}$. Tämän perusteella on ennustettava jonkin satunnaismuuttujan Z arvoa, kun Z:n jakauma riippuu parametrista $\boldsymbol{\theta}$ mutta sen arvoa ei voida (vielä) havaita. Esimerkiksi Z voisi olla aikasarjassa seuraavana tuleva muuttuja Y_{n+1}.

Mallin sopivuuden ja riittävyyden arviointi. Kun aineisto on analysoitu mallia $f_{\boldsymbol{Y}}$ käyttäen, on tutkittava, onko kyseinen malli riittävän hyvä tai lainkaan sopiva kuvaamaan aineistoa. Esimerkiksi yhden selittäjän lineaarisen regressiomallin tapauksessa, jossa rakenneoletus on muotoa $E(Y_i) = \alpha + \beta x_i$, on varmistuttava siitä, että selittävän muuttujan ja vastemuuttujan välinen riippuvuus todella on keskimäärin lineaarista. Mallin valintaa ohjaavat tällöin sekä tilastolliset että asialoogiset perusteet. Tässä yhteydessä on muistettava myös ns. säästäväisyysperiaate, jonka mukaan tieteellisen selityksen on aina oltava mahdollisimman yksinkertainen; siten on katsottava eduksi, jos $d-1$ parametria sisältävällä mallilla voidaan kuvata tutkittava ilmiö (lähes) yhtä hyvin kuin d parametria sisältävällä.

Käytettävien mallien sopivuuden ja riittävyyden arvioinnin tulisi olla jokaisen tilastolliseen päättelyyn perustuvan tutkimuksen osa. Siihen liittyy myös eräs tilastollisen päättelyn ja mallittamisen keskeinen dilemma: aineiston analysointi ja siihen perustuva päättely edellyttää osaltaan tilastollisen mallin spesifiointia, mutta toisaalta mallin sopivuuden tilastollinen arviointi on täysin mahdollista vasta analyysin jälkeen.

1.4 Todennäköisyyskäsitteen tulkinnasta: frekventistinen ja bayesläinen päättely

Kuten luvun alusta käy ilmi, aineiston $\boldsymbol{y}$ ymmärtäminen satunnaisvektorin toteutuneena arvona perustuu siihen usein sangen hypoteettiseen ajatukseen, että aineistonkeruu voitaisiin samoissa olosuhteissa toistaa riippumattomasti uudelleen ja uudelleen, ja tarkasteltava tilastollinen malli $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ kuvaa näin syntyvään empiiriseen jakaumaan liittyviä todennäköisyyksiä. Todennäköisyyskäsitettä tulkitaan siis *frekventistisesti*, viime kädessä nojautuen suurten lukujen lakiin. Parametrin $\boldsymbol{\theta}$ ajatellaan olevan kiinteä mutta tuntematon piste parametriavaruudessa eikä siihen koskaan liitetä mitään todennäköisyyslausumia. Tähän paradigmaan perustuvaa päättelyä kutsutaan *frekventistiseksi* tai joskus myös *klassiseksi* päättelyksi.

Bayesläisessä päättelyssä myös mallin parametri ajatellaan satunnaisvektorin $\boldsymbol{\Theta}$ arvoksi. Tutkija määrittelee, periaatteessa jo ennen aineistonkeruuta, parametriavaruuteen ns. priorijakauman $f_{\boldsymbol{\Theta}}$, johon hän kvantifioi omat ennakkotietonsa ja -uskomuksensa mallin parametrista. Tällöin todennäköisyyttä tulkitaan useimmiten subjektiivisesti eli uskomusasteen mittana. Malli $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ puolestaan edustaa ehdollista jakaumaa $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$. Kun aineisto $\boldsymbol{y}$ on kerätty, tutkija soveltaa Bayesin kaavaa ja

päätyy ehdolliseen jakaumaan

$$f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})}{f_{\boldsymbol{Y}}(\boldsymbol{y})},$$

jossa

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_{\Omega} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

jatkuvassa tapauksessa ja

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{\theta}\in\Omega} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$$

diskreetissä tapauksessa. Tämä ns. posteriorijakauma kuvaa mallin parametriin liittyvää tietoa sen jälkeen, kun aineiston antama informaatio on otettu huomioon, ja se toimii päätelmien perustana.

Frekventistisen ja bayesläisen koulukunnan välinen näkemysero on toisinaan aiheuttanut hyvinkin kiivaita kiistoja. Bayesläistä päättelyä voidaan syyttää priorijakauman valinnan mukanaan tuomasta subjektiivisuudesta. Bayesläiset puolestaan huomauttavat frekventistiseen päättelyyn liittyvästä tulkinnallisesta vaikeudesta ja nurinkurisuudesta: empiirisen tutkijan kannaltahan aineisto $\boldsymbol{y}$ on kiinteä suure ja epävarmuus liittyy nimenomaan parametriin $\boldsymbol{\theta}$ eikä päinvastoin. Bayesläinen paradigma johtaa yhtenäisempään ja ehkä kauniimpaan päättelyn teoriaan, mutta käytännön johtopäätöksissä harvoin on merkittäviä eroja klassisen päättelyn antamiin tuloksiin verrattuna. Tällä kurssilla bayesläistä päättelyä ei käsitellä.

Harjoitustehtäviä

1.1. Halutaan selvittää, montako kolibakteeria erään järven vedessä on keskimäärin senttilitraa kohti. Tätä varten otetaan n riippumatonta senttilitran suuruista vesinäytettä satunnaisesti eri puolilta järveä ja mitataan niistä kolibakteerien määrät $y_1, \ldots, y_n$. Oletetaan, että lukumäärä kussakin näyteessä noudattaa Poisson-jakaumaa $P(\mu)$. Muodosta asetelmaa kuvaava malli. Mikä on parametrin μ tulkinta?

1.2. Tehtävän 1.1 tilanteessa osoittautuu, että bakteerien lukumäärän mittaaminen kustakin näytteestä on liian kallista. Käytetäänkin vain testiä, joka kertoo, onko näytteessä lainkaan kolibakteereita. Saadaan siis aineisto $\boldsymbol{y} = (y_1, \ldots, y_n)$, jossa $y_i = 0$, jos näytteessä i ei ole bakteereita, ja $y_i = 1$, jos bakteereita on. Muodosta asetelmaa kuvaava malli parametrille μ.

Vihje. Ajattele asetelmaa toistokokeena, jonka onnistumistodennäköisyys määräytyy jakaumasta $P(\mu)$ parametrin μ funktiona.

1.3. Johda esimerkissä 1.2.5 jatkuvien jakaumien muunnosteorian avulla satunnaisvektorin $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ yhteistiheysfunktio $f_{\boldsymbol{Y}}$. Voit olettaa yksinkertaisuuden vuoksi $Y_0 = y_0 = 0$.

Ohje. Vektorin $\boldsymbol{U} = (U_1, \ldots, U_n)$ ytf on helppo muodostaa. Palautuskaavasta voi ratkaista U_i:n Y_{i-1}:n ja Y_i:n funktiona, jolloin $\boldsymbol{U}$ tulee esitettyä melko yksinkertaisena lineaarimuunnoksena $\boldsymbol{Y}$:stä. Muista, että kolmiomatriisin determinantti on diagonaalialkioiden tulo.

1.4. Tarkastellaan esimerkin 1.2.1 koeasetelmaa, jossa tutkittiin rikkinäisten lamppujen suhteellista osuutta. Eläydytään bayesläisen tilastotieteilijän ajatusmalliin. Oletetaan, että hänellä ei ole käytössään mitään erityistä ennakkotietoa rikkinäisten suhteellisesta osuudesta θ, joten hän asettaa tasaisen priorijakauman: $f_{\Theta}(\theta) = 1$, kun $0 < \theta < 1$. Sitten hän poimii otoksen kokoa n, analysoi sen ja laskee lopuksi Bayesin kaavaa käyttäen posteriorijakauman $f_{\Theta|\boldsymbol{Y}}(\theta|\boldsymbol{y})$. Totea, että kyseessä on eräs beetajakauma, ja laske sen odotusarvo sekä moodi (so.

kohta, jossa tiheysfunktio on suurimmillaan). Hahmottele myös sen tiheysfunktion kuvaajaa; kiinnitä erityisesti huomiota siihen, mitä tapahtuu jos otoskoko n kasvaa ja rikkinäisten suhteellinen osuus otoksessa pysyy samana. Millaisia johtopäätöksiä tekisit parametrista θ tämän posteriorijakauman perusteella?

Apu. Beetajakauman tiheysfunktio on muotoa $f(\theta) = c \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$, $0 < \theta < 1$, jossa $\alpha, \beta > 0$ ovat jakauman parametrit ja c on niistä määräytyvä vakio, joka ei riipu θ:sta ja jonka arvolla ei tässä ole merkitystä. Jakauman odotusarvo on $\alpha/(\alpha+\beta)$.

2 Uskottavuus ja informaatio

2.1 Uskottavuusfunktio

2.1.1 Perusmääritelmä ja tulkinta. Tarkastellaan tilastollista mallia, jonka yptf/ytf on $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ ja jonka parametriavaruus on $\Omega \subset \mathbb{R}^d$. Tällöin funktiota

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta};\boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \Omega,$$

kutsutaan aineistoon $\boldsymbol{y}$ liittyväksi *uskottavuusfunktioksi* parametrille $\boldsymbol{\theta}$. Uskottavuusfunktio on siis oikeastaan sama asia kuin mallin määrittelevä yptf/ytf, mutta siinä aineistoa edustava vektori $\boldsymbol{y}$ ajatellaan kiinteäksi ja parametri $\boldsymbol{\theta}$ muuttujaksi. Tämä oivallus tekee uskottavuusfunktiosta tilastollisen päättelyn kenties keskeisimmän työkalun.

Tarkastellaan erityisesti diskreettiä tapausta. Tällöin

$$L(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}\{\boldsymbol{Y} = \boldsymbol{y}\},$$

jossa alaindeksi $\boldsymbol{\theta}$ viittaa siihen, että tapahtuman $\{\boldsymbol{Y} = \boldsymbol{y}\}$ todennäköisyyden laskennassa käytetään nimenomaan parametriarvoa $\boldsymbol{\theta}$. Siis luku $L(\boldsymbol{\theta})$ ilmaisee todennäköisyyden sille, että parametriarvo $\boldsymbol{\theta}$ ”tuottaa” juuri sellaisen havaintoaineiston $\boldsymbol{y}$ kuin nyt on saatu. Jos esimerkiksi $\boldsymbol{\theta}'$ ja $\boldsymbol{\theta}''$ ovat Ω:n pisteitä, joille

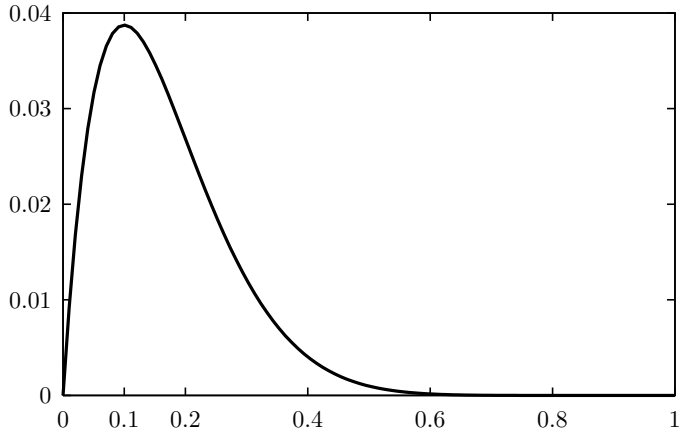
$$L(\boldsymbol{\theta}') < L(\boldsymbol{\theta}'') \qquad \text{eli} \qquad P_{\boldsymbol{\theta}'}\{\boldsymbol{Y} = \boldsymbol{y}\} < P_{\boldsymbol{\theta}''}\{\boldsymbol{Y} = \boldsymbol{y}\},$$

todennäköisyys saada aineisto $\boldsymbol{y}$ on suurempi tapauksessa $\boldsymbol{\theta} = \boldsymbol{\theta}''$ kuin tapauksessa $\boldsymbol{\theta} = \boldsymbol{\theta}'$. On tapana sanoa, että havaitun aineiston valossa parametriarvo $\boldsymbol{\theta}''$ on *uskottavampi* kuin $\boldsymbol{\theta}'$.

Jatkuvassa mallissa uskottavuusfunktion tulkinta on samanlainen. Jos nimittäin A on pisteen $\boldsymbol{y}$ pieni ympäristö avaruudessa $\mathbb{R}^n$, jonka n-ulotteinen mitta on $m(A)$, niin approksimatiivisesti

$$P_{\boldsymbol{\theta}}\{\boldsymbol{Y} \in A\} = \int_A f_{\boldsymbol{Y}}(\cdot;\boldsymbol{\theta}) \approx L(\boldsymbol{\theta})m(A)$$

ainakin silloin kun $f_{\boldsymbol{Y}}(\cdot;\boldsymbol{\theta})$ on jatkuva pisteessä $\boldsymbol{y}$. Luku $L(\boldsymbol{\theta})$ on siis approksimatiivisesti suoraan verrannollinen todennäköisyyteen saada aineisto, joka on ”lähellä” $\boldsymbol{y}$:tä. Vaihtoehtoisesti $L(\boldsymbol{\theta})$ voidaan ajatella ”todennäköisyystiheytenä”.



Kuva 2.1. Uskottavuusfunktion $L(\theta) = \theta(1-\theta)^9$ kuvaaja.

2.1.2 Esimerkki: toistokoemalli. Palataan kohdan 1.2.1 esimerkkiin, jossa tarkasteltiin rikkinäisten hehkulamppujen suhteellista osuutta θ erään tehtaan tuotannossa. Tilastollinen malli oli

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) = \theta^k(1-\theta)^{n-k},$$

jossa $k = k(\boldsymbol{y}) = y_1 + \cdots + y_n$ on rikkinäisten lukumäärä n lampun otoksessa. Oletetaan nyt, että otoksen koko on $n = 10$ ja siinä toisena oleva lamppu havaitaan rikkinäiseksi, muut ehjiksi. Aineisto on siis $\boldsymbol{y} = (0,1,0,0,0,0,0,0,0,0)$, ja $k = 1$. Uskottavuusfunktio on tällöin

$$L(\theta) = P_\theta\{\boldsymbol{Y} = \boldsymbol{y}\} = \theta(1-\theta)^9, \qquad 0 < \theta < 1,$$

jonka kuvaaja on kuvassa 2.1.

Uskottavuusfunktio on suurimmillaan pisteessä $\theta = 0.1$, jossa $L(0.1) \approx 0.039$. Tämä on siis aineiston valossa uskottavin θ:n arvo, ja sitä tullaan kutsumaan θ:n *suurimman uskottavuuden estimaatiksi*. Myös monet muut parametriarvot ovat varsin uskottavia. Esimerkiksi $L(0.25) \approx 0.018$ on lähes puolet maksimiarvosta $L(0.1)$, joten siinä tapauksessa, että todellinen parametriarvo on 0.25, tällaisia otoksia saadaan (otantaa toistettaessa) vain noin kaksi kertaa harvemmin kuin siinä tapauksessa, että todellinen arvo on 0.1. Aineiston valossa ei ole siis lainkaan poissuljettua se, että peräti joka neljäs tuotetuista lampuista olisi rikki. Toisaalta myös arvo 0.01 (joka sadas lamppu rikki) on melko uskottava, sillä $L(0.01) \approx 0.009$. Sen sijaan $L(0.002) \approx 0.002$ on vain kahdeskymmenesosa uskottavuusfunktion maksimiarvosta, joten vaikuttaa aineiston perusteella jokseenkin epäuskottavalta, jos valmistaja väittää, että keskimäärin vain joka viidessadas tuotetuista lampuista olisi rikki. Yleisvaikutelmaksi kuitenkin jää, että kovin spesifisiä ja luotettavia päätelmiä parametrista θ ei voi tämän aineiston perusteella tehdä, mikä ei olekaan yllättävää otoksen pienuudesta johtuen.

2.1.3 Yleinen määritelmä. Kuten edeltä käy ilmi, yleensä uskottavuusfunktion absoluuttisilla arvoilla ei ole sinänsä merkitystä vaan tärkeintä on voida verrata sen eri kohdissa saamia arvoja eli tarkastella suhteita $L(\boldsymbol{\theta}')/L(\boldsymbol{\theta}'')$. Nämä eivät muutu, jos uskottavuusfunktio kerrotaan jollakin (mahdollisesti aineistosta riippuvalla) vakiolla. Uskottavuusfunktion määritelmä onkin tapana laajentaa seuraavasti:

Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ tilastollinen malli, jonka parametriavaruus on Ω. Tällöin jokainen

muotoa

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \boldsymbol{y}) = c(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \Omega, \tag{2.1}$$

oleva funktio L on aineistoon $\boldsymbol{y}$ liittyvä *uskottavuusfunktio*, kun $c(\boldsymbol{y}) > 0$ on mahdollisesti aineistosta (mutta ei parametrista) riippuva vakio. Tämä vakio kannattaa yleensä valita siten, että uskottavuusfunktion lauseke tulee mahdollisimman yksinkertaiseksi.

2.1.4 Esimerkki: normaalimalli. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa parametri on kaksiulotteinen (μ, σ^2). Kohdan 1.2.3 esimerkissä muodostettiin tämän mallin ytf

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\Big\}.$$

Kaavan (2.1) vakioksi kannattaa nyt valita $c(\boldsymbol{y}) = (2\pi)^{n/2}$. Käytetään lisäksi eksponentissa hajotelmaa

$$\sum_{i=1}^{n}(y_i - \mu)^2 = (n-1)s^2 + n(\overline{y} - \mu)^2,$$

jossa

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

ovat aineistosta lasketut *otoskeskiarvo* ja *otosvarianssi* (ks. teht. 2.2). Tällöin nähdään, että mallin uskottavuusfunktio riippuu aineistosta ainoastaan kaksiulotteisen tunnusluvun† $(\overline{y}, s^2)$ välityksellä:

$$L(\mu, \sigma^2) = \frac{1}{(\sigma^2)^{n/2}} \exp\Big\{-\frac{(n-1)s^2 + n(\overline{y} - \mu)^2}{2\sigma^2}\Big\}.$$

(Muista, että n on kiinteä ja tunnettu tilastolliseen malliin liittyvä luku eikä sitä ajatella aineistosta riippuvaksi.)

Tarkastellaan vielä tapausta, jossa varianssi onkin eräs tunnettu luku $\sigma^2 = \sigma_0^2 > 0$ ja mallin varsinainen parametri on vain odotusarvo μ. Tällöin kerroin $1/(\sigma_0^2)^{n/2}$ samoin kuin eksponenttilausekkeesta tuleva tekijä $\exp\{-(n-1)s^2/2\sigma_0^2\}$ ovat vain aineistosta (sekä tunnetuista luvuista n ja σ_0^2) riippuvia, joten ne voidaan jakaa pois uskottavuusfunktion lausekkeesta. Uskottavuusfunktio parametrille μ saa siis yksinkertaisen muodon

$$L(\mu) = \exp\Big\{-\frac{n(\overline{y} - \mu)^2}{2\sigma_0^2}\Big\}.$$

On mielenkiintoista havaita, että nyt uskottavuusfunktio riippuu aineistosta vain yksiulotteisen tunnusluvun $\overline{y}$ kautta.

2.1.5 Esimerkki: toistokokeen binomimalli. Palataan äskeiseen lamppuesimerkkiin. Oletetaan, että tutkijalla on tiedossaan n lampun otoksesta ainoastaan rikkinäisten lukumäärä k mutta ei järjestystä, jossa ehjät ja rikkinäiset esiintyvät otoksessa.

† *Tunnusluvulla* tarkoitetaan mitä tahansa aineiston $\boldsymbol{y}$ muunnosta eli funktiota $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$ tai vastaavaa satunnaismuuttujaa $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$. Tunnusluvut voivat olla reaali- tai vektoriarvoisia.

Aineisto on siis yksiulotteinen suure k, joka voidaan tulkita "onnistumisten" lukumääräksi n-kertaisessa toistokokeessa. Siispä vastaava satunnaismuuttuja K noudattaa binomijakaumaa $Bin(n, \theta)$, jonka ptf

$$f_K(k;\theta) = P_\theta\{K = k\} = \binom{n}{k}\theta^k(1-\theta)^{n-k}, \qquad k = 0, \ldots, n,$$

spesifioi tilannetta kuvaavan mallin. Mallin parametri on rikkinäisten lamppujen suhteellinen osuus θ, sillä otoskokoa n pidetään tunnettuna lukuna. Valitsemalla uskottavuusfunktion määritelmässä (2.1) esiintyväksi vakioksi $1/\binom{n}{k}$ saadaan mallia vastaavaksi uskottavuusfunktioksi

$$L(\theta) = \theta^k(1-\theta)^{n-k}, \qquad 0 < \theta < 1,$$

aivan kuten esimerkissä 2.1.2.

Tuntuu intuitiivisesti selvältä, että tuntematonta parametria θ koskeviin päätelmiin ei pitäisi mitenkään vaikuttaa sen, missä järjestyksessä ehjät ja rikkinäiset lamput otoksessa esiintyvät, vaan että kaikki oleellinen informaatio otoksesta sisältyy lukuun k. Tätä seikkaa heijastaa osaltaan se, että kumpikin malli johtaa samaan uskottavuusfunktioon. Toisinaan tilastollisen päättelyn teoriassa yleisestikin vedotaan ns. *uskottavuusperiaatteeseen*, jonka mukaan kahden eri mallin pohjalta tehtävien päätelmien tulisi aina olla samoja, mikäli havaittuihin aineistoihin liittyvät uskottavuusfunktiot ovat malleissa identtiset. Monet tilastolliset menetelmät eivät tätä periaatetta kuitenkaan noudata.

2.1.6 Logaritminen uskottavuusfunktio. Sekä päättelyn teorian että käytännön laskujen kannalta osoittautuu hyödylliseksi tarkastella uskottavuusfunktion logaritmia. Sitä kutsutaan *logaritmiseksi uskottavuusfunktioksi* (lyh. *log-uskottavuusfunktio*) ja merkitään

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \boldsymbol{y}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}).$$

Tässä log on luonnollinen eli e-kantainen logaritmi. Huomaa, että koska logaritmi on aidosti kasvava funktio, eri parametriarvojen uskottavuuden vertailu voidaan suorittaa myös log-uskottavuusfunktion avulla. Koska uskottavuusfunktiota oli sallittua kertoa (tai yhtä hyvin jakaa) aineistosta riippuvilla positiivisilla vakioilla, on log-uskottavuusfunktioon lupa lisätä (tai siitä vähentää) mikä tahansa aineistosta riippuva vakio.

Logaritmisen uskottavuusfunktion teoreettinen merkitys paljastuu myöhemmin informaatiokäsitteen ja testiteorian tarkastelun yhteydessä. Sen käytännöllinen merkitys puolestaan selittyy sillä, että funktio l on usein yksinkertaisempaa tyyppiä kuin L. Jos esimerkiksi malliin liittyvät havainnot $Y_1, \ldots, Y_n$ ovat riippumattomia, niin

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{Y_i}(y_i; \boldsymbol{\theta})$$

kun taas

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_{Y_i}(y_i; \boldsymbol{\theta}).$$

Summamuotoista funktiota on yleensä mukavampi käsitellä, esimerkiksi derivoida, kuin tulomuotoista.

2.1.7 Esimerkkejä. a) Toistokokeen (ks. 2.1.2 ja 2.1.5) log-uskottavuusfunktio on

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

b) Normaalijakaumamallin $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$ (ks. 2.1.4) log-uskottavuusfunktio on

$$l(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{(n-1)s^2 + n(\overline{y} - \mu)^2}{2\sigma^2}.$$

Siinä tapauksessa, että varianssi on tunnettu luku $\sigma^2 = \sigma_0^2$, saadaan parametrin μ log-uskottavuusfunktioksi toisen asteen polynomifunktio

$$l(\mu) = -\frac{n(\overline{y} - \mu)^2}{2\sigma_0^2}.$$

2.2 Suurimman uskottavuuden estimointi

2.2.1 Estimaatti ja estimaattori. Kuten pykälästä 1.3 kävi ilmi, yksi parametrisen päättelyn tavoitteista on piste-estimointi: kun mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ puitteissa on havaittu aineisto $\boldsymbol{y}$, on etsittävä parametriavaruuden Ω piste, joka on hyvä arvio tuntemattomalle parametrille $\boldsymbol{\theta}$. Tätä varten aineistosta laskettua tunnuslukua $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$ sanotaan parametrin $\boldsymbol{\theta}$ *estimaatiksi*. Vastaava satunnaismuuttuja $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on nimeltään *estimaattori*.

Estimaattien johtamiseen on kehitetty useita eri tapoja. Tällä kurssilla käsitellään lähinnä ns. suurimman uskottavuuden estimointia, joka on tilastotieteessä yleisimmin käytetty estimointimenetelmä.

2.2.2 Su-estimaatin määritelmä. Olkoon $L(\boldsymbol{\theta}; \boldsymbol{y})$ yo. mallia ja aineistoa $\boldsymbol{y}$ vastaava uskottavuusfunktio. Tällöin mikä tahansa parametriavaruuden Ω piste $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$, jossa uskottavuusfunktio saa suurimman arvonsa eli jossa

$$L(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) \geq L(\boldsymbol{\theta}; \boldsymbol{y}) \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega,$$

on parametrin $\boldsymbol{\theta}$ *suurimman uskottavuuden estimaatti* (lyh. *su-estimaatti*). Vastaava satunnaismuuttuja $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ on *su-estimaattori*, ja sitäkin voidaan lyhyesti merkitä symbolilla $\hat{\boldsymbol{\theta}}$.

Suurimman uskottavuuden menetelmä tuntuu intuitiivisesti luontevalta estimointiperiaatteelta: valitaan sellainen parametriavaruuden piste, joka tekee juuri havaitun aineiston esiintymisen kaikkein todennäköisimmäksi. On myös osoittautunut, että su-estimaattorit ovat yleensä varsin toimivia ja hyviä teoreettisiltakin ominaisuuksiltaan, joskaan eivät aina optimaalisia. Huomaa, että uskottavuusfunktion määritelmässä esiintyvä vapaavalintainen vakio ei vaikuta kohtaan, jossa funktio saa suurimman arvonsa eli globaalin maksimin. Su-estimaatin voi määrittää yhtä hyvin myös etsimällä log-uskottavuusfunktion globaalin maksimikohdan, mikä onkin usein laskuteknisesti mukavampaa.

2.2.3 Esimerkki: normaalimalli kun varianssi tunnettu. Mallina on $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jossa $\sigma_0^2 > 0$ on tunnettu luku. Tämän mallin log-uskottavuusfunktion $l(\mu) = -n(\overline{y} - \mu)^2 / 2\sigma_0^2$ (ks. 2.1.7) kuvaaja on alasaukeava paraabeli, jonka huippu on kohdassa $\overline{y}$. Siis odotusarvon μ su-estimaatti on $\hat{\mu} = \overline{y} = (y_1 + \cdots + y_n)/n$.

2.2.4 Uskottavuusyhtälöt. Tavallisesti log-uskottavuusfunktiot (ja uskottavuusfunktiot) ovat määrittelyjoukossaan derivoituvia, joten niiden maksimikohtia voi etsiä derivaattatarkastelun avulla. Differentiaalilaskennasta tiedetään, että parametriavaruuden sisäpisteistä funktion l ainoat mahdolliset ääriarvokohdat ovat ne pisteet $\boldsymbol{\theta}$, joissa ensimmäisen kertaluvun osittaisderivaatat häviävät eli

$$\frac{\partial}{\partial \theta_j} l(\boldsymbol{\theta}; \boldsymbol{y}) = 0, \qquad j = 1, \ldots, d. \tag{2.2}$$

Yksiulotteisen parametrin tapauksessa ($d = 1$) tämä tietysti redusoituu yhdeksi yhtälöksi

$$l'(\theta; \boldsymbol{y}) = 0. \tag{2.3}$$

Yhtälöitä (2.2) ja (2.3) kutsutaan *uskottavuusyhtälöiksi*. Lisäksi maksimikohtia voi olla parametriavaruuden reunapisteiden joukossa, jos reunapisteitä ylipäätään parametriavaruuteen kuuluu, esim. mikäli yksiulotteisessa tapauksessa parametriavaruus on suljettu väli.

Uskottavuusyhtälöiden ratkaisuina saatavien pisteiden laatua voi tutkia toisen kertaluvun derivaatan tai osittaisderivaattojen avulla, mikäli nämä ovat olemassa (ja jatkuvia). Yksiulotteisessa tapauksessa lokaaleja maksimikohtia ovat ainakin ne yhtälön (2.3) ratkaisut, joissa $l''(\theta; \boldsymbol{y}) < 0$. Moniulotteisen parametrin tapauksessa vastaava ehto on, että log-uskottavuusfunktion toisen kertaluvun osittaisderivaatoista

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta}; \boldsymbol{y}), \qquad j, k = 1, \ldots, d,$$

muodostuva symmetrinen $d \times d$ -matriisi (ns. *Hessen matriisi*) on negatiivisesti definiitti. Useiden perusmallien kohdalla tilanne on niin yksinkertainen, että parametriavaruus on yhtenäinen ja avoin (ei reunapisteitä) joukko, esim. avoin väli, ja uskottavuusyhtälöillä (2.2) tai (2.3) on vain yksi ratkaisu. Jos tämä ratkaisu toteuttaa em. toisen derivaatan negatiivisuusehdon, niin kyseessä on välttämättä log-uskottavuusfunktion globaali maksimikohta eli siis su-estimaatti $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$.

Tällä kurssilla tarkasteltavien mallien uskottavuusyhtälöt osataan tavallisesti ratkaista analyyttisesti ja su-estimaatit näin ollen ilmoittaa ”suljetussa muodossa” siisteinä lausekkeina. Monimutkaisemmissa malleissa tämä ei yleensä onnistu vaan uskottavuusfunktion maksimikohtien etsinnässä joudutaan turvautumaan tietokoneeseen ja numeerisiin optimointimenetelmiin.

2.2.5 Esimerkki: toistokoemalli. Palataan hehkulamppuesimerkistä tuttuun toistokoemalliin, jossa havaintoja vastaavat satunnaismuuttujat olivat $Y_1, \ldots, Y_n \sim B(\theta)$ $\perp\!\!\!\perp$ (ks. 1.2.1) tai vaihtoehtoisesti vain $K \sim Bin(n, \theta)$ (ks. 2.1.5). Kummassakin tapauksessa log-uskottavuusfunktio parametrille θ eli onnistumistodennäköisyydelle on (ks. 2.1.7)

$$l(\theta) = k \log \theta + (n - k) \log(1 - \theta),$$

kun n toistossa on havaittu k onnistumista. Derivoimalla saadaan

$$l'(\theta) = \frac{k}{\theta} - \frac{n - k}{1 - \theta} = \frac{k - n\theta}{\theta(1 - \theta)},$$

joten uskottavuusyhtälön $l'(\theta) = 0$ ainoa ratkaisu on $\theta = k/n$. Tämän laatua voi tutkia toisen derivaatan avulla, mutta voi myös suoraan havaita, että $l'(\theta)$ on samanmerkkinen kuin osoittajansa, joka vaihtaa merkkinsä kohdassa $\theta = k/n$ plussasta

miinukseen. Siten su-estimaatti on $\hat{\theta} = k/n$ eli onnistumisten suhteellinen osuus suoritetuissa toistoissa, mikä tuntuu järkevältä.

Tarkkaan ottaen yo. päättely oli pätevä vain kun $0 < k < n$; ääritapauksissa $k = 0$ ja $k = n$ su-estimaattia ei ole olemassa, jollei sitten parametriavaruudeksi valita suljettu väli $[0, 1]$ avoimen $(0, 1)$ sijasta.

2.2.6 Esimerkki: normaalimalli. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa estimoitava parametri on kaksiulotteinen (μ, σ^2) ja parametriavaruus ylempi puolitaso eli $\mu \in \mathbb{R}$ ja $\sigma^2 > 0$. Nyt log-uskottavuusfunktio on (ks. 2.1.7)

$$l(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{(n-1)s^2 + n(\overline{y} - \mu)^2}{2\sigma^2}.$$

Pannaan aluksi merkille aivan kuten esimerkissä 2.2.3, että koska aina $-n(\overline{y}-\mu)^2 \leq 0$, niin μ:n suhteen l saa suurimman arvonsa täsmälleen kohdassa $\mu = \hat{\mu} = \overline{y}$. Sijoitetaan tämä l:n lausekkeeseen ja etsitään yhden muuttujan funktion

$$l_{\mathrm{P}}(\sigma^2) = l(\hat{\mu}, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{(n-1)s^2}{2\sigma^2}$$

globaali maksimikohta derivaattatarkastelun avulla. Saadaan

$$l'_{\mathrm{P}}(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{(n-1)s^2}{2\sigma^4},$$

jonka ainoa nollakohta on $\sigma^2 = \hat{\sigma}^2 = (n-1)s^2/n$. Sijoittamalla tämä toisen derivaatan

$$l''_{\mathrm{P}}(\sigma^2) = \frac{n}{2\sigma^4} - \frac{(n-1)s^2}{\sigma^6}$$

lausekkeeseen nähdään sievennyksen jälkeen $l''_{\mathrm{P}}(\hat{\sigma}^2) = -n/2\hat{\sigma}^4 < 0$, joten $\hat{\sigma}^2$ todella on globaali maksimikohta.

Kaiken kaikkiaan voidaan todeta, että parametriparin (μ, σ^2) su-estimaatit ovat

$$\hat{\mu} = \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Huomionarvoista on, että varianssin su-estimaatti poikkeaa tavallisesti käytetystä otosvarianssista

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Jälkimmäinen on suositeltavampi siksi, että sitä vastaava estimaattori S^2 on *harhaton* eli $E(S^2) = \sigma^2$. Huomaa kuitenkin, että jos n ei ole aivan pieni, niin $(n-1)/n \approx 1$ ja ero näiden kahden välillä on käytännössä merkityksetön.

2.2.7 Esimerkki: yhden selittäjän lineaarinen regressio. Yhden selittäjän lineaarisessa regressiomallissa (ks. 1.2.4) oletettiin, että $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$ ja

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2), \qquad i = 1, \ldots, n.$$

Tämän mallin log-uskottavuusfunktio on

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

Parametrivektorin $(\alpha, \beta, \sigma^2)$ su-estimaatti on (ks. teht. 2.9)

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}, \qquad \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2,$$

jossa $\overline{x}$ on lukujen x_i keskiarvo ja $\overline{y}$ lukujen y_i keskiarvo.

Parametrien α ja β osalta su-estimaatit ovat samat kuin tilastotieteen johdantokurssilla esitetyt pienimmän neliösumman estimaatit. Varianssin estimointiin käytetään yleensä $\hat{\sigma}^2$:n sijasta harhatonta estimaattia

$$s^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2.$$

Lineaaristen regressiomallien estimoinnista puhutaan perusteellisemmin ja yleisemmin lineaaristen mallien kurssilla.

2.2.8 Su-estimaatin olemassaolo ja yksikäsitteisyys. Edellä tarkastelluissa esimerkeissä on aina löytynyt yksikäsitteinen parametriavaruuden piste $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$, jossa uskottavuusfunktio saa suurimman arvonsa, paitsi ehkä joidenkin yksittäisten aineistojen $\boldsymbol{y}$ tapauksessa (vrt. 2.2.5). Joissakin malleissa näin ei ole.

Tarkastellaan esimerkkinä riippumatonta otosta välin $(0, \theta)$ tasajakaumasta, jossa $\theta > 0$ on estimoitava parametri. Siis $Y_1, \ldots, Y_n \sim Tas(0, \theta)$ ⫫. Käytännössä tämän voisi ajatella mallittavan esimerkiksi seuraavaa tilannetta: Metrojuna kulkee säännöllisesti tasaisin väliajoin. Henkilö menee n eri kertaa satunnaiseen aikaan metroasemalle, mittaa joka kerta ajan, jonka hän joutuu junaa odottamaan, ja näiden mittausten perusteella pyrkii tekemään arvion junien väliajasta θ.

Tasajakauman tiheysfunktioksi on todennäköisyyslaskennassa yleensä valittu

$$f(y; \theta) = \begin{cases} 1/\theta, & \text{kun } 0 < y < \theta, \\ 0, & \text{muulloin.} \end{cases} \tag{2.4}$$

Satunnaisvektorin $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ ytf on siten

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \prod_{i=1}^n f(y_i; \theta) = \begin{cases} 1/\theta^n, & \text{kun } 0 < y_i < \theta \text{ jokaisella } i, \\ 0, & \text{muulloin.} \end{cases}$$

Havaitaan aineisto $\boldsymbol{y} = (y_1, \ldots, y_n)$. Jos merkitään $y_{(n)} = \max(y_1, \ldots, y_n)$, niin aineistoon liittyvän uskottavuusfunktion lauseke voidaan kirjoittaa muodossa

$$L(\theta; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \begin{cases} 0, & \text{kun } 0 < \theta \le y_{(n)}, \\ 1/\theta^n, & \text{kun } \theta > y_{(n)}. \end{cases}$$

Huomaa, että funktio $\theta \mapsto 1/\theta^n$ on aidosti vähenevä. Siten uskottavuusfunktio saa arvoja, jotka ovat mielivaltaisen lähellä lukua $1/y_{(n)}^n$, mutta se ei kuitenkaan saavuta kyseistä arvoa. Näin ollen uskottavuusfunktiolla ei ole suurinta arvoa, eli su-estimaattia ei ole olemassa.

Ongelmasta voidaan tässä esimerkissä helposti vapautua. Valitaan nimittäin jakauman $Tas(0, \theta)$ tiheysfunktioksi (2.4):n sijasta

$$f(y; \theta) = \begin{cases} 1/\theta, & \text{kun } 0 \le y \le \theta, \\ 0, & \text{muulloin.} \end{cases} \tag{2.5}$$

(Funktion arvojen muuttaminen äärellisessä pistejoukossa ei vaikuta funktiosta laskettaviin integraaleihin, joten (2.5) määrittelee täysin saman jakauman kuin (2.4).) Nyt uskottavuusfunktio saa muodon

$$L(\theta; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \begin{cases} 0, & \text{kun } 0 < \theta < y_{(n)}, \\ 1/\theta^n, & \text{kun } \theta \geq y_{(n)}, \end{cases}$$

josta nähdään, että parametrin θ su-estimaatti on $\hat{\theta}(\boldsymbol{y}) = y_{(n)}$. Tämä esimerkki kertoo lähinnä siitä, että jatkuvaan jakaumaan perustuvassa mallissa tiheysfunktioon liittyvä monikäsitteisyys aiheuttaa periaatteessa epämääräisyyttä su-estimaatin määrittelyyn. Käytännössä tästä ei yleensä aiheudu ongelmia, kun käytetään jotain tiettyä vakiintunutta tiheysfunktion versiota.

Joissakin malleissa voi käydä niin, että uskottavuusfunktio saa suurimman arvonsa useassa pisteessä, jopa äärettömän monessa. Tällöin su-estimaatti ei olekaan aineiston perusteella yksikäsitteisesti määrätty. Tällainen esimerkki saadaan poimimalla otos jakaumasta $Tas(\theta, \theta + 1)$, jossa θ on tuntematon reaalinen parametri (ks. teht. 2.6).

2.3 Su-estimaatin invarianssiominaisuus

2.3.1 Uudelleenparametrointi. Toisinaan on tarpeellista parametroida tarkasteltava tilastollinen malli $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ uudelleen jollakin toisella parametrilla $\boldsymbol{\phi}$, joka voi olla esimerkiksi tulkinnallisesti luontevampi kuin parametri $\boldsymbol{\theta}$. Tämä tarkoittaa seuraavaa: Jos mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ parametriavaruus on Ω, niin spesifioidaan jokin toinen parametriavaruus Ω^* ja bijektiivinen eli kääntäen yksikäsitteinen funktio $\boldsymbol{g}\colon \Omega \to \Omega^*$, jonka välityksellä Ω:n pisteet $\boldsymbol{\theta}$ ja Ω^*:n pisteet $\boldsymbol{\phi}$ vastaavat toisiaan, ts. $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ ja $\boldsymbol{\theta} = \boldsymbol{g}^{-1}(\boldsymbol{\phi})$. Mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ sijasta voidaan nyt tarkastella mallia

$$f^*_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\phi}) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{g}^{-1}(\boldsymbol{\phi})).$$

Merkitään parametrin $\boldsymbol{\theta}$ uskottavuusfunktiota symbolilla L ja parametrin $\boldsymbol{\phi}$ uskottavuusfunktiota symbolilla L^*. Äskeisen muunnoskaavan mukaan pätee

$$L^*(\boldsymbol{\phi}) = L(\boldsymbol{g}^{-1}(\boldsymbol{\phi})), \qquad \boldsymbol{\phi} \in \Omega^*.$$

Koska $L(\boldsymbol{\theta})$ saa suurimman arvonsa pisteessä $\hat{\boldsymbol{\theta}}$, nähdään tästä, että $L^*(\boldsymbol{\phi})$ saa suurimman arvonsa pisteessä $\boldsymbol{g}(\hat{\boldsymbol{\theta}})$. Näin ollen parametrin $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ suurimman uskottavuuden estimaatti on $\boldsymbol{g}(\hat{\boldsymbol{\theta}})$.

Saatu tulos on varsin tärkeä ja hyödyllinen *su-estimaatin invarianssiominaisuus*. Muotoillaan se vielä tiivistetysti:

Lause. *Olkoon mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ parametriavaruus Ω, ja olkoon $\boldsymbol{g}\colon \Omega \to \Omega^*$ bijektiivinen funktio. Jos parametrin $\boldsymbol{\theta}$ su-estimaatti on $\hat{\boldsymbol{\theta}}$, niin parametrin $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ su-estimaatti on $\hat{\boldsymbol{\phi}} = \boldsymbol{g}(\hat{\boldsymbol{\theta}})$.*

2.3.2 Esimerkki: eksponenttimalli. Esimerkissä 1.2.2 johdettiin kestoikiä kuvaava malli $Y_1, \ldots, Y_n \sim Exp(\lambda)$ ⫫ ja sen uudelleenparametrointi $\mu = 1/\lambda$. Koska $\hat{\lambda} = 1/\overline{y}$ (ks. teht. 2.3), niin $\hat{\mu} = \overline{y}$ (otoskeskiarvo).

2.3.3 Esimerkki: toistopyydystysotanta. Halutaan estimoida järvessä elävien kalojen kokonaislukumäärä ν. Tätä varten pyydystetään m kalaa, merkitään ne jotenkin ja palautetaan järveen. Jonkin ajan kuluttua, kun merkittyjen kalojen voidaan olettaa sekoittuneen merkitsemättömien kanssa, pyydystetään n kalaa ja lasketaan niistä merkittyjen lukumäärä k. Tätä menetelmää kutsutaan toistopyydystysotannaksi (engl. *capture-recapture sampling*).

Jos oletetaan, että otoskoko n on pieni verrattuna kalojen kokonaismäärään ν, voidaan katsoa, että k on "onnistumisten" lukumäärä n-kertaisessa riippumattomassa toistokokeessa, jossa onnistumistodennäköisyys on m/ν eli merkittyjen kalojen suhteellinen osuus järvessä. Vastaava tilastollinen malli on siten

$$K \sim Bin(n, \theta), \qquad \text{jossa } \theta = m/\nu \text{ eli } \nu = m/\theta.$$

Esimerkistä 2.2.5 tiedetään, että $\hat{\theta} = k/n$, joten kokonaismäärän ν su-estimaatti on $\hat{\nu} = m/\hat{\theta} = mn/k$.

2.3.4 Parametrin funktion su-estimointi. Monesti käytännön sovelluksissa halutaan esittää arvioita jostakin parametrin $\boldsymbol{\theta}$ funktiosta $\boldsymbol{g}(\boldsymbol{\theta})$ siinäkin tapauksessa, että $\boldsymbol{g}$ ei ole bijektio eikä näin ollen ole kyse mallin uudelleenparametroinnista. Edellä todetun invarianssiominaisuuden valossa tuntuu luontevalta sopia yleisesti, että funktion $\boldsymbol{g}(\boldsymbol{\theta})$ su-estimaatti on $\boldsymbol{g}(\hat{\boldsymbol{\theta}})$, jossa $\hat{\boldsymbol{\theta}}$ on $\boldsymbol{\theta}$:n su-estimaatti.

2.3.5 Esimerkki: normaalimalli. Tarkastellaan mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jolle $(\hat{\mu}, \hat{\sigma}^2)$ laskettiin esimerkissä 2.2.6. Toisen (origo)momentin $E(Y_i^2) = \mu^2 + \sigma^2$ su-estimaatti on äskeisen sopimuksen mukaan $\hat{\mu}^2 + \hat{\sigma}^2 = \sum_{i=1}^n y_i^2/n$. Jos taas estimoitavana on todennäköisyys $p = P\{Y_i > 0\}$, voidaan kirjoittaa standardinormaalijakauman kertymäfunktiota Φ käyttäen $p = \Phi(\mu/\sigma)$ ja siten su-estimaatiksi saadaan $\hat{p} = \Phi(\hat{\mu}/\hat{\sigma})$, jossa $\hat{\sigma}$ on tietysti $\hat{\sigma}^2$:n neliöjuuri.

2.4 Informaation käsite, tapaus $d = 1$

2.4.1 Motivoiva esimerkki. Palataan jälleen asetelmaan, jossa tutkittiin rikkinäisten lamppujen osuutta tehtaan tuotannosta (ks. 1.2.1). Mallina on tällöin $Y_1, \ldots, Y_n \sim B(\theta)$ $\perp\!\!\!\perp$, ja aineistoa $\boldsymbol{y} = (y_1, \ldots, y_n)$ vastaava log-uskottavuusfunktio on

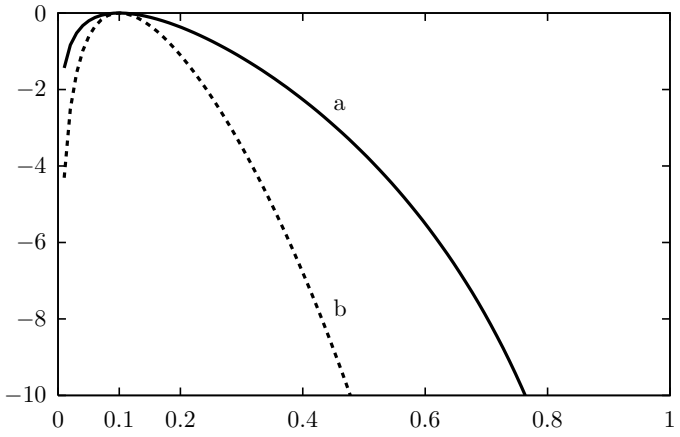
$$l(\theta) = k \log \theta + (n - k) \log(1 - \theta),$$

jossa $k = y_1 + \cdots + y_n$. Parametrin θ su-estimaatti on $\hat{\theta} = k/n$ (ks. 2.2.5).

Tarkastellaan otoskoon n vaikutusta log-uskottavuusfunktion muotoon. Kuvaan 2.2 on piirretty vertailun helpottamiseksi funktion

$$l_0(\theta) = l(\theta) - l(\hat{\theta}) \tag{2.6}$$

kuvaajat, kun a) $n = 10$, $k = 1$ ja b) $n = 40$, $k = 4$. Kummassakin tapauksessa on siis $\hat{\theta} = 0.1$. Suurempi otoskoko näkyy funktion kuvaajan "huipukkuutena": aineiston valossa uskottavat parametriarvot sijaitsevat b-tapauksessa selvästi kapeammalla välillä kuin a-tapauksessa. Jossakin mielessä b-aineisto näyttäisi siis sisältävän enemmän informaatiota parametria θ koskevien päätelmien tekoon.



Kuva 2.2. Log-uskottavuusfunktion (2.6) kuvaajat, kun a) $n = 10$, $k = 1$ ja b) $n = 40$, $k = 4$.

Maksimipisteen $\hat{\theta}$ läheisyydessä log-uskottavuusfunktion kulkua voi kuvata sen toisen derivaatan $l''(\hat{\theta})$ avulla (muista, että $l'(\hat{\theta}) = 0$, koska kyseessä on maksimikohta). Derivoimalla saadaan

$$l''(\theta) = -\frac{k}{\theta^2} - \frac{n-k}{(1-\theta)^2}$$

ja sijoittamalla tähän $\theta = \hat{\theta} = k/n$ päädytään lausekkeeseen

$$l''(\hat{\theta}) = -\frac{n^3}{k(n-k)} = -\frac{n}{\hat{\theta}(1-\hat{\theta})}.$$

Jos suhde $\hat{\theta} = k/n$ pysyy vakiona, on toinen derivaatta $l''(\hat{\theta})$ siis suoraan verrannollinen aineiston kokoon. Mitä suurempi $l''(\hat{\theta})$ on itseisarvoltaan, sitä jyrkemmin l:n kuvaaja kaareutuu pisteessä $\hat{\theta}$ ja sitä tarkempia päätelmiä parametrista θ voidaan uskottavuuspohjaisesti tehdä.

2.4.2 Havaittu informaatio. Tarkastellaan parametrista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$, jonka parametri $\theta \in \Omega$ on reaalinen (eli $d = 1$). Oletetaan lisäksi, että log-uskottavuusfunktio $l(\theta;\boldsymbol{y}) = c(\boldsymbol{y}) + \log f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ on ainakin kahdesti derivoituva. Tällöin määritellään, että aineistosta $\boldsymbol{y}$ *havaittu informaatio* on

$$j(\theta;\boldsymbol{y}) = -l''(\theta;\boldsymbol{y}), \qquad \theta \in \Omega.$$

Pääasiassa kiinnitetään huomiota tämän arvoon vain suurimman uskottavuuden pisteessä $\hat{\theta}$.

Havaitun informaation merkitystä voidaan havainnollistaa differentiaalilaskennassa opitun Taylorin kaavan avulla. Muodostamalla log-uskottavuusfunktion toisen asteen Taylorin polynomi pisteessä $\hat{\theta}$ saadaan nimittäin approksimaatio

$$l(\theta;\boldsymbol{y}) \approx l(\hat{\theta};\boldsymbol{y}) + l'(\hat{\theta};\boldsymbol{y})(\theta-\hat{\theta}) + \tfrac{1}{2}l''(\hat{\theta};\boldsymbol{y})(\theta-\hat{\theta})^2,$$

joka on voimassa $\hat{\theta}$:n ympäristössä. Kun otetaan huomioon havaitun informaation määritelmä ja se, että $l'(\hat{\theta};\boldsymbol{y}) = 0$, tämä voidaan kirjoittaa muotoon

$$l(\theta;\boldsymbol{y}) - l(\hat{\theta};\boldsymbol{y}) \approx -\tfrac{1}{2}j(\hat{\theta};\boldsymbol{y})(\theta-\hat{\theta})^2. \tag{2.7}$$

Vasemmalla esiintyvää funktiota $l_0(\theta; \boldsymbol{y}) = l(\theta; \boldsymbol{y}) - l(\hat{\theta}; \boldsymbol{y})$ kutsutaan *normitetuksi log-uskottavuusfunktioksi*: kyseessä on se log-uskottavuusfunktion versio, jonka suurin arvo eli arvo pisteessä $\hat{\theta}$ on nolla. Approksimaatio (2.7) kertoo, että $\hat{\theta}$:n ympäristössä tämän funktion kuvaaja on likimain alasaukeava paraabeli, jonka huippu on kohdassa $\hat{\theta}$ ja jonka leveyden määrää havaittu informaatio $j(\hat{\theta}; \boldsymbol{y})$. Mitä suurempi havaittu informaatio on, sitä kapeampi tämä paraabeli on ja sitä epäuskottavampia ovat vähänkin $\hat{\theta}$:sta poikkeavat parametriarvot θ.

2.4.3 Esimerkki: normaalimalli. Oletetaan, että $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jossa $\sigma_0^2 > 0$ on tunnettu luku ja μ on varsinainen parametri. Tämän mallin log-uskottavuusfunktio voidaan kirjoittaa muodossa (ks. 2.1.7)

$$l(\mu; \boldsymbol{y}) = -\frac{n}{2\sigma_0^2}(\mu - \overline{y})^2.$$

Derivoimalla kahdesti nähdään, että

$$j(\mu; \boldsymbol{y}) = -l''(\mu; \boldsymbol{y}) = \frac{n}{\sigma_0^2}.$$

Havaittu informaatio on siis suoraan verrannollinen otoskokoon n aivan kuten esimerkissä 2.4.1. Varianssilla σ_0^2 on sen sijaan käänteinen vaikutus: jos σ_0^2 on suuri, havaintoihin liittyy voimakas satunnaisvaihtelu ja ne sisältävät vähän informaatiota parametrista μ.

Huomaa, että normaalimallin tapauksessa log-uskottavuusfunktio on jo itsessään toisen asteen polynomi ja siten kaavassa (2.7) päteekin yhtäsuuruus. Approksimaatiota (2.7) voidaankin yleisesti kutsua log-uskottavuusfunktion ”normaaliapproksimaatioksi”, ja sillä on käyttöä myöhemmin ns. asymptoottista päättelyn teoriaa kehitettäessä.

2.4.4 Fisherin informaatio. Havaittu informaatio $j(\theta; \boldsymbol{y})$ riippuu nimensä mukaisesti havaitusta aineistosta $\boldsymbol{y}$. Osoittautuu tarpeelliseksi tutkia myös vastaavan satunnaismuuttujan $j(\theta; \boldsymbol{Y})$ odotusarvoa eli sitä, millaisia arvoja havaittu informaatio keskimäärin saa toistetussa aineistonkeruussa. Tätä kutsutaan mallin *odotetuksi informaatioksi* eli *Fisherin informaatioksi* ja se siis määritellään kaavalla

$$i(\theta) = E[j(\theta; \boldsymbol{Y})] = E[-l''(\theta; \boldsymbol{Y})], \qquad \theta \in \Omega.$$

Huomaa, että tässä odotusarvo on laskettava nimenomaan parametriarvoa θ käyttäen. Tarvittaessa tämän voi osoittaa alaindeksillä tyyliin $E = E_\theta$.

Myöhemmin osoittautuu, että Fisherin informaatiolla on varsin syvällinen rooli päättelyn teoriassa. Käytännön sovellusten kannalta on hyvä huomata, että $i(\theta)$ on pelkän mallin määräämä suure ja siis laskettavissa jo ennen kuin aineistonkeruuta on edes suoritettu. Tätä voi käyttää koesuunnittelun apuna kuten esimerkissä 2.4.6 alla ja tehtävässä 2.17.

2.4.5 Esimerkki: toistokoemalli. Tarkastellaan taas mallia $Y_1, \ldots, Y_n \sim B(\theta)$ $\perp\!\!\!\perp$. Esimerkin 2.4.1 laskuista nähdään, että havaittu informaatio on

$$j(\theta; \boldsymbol{y}) = \frac{k}{\theta^2} + \frac{n-k}{(1-\theta)^2}$$

ja erityisesti pisteessä $\hat{\theta} = k/n$

$$j(\hat{\theta}; \boldsymbol{y}) = \frac{n^3}{k(n-k)} = \frac{n}{\hat{\theta}(1-\hat{\theta})},$$

jossa $k = y_1 + \cdots + y_n$. Koska $K = Y_1 + \cdots + Y_n \sim Bin(n, \theta)$, niin $E(K) = n\theta$ ja Fisherin informaatioksi saadaan

$$i(\theta) = E[j(\theta; \boldsymbol{Y})] = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.$$

2.4.6 Esimerkki: mallien informaatioiden vertailu. Eräästä geenistä on olemassa alleelit r ja R, jolloin populaatio jakautuu genotyyppeihin rr, rR ja RR. Näiden suhteelliset osuudet ovat Hardyn–Weinbergin lain mukaan muotoa θ^2, $2\theta(1-\theta)$ ja $(1-\theta)^2$, jossa $0 < \theta < 1$ halutaan estimoida. Käytettävissä on kaksi vaihtoehtoista koeasetelmaa:

a) Poimitaan n yksilön otos ja testataan kunkin genotyyppi. Aineistoksi saadaan tyyppien rr, rR ja RR lukumäärät y_1, y_2 ja y_3, jossa $y_1 + y_2 + y_3 = n$.
b) Käytetään halvempaa testiä, joka ei erota tyyppejä rR ja RR toisistaan vaan ainoastaan erottaa tyypin rr näistä kahdesta. Otoskoko on kolminkertainen eli $3n$. Aineisto on luku k, joka kertoo tyypin rr lukumäärän; tyyppejä rR ja RR on siis otoksessa yhteensä $3n - k$.

Kumpi asetelma (malli) tuottaa enemmän informaatiota parametrista θ? Tutkitaan asiaa vertaamalla mallien Fisherin informaatioita $i_{\mathrm{a}}(\theta)$ ja $i_{\mathrm{b}}(\theta)$.

Mallissa a aineistoa vastaava satunnaisvektori $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ noudattaa ns. multinomijakaumaa, jonka yptf on muotoa

$$\begin{aligned} P\{Y_1 = y_1,\, Y_2 = y_2,\, Y_3 = y_3\} &= c(\boldsymbol{y})\,(\theta^2)^{y_1}\,[2\theta(1-\theta)]^{y_2}\,[(1-\theta)^2]^{y_3} \\ &= 2^{y_2} c(\boldsymbol{y})\,\theta^{2y_1+y_2}(1-\theta)^{2y_3+y_2}, \end{aligned}$$

kun $y_1 + y_3 + y_3 = n$. Tässä $c(\boldsymbol{y})$ on multinomikerroin, joka kertoo, kuinka monella eri tavalla n alkion joukko voidaan jakaa kolmeen osaan siten, että osissa on y_1, y_2 ja y_3 alkiota. Sen arvolla ei ole jatkossa merkitystä, koska se voidaan luvun 2^{y_2} tavoin jättää pois uskottavuus- ja log-uskottavuusfunktion lausekkeista. Mallin a log-uskottavuusfunktioksi saadaan näin

$$l_{\mathrm{a}}(\theta; \boldsymbol{y}) = (2y_1 + y_2)\log\theta + (2y_3 + y_2)\log(1-\theta).$$

Tämän toinen derivaatta on

$$l''_{\mathrm{a}}(\theta; \boldsymbol{y}) = -\frac{2y_1 + y_2}{\theta^2} - \frac{2y_3 + y_2}{(1-\theta)^2}.$$

Koska $E(Y_1) = \theta^2 n$, $E(Y_2) = 2\theta(1-\theta)n$ ja $E(Y_3) = (1-\theta)^2 n$, malliin liittyvä Fisherin informaatio on

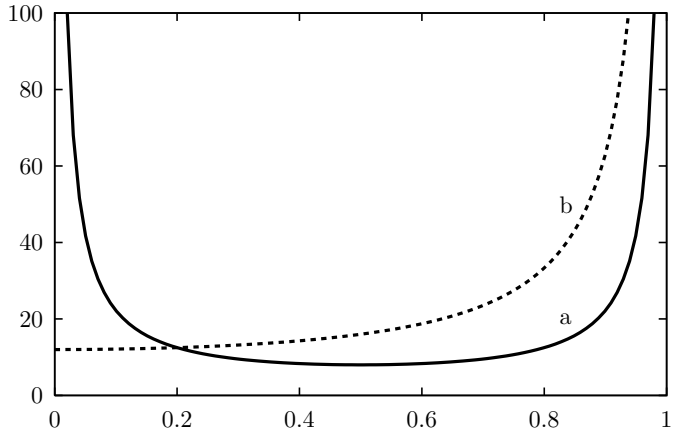
$$i_{\mathrm{a}}(\theta) = E[-l''_{\mathrm{a}}(\theta; \boldsymbol{Y})] = \frac{2\theta^2 n + 2\theta(1-\theta)n}{\theta^2} + \frac{2(1-\theta)^2 n + 2\theta(1-\theta)n}{(1-\theta)^2} = \frac{2n}{\theta(1-\theta)}.$$

Mallissa b havaitaan satunnaismuuttuja $K \sim Bin(3n, \theta^2)$, jonka pistetodennäköisyysfunktio on

$$P\{K = k\} = \binom{3n}{k} (\theta^2)^k (1-\theta^2)^{3n-k}.$$

Siten log-uskottavuusfunktio voidaan kirjoittaa muotoon

$$\begin{aligned} l_{\mathrm{b}}(\theta; k) &= k\log\theta^2 + (3n-k)\log(1-\theta^2) \\ &= 2k\log\theta + (3n-k)\log(1+\theta) + (3n-k)\log(1-\theta). \end{aligned}$$



Kuva 2.3. Fisherin informaatioiden kuvaajat esimerkin 2.4.6 malleissa a ja b.

Tällöin

$$l''_{\mathrm{b}}(\theta; k) = -\frac{2k}{\theta^2} - \frac{3n-k}{(1+\theta)^2} - \frac{3n-k}{(1-\theta)^2},$$

ja koska $E(K) = 3n\theta^2$, niin mallin Fisherin informaatioksi saadaan

$$i_{\mathrm{b}}(\theta) = E[-l''_{\mathrm{b}}(\theta; K)] = \frac{6n\theta^2}{\theta^2} + \frac{3n-3n\theta^2}{(1+\theta)^2} + \frac{3n-3n\theta^2}{(1-\theta)^2} = \frac{12n}{(1+\theta)(1-\theta)}.$$

Funktioiden i_{a} ja i_{b} kuvaajat on piirretty kuvaan 2.3. Niiden perusteella voitaneen sanoa seuraavaa: Jos on syytä epäillä, että $\theta < 0.2$ (eli rr-tyypin osuus on alle 4 %), kannattaa käyttää asetelmaa a. Muussa tapauksessa b on jonkin verran parempi.

2.5 Pistemäärä ja säännölliset mallit, tapaus $d = 1$

Päättelyn teoriassa log-uskottavuusfunktion ensimmäisellä derivaatalla on syvällisempikin merkitys, kuin mitä pelkkä uskottavuusyhtälö $l'(\hat{\theta}; \boldsymbol{y}) = 0$ antaa ymmärtää. Erityisesti sillä on mielenkiintoinen yhteys Fisherin informaation käsitteeseen silloin, kun tarkasteltava malli täyttää tietyt säännöllisyysehdot.

2.5.1 Pistemääräfunktio. Tarkastellaan mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$, jonka parametriavaruus on $\Omega \subset \mathbb{R}$; tyypillisesti Ω on avoin väli. Oletetaan, että uskottavuusfunktio $\theta \mapsto f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ on derivoituva Ω:ssa kaikilla $\boldsymbol{y}$. Log-uskottavuusfunktion ensimmäistä derivaattaa

$$l'(\theta) = l'(\theta; \boldsymbol{y}) = \frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)}, \qquad \theta \in \Omega,$$

kutsutaan tällöin aineistoon $\boldsymbol{y}$ liittyväksi *pistemääräksi* tai *pistemääräfunktioksi*.

2.5.2 Säännölliset mallit. Jatkuva malli (eli ytf) $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ on *säännöllinen*, jos seuraavat ehdot ovat voimassa:

a) jakauman *alusta* eli joukko $A = \{\boldsymbol{y} : f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) > 0\}$ ei riipu θ:sta,

b) funktio $\theta \mapsto f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ on kaksi kertaa jatkuvasti derivoituva jokaisella $\boldsymbol{y}$,

c) aina kun $T = t(\boldsymbol{Y})$ on tunnusluku, jolle $E_\theta(T)$ on olemassa kaikilla θ, pätee

$$\frac{d}{d\theta} \int_A t(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\, d\boldsymbol{y} = \int_A t(\boldsymbol{y}) \frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\, d\boldsymbol{y},$$

d) ja lisäksi

$$\frac{d^2}{d\theta^2}\int_A f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} = \int_A \frac{\partial^2}{\partial\theta^2} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y}.$$

Diskreetin mallin eli yptf:n $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ säännöllisyys määritellään samalla tavalla, kunhan c- ja d-ehdossa integraalit korvataan summilla.

Emme tällä kurssilla syvenny kovinkaan tarkasti analysoimaan yo. ehtoja, vaan jatkossa yleensä oletetaan, että tarkasteltavat mallit ovat riittävän säännöllisiä, jotta tarvittavat operaatiot voidaan niille suorittaa. On syytä huomata, että ehto a sulkee muutamia yksinkertaisiakin malleja säännöllisten mallien luokan ulkopuolelle: tällainen on esimerkiksi otos jakaumasta $Tas(0,\theta)$ (ks. 2.2.8). Tekniset ehdot c ja d puolestaan kertovat, että asianomainen ytf (tai yptf) on niin ”hyvin” integroituva (tai summautuva), että derivoinnin ja integroinnin (tai summauksen) järjestys voidaan vaihtaa. Koska tiheysfunktion määritelmän perusteella $\int_A f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} = 1$ kaikilla θ, nähdään, että c:n erikoistapaus $t(\boldsymbol{y}) = 1$ ja d merkitsevät, että

$$\int_A \frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} = 0 \qquad \text{ja} \qquad \int_A \frac{\partial^2}{\partial\theta^2} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} = 0. \tag{2.8}$$

Näitä yhtälöitä käytetään hetken päästä.

Käytännössä tärkeintä on tietää, että lähes kaikki tavallisesti käytettävät mallit ovat säännöllisiä. Yleisesti voidaan sanoa, että tällaisia ovat kaikki ns. *eksponenttiperheen* jakaumiin pohjautuvat mallit. Eksponenttiperhe pitää sisällään mm. Bernoulli-, binomi-, Poisson-, normaali-, gamma- ja eksponenttijakaumat (ks. teht. 2.20).

2.5.3 Pistemäärän odotusarvo ja varianssi säännölliselle mallille. Seuraavassa apulauseessa tarkastellaan säännöllisen mallin pistemääräfunktiota satunnaismuuttujana ja todetaan sen yhteys Fisherin informaation käsitteeseen:

Apulause. *Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ säännöllinen malli ja $l(\theta;\boldsymbol{y}) = \log f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ sen log-uskottavuusfunktio. Tällöin*

a) $E[l'(\theta;\boldsymbol{Y})] = 0$

b) $E[l'(\theta;\boldsymbol{Y})^2] = \operatorname{var}[l'(\theta;\boldsymbol{Y})] = i(\theta)$.

Todistus. a) Odotusarvon määritelmän mukaan

$$E[l'(\theta;\boldsymbol{Y})] = \int_A l'(\theta;\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} = \int_A \frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y}.$$

Väite seuraa nyt yhtälöstä (2.8).

b) Osamäärän derivoimissääntöä käyttämällä saadaan

$$\begin{aligned} l''(\theta;\boldsymbol{y}) &= \frac{\partial}{\partial\theta}\frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)}{f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)} \\ &= \frac{\left[\frac{\partial^2}{\partial\theta^2} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\right] f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) - \left[\frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\right]^2}{f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)^2} \\ &= \frac{\frac{\partial^2}{\partial\theta^2} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)}{f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)} - l'(\theta;\boldsymbol{y})^2. \end{aligned}$$

Korvaamalla $\boldsymbol{y}$ vastaavalla satunnaismuuttujalla $\boldsymbol{Y}$ ja ottamalla odotusarvot saadaan tästä

$$i(\theta) = E[-l''(\theta;\boldsymbol{Y})] = -\int_A \frac{\partial^2}{\partial\theta^2} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y} + E[l'(\theta;\boldsymbol{Y})^2].$$

Koska yhtälön (2.8) mukaan tässä esiintyvä integraali on nolla, niin väite seuraa. □

Monissa päättelyn kirjoissa kaava $i(\theta) = \mathrm{var}[l'(\theta; \boldsymbol{Y})]$ otetaan Fisherin informaation määritelmäksi, mutta käytännön tehtävissä määritelmä $i(\theta) = E[-l''(\theta; \boldsymbol{Y})]$ yleensä johtaa helpompiin laskuihin. Tulos b on silti teoreettisesti tärkeä; sitä tarvitaan pykälässä 3.4 johdettaessa estimaattorien variansseja koskeva ns. informaatioepäyhtälö.

2.6 Informaatio ja pistemäärä, tapaus $d > 1$

Tässä pykälässä yleistetään pykälien 2.4 ja 2.5 käsitteet ja tulokset mailleihin, joiden parametri on vektori. Periaatteellisella tasolla tilanne ei ole yhtään sen vaikeampi kuin reaalisen parametrin tapaus $d = 1$; ainoastaan merkinnät ovat raskaampia.

2.6.1 Informaatiokäsitteet. Tarkastellaan mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jonka parametri on vektori $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Oletetaan, että log-uskottavuusfunktiolla

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \boldsymbol{y}) = c(\boldsymbol{y}) + \log f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$$

on jatkuvat toisen kertaluvun osittaisderivaatat parametriavaruudessa $\Omega \subset \mathbb{R}^d$. Aineistosta $\boldsymbol{y}$ *havaittu informaatio* määritellään $d \times d$ -matriisina $\boldsymbol{j}(\boldsymbol{\theta}; \boldsymbol{y})$, jonka alkiot $j_{a,b}(\boldsymbol{\theta}; \boldsymbol{y})$ ovat funktion $-l$ toisen kertaluvun osittaisderivaatat. Kyseessä on siis funktion l Hessen matriisi kerrottuna luvulla -1 eli

$$\boldsymbol{j}(\boldsymbol{\theta}; \boldsymbol{y}) = \begin{bmatrix} j_{1,1}(\boldsymbol{\theta}; \boldsymbol{y}) & \cdots & j_{1,d}(\boldsymbol{\theta}; \boldsymbol{y}) \\ \vdots & & \vdots \\ j_{d,1}(\boldsymbol{\theta}; \boldsymbol{y}) & \cdots & j_{d,d}(\boldsymbol{\theta}; \boldsymbol{y}) \end{bmatrix},$$

jossa

$$j_{a,b}(\boldsymbol{\theta}; \boldsymbol{y}) = -\frac{\partial^2}{\partial\theta_a \partial\theta_b} l(\boldsymbol{\theta}; \boldsymbol{y}),$$

kun $a, b = 1, \ldots, d$.

Fisherin informaatio määritellään luonnollisesti korvaamalla yllä havaittu aineisto $\boldsymbol{y}$ sitä vastaavalla satunnaismuuttujalla $\boldsymbol{Y}$ ja ottamalla odotusarvot alkioittain. Siis

$$\boldsymbol{i}(\boldsymbol{\theta}) = \begin{bmatrix} i_{1,1}(\boldsymbol{\theta}) & \cdots & i_{1,d}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ i_{d,1}(\boldsymbol{\theta}) & \cdots & i_{d,d}(\boldsymbol{\theta}) \end{bmatrix},$$

jossa

$$i_{a,b}(\boldsymbol{\theta}) = E[j_{a,b}(\boldsymbol{\theta}; \boldsymbol{Y})],$$

kun $a, b = 1, \ldots, d$. Huomaa, että $\boldsymbol{j}(\boldsymbol{\theta}; \boldsymbol{y})$ ja $\boldsymbol{i}(\boldsymbol{\theta})$ ovat symmetrisiä matriiseja osittaisderivoinnin järjestyksen vaihdannaisuuden vuoksi.

2.6.2 Parametrien ortogonaalisuus. Oletetaan, että mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ parametrivektori $\boldsymbol{\theta}$ voidaan osittaa kahteen osaan – olkoot ne vaikkapa $\boldsymbol{\psi} = (\theta_1, \ldots, \theta_q)$ ja $\boldsymbol{\lambda} = (\theta_{q+1}, \ldots, \theta_d)$ – siten, että

$$i_{a,b}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2}{\partial\theta_a \partial\theta_b} l(\boldsymbol{\theta}; \boldsymbol{Y})\right] = 0, \qquad \boldsymbol{\theta} \in \Omega,$$

aina kun $a = 1, \ldots, q$ ja $b = q+1, \ldots, d$. Toisin sanoen Fisherin informaatiomatriisi oletetaan lohkodiagonaaliseksi. Tällöin osia $\boldsymbol{\psi}$ ja $\boldsymbol{\lambda}$ sanotaan *ortogonaalisiksi*. Ääritapauksessa $\boldsymbol{i}(\boldsymbol{\theta})$ voi olla diagonaalimatriisi, jolloin jokainen parametrin komponentti θ_j on ortogonaalinen kaikkia muita kohtaan.

Parametrien ortogonaalisuus on yleensä erittäin hyödyllinen mallin ominaisuus. Kuten myöhemmin nähdään, se merkitsee tietyssä mielessä sitä, että yhdestä parametrin osasta tehtävät päätelmät eivät vaikuta toisesta osasta tehtäviin päätelmiin ja kääntäen.

2.6.3 Esimerkki: normaalimalli. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa parametri on kaksiulotteinen (μ, σ^2). Tämän mallin log-uskottavuusfunktioksi on aikaisemmin (ks. 2.1.7) todettu

$$l(\mu, \sigma^2; \boldsymbol{y}) = -\frac{n}{2}\log\sigma^2 - \frac{(n-1)s^2 + n(\mu - \overline{y})^2}{2\sigma^2}.$$

Suoraviivaisten derivointien jälkeen saadaan

$$\begin{aligned}
\frac{\partial^2}{\partial\mu^2} l &= -\frac{n}{\sigma^2}, \\
\frac{\partial^2}{\partial\mu\partial(\sigma^2)} l &= \frac{n(\mu - \overline{y})}{\sigma^4}, \\
\frac{\partial^2}{\partial(\sigma^2)^2} l &= \frac{n}{2\sigma^4} - \frac{(n-1)s^2 + n(\mu - \overline{y})^2}{\sigma^6},
\end{aligned}$$

joten havaittu informaatio on

$$\boldsymbol{j}(\mu, \sigma^2; \boldsymbol{y}) = \begin{bmatrix} n/\sigma^2 & -n(\mu - \overline{y})/\sigma^4 \\ -n(\mu - \overline{y})/\sigma^4 & -n/2\sigma^4 + [(n-1)s^2 + n(\mu - \overline{y})^2]/\sigma^6 \end{bmatrix}.$$

Erityisesti suurimman uskottavuuden pisteessä $\hat{\mu} = \overline{y}$, $\hat{\sigma}^2 = (n-1)s^2/n$ tämä saa arvon

$$\boldsymbol{j}(\hat{\mu}, \hat{\sigma}^2; \boldsymbol{y}) = \begin{bmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & n/2\hat{\sigma}^4 \end{bmatrix}.$$

Lasketaan vielä odotettu eli Fisherin informaatio. Selvästi $i_{1,1}(\mu, \sigma^2) = n/\sigma^2$ ja koska $E(\overline{Y}) = \mu$, niin $i_{1,2}(\mu, \sigma^2) = i_{2,1}(\mu, \sigma^2) = 0$. Alkion $i_{2,2}(\mu, \sigma^2)$ laskemiseksi lienee yksinkertaisinta muistaa, että

$$(n-1)S^2 + n(\mu - \overline{Y})^2 = \sum_{i=1}^{n} (Y_i - \mu)^2,$$

jonka odotusarvo on $n\sigma^2$. Siispä

$$\boldsymbol{i}(\mu, \sigma^2) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}.$$

Erityisesti nähdään, että parametrit μ ja σ^2 ovat ortogonaaliset.

2.6.4 Pistemäärä ja säännölliset mallit. Useampiulotteisen parametrin tapauksessa *pistemäärä(funktio)* on log-uskottavuusfunktion gradientti eli sen ensimmäisen kertaluvun osittaisderivaatoista muodostuva vektori

$$\nabla l(\boldsymbol{\theta}) = \nabla l(\boldsymbol{\theta}; \boldsymbol{y}) = \left(\frac{\partial}{\partial\theta_1} l(\boldsymbol{\theta}; \boldsymbol{y}), \ldots, \frac{\partial}{\partial\theta_d} l(\boldsymbol{\theta}; \boldsymbol{y}) \right).$$

Mallin säännöllisyys määritellään aivan samalla tavalla kuin kohdassa 2.5.2 tehtiin. Ehtojen c ja d on oltava nyt voimassa kaikille funktion $\boldsymbol{\theta} \mapsto f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ ensimmäisen ja vastaavasti toisen kertaluvun osittaisderivaatoille. Periaatteessa täysin samanlaiset laskelmat kuin aiemmin osoittavat, että säännöllisen mallin pistemäärän odotusarvovektori on nolla ja sen kovarianssimatriisi yhtyy Fisherin informaatioon:

$$E[\nabla l(\boldsymbol{\theta};\boldsymbol{Y})] = \mathbf{0}, \qquad \operatorname{cov}[\nabla l(\boldsymbol{\theta};\boldsymbol{Y})] = \boldsymbol{i}(\boldsymbol{\theta}).$$

Komponenttimuodossa nämä yhtälöt merkitsevät, että

$$E\left[\frac{\partial}{\partial\theta_a} l(\boldsymbol{\theta};\boldsymbol{Y})\right] = 0,$$

$$\operatorname{cov}\left[\frac{\partial}{\partial\theta_a} l(\boldsymbol{\theta};\boldsymbol{Y}), \frac{\partial}{\partial\theta_b} l(\boldsymbol{\theta};\boldsymbol{Y})\right] = i_{a,b}(\boldsymbol{\theta}),$$

kun $a, b = 1, \ldots, d$. Huomaa, että keskenään ortogonaalisten parametrin komponenttien suhteen laskettujen osittaisderivaattojen kovarianssi on nolla eli kyseiset pistemääräfunktion komponentit ovat korreloimattomat.

Harjoitustehtäviä

2.1. Eräs leijona viettää yönsä jossakin kolmesta tilasta: se on koko yön joko hyvin aktiivinen ($\theta = 1$), kohtalaisen aktiivinen ($\theta = 2$) tai unelias ($\theta = 3$). Leijona syö yön aikana Y ihmistä. Satunnaismuuttujan Y pistetodennäköisyydet riippuvat leijonan tilasta ja käyvät ilmi taulukosta alla.

y	0	1	2	3	4
$f_Y(y;1)$	.00	.05	.05	.80	.10
$f_Y(y;2)$	.05	.05	.80	.10	.00
$f_Y(y;3)$	.90	.08	.02	.00	.00

Eräänä aamuna havaittiin, että yön aikana leijona oli syönyt a) $y = 1$, b) $y = 3$ ihmistä. Esitä graafisesti vastaavat uskottavuusfunktiot. Millaisia päätelmiä tekisit leijonan tilasta? Pohdi päätelmien luotettavuutta.

2.2. Olkoot $y_1, \ldots, y_n$ ja a mielivaltaisia reaalilukuja. Tarkista, että

$$\sum_{i=1}^{n}(y_i - a)^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - a)^2,$$

kun $\overline{y} = \sum_{i=1}^{n} y_i/n$. Mitä tekemistä tällä on Pythagoraan lauseen kanssa? (Tulosta käytettiin esimerkissä 2.1.4.)

2.3. Oletetaan, että $Y_1, \ldots, Y_n \sim Exp(\lambda)$ ⫫. Muodosta aineistoa $\boldsymbol{y} = (y_1, \ldots, y_n)$ vastaavat uskottavuus- ja log-uskottavuusfunktio sekä määritä huolellisesti perustellen parametrin λ suurimman uskottavuuden estimaatti. Hahmottele log-uskottavuusfunktion kuvaajaa.

2.4. Totea, että Poisson-mallin $Y_1, \ldots, Y_n \sim P(\mu)$ ⫫ log-uskottavuusfunktio voidaan saattaa muotoon

$$l(\mu) = -n\mu + n\overline{y}\log\mu,$$

jossa $\overline{y}$ on otoskeskiarvo. Ratkaise uskottavuusyhtälö $l'(\mu) = 0$ ja määritä μ:n suurimman uskottavuuden estimaatti. Eräässä ääritapauksessa su-estimaattia ei ole olemassa; missä?

2.5. Olkoon θ positiivinen parametri, ja asetetaan

$$f(y;\theta) = 2\theta^{-1} y \exp(-y^2/\theta), \qquad \text{kun } y > 0,$$

ja $f(y;\theta) = 0$, kun $y \le 0$.

a) Tarkista integroimalla, että tämä kelpaa erään jatkuvan jakauman tiheysfunktioksi.

b) Oletetaan, että $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat kukin em. jakaumaa. Muodosta tämän mallin uskottavuusfunktio ja log-uskottavuusfunktio sekä määritä suurimman uskottavuuden estimaatti $\hat{\theta}$, kun aineisto on $\boldsymbol{y} = (y_1, \ldots, y_n)$.

2.6. Olkoon mallina $Y_1, \ldots, Y_n \sim \mathit{Tas}(\theta, \theta+1)$ $\perp\!\!\!\perp$. Johda aineistoa $\boldsymbol{y} = (y_1, \ldots, y_n)$ vastaava uskottavuusfunktio ja totea, että se saa suurimman arvonsa jokaisessa välin $\big(y_{(n)} - 1, y_{(1)}\big)$ pisteessä, kun merkitään $y_{(1)} = \min(y_1, \ldots, y_n)$ ja $y_{(n)} = \max(y_1, \ldots, y_n)$. Siten su-estimaatti $\hat{\theta}(\boldsymbol{y})$ ei ole yksikäsitteinen (todennäköisyydellä yksi).

2.7. Laatikossa on arpaliput, jotka on numeroitu luvuilla $1, 2, \ldots, \nu$. Lippujen lukumäärä ν on tuntematon positiivinen kokonaisluku. Laatikosta poimitaan umpimähkään ja palauttaen viisi lippua. Ne ovat 21, 2, 13, 30 ja 11. Muotoile tätä koeasetelmaa kuvaava tilastollinen malli. Ilmoita sitten havaintoja vastaava uskottavuusfunktio ja perustele, että $\hat{\nu} = 30$.

Vihje. Kyseessä on diskreetti versio kohdassa 2.2.8 esitetystä esimerkistä.

2.8. Tarkastellaan gammajakaumamallia $Y_1, \ldots, Y_n \sim G(\alpha, 1/\beta)$ $\perp\!\!\!\perp$, jossa $\alpha, \beta > 0$.

a) Parametrina on (α, β). Totea, että uskottavuusyhtälöt voidaan saattaa muotoon

$$\begin{cases} \log\alpha - \psi(\alpha) = \log\overline{y} - n^{-1}\sum_{i=1}^n \log y_i, \\ \beta = \overline{y}/\alpha, \end{cases}$$

jossa $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$, ja järkeile, ettei niitä voi ratkaista suljetussa muodossa.

b) Parametrina on vain β, ja α on tunnettu luku. Mikä on β:n su-estimaatti?

2.9. Tutkitaan yhden selittäjän lineaarista regressiomallia $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, $i = 1, \ldots, n$, jossa $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$. Totea, että sen log-uskottavuusfunktio voidaan saattaa muotoon

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

Osoita sitten, että parametrin $(\alpha, \beta, \sigma^2)$ suurimman uskottavuuden estimaatit ovat

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}, \qquad \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2,$$

jossa $\overline{y} = (y_1 + \cdots + y_n)/n$ ja $\overline{x} = (x_1 + \cdots + x_n)/n$.

Ehdotus. On siis määritettävä piste, jossa l saa suurimman arvonsa. Etsi ensin piste $(\hat{\alpha}, \hat{\beta})$, jossa neliösumma $S(\alpha, \beta) = \sum(y_i - \alpha - \beta x_i)^2$ saa pienimmän arvonsa. Tätä varten määritä S:n osittaisderivaattojen nollakohdat ja perustele, että löydetty piste on todella globaali minimikohta, esim. tutkimalla toisen kertaluvun osittaisderivaattojen muodostamaa matriisia. Vaihtoehtoisesti minimoi S aluksi α:n suhteen (pitäen β kiinteänä) ja sitten β:n suhteen (kun α:n paikalle on sijoitettu äsken saatu β:sta riippuva lauseke). Lopuksi etsi piste $\hat{\sigma}^2$, jossa $l(\hat{\alpha}, \hat{\beta}, \sigma^2)$ (σ^2:n funktiona) saa suurimman arvonsa.

2.10. Vuonna 1898 ilmestyneessä kuuluisassa tilastossa oli raportoitu hevosenpotkuun kuolleiden miesten vuosittaiset lukumäärät neljässätoista Preussin armeijan yksikössä kahdenkymmenen vuoden ajalta, yhteensä siis 280 havaintoa. Yhteenveto tuloksista on alla.

Kuolleita	0	1	2	3	4	≥ 5
Havaintoja	144	91	32	11	2	0

Oletetaan, että kuolleiden lukumäärä yhtenä vuonna yhdessä yksikössä noudattaa Poisson-jakaumaa ja on riippumaton sekä muiden yksiköiden että muiden vuosien lukumääristä. Olkoon μ kyseisen Poisson-jakauman odotusarvo. Muodosta aineistoa vastaavan log-uskottavuusfunktion lauseke ja etsi μ:n suurimman uskottavuuden estimaatti. Mikä on su-estimaatti todennäköisyydelle, että yhtään miestä ei kuole tietyssä yksikössä vuoden aikana?

2.11. Olkoot $Y_1, \ldots, Y_n \sim \mathit{Exp}(\lambda)$ $\perp\!\!\!\perp$ (ks. teht. 2.3). Laske havaittu informaatio $j(\hat{\lambda}; \boldsymbol{y})$, Fisherin informaatio $i(\lambda)$ ja odotusarvo $E[l'(\lambda; \boldsymbol{Y})^2]$.

2.12. Laske havaittu informaatio $j(\hat{\mu}; \boldsymbol{y})$ ja Fisherin informaatio $i(\mu)$ mallissa $Y_1, \ldots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$ (ks. teht. 2.4).

2.13. Tarkastellaan mallia, jossa havaintoja vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomat. Mallin parametri on θ. Totea, että mallin havaittu informaatio ja Fisherin informaatio ovat

$$j(\theta; \boldsymbol{y}) = j_1(\theta; y_1) + \cdots + j_n(\theta; y_n), \qquad i(\theta) = i_1(\theta) + \cdots + i_n(\theta),$$

jossa $j_k(\theta; y_k)$ on pelkästään yhteen havaintoon y_k perustuva havaittu informaatio ja $i_k(\theta) = E[j_k(\theta; Y_k)]$ on vastaava Fisherin informaatio. Miten tulkitset tämän tuloksen?

2.14. Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ tilastollinen malli, jonka parametri θ on yksiulotteinen. Olkoon $\phi = \phi(\theta)$ kääntäen yksikäsitteinen parametrimuunnos, jonka käänteismuunnos on $\theta = \theta(\phi)$. Tarkastellaan uudelleenparametroitua mallia $f^*_{\boldsymbol{Y}}(\boldsymbol{y}; \phi) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta(\phi))$. Näytä, että sen havaittu informaatio ja Fisherin informaatio saadaan alkuperäisen mallin informaatioista kaavoilla

$$j^*(\hat{\phi}; \boldsymbol{y}) = j(\hat{\theta}; \boldsymbol{y})\, \theta'(\hat{\phi})^2, \qquad i^*(\phi) = i(\theta(\phi))\, \theta'(\phi)^2.$$

Oletetaan, että malli täyttää kaikki tarpeelliset säännöllisyysehdot ja että parametrimuunnos on riittävän monta kertaa derivoituva. *Muista.* $l'(\hat{\theta}; \boldsymbol{y}) = 0$ ja $E[l'(\theta; \boldsymbol{Y})] = 0$.

2.15. Tarkastellaan parametrin $\theta \in (0, 1)$ *logit-muunnosta* $\phi = \phi(\theta) = \log\{\theta/(1-\theta)\}$.

a) Totea, että kyseessä on kääntäen yksikäsitteinen eli bijektiivinen funktio $(0, 1) \to \mathbb{R}$, ja määritä käänteisfunktio $\theta = \theta(\phi)$. Hahmottele myös kuvaaja (θ, ϕ)-koordinaatistoon.

b) Toistokoemalli $Y_1, \ldots, Y_n \sim B(\theta)$ $\perp\!\!\!\perp$ uudelleenparametroidaan logit-muunnoksella. Johda syntyvän mallin log-uskottavuusfunktio $l^*(\phi)$, su-estimaatti $\hat{\phi}$ ja Fisherin informaatio $i^*(\phi)$.

Lisätieto. Ns. logistisessa regressiossa mallitetaan toistokokeen onnistumistodennäköisyyden logit-muunnosta joidenkin selittävien muuttujien avulla. Esimerkiksi $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$, jossa $Y_i \sim B(\theta_i)$ ja $\phi(\theta_i) = \alpha + \beta x_i$, jolloin mallin parametri olisi pari (α, β).

2.16. Sähkölaitteen kestoikä noudattaa jakaumaa $\mathit{Exp}(\lambda)$, jossa $\lambda > 0$ halutaan estimoida. Tätä varten poimitaan n laitteen satunnaisotos, pannaan laitteet käyntiin ja ajan $t > 0$ (kiinteä annettu luku) kuluttua lasketaan yhä toiminnassa olevien laitteiden lukumäärä k.

a) Muodosta asetelmaa kuvaava malli parametrille λ ja määritä sen Fisherin informaatio $i(\lambda; t)$. Argumentti t viittaa tässä siihen, että informaatio riippuu valitusta ajasta t.

b) Totea, että $i(\lambda; t) \to 0$, kun $t \to 0+$ tai $t \to \infty$. Siis on odotettavissa, että päätelmät λ:sta ovat epätarkkoja, jos t on hyvin pieni tai suuri. Pohdi miksi.

Ohje. k on "onnistumisten" lukumäärä toistokokeessa, jonka onnistumistodennäköisyys θ määräytyy eksponenttijakaumasta λ:n lausekkeena. Informaation laskemisessa voi käyttää tehtävää 2.14 ja esimerkissä 2.4.5 laskettua informaatiota θ:lle.

2.17. Miten edellisessä tehtävässä aika t olisi pyrittävä valitsemaan (λ:n funktiona), jotta odotettu informaatio $i(\lambda;t)$ olisi suurin mahdollinen? Totea derivaattaa tutkimalla, että kyseinen maksimikohta $t = t_\lambda$ määräytyy yhtälöstä $2 - 2e^{-\lambda t} - \lambda t = 0$. Numeerisesti ratkaisemalla saadaan $e^{-\lambda t_\lambda} \approx 0.203$. Kuinka monta prosenttia pienempi on luku $i(\lambda;t_\lambda)$ kuin Fisherin informaatio kokeesta, jossa olisi mitattu jokaisen otosyksikön kestoikä (ks. teht. 2.11)?

2.18. Eräällä järjestelmällä on kaksi mahdollista tilaa: 0 ja 1. Satunnaismuuttuja Y_i kertoo järjestelmän tilan ajanhetkellä $i = 1,2,3,4$. Tila hetkellä i riippuu tilasta hetkellä $i-1$ (mutta ei aikaisemmista) seuraavien todennäköisyyksien kuvaamalla tavalla:

$$P\{Y_i = 1 \,|\, Y_{i-1} = 1\} = \alpha, \qquad P\{Y_i = 0 \,|\, Y_{i-1} = 1\} = 1-\alpha,$$
$$P\{Y_i = 1 \,|\, Y_{i-1} = 0\} = 1-\beta, \qquad P\{Y_i = 0 \,|\, Y_{i-1} = 0\} = \beta,$$

jossa $0 < \alpha < 1$ ja $0 < \beta < 1$. Oletetaan, että järjestelmä lähtee aina tilasta 0: $Y_0 = 0$.

a) Aineisto on $\boldsymbol{y} = (y_1, y_2, y_3, y_4) = (0,1,1,0)$. Muodosta vastaava log-uskottavuusfunktio $l(\alpha, \beta; \boldsymbol{y})$ ja määritä suurimman uskottavuuden estimaatti $(\hat{\alpha}, \hat{\beta})$.

b) Ovatko parametrit α ja β ortogonaaliset?

2.19. Satunnaismuuttujan Y tiheysfunktio on muotoa $f_Y(y;\theta) = c(y)\exp\{\theta y - h(\theta)\}$ ja se on säännöllinen (ks. 2.5.2). Laske $E_\theta(Y)$ ja $\mathrm{var}_\theta(Y)$ funktion h ensimmäisen ja toisen derivaatan avulla.

2.20. Vektorilla $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^d$ parametroitu jakaumaperhe on d-ulotteinen *eksponenttiperhe*, jos sen yptf/ytf voidaan kirjoittaa muotoon

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\boldsymbol{y})\exp\left\{\sum_{j=1}^{d}\phi_j(\boldsymbol{\theta})t_j(\boldsymbol{y})\right\},$$

jossa $c(\boldsymbol{\theta})$ sekä $\phi_j(\boldsymbol{\theta})$:t ovat reaalisia ja riippuvat vain parametrista $\boldsymbol{\theta}$ ja $h(\boldsymbol{y})$ sekä $t_j(\boldsymbol{y})$:t ovat reaalisia vain aineistosta $\boldsymbol{y}$ riippuvia tunnuslukuja. Vektoria $\boldsymbol{\phi} = \big(\phi_1(\boldsymbol{\theta}), \ldots, \phi_d(\boldsymbol{\theta})\big)$ kutsutaan perheen *luonnolliseksi parametriksi*.

a) Osoita, että mallia $K \sim Bin(n,\theta)$ vastaava jakaumaperhe $f_K(k;\theta)$ (ks. 2.1.5) on yksiulotteinen eksponenttiperhe, luonnollisena parametrina θ:n logit-muunnos $\log\{\theta/(1-\theta)\}$ (vrt. teht. 2.15).

b) Osoita, että mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$ vastaava jakaumaperhe $f_{\boldsymbol{Y}}(\boldsymbol{y};\mu,\sigma^2)$ (ks. 1.2.3) on kaksiulotteinen eksponenttiperhe, luonnollisena parametrina $(\mu/\sigma^2, 1/\sigma^2)$. *Ehdotus.* Aloita kirjoittamalla $(y_i - \mu)^2 = y_i^2 - 2\mu y_i + \mu^2$.

c) Perustele, että eksponenttiperheen määritelmässä yllä vakio $c(\boldsymbol{\theta})$ riippuu $\boldsymbol{\theta}$:sta vain $\boldsymbol{\phi}$:n kautta. Siten perhe voidaan aina parametroida myös luonnollisen parametrinsa avulla.

3 Yleistä estimointiteoriaa

3.1 Johdanto

Ajatellaan yleisellä tasolla sitä piste-estimointitehtävää, joka kurssin alussa (ks. 1.3) asetettiin: havaitun aineiston $\boldsymbol{y}$ perusteella on määritettävä parametriavaruuden piste, joka on mahdollisimman hyvä arvio eli estimaatti annetun mallin tuntemattomalle parametrille $\boldsymbol{\theta}$. Yleisemmin voidaan tarkastella tilannetta, jossa estimoitavana on jokin parametrin funktio $\boldsymbol{g}(\boldsymbol{\theta})$ (ei välttämättä kääntäen yksikäsitteinen), esimerkiksi vain jokin $\boldsymbol{\theta}$:n komponentti.

Edellisessä luvussa esitetty suurimman uskottavuuden menetelmä on varsin luonteva ja yleispätevä ratkaisu tähän tehtävään. Herää kuitenkin eräitä kysymyksiä:

a) Millaisia muita menetelmiä on konstruoida estimaatteja?

b) Millä kriteereillä estimaattien hyvyyttä mitataan? Jos estimointitehtävään on tarjolla erilaisia ratkaisuja, mikä niistä on optimaalinen?

c) Onko suurimman uskottavuuden menetelmä hyvä tai jopa optimaalinen estimointimenetelmä?

Tässä luvussa otetaan kantaa näihin kysymyksiin. Kysymystä a sivutaan tosin vain hyvin lyhyesti, kun pykälässä 3.3 esitellään momenttimenetelmän nimellä tunnettu estimointimenetelmä. Kysymyksiä b ja c sen sijaan tarkastellaan monipuolisemmin.

Estimaattien hyvyyden arviointi perustuu toistetun aineistonkeruun ajatukseen (vrt. myös 1.4). Oletetaan, että $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$ on aineistosta $\boldsymbol{y}$ tavalla tai toisella muodostettu estimaatti funktiolle $\boldsymbol{g}(\boldsymbol{\theta})$. Kuvitellaan, että aineiston tuottanut satunnaiskoe, joka noudattaa annettua tilastollista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, toistettaisiin samoissa olosuhteissa yhä uudelleen ja uudelleen. Saatava aineisto vaihtelisi tällöin satunnaisesti kyseisen mallin eli satunnaisvektorin $\boldsymbol{Y}$ jakauman kuvaamalla tavalla. Toivottavaa ilmeisestikin olisi, että tällöin vastaavat estimaatit $\boldsymbol{t}$ osuisivat "mahdollisimman usein mahdollisimman lähelle" funktion todellista arvoa $\boldsymbol{g}(\boldsymbol{\theta})$. Toisin sanoen vastaavan satunnaismuuttujan tai -vektorin $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ eli *estimaattorin* jakauman tulisi olla keskittynyt mahdollisimman tiiviisti $\boldsymbol{g}(\boldsymbol{\theta})$:n läheisyyteen.

Jatkossa puhutaankin pääasiassa estimaattoreista ja niiden ominaisuuksista eikä niinkään yksittäisen aineiston perusteella lasketuista estimaateista. Pykälissä 3.2, 3.4 ja 3.5 esitetään eräitä perinteisiä kriteerejä, joita estimaattorien toivotaan toteuttavan. Jo tässä yhteydessä on kuitenkin syytä huomauttaa, että kaikkein yksinkertaisimpia estimointiongelmia lukuunottamatta näitä kaikkia ehtoja ei yleensä voida samanaikaisesti täyttää. Luvun lopuksi pykälässä 3.6 todetaan, että suurimman uskottavuuden menetelmän tuottamat estimaattorit ovat yleensä melko hyviä, ainakin jos aineiston koko on kyllin suuri.

3.2 Harhattomuus

3.2.1 Määritelmät. Tarkastellaan tilastollista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, jonka parametriavaruus on $\Omega \subset \mathbb{R}^d$. Funktion $\boldsymbol{g}(\boldsymbol{\theta})$ estimaattori $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on *harhaton*, jos on voimassa

$$E_{\boldsymbol{\theta}}(\boldsymbol{T}) = \boldsymbol{g}(\boldsymbol{\theta}) \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega.$$

Muussa tapauksessa estimaattori on *harhainen* ja sen *harha*

$$\boldsymbol{b}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\boldsymbol{T}) - \boldsymbol{g}(\boldsymbol{\theta})$$

on nollasta poikkeava ainakin yhdessä pisteessä $\boldsymbol{\theta} \in \Omega$. Tässä odotusarvo-operaattorin symboliin on liitetty alaindeksi $\boldsymbol{\theta}$ sen korostamiseksi, että odotusarvo on laskettava juuri kyseistä parametriarvoa käyttäen. Sen voi jättää pois, jos tämän käytännön voi katsoa olevan asiayhteydestä selvän.

Tärkein erikoistapaus on tietysti se, jossa estimoitavana on itse parametri $\boldsymbol{\theta}$. Sen estimaattori $\boldsymbol{T}$ on siis harhaton, jos

$$E_{\boldsymbol{\theta}}(\boldsymbol{T}) = \boldsymbol{\theta} \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega,$$

ja harhaisen estimaattorin harha saadaan kaavasta

$$\boldsymbol{b}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\boldsymbol{T}) - \boldsymbol{\theta}.$$

Huomaa erityisesti, että jos $d > 1$ eli parametri on vektori, niin $E_{\boldsymbol{\theta}}(\boldsymbol{T})$ tarkoittaa odotusarvovektoria eli vektoria, jonka komponentit ovat $\boldsymbol{T}$:n komponenttien T_j odotusarvot. Tällöin harhattomuusehto merkitsee, että

$$E_{\boldsymbol{\theta}}(T_j) = \theta_j, \qquad \text{kun } \boldsymbol{\theta} \in \Omega \text{ ja } j = 1, \ldots, d.$$

Samoin harha $\boldsymbol{b}(\boldsymbol{\theta})$ on estimaattorin komponenttien harhoista koostuva d-ulotteinen vektori.

Tulkinnallisesti harhattomuus merkitsee sitä, että estimaattorin arvot eli estimaatit tuottavat toistetussa aineistonkeruussa keskimäärin oikean tuloksen.

3.2.2 Esimerkki: jakauman odotusarvon harhaton estimointi. Oletetaan, että havainnot $Y_1, \ldots, Y_n$ noudattavat jakaumaa, jonka odotusarvo on μ. Niiden ei tarvitse olla riippumattomia. Tällöin otoskeskiarvo

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

on μ:n harhaton estimaattori, sillä odotusarvon lineaarisuuden perusteella

$$E(\overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

Tästä yleisestä havainnosta seuraa, että jos $Y_1, \ldots, Y_n$ noudattavat jotakin tutuista jakaumista $B(\mu)$, $P(\mu)$, $Exp(1/\mu)$ tai $N(\mu, \sigma^2)$, niin $\overline{Y}$ on parametrin μ harhaton estimaattori. Jos $Y_1, \ldots, Y_n$ ovat riippumattomat, on kaikissa näissä tapauksissa todettu, että $\overline{Y}$ on myös μ:n su-estimaattori eli $\hat{\mu} = \overline{Y}$ (ks. 2.2.5, 2.2.6 ja 2.3.2 sekä teht. 2.4).

3.2.3 Esimerkki: jakauman varianssin harhaton estimointi. Olkoot $Y_1, \ldots, Y_n$ riippumattomia ja samoin jakautuneita, yhteisenä varianssinaan σ^2. Määritellään *otosvarianssi* S^2 kaavalla

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Tehtävässä 3.3 todetaan, että $E(S^2) = \sigma^2$, joten S^2 on σ^2:n harhaton estimaattori.

Mallin $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$ osalta on aikaisemmin todettu (ks. 2.2.6), että varianssin su-estimaattori on

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Tämä on harhainen, ja yhtälöstä $\hat{\sigma}^2 = (n-1)S^2/n$ nähdään, että

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \qquad \text{eli} \qquad b(\sigma^2) = -\frac{\sigma^2}{n}.$$

Estimaattorin $\hat{\sigma}^2$ antamat arvot ovat siis keskimäärin hieman alle todellisen parametriarvon σ^2. Varianssin estimointiin käytetäänkin yleensä estimaattoria S^2.

3.2.4 Harhattomuuden merkityksestä ja ongelmista. Harhattomuus tuntuu perin luonnolliselta ja haluttavalta estimaattorin ominaisuudelta – kukapa tutkija haluaisi perustaa esittämänsä tutkimustulokset menetelmään, jonka tiedetään tuottavan keskimäärin vääriä tuloksia. Harhattomuus onkin ollut vanhastaan keskeisessä asemassa piste-estimointia käsittelevässä tilastollisen päättelyn kirjallisuudessa, ja vuosikymmenten saatossa on kehitetty laaja teoria optimaalisten harhattomien estimaattorien konstruoimiseksi. Perustavoitteena on ollut kussakin estimointitehtävässä etsiä harhaton estimaattori, jolla olisi pienin mahdollinen varianssi, eli ns. UMVU-estimaattori (*uniformly minimum variance unbiased*).

Harhattomuuden merkitys estimointikriteerinä on kuitenkin viime aikoina vähentynyt. Monissa käytännön sovelluksissa se onkin liian rajoittava vaatimus, ja lisäksi sitä kohtaan voidaan esittää voimakasta periaatteellista kritiikkiä:

a) Toisinaan harhattomia estimaattoreita ei ole lainkaan tai ne ovat muuten ilmeisen epätyydyttäviä. Katso esimerkki 3.2.6 ja tehtävä 3.8.

b) Harhattomuus ei ole invariantti parametrimuunnosten suhteen. Invarianssiperiaatteen mukaan estimointitulos ei saisi riippua käytetystä mallin parametroinnista: jos siis $\boldsymbol{T}$ on parametrin $\boldsymbol{\theta}$ estimointiin käytetty estimaattori ja $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ on mallin uudelleenparametrointi, tulisi $\boldsymbol{\phi}$:n estimaattorina käyttää vastaavaa muunnosta $\boldsymbol{g}(\boldsymbol{T})$. Harhattomuus ei kuitenkaan yleensä säily kuin vain lineaarisissa muunnoksissa. Muista, että su-estimaattorit ovat aina invariantteja (ks. 2.3). Katso esimerkki 3.2.5.

c) Harhattomuusvaatimus ei ole sopusoinnussa uskottavuusperiaatteen kanssa. Uskottavuusperiaate edellyttää, että mallin parametrista tehtävien päätelmien tulisi perustua vain uskottavuusfunktioon, ts. kahden eri mallin puitteissa tehtävien päätelmien tulisi olla samoja, mikäli havaittuihin aineistoihin liittyvät uskottavuusfunktiot ovat identtiset mahdollisesti aineistosta riippuvaa vakiokerrointa vaille. Harhattomuus sen sijaan vahvasti riippuu spesifioidusta mallista. Katso esimerkki 3.2.6.

Yleisohjeena voidaan sanoa, että harhattomuus on hyvä ominaisuus muttei ollenkaan ehdoton vaatimus käytettävälle estimointimenetelmälle. Pienen harhan vuoksi ei estimaattoria pidä hylätä, jos se muilta ominaisuuksiltaan on hyvä.

3.2.5 Esimerkki: eksponenttijakauman kaksi parametrointia. Edellä (ks. 3.2.2) todettiin, että mallissa $Y_1, \ldots, Y_n \sim Exp(1/\mu)$ on $\hat{\mu} = \overline{Y}$ parametrin μ harhaton estimaattori. Mallin vaihtoehtoinen parametrointi on $\lambda = 1/\mu$, jonka su-estimaattori on $\hat{\lambda} = 1/\overline{Y}$ invarianssiominaisuuden mukaisesti. Lasketaan seuraavaksi $\hat{\lambda}$:n odotusarvo ja todetaan, ettei kyseessä ole harhaton estimaattori.

Todennäköisyyslaskennasta tiedetään, että riippumattomien eksponenttihavaintojen summana muuttuja $S = Y_1 + \cdots + Y_n$ noudattaa gammajakaumaa $G(n, \lambda)$. Sen tiheysfunktio on siis

$$f_S(s) = \frac{\lambda^n}{(n-1)!} s^{n-1} e^{-\lambda s}, \qquad s > 0.$$

Koska $\hat{\lambda} = n/S$, saadaan

$$\begin{aligned} E(\hat{\lambda}) &= \int_0^\infty \frac{n}{s} f_S(s)\, ds = n \int_0^\infty \frac{\lambda^n}{(n-1)!} s^{n-2} e^{-\lambda s}\, ds \\ &= \frac{n\lambda}{n-1} \int_0^\infty \frac{\lambda^{n-1}}{(n-2)!} s^{n-2} e^{-\lambda s}\, ds = \frac{n\lambda}{n-1}. \end{aligned}$$

Viimeinen yhtäsuuruus perustui siihen, että integroitava funktio on $G(n-1, \lambda)$-jakauman tiheysfunktio, joten integraalin arvo on 1. Näin ollen $\hat{\lambda}$ on harhainen, ja sen harha on $b(\lambda) = n\lambda/(n-1) - \lambda = \lambda/(n-1)$. Tässä päättelyssä on selvästikin oletettava, että $n \geq 2$; jos $n = 1$, on helppoa todeta, että muuttujalla $\hat{\lambda} = 1/Y_1$ ei ole (äärellistä) odotusarvoa lainkaan.

Harhainen su-estimaattori $\hat{\lambda}$ on tässä mallissa helposti korjattavissa harhattomaksi: määritellään $T = (n-1)\hat{\lambda}/n = (n-1)/S$, jolloin äskeinen lasku osoittaa, että $E(T) = \lambda$, jos $n \geq 2$.

3.2.6 Esimerkki: geometrinen jakauma. Ajatellaan riippumatonta toistokoetta, jossa onnistumistodennäköisyys on θ. Olkoon N sen kokeen järjestysnumero, jolloin ensimmäinen onnistuminen sattuu. Esimerkiksi lantinheitossa kruunun todennäköisyys on θ ja ensimmäinen kruunu saadaan N:nnellä heitolla. Tällöin N on satunnaismuuttuja, joka noudattaa geometrista jakaumaa: sen ptf on

$$f_N(n; \theta) = P_\theta\{N = n\} = \theta(1-\theta)^{n-1}, \qquad n = 1, 2, \ldots. \tag{3.1}$$

Oletetaan, että $T = t(N)$ on jokin parametrin θ harhaton estimaattori. Tämä tarkoittaa, että

$$\theta = E(T) = \sum_{n=1}^{\infty} t(n) f_N(n; \theta) = \theta \sum_{n=1}^{\infty} t(n)(1-\theta)^{n-1}$$

eli siis

$$\sum_{n=1}^{\infty} t(n)(1-\theta)^{n-1} = 1.$$

Jos merkitään $\xi = 1 - \theta$, tämä yhtälö voidaan kirjoittaa muodossa

$$(t(1) - 1) + t(2)\xi + t(3)\xi^2 + t(4)\xi^3 + \cdots = 0.$$

Koska θ voi olla mikä tahansa luku välillä $(0, 1)$, on tämäkin voimassa kaikilla $\xi \in (0, 1)$. Kuten differentiaalilaskennasta tiedetään, suppenevan potenssisarjan summa on identtisesti nolla vain siinä tapauksessa, että kaikki sen kertoimet häviävät. Siten $t(1) = 1$ ja $t(n) = 0$ kaikilla $n \geq 2$.

Tarkasteltavassa mallissa on siis parametrilla θ vain yksi harhaton estimaattori. Sen mukaan on estimaatiksi valittava $t = 1$, jos havaittiin $n = 1$ (heti ensimmäinen koe onnistui), ja $t = 0$, jos havaittiin $n \geq 2$. Tämä on ilmeisen epätyydyttävä estimointimenetelmä. Erityisen järjettömältä tuntuisi estimoida todennäköisyys θ nollaksi, kun kuitenkin tiedetään, että yksi kokeista on onnistunut.

Kaavasta (3.1) voidaan lukea, että havaintoa n vastaava uskottavuusfunktio on $L(\theta) = \theta(1-\theta)^{n-1}$. Parametrin θ suurimman uskottavuuden estimaatiksi saadaan siitä $\hat{\theta} = 1/n$. Tämä on järkevä estimaatti, vaikkakin edellisen tarkastelun perusteella vastaavan su-estimaattorin $1/N$ täytyy olla harhainen. Huomaa lopuksi, että uskottavuusfunktio ja su-estimaatti ovat samat kuin tutussa toistokokeen binomimallissa $K \sim Bin(n, \theta)$ havaittaessa $k = 1$ (ks. 2.1.5 ja 2.2.5). Binomimallin su-estimaattori K/n on kuitenkin harhaton. Tämä merkillisyys osoittaa sen, että harhattomuus riippuu hyvin voimakkaasti mallin määrittelystä toisin kuin su-menetelmä, joka käyttää hyväkseen vain havaintoihin liittyvää uskottavuusfunktiota.

3.2.7 Asymptoottinen harhattomuus. Oletetaan, että $\boldsymbol{T}^{(n)}$ on funktion $\boldsymbol{g}(\boldsymbol{\theta})$ estimaattori, joka perustuu kokoa n olevaan aineistoon $(Y_1, \ldots, Y_n)$. Jos estimaattorin $\boldsymbol{T}^{(n)}$ harha lähestyy nollaa havaintojen lukumäärän n kasvaessa, ts.

$$\lim_{n\to\infty} E_{\boldsymbol{\theta}}(\boldsymbol{T}^{(n)}) = \boldsymbol{g}(\boldsymbol{\theta}) \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega,$$

niin sanotaan, että $\boldsymbol{T}^{(n)}$ – tai oikeammin jono $(\boldsymbol{T}^{(1)}, \boldsymbol{T}^{(2)}, \ldots)$ – on *asymptoottisesti harhaton*. Tämä on estimaattorilta varsin toivottava ominaisuus. Huomaa, että esimerkkien 3.2.3 ja 3.2.5 harhaiset estimaattorit $\hat{\sigma}^2$ ja $\hat{\lambda}$ ovat asymptoottisesti harhattomia. Eräissä kirjoissa asymptoottinen harhattomuus määritellään hieman eri tavalla.

3.3 Momenttimenetelmä

Momenttimenetelmä on estimointimenetelmä, joka juontaa juurensa jo 1800-luvun lopulle. Se on varsin yksinkertainen käyttää ja lähes aina tuottaa jonkinlaisen estimaattorin. Tuloksen optimaalisuudessa voi tosin olla toivomisen varaa.

3.3.1 Määritelmä. Oletetaan, että $Y_1, \ldots, Y_n$ noudattavat kaikki samaa jakaumaa, jolla on äärellinen odotusarvo. Esimerkissä 3.2.2 todettiin, että tällöin otoskeskiarvo $\overline{Y}$ on aina odotusarvon eli ensimmäisen momentin harhaton estimaattori. Vastaava lasku osoittaa yleisemmin, että k:s *otosmomentti*

$$m_k(\boldsymbol{Y}) = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$$

on jakauman k:nnen (origo)momentin $\mu_k = E(Y_1^k)$ harhaton estimaattori, mikäli tarkasteltavalla jakaumalla on kyseinen momentti olemassa. Siinä mielessä siis m_k on varsin luonteva empiirinen vastine jakauman teoreettiselle tunnusluvulle μ_k.

Oletetaan, että jakauma riippuu d-ulotteisesta parametrista $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Tällöin myös jakauman momentit μ_k riippuvat $\boldsymbol{\theta}$:sta eli

$$\mu_k = E(Y_1^k) = \mu_k(\theta_1, \ldots, \theta_d).$$

Momenttimenetelmässä kirjoitetaan parametrin $\boldsymbol{\theta}$ piste-estimoinnin lähtökohdaksi yhtälöryhmä

$$\begin{cases} m_1 = \mu_1(\theta_1, \ldots, \theta_d) \\ \qquad \vdots \\ m_d = \mu_d(\theta_1, \ldots, \theta_d), \end{cases}$$

jossa $m_k = m_k(\boldsymbol{y})$, $k = 1, \ldots, d$, ovat aineistosta $\boldsymbol{y}$ lasketut otosmomenttien arvot. Yhtälöitä muodostetaan siis yhtä monta kuin parametrilla on komponentteja. Tämä yhtälöryhmä pyritään nyt ratkaisemaan muuttujien $(\theta_1, \ldots, \theta_d)$ suhteen. Monesti sillä on yksikäsitteinen ratkaisu $(\tilde{\theta}_1, \ldots, \tilde{\theta}_d)$, joka riippuu aineistosta otosmomenttien $m_1, \ldots, m_d$ kautta:

$$\begin{cases} \tilde{\theta}_1 = \tilde{\theta}_1(m_1, \ldots, m_d) \\ \qquad \vdots \\ \tilde{\theta}_d = \tilde{\theta}_d(m_1, \ldots, m_d). \end{cases}$$

Näiden muodostama vektori $\tilde{\boldsymbol{\theta}}$ on parametrin $\boldsymbol{\theta}$ momenttimenetelmän mukainen estimaatti.

3.3.2 Esimerkki: normaalimalli. Tarkastellaan mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Jakauman $N(\mu, \sigma^2)$ ensimmäinen momentti on μ ja toinen momentti $\mu^2 + \sigma^2$, joten momenttimenetelmän mukaan on ratkaistava yhtälöpari

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i = \mu \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i^2 = \mu^2 + \sigma^2. \end{cases}$$

Tämän yksikäsitteinen ratkaisu on selvästi $(\mu, \sigma^2) = (\tilde{\mu}, \tilde{\sigma}^2)$, jossa

$$\begin{cases} \tilde{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i = \overline{y} \\ \tilde{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i^2 - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i\right)^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} (y_i - \overline{y})^2. \end{cases}$$

Normaalimallin kohdalla momenttimenetelmä johtaa siis samaan tulokseen kuin suurimman uskottavuuden menetelmä.

3.3.3 Esimerkki: välin $(0, \theta)$ tasajakauma. Olkoot $Y_1, \ldots, Y_n \sim Tas(0, \theta)$ $\perp\!\!\!\perp$. Kohdassa 2.2.8 todettiin pienten määrittelyvaikeuksien jälkeen, että parametrin θ su-estimaatti on suurin havainto eli

$$\hat{\theta} = y_{(n)} = \max(y_1, \ldots, y_n).$$

Koska $E(Y_1) = \theta/2$, niin momenttimenetelmän mukainen estimaatti on yhtälön $\sum_{i=1}^{n} y_i/n = \theta/2$ ratkaisu eli

$$\tilde{\theta} = \frac{2}{n}\sum_{i=1}^{n} y_i = 2\overline{y}.$$

Vastaava estimaattori on harhaton toisin kuin su-estimaattori. Tarkemmin näitä vertaillaan tehtävässä 3.10.

3.4 Tehokkuus ja informaatioepäyhtälö

3.4.1 Estimaattorin keskineliövirhe. Tarkastellaan estimointiongelmaa, jossa estimoitavana on reaalinen funktio $g(\boldsymbol{\theta})$ annetun mallin parametrista $\boldsymbol{\theta}$, esimerkiksi jokin $\boldsymbol{\theta}$:n komponentti tai tapauksessa $d = 1$ itse parametri θ. Jos T on $g(\boldsymbol{\theta})$:n harhaton estimaattori, niin T:n saamat arvot ovat keskimäärin oikeita eli sen todennäköisyysjakauman painopiste on sijoittunut todellisen parametriarvon kohdalle. Tämä ei vielä merkitse sitä, että T olisi erityisen ”tarkka” $g(\boldsymbol{\theta})$:n estimaattori, koska sen arvoissa voi olla hyvinkin paljon hajontaa eli T voi poiketa $g(\boldsymbol{\theta})$:sta huomattavastikin suurella todennäköisyydellä. Harhattoman estimaattorin hyvyyttä tässä suhteessa voidaankin mitata sen varianssilla $\mathrm{var}_{\boldsymbol{\theta}}(T)$: mitä pienempi varianssi on, sitä tiiviimmin T:n todennäköisyysmassa on keskittynyt $g(\boldsymbol{\theta})$:n ympärille, mikä merkitsee, että T:n arvot osuvat todennäköisesti lähelle odotusarvoa $g(\boldsymbol{\theta})$.

Mikäli T ei ole harhaton estimaattori, ei ole järkevää kiinnittää huomiota niinkään sen varianssiin vaan *keskineliövirheeseen* $E_{\boldsymbol{\theta}}[(T - g(\boldsymbol{\theta}))^2]$. Harhattoman estimaattorin tapauksessa nämä ovat tietysti yksi ja sama asia. Yleisesti sen sijaan on voimassa hajotelma

$$E_{\boldsymbol{\theta}}[(T - g(\boldsymbol{\theta}))^2] = \mathrm{var}_{\boldsymbol{\theta}}(T) + b(\boldsymbol{\theta})^2,$$

jossa $b(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(T) - g(\boldsymbol{\theta})$ on estimaattorin harha (ks. teht. 3.1). Estimaattori T on keskineliövirheen mielessä parempi kuin T', jos

$$E_{\boldsymbol{\theta}}[(T - g(\boldsymbol{\theta}))^2] \le E_{\boldsymbol{\theta}}[(T' - g(\boldsymbol{\theta}))^2] \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega,$$

jossa Ω on mallin parametriavaruus. Jos T ja T' ovat kumpikin harhattomia, tämä voidaan yhtäpitävästi kirjoittaa muodossa

$$\mathrm{var}_{\boldsymbol{\theta}}(T) \le \mathrm{var}_{\boldsymbol{\theta}}(T') \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega$$

ja sanotaan, että T on *tehokkaampi* kuin T'.

3.4.2 Esimerkki: normaalimallin odotusarvo. Pohditaan odotusarvon μ estimointia tutussa mallissa $Y_1 \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jossa σ_0^2 on tunnettu luku. On todettu, että $\hat{\mu} = \overline{Y}$ on harhaton. Muita harhattomia estimaattoreita olisivat esimerkiksi $S = Y_1$ ja $T = (Y_1 + Y_2)/2$. Näiden kolmen varianssit ovat σ_0^2/n, σ_0^2 ja $\sigma_0^2/2$, joten $\hat{\mu}$ on selvästi tehokkain (jos $n > 2$).

3.4.3 Informaatioepäyhtälö tapauksessa $d = 1$. Edellisen pohjalta herää kysymys, onko yleensä olemassa estimaattoreita, joiden keskineliövirhe on mielivaltaisen pieni. Olisiko vaikkapa yo. esimerkissä mahdollista löytää μ:n harhaton estimaattori, jonka varianssi olisi vieläkin pienempi kuin σ_0^2/n? Seuraava yleinen tulos kertoo, että näin ei ole ainakaan säännöllisten mallien kohdalla, vaan estimaattorin varianssilla ja siten keskineliövirheellä on alaraja, jota ei voi ohittaa. Tarkastellaan tässä vaiheessa vain reaaliparametrisen mallin ($d = 1$) tapausta.

Lause. *Oletetaan, että $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ on säännöllinen malli (ks. 2.5.2), ja olkoon $i(\theta)$ sen Fisherin informaatio. Olkoon lisäksi $T = t(\boldsymbol{Y})$ jokin funktion $g(\theta)$ estimaattori.*

Jos estimaattorin T harha on $b(\theta)$, niin

$$\mathrm{var}_{\theta}(T) \ge \frac{[g'(\theta) + b'(\theta)]^2}{i(\theta)}. \tag{3.2a}$$

Jos T on $g(\theta)$:n harhaton estimaattori, pätee

$$\operatorname{var}_\theta(T) \geq \frac{g'(\theta)^2}{i(\theta)}, \tag{3.2b}$$

ja erityisesti jos T on θ:n harhaton estimaattori,

$$\operatorname{var}_\theta(T) \geq \frac{1}{i(\theta)}. \tag{3.2c}$$

Todistus. Todistetaan lause jatkuvan mallin tapauksessa; diskreetissä tapauksessa on seuraavassa integraalit korvattava summilla.

Todistus perustuu todennäköisyyslaskennasta tuttuun Schwarzin epäyhtälöön

$$\operatorname{cov}(U, V)^2 \leq \operatorname{var}(U) \operatorname{var}(V),$$

jota sovelletaan satunnaismuuttujiin $U = T = t(\boldsymbol{Y})$ ja $V = l'(\theta; \boldsymbol{Y})$. Palautetaan mieleen, että mallin log-uskottavuusfunktio on $l(\theta; \boldsymbol{y}) = \log f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ ja

$$l'(\theta; \boldsymbol{y}) = \frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)}$$

on siihen liittyvä pistemääräfunktio (ks. 2.5.1). Säännöllisyydestä puolestaan seuraa (ks. 2.5.3), että $E[l'(\theta; \boldsymbol{Y})] = 0$ ja $\operatorname{var}[l'(\theta; \boldsymbol{Y})] = i(\theta)$.

Lasketaan mainittu kovarianssi:

$$\begin{aligned}
\operatorname{cov}[T, l'(\theta; \boldsymbol{Y})] &= E[t(\boldsymbol{Y}) l'(\theta; \boldsymbol{Y})] \\
&= \int_A t(\boldsymbol{y}) l'(\theta; \boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\, d\boldsymbol{y} \\
&= \int_A t(\boldsymbol{y}) \frac{\partial}{\partial\theta} f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\, d\boldsymbol{y} \\
&= \frac{d}{d\theta} \int_A t(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\, d\boldsymbol{y} \\
&= \frac{d}{d\theta} E(T),
\end{aligned}$$

jossa integroinnit suoritetaan yli jakauman alustan $A = \{\boldsymbol{y} : f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) > 0\}$. Toiseksi viimeinen yhtälö perustui jälleen mallin säännöllisyyteen. Huomaa myös, että tässä on jätetty merkitsemättä näkyviin se, että laskettavat odotusarvot ym. riippuvat θ:sta. Soveltamalla Schwarzin epäyhtälöä kuten alussa kerrottiin, saadaan siis

$$\left(\frac{d}{d\theta} E(T)\right)^2 \leq \operatorname{var}(T)\, i(\theta).$$

Väite (3.2a) seuraa tästä jakamalla puolittain $i(\theta)$:lla, sillä $E(T) = g(\theta) + b(\theta)$. Väite (3.2b) saadaan erikoistapauksena $b(\theta) = 0$ ja (3.2c) valitsemalla lisäksi $g(\theta) = \theta$. □

Esitetty lause on varsin voimakas tulos, ja samalla se ilmaisee Fisherin informaation syvällisen merkityksen piste-estimoinnin teoriassa. Mitä vähemmän informaatiota on, sitä vaikeampaa parametrin estimointi on eli sitä suurempi käytettävän estimaattorin varianssi pakosti on, ainakin kun malli on säännöllinen. Epäyhtälöitä (3.2) kutsutaan *informaatioepäyhtälöiksi*, vanhemmassa kirjallisuudessa usein *Cramér–Rao-epäyhtälöiksi*. Tärkein ja samalla helpoin muistaa on parametrin θ harhattomia estimaattoreita koskeva (3.2c).

Sellaista harhatonta estimaattoria, jonka varianssi saavuttaa informaatioepäyhtälön antaman alarajan jokaisessa pisteessä θ, sanotaan *täystehokkaaksi*. Yleisesti harhattoman estimaattorin T *tehokkuus* voidaan määritellä vertaamalla estimaattorin varianssia vastaavaan alarajaan, esim. prosenttilukuna

$$T\text{:n tehokkuus} = \frac{g'(\theta)^2/i(\theta)}{\mathrm{var}_\theta(T)} \cdot 100\ \%,$$

joka voi riippua θ:sta. Tämä käsite on varsinaisesti mielekäs vain silloin kun malli on säännöllinen.

Informaatioepäyhtälöllä on seuraava käytännön merkitys: jos säännöllisyysehdot täyttävässä mallissa on löydetty estimaattori T, joka on funktion $g(\theta)$ harhaton estimaattori ja täystehokas eli jolle $E_\theta(T) = g(\theta)$ ja $\mathrm{var}_\theta(T) = g'(\theta)^2/i(\theta)$ kaikilla θ, niin voidaan olla varmoja siitä, että T on $g(\theta)$:n *paras* (so. tehokkain mahdollinen) *harhaton estimaattori*. Tämän päättelyn sovellettavuutta vähentää kuitenkin se, että läheskään kaikissa estimointitehtävissä ei täystehokkaita estimaattoreita ole lainkaan olemassa.

3.4.4 Esimerkki: normaalimallin odotusarvo. Edellä esimerkissä 3.4.2 pohdittiin parametrin μ estimointia mallissa $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ ⫫. Luonnollinen harhaton estimaattori on $\hat{\mu} = \overline{Y}$, jonka varianssi on σ_0^2/n. Koska $i(\mu) = n/\sigma_0^2$ (mikä näkyy esimerkistä 2.4.3 tai 2.6.3), nähdään, että $\hat{\mu}$ saavuttaa informaatioepäyhtälön (3.2c) antaman alarajan eli on täystehokas. Koska kyseessä on säännöllinen malli (todistus sivuutetaan), $\hat{\mu}$ on siis μ:n paras harhaton estimaattori.

3.4.5 Esimerkki: eksponenttimalli. Tarkastellaan mallia $Y_1, \ldots, Y_n \sim \mathit{Exp}(\lambda)$ ja oletetaan, että estimoitavana on odotusarvo $\mu = 1/\lambda$. Su-estimaattori on $\hat{\mu} = \overline{Y}$, joka on myös harhaton (ks. 3.2.2 ja 3.2.5). Koska $\mathit{Exp}(\lambda)$-jakauman varianssi on $1/\lambda^2$, niin

$$\mathrm{var}(\hat{\mu}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(Y_i) = \frac{1}{n\lambda^2}.$$

Tehtävässä 2.11 on laskettu tämän mallin Fisherin informaatio $i(\lambda) = n/\lambda^2$. Laskemalla epäyhtälön (3.2b) oikea puoli tapauksessa $g(\lambda) = 1/\lambda$ saadaan siten varianssin alarajaksi

$$\frac{g'(\lambda)^2}{i(\lambda)} = \frac{(-1/\lambda^2)^2}{n/\lambda^2} = \frac{1}{n\lambda^2}.$$

Siispä $\hat{\mu}$ on täystehokas ja sen johdosta paras (eli varianssiltaan pienin mahdollinen) harhaton estimaattori parametrille $\mu = 1/\lambda$.

3.4.6 Informaatioepäyhtälö tapauksessa $d > 1$. Tarkastellaan nyt mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jonka parametri on vektori $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Oletetaan, että estimoitavana on jokin reaaliarvoinen funktio $g(\boldsymbol{\theta})$ parametrista $\boldsymbol{\theta}$, esimerkiksi jokin $\boldsymbol{\theta}$:n komponentti.

Palautetaan mieleen pykälästä 2.6, että Fisherin informaatio on $d \times d$ -matriisi $\boldsymbol{i}(\boldsymbol{\theta})$, jonka alkiot ovat

$$i_{a,b}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2}{\partial\theta_a \partial\theta_b} l(\boldsymbol{\theta}; \boldsymbol{Y})\right].$$

Reaaliparametrisen mallin informaatioepäyhtälöt yleistyvät varsin suoraviivaisesti tähän tilanteeseen, kunhan vain informaatiolla $i(\theta)$ jakaminen korvataan informaatiomatriisin $\boldsymbol{i}(\boldsymbol{\theta})$ käänteismatriisilla $\boldsymbol{i}^{-1}(\boldsymbol{\theta})$ kertomisella. Merkitään käänteismatriisin

alkioita tarpeen vaatiessa yläindeksoiduilla symboleilla $i^{a,b}(\boldsymbol{\theta})$, $a, b = 1, \ldots, d$. Siis

$$\boldsymbol{i}^{-1}(\boldsymbol{\theta}) = \boldsymbol{i}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} i^{1,1}(\boldsymbol{\theta}) & \cdots & i^{1,d}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ i^{d,1}(\boldsymbol{\theta}) & \cdots & i^{d,d}(\boldsymbol{\theta}) \end{bmatrix}.$$

Seuraavassa ∇ viittaa funktion gradienttiin eli osittaisderivaattojen muodostamaan vektoriin, esimerkiksi

$$\nabla g(\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial \theta_1} g(\boldsymbol{\theta}), \ldots, \frac{\partial}{\partial \theta_d} g(\boldsymbol{\theta}) \Big).$$

Matriisilaskuissa tämä ajatellaan pystyvektoriksi eli $d \times 1$ -matriisiksi. Vektorin transpoosia merkitään yläpilkulla.

Lause. *Oletetaan, että $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ on säännöllinen malli. Olkoon $T = t(\boldsymbol{Y})$ jokin funktion $g(\boldsymbol{\theta})$ estimaattori.*

Jos estimaattorin T harha on $b(\boldsymbol{\theta})$, niin

$$\operatorname{var}_{\boldsymbol{\theta}}(T) \geq [\nabla g(\boldsymbol{\theta}) + \nabla b(\boldsymbol{\theta})]' \, \boldsymbol{i}^{-1}(\boldsymbol{\theta}) \, [\nabla g(\boldsymbol{\theta}) + \nabla b(\boldsymbol{\theta})]. \tag{3.3a}$$

Jos T on $g(\boldsymbol{\theta})$:n harhaton estimaattori, pätee

$$\operatorname{var}_{\boldsymbol{\theta}}(T) \geq [\nabla g(\boldsymbol{\theta})]' \, \boldsymbol{i}^{-1}(\boldsymbol{\theta}) \, [\nabla g(\boldsymbol{\theta})], \tag{3.3b}$$

ja erityisesti jos T on θ_a:n harhaton estimaattori,

$$\operatorname{var}_{\boldsymbol{\theta}}(T) \geq i^{a,a}(\boldsymbol{\theta}). \tag{3.3c}$$

Todistus on perusidealtaan samanlainen kuin tapauksessa $d = 1$, ja se sivuutetaan. On syytä huomata, että epäyhtälön (3.3c) antama alaraja on yleensä erisuuri kuin $1/i_{a,a}(\boldsymbol{\theta})$. Itse asiassa voidaan osoittaa, että

$$i^{a,a}(\boldsymbol{\theta}) \geq \frac{1}{i_{a,a}(\boldsymbol{\theta})}$$

ja yhtäsuuruus pätee vain siinä tapauksessa, että θ_a on ortogonaalinen kaikkiin muihin parametrivektorin komponentteihin θ_b $(b \neq a)$ nähden (ks. 2.6.2). Luku $1/i_{a,a}(\boldsymbol{\theta})$ on se varianssin alaraja, joka saataisiin epäyhtälöstä (3.2c) siinä tapauksessa, että mallin tuntematon parametri olisi yksiulotteinen θ_a ja muut komponentit θ_b ajateltaisiin tunnetuiksi vakioiksi. Epäyhtälöstä (3.3c) siis nähdään, että varianssin alaraja kasvaa eli estimointiin liittyvä epävarmuus lisääntyy, kun malliin otetaan θ_a:n lisäksi muitakin tuntemattomia parametreja, paitsi jos nämä ovat ortogonaalisia θ_a:han nähden.

Harhattoman estimaattorin tehokkuuden ja täystehokkuuden käsitteet määritellään aivan samoin kuin edellä tapauksessa $d = 1$ vertaamalla estimaattorin varianssia yo. lauseen antamaan alarajaan.

3.4.7 Esimerkki: normaalimalli. Tarkastellaan mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa sekä μ että σ^2 ovat tuntemattomia parametreja. Esimerkissä 2.6.3 laskettiin tämän mallin Fisherin informaatio

$$\boldsymbol{i}(\mu, \sigma^2) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}.$$

Sen käänteismatriisi on

$$\boldsymbol{i}^{-1}(\mu, \sigma^2) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

Epäyhtälön (3.3c) antamat varianssin alarajat μ:n ja σ^2:n harhattomille estimaattoreille ovat $i^{1,1}(\mu, \sigma^2) = \sigma^2/n$ ja vastaavasti $i^{2,2}(\mu, \sigma^2) = 2\sigma^4/n$.

Su-estimaattori $\hat{\mu} = \overline{Y}$ on μ:n harhaton estimaattori ja sen varianssi on

$$\operatorname{var}(\hat{\mu}) = \frac{\sigma^2}{n},$$

joten se on täystehokas. Kuten on aiemminkin todettu, varianssin σ^2 harhattomana estimaattorina käytetään yleisesti muuttujaa

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

Sille pätee (ks. teht. 3.5)

$$\operatorname{var}(S^2) = \frac{2\sigma^4}{n-1},$$

joten

$$S^2\text{:n tehokkuus} = \frac{2\sigma^4}{n} : \frac{2\sigma^4}{n-1} = \frac{n-1}{n} \cdot 100\ \%.$$

Voidaan kuitenkin osoittaa, että S^2 on paras (eli varianssiltaan pienin) harhaton estimaattori σ^2:lle.

3.4.8 Informaatioepäyhtälön toinen muotoilu tapauksessa $d > 1$. Toisinaan informaatioepäyhtälö säännöllisen vektoriparametrisen mallin tapauksessa esitetään niin, että siinä verrataan parametrivektorin estimaattorin kovarianssimatriisia Fisherin informaation käänteismatriisiin.

Olkoon mallin parametrina vektori $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ kuten edellä, ja oletetaan yksinkertaisuuden vuoksi, että $\boldsymbol{T} = (T_1, \ldots, T_d)$ on sen harhaton estimaattori. Palautetaan mieleen, että satunnaisvektorin $\boldsymbol{T}$ kovarianssimatriisilla $\operatorname{cov}(\boldsymbol{T})$ tarkoitetaan symmetristä $d \times d$ -matriisia $(\sigma_{a,b})$, jossa $\sigma_{a,b} = \operatorname{cov}(T_a, T_b)$. Olkoon $\boldsymbol{a} = (a_1, \ldots, a_d)$ jokin reaalisista vakioista koostuva vektori, ja tarkastellaan satunnaismuuttujaa

$$\boldsymbol{a}'\boldsymbol{T} = a_1 T_1 + \cdots + a_d T_d.$$

Sen odotusarvo on

$$E(\boldsymbol{a}'\boldsymbol{T}) = \boldsymbol{a}'\boldsymbol{\theta} = a_1\theta_1 + \cdots + a_d\theta_d$$

ja varianssi (perustelu jätetään harjoitustehtäväksi)

$$\operatorname{var}(\boldsymbol{a}'\boldsymbol{T}) = \boldsymbol{a}' \operatorname{cov}(\boldsymbol{T})\boldsymbol{a}.$$

Siis $\boldsymbol{a}'\boldsymbol{T}$ on funktion $\boldsymbol{a}'\boldsymbol{\theta}$ harhaton estimaattori. Tämän funktion gradientti on $\boldsymbol{a}$, joten epäyhtälöstä (3.3b) saadaan

$$\boldsymbol{a}' \operatorname{cov}(\boldsymbol{T})\boldsymbol{a} \geq \boldsymbol{a}'\boldsymbol{i}^{-1}(\boldsymbol{\theta})\boldsymbol{a}$$

eli

$$\boldsymbol{a}'[\operatorname{cov}(\boldsymbol{T}) - \boldsymbol{i}^{-1}(\boldsymbol{\theta})]\boldsymbol{a} \geq 0.$$

Huomaa, että tämä pätee kaikilla mahdollisilla d-ulotteisilla vektoreilla $\boldsymbol{a}$. Matriisilaskennan käsitteitä käyttäen voidaan sanoa, että erotusmatriisi $\operatorname{cov}(\boldsymbol{T}) - \boldsymbol{i}^{-1}(\boldsymbol{\theta})$ on *positiivisesti semidefiniitti*, mitä merkitään

$$\operatorname{cov}(\boldsymbol{T}) - \boldsymbol{i}^{-1}(\boldsymbol{\theta}) \geq 0. \tag{3.4}$$

Tämä on ilmeinen moniulotteinen vastine epäyhtälölle (3.2c).

3.5 Tarkentuvuus

Tässä ja seuraavassa pykälässä puhutaan estimaattorien asymptoottisista ominaisuuksista. Sana ”asymptotiikka” viittaa siihen, miten estimaattori tai jokin muu tilastollinen menetelmä käyttäytyy, kun havaintojen lukumäärä kasvaa rajatta. Asymptoottisten tulosten käytännön merkitys on siinä, että niiden voidaan ajatella toteutuvan likimäärin, kun havaintoja on tarpeeksi paljon – mitä tämä sitten tarkoittaakaan.

3.5.1 Tarkentuvuuden määritelmä. Tarkastellaan jälleen mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ ja siihen liittyvää estimointitehtävää, jossa estimoitavana on mallin parametrin jokin reaaliarvoinen funktio $g(\boldsymbol{\theta})$. Olkoon $T^{(n)}$ kokoa n olevaan aineistoon $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ perustuva estimaattori:

$$T^{(n)} = t(Y_1, \ldots, Y_n).$$

Estimaattoria $T^{(n)}$ – tai oikeammin jonoa $(T^{(1)}, T^{(2)}, \ldots)$ – sanotaan *tarkentuvaksi*, mikäli se suppenee stokastisesti kohti estimoitavan todellista arvoa eli lukua $g(\boldsymbol{\theta})$, ts.

$$\lim_{n\to\infty} P\{|T^{(n)} - g(\boldsymbol{\theta})| > \varepsilon\} = 0 \qquad \text{kaikilla } \varepsilon > 0\,. \tag{3.5}$$

Tällöin merkitään $T^{(n)} \xrightarrow{P} g(\boldsymbol{\theta})$.

Monesti jätetään aineiston kokoon viittaava indeksi n merkitsemättä estimaattorin symboliin (kuten vaikkapa esimerkissä 3.5.2 alla); tarkentuvuudesta tai muista asymptoottisista ominaisuuksista puhuttaessa on kuitenkin aina muistettava, että tarkasteltavana on oikeastaan kokonainen estimaattorijono eikä vain yksi ainoa estimaattori. Tarkentuvuus voidaan luonnollisesti määritellä myös siinä tapauksessa, että estimoitavana on parametrivektori $\boldsymbol{\theta}$ itse tai jokin sen vektoriarvoinen muunnos $\boldsymbol{g}(\boldsymbol{\theta})$. Tällöin vaaditaan tarkentuvuus komponenteittain: käytetyn estimaattorivektorin jokaisen komponentin tulee supeta stokastisesti kohti vastaavaa estimoitavan komponenttia.

Stokastinen suppeneminen kuvaa ”todennäköisyysmassan” suppenemista. Tarkentuvuusehto (3.5) merkitsee sitä, että mielivaltaisen pienetkin poikkeamat $|T^{(n)} - g(\boldsymbol{\theta})|$ tulevat yhä epätodennäköisemmiksi, kun havaintojen lukumäärä n kasvaa. Tämä on hyvin luonnollinen estimaattorille asetettava vaatimus, ja tarkentumattomuus viittaa yleensä ongelmaan mallin spesifioinnissa tai käytetyssä estimointimenetelmässä.

Tavallisesti tarkentuvuuden taustalla on jokin suurten lukujen laki. Seuraavassa on eräs todennäköisyyslaskennan kurssilla opittu perusversio siitä:

Heikko suurten lukujen laki. *Olkoot $X_1, \ldots, X_n$ riippumattomia satunnaismuuttujia, joilla on sama odotusarvo μ ja (äärellinen) varianssi σ^2. Tällöin*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu,$$

kun $n \to \infty$.

3.5.2 Esimerkki: otoskeskiarvo on tarkentuva. Olkoon $Y_1, \ldots, Y_n$ riippumaton otos jakaumasta, jonka odotusarvo on μ ja jolla on äärellinen varianssi tai yhtäpitävästi toinen momentti. Silloin heikon suurten lukujen lain nojalla

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} \mu.$$

Siis otoskeskiarvo $\overline{Y}$ on odotusarvon μ tarkentuva estimaattori. Muista, että $\overline{Y}$ on μ:n su-estimaattori, kun tarkasteltava jakauma on jokin seuraavista: $B(\mu)$, $P(\mu)$, $Exp(1/\mu)$ tai $N(\mu, \sigma^2)$. Esimerkissä 3.2.2 todettiin, että se on myös harhaton, jopa ilman riippumattomuusoletusta.

3.5.3 Eräs riittävä ehto tarkentuvuudelle. Tarkentuvuuden toteamiseen on usein mukava käyttää seuraavan lauseen antamaa kriteeriä. Lauseen todistus perustuu todennäköisyyslaskennassa opittuun Markovin epäyhtälöön: jos X on ei-negatiivinen satunnaismuuttuja, jolla on odotusarvo, niin

$$P\{X \geq x\} \leq \frac{E(X)}{x}, \qquad x > 0.$$

Lause. *Oletetaan, että $g(\boldsymbol{\theta})$:n estimaattorille $T^{(n)}$ pätee kaikilla $\boldsymbol{\theta}$:n arvoilla*

a) $\lim_{n\to\infty} E(T^{(n)}) = g(\boldsymbol{\theta})$ *(asymptoottinen harhattomuus)*

b) $\lim_{n\to\infty} \operatorname{var}(T^{(n)}) = 0$.

Tällöin $T^{(n)}$ on tarkentuva.

Todistus. Lähdetään yhtälöstä

$$E\big[(T^{(n)} - g(\boldsymbol{\theta}))^2\big] = \operatorname{var}(T^{(n)}) + \big[E(T^{(n)}) - g(\boldsymbol{\theta})\big]^2,$$

joka perustellaan tehtävässä 3.1. Oletuksista seuraa nyt, että tämä odotusarvo lähestyy nollaa, kun $n \to \infty$. Olkoon $\varepsilon > 0$, ja sovelletaan Markovin epäyhtälöä kyseiseen satunnaismuuttujaan $(T^{(n)} - g(\boldsymbol{\theta}))^2$:

$$P\big\{|T^{(n)} - g(\boldsymbol{\theta})| > \varepsilon\big\} = P\big\{(T^{(n)} - g(\boldsymbol{\theta}))^2 > \varepsilon^2\big\} \leq \frac{E\big[(T^{(n)} - g(\boldsymbol{\theta}))^2\big]}{\varepsilon^2}.$$

Äsken todetun nojalla tämä $\to 0$, kun $n \to \infty$. □

3.5.4 Esimerkki: normaalimallin varianssi. Malli olkoon $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ ⫫. Tarkastellaan varianssin σ^2 estimaattoria

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Osoitetaan yo. lauseen avulla, että se on tarkentuva. Koska S^2 on harhaton, ehto a on automaattisesti voimassa. Lisäksi $\operatorname{var}(S^2) = 2\sigma^4/(n-1) \to 0$, kun $n \to \infty$ (ks. teht. 3.5), joten myös b toteutuu.

Su-estimaattori

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

on yhtä lailla tarkentuva; tämän voi perustella yo. lauseen avulla tai jo siitä havainnosta, että $\hat{\sigma}^2 = [(n-1)/n]S^2$, jossa $(n-1)/n \to 1$.

3.6 Su-estimaattorien asymptotiikka

3.6.1 Tarkasteltava tilanne. Edellä on esitelty optimaalisuuskriteerejä, joita käytettävien estimaattorien toivotaan toteuttavan. Suurimman uskottavuuden menetelmän tuottamat estimaattorit eivät suinkaan aina ole tässä mielessä optimaalisia (esim. harhattomia tai täystehokkaita). Niin kuin tässä pykälässä todetaan, ne kuitenkin ovat useimmissa tapauksissa *asymptoottisilta* ominaisuuksiltaan varsin hyviä: tarkentuvia, asymptoottisesti harhattomia ja täystehokkaita sekä likimain normaalisti jakautuneita. Suurimman uskottavuuden menetelmä on osittain tästä syystä saavuttanut niin suuren suosion tilastotieteessä.

Tarkasteltava perustilanne on seuraava: Havaintoja vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomia ja samoin jakautuneita, pistetodennäköisyys- tai tiheysfunktionaan $f(y;\boldsymbol{\theta})$, jossa parametri $\boldsymbol{\theta}$ on d-ulotteinen. Tilastollinen malli on tällöin siis

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}),$$

ja sen log-uskottavuusfunktio on

$$l(\boldsymbol{\theta};\boldsymbol{y}) = \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i;\boldsymbol{\theta}). \tag{3.6}$$

Huomaa, että tähän viitekehykseen ei sovi esimerkiksi lineaarinen regressiomalli (ks. 1.2.4), koska siinä havainnot eivät ole samoin jakautuneita, eikä myöskään autoregressiivinen aikasarjamalli (ks. 1.2.5), koska siinä havainnot riippuvat toisistaan. Esitettävät tulokset yleistyvät kyllä näihinkin tilanteisiin, mutta jatkossa tarkastellaan vain yllä kuvattua yksinkertaista satunnaisotosta.

Merkitään parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattoria lyhyesti

$$\hat{\boldsymbol{\theta}}^{(n)} = \hat{\boldsymbol{\theta}}(Y_1, \ldots, Y_n).$$

Yläindeksi n viittaa tässä siihen, että estimaattori on muodostettu n:n havainnon perusteella, ja jatkossa pohditaan, mitä tapahtuu, kun $n \to \infty$.

3.6.2 Su-estimaattorin tarkentuvuus. Tarkastellaan edellä esiteltyä mallia. Tällöin on voimassa seuraava yleinen tulos:

Lause. *Riittävien säännöllisyysehtojen vallitessa estimaattori $\hat{\boldsymbol{\theta}}^{(n)}$ on tarkentuva.*

Tarvittavien säännöllisyysehtojen olemukseen emme syvenny tarkemmin. Lähinnä ne edellyttävät, että ptf/tf $f(y;\boldsymbol{\theta})$ on kyllin "siisti" funktio. Erityisesti vaaditaan, että jakauman alusta eli joukko $\{y : f(y;\boldsymbol{\theta}) > 0\}$ on riippumaton parametrista $\boldsymbol{\theta}$ (vrt. 2.5.2). Tämä sulkee pois esimerkiksi jakauman $Tas(0,\theta)$ (ks. 2.2.8).

Lisäksi yo. lauseessa on vaadittava, että parametri $\boldsymbol{\theta}$ on *identifioituva*: jos $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, niin $f(y;\boldsymbol{\theta})$ ja $f(y;\boldsymbol{\theta}')$ ovat eri jakaumien pistetodennäköisyys- tai tiheysfunktioita. Muussa tapauksessahan on olemassa kaksi eri parametriarvoa, jotka johtavat täysin samaan havaintojen yhteisjakaumaan, eikä tällöin voi havaintoaineiston perusteella tilastollisin menetelmin päätellä, kumpi parametriarvo on "oikeampi". Esimerkkinä epäidentifioituvasta parametroinnista voisi olla regressiomalli $Y_i \sim N(\alpha+\beta x_i, \sigma_0^2)$, $i = 1, \ldots, n$, siinä tapauksessa, että selittäjien x_i arvot ovat samat eli $x_1 = \cdots = x_n = c$; tällöin kunkin suoran $\alpha + c\beta =$ vakio pisteet (α, β) johtavat samaan jakaumaan.

Useimmat tällä kurssilla tarkasteltavat mallit toteuttavat lauseen oletukset. Tällaisia malleja ovat muun muassa ns. eksponenttiperheen jakaumista muodostetut mallit (vrt. teht. 2.20).

**Lauseen todistuksen idea.* Aluksi on selvitettävä eräs merkinnällinen ongelma. Symbolia $\boldsymbol{\theta}$ on tällä kurssilla käytetty nimittäin kahdessa merkityksessä: yhtäältä merkitsemään sitä ”todellista” parametrinarvoa, joka on tuottanut havainnot, ja toisaalta, esim. uskottavuusfunktiota ja sen johdannaisia tarkasteltaessa, ”vapaana muuttujana” parametriavaruudessa. Tässä todistuksessa on erotettava nämä merkitykset toisistaan. Merkitään ”todellista” arvoa siksi $\boldsymbol{\theta}_0$:lla. Sen määrittelemää jakaumaa $f(y;\boldsymbol{\theta}_0)$ havainnot siis noudattavat, ja sitä kohti estimaattorien $\hat{\boldsymbol{\theta}}^{(n)}$ halutaan suppenevan.

Palautetaan mieleen, että havaitusta aineistosta $\boldsymbol{y} = (y_1, \ldots, y_n)$ laskettava suestimaatti $\hat{\boldsymbol{\theta}}^{(n)} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ on se parametriavaruuden piste, jossa log-uskottavuusfunktio (3.6) saa suurimman arvonsa. Yhtäpitävästi kyseessä on se piste, jossa funktio

$$\frac{1}{n} l(\boldsymbol{\theta}; \boldsymbol{y}) = \frac{1}{n} \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}) \tag{3.7}$$

maksimoituu. Korvataan tässä lausekkeessa kukin y_i vastaavalla satunnaismuuttujalla Y_i. Tällöin saadaan aritmeettinen keskiarvo muuttujista $\log f(Y_i; \boldsymbol{\theta})$, jotka oletusten perusteella ovat samoin jakautuneita ja riippumattomia. Heikon suurten lukujen lain (ks. 3.5.1) nojalla tämä keskiarvo suppenee stokastisesti kohti odotusarvoa

$$z(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}[\log f(Y_1; \boldsymbol{\theta})],$$

kun $n \to \infty$. Huomaa, että tämä raja-arvo riippuu parametrista $\boldsymbol{\theta}$ mutta odotusarvon laskennassa käytetään ”todellista” parametriarvoa $\boldsymbol{\theta}_0$.

Voidaan osoittaa, että funktio $z(\boldsymbol{\theta})$ saa suurimman arvonsa täsmälleen pisteessä $\boldsymbol{\theta}_0$ (ks. teht. 3.14). Todistuksen loppuosa muodostuu nyt seuraavanlaisesta järkeilystä: koska funktiot (3.7) lähestyvät stokastisesti funktiota $z(\boldsymbol{\theta})$, niin myös niiden globaalit maksimikohdat $\hat{\boldsymbol{\theta}}^{(n)}$ lähestyvät stokastisesti $z(\boldsymbol{\theta})$:n maksimikohtaa $\boldsymbol{\theta}_0$. Tämän seikan täsmällinen todistus on jonkin verran tekninen ja vaatii oletettujen säännöllisyysominaisuuksien käyttöä; sivuutamme sen. □

3.6.3 Estimaattorin jakauma. Unohdetaan hetkeksi asymptoottinen näkökulma ja tarkastellaan estimointiongelmaa, jossa on annettu malli $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ (yksinkertaisuuden vuoksi reaaliparametrinen). Olkoon lisäksi T jokin θ:n estimaattori. Edellä pykälissä 3.2 ja 3.4 on esitetty vaatimuksia koskien T:n ensimmäistä ja toista momenttia: harhattomuus merkitsee, että T:n odotusarvon tulisi olla θ, ja tehokkuus merkitsee, että T:n varianssin tulisi olla mahdollisimman pieni. Monia tarkoituksia varten olisi kuitenkin hyödyllistä tuntea estimaattorin T jakauma täydellisesti. Tämä mahdollistaisi parametriin θ liittyvien luottamusväliarvioiden esittämisen ja hypoteesien testauksen, joihin paneudutaan kurssin loppupuolella.

Esimerkki: Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jossa μ on estimoitava ja $\sigma_0^2 > 0$ on tunnettu luku. Tällöin estimaattori $\hat{\mu} = \overline{Y} = (Y_1 + \cdots + Y_n)/n$ on harhaton ja täystehokas, mutta todennäköisyyslaskennan tulosten perusteella tiedetään myös, että se on normaalisti jakautunut: $\hat{\mu} \sim N(\mu, \sigma_0^2/n)$. Tällöin voidaan normaalijakauman taulukoista etsiä esimerkiksi luku $a > 0$ siten, että

$$P\{|\hat{\mu} - \mu| < a\} = 0.95,$$

jolloin väliä $(\hat{\mu} - a, \hat{\mu} + a)$ tullaan kutsumaan 95 %:n luottamusväliksi parametrille μ.

Yleensä monimutkaisempien mallien tapauksessa estimaattorien jakaumaa ei ole mahdollista täysin selvittää. Vaikka aineiston $\boldsymbol{Y}$ jakauma onkin tunnettu, esimerkiksi suurimman uskottavuuden estimaattori $\hat{\theta}(\boldsymbol{Y})$ voi riippua $\boldsymbol{Y}$:stä niin monimutkaisella tavalla, että sen jakaumaa ei täysin tunneta. Monesti ei ole edes mahdollista lausua $\hat{\theta}(\boldsymbol{Y})$:tä suljetussa muodossa olevana lausekkeena aineistosta; tällainen esimerkki saadaan vaikkapa gammajakaumaan perustuvasta mallista (ks. teht. 2.8). Siksi joudutaankin turvautumaan asymptoottisiin tuloksiin: osoittautuu, että varsin yleisin ehdoin su-estimaattorit ovat asymptoottisesti normaalisia.

3.6.4 Keskeinen raja-arvolause ja heikko suppeneminen. Asymptoottinen normaalisuus perustuu todennäköisyyslaskennan keskeiseen raja-arvolauseeseen. Tämän perusversio on seuraava:

Keskeinen raja-arvolause. *Olkoot $X_1, \ldots, X_n$ riippumattomia samoin jakautuneita satunnaismuuttujia, joilla on odotusarvo μ ja (äärellinen) varianssi σ^2. Tällöin*

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{w} N(0,1),$$

kun $n \to \infty$.

Tässä $\xrightarrow{w}$ viittaa ns. heikkoon eli jakaumasuppenemiseen. Yo. raja-arvotulos merkitsee siis määritelmän mukaan sitä, että muuttujien Z_n kertymäfunktioille F_{Z_n} pätee

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z) \qquad \text{kaikilla } z \in \mathbb{R},$$

jossa Φ on standardinormaalijakauman $N(0,1)$ kertymäfunktio. Käytännössä tämä merkitsee, että suurilla n:n arvoilla muuttujan Z_n jakauma on likimain standardinormaalijakauma. Vastaavasti tällöin keskiarvo $\sum_{i=1}^{n} X_i/n$ on likimain $N(\mu, \sigma^2/n)$-jakautunut, mitä merkitään lyhyesti

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{\text{as}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right).$$

3.6.5 Su-estimaattorin asymptoottinen normaalisuus, kun $d = 1$. Palataan kohdassa 3.6.1 esiteltyyn tilanteeseen ja oletetaan aluksi, että mallin parametri on yksiulotteinen θ.

Lause. *Riittävien säännöllisyysehtojen vallitessa*

$$\hat{\theta}^{(n)} \underset{\text{as}}{\sim} N\left(\theta, \frac{1}{i(\theta)}\right). \tag{3.8}$$

Tässä $i(\theta)$ on Fisherin informaatio n havainnon mallissa. Koska havainnot ovat riippumattomia ja samoin jakautuneita, pätee $i(\theta) = n i_1(\theta)$, jossa $i_1(\theta)$ on yhden havainnon antama Fisherin informaatio (ks. teht. 2.13). Heikon suppenemisen avulla ilmaistuna lause toteaa, että

$$\sqrt{i(\theta)}\big(\hat{\theta}^{(n)} - \theta\big) \xrightarrow{w} N(0,1),$$

kun $n \to \infty$.

Suurissa otoksissa estimaattori $\hat{\theta}^{(n)}$ noudattaa siis likimain normaalijakaumaa, jonka odotusarvo on θ ja varianssi $1/i(\theta)$ eli informaatioepäyhtälöstä saatava alaraja

(ks. 3.4.3). Tässä mielessä voidaan siis sanoa, että kyllin säännöllisen mallin tapauksessa su-estimaattorit ovat asymptoottisesti harhattomia ja täystehokkaita, mikä on tietysti varsin rohkaiseva tulos. Lauseen käytännön merkitystä pohdittaessa on kuitenkin otettava huomioon, että ajatus havaintojen lukumäärän kasvattamisesta rajatta on täysin vailla todellisuuspohjaa eikä voida antaa mitään yleisohjetta siitä, milloin havaintoja on tarpeeksi paljon, jotta approksimatiivinen jakaumatulos (3.8) olisi kyllin tarkka. Lisäksi on muistettava, että puhe estimaattorista satunnaismuuttujana ja sen jakaumasta viittaa toistetun aineistonkeruun ajatukseen, joka sekin on usein täysin hypoteettinen. Silti monet käytännön tilastotieteessä käytettävät menetelmät pohjautuvat tulokseen (3.8), ja joitakin niistä käsitellään myöhemmin tällä kurssilla luottamusjoukkojen ja testiteorian yhteydessä.

**Lauseen todistuksen idea.* Merkitään jälleen selvyyden vuoksi ”todellista” parametriarvoa symbolilla θ_0. Oletetaan mallin log-uskottavuusfunktio $l(\theta; \boldsymbol{y})$ niin säännölliseksi, että sen derivaatta voidaan esittää θ_0:n ympäristössä Taylorin kehitelmänä

$$l'(\theta; \boldsymbol{y}) = l'(\theta_0; \boldsymbol{y}) + l''(\theta_0; \boldsymbol{y})(\theta - \theta_0) + \tfrac{1}{2} l'''(\theta^*; \boldsymbol{y})(\theta - \theta_0)^2,$$

jossa θ^* on θ_0:n ja θ:n välissä oleva piste.

Siirrytään tässä aineistoa vastaavaan satunnaisvektoriin $\boldsymbol{Y}$ ja sijoitetaan θ:n paikalle su-estimaattori $\hat{\theta}^{(n)}$. Viimeinen termi saa tällöin muodon $\frac{1}{2} l'''(\theta^*; \boldsymbol{Y})(\hat{\theta}^{(n)} - \theta_0)^2$. Koska $\hat{\theta}^{(n)}$ on tarkentuva, voidaan näyttää, että tämä termi on suurilla n:n arvoilla häviävän pieni suhteessa summan muihin termeihin. Näin päädytään arvioon

$$l'(\hat{\theta}^{(n)}; \boldsymbol{Y}) \approx l'(\theta_0; \boldsymbol{Y}) + l''(\theta_0; \boldsymbol{Y})(\hat{\theta}^{(n)} - \theta_0).$$

Oletetaan, että su-estimaattori saadaan uskottavuusyhtälön ratkaisuna (eli log-uskottavuusfunktion derivaatan nollakohtana), jolloin tässä vasen puoli on nolla. Siispä saadaan

$$\hat{\theta}^{(n)} - \theta_0 \approx \frac{l'(\theta_0; \boldsymbol{Y})}{-l''(\theta_0; \boldsymbol{Y})}$$

ja edelleen

$$\sqrt{n}(\hat{\theta}^{(n)} - \theta_0) \approx \frac{l'(\theta_0; \boldsymbol{Y})/\sqrt{n}}{-l''(\theta_0; \boldsymbol{Y})/n}. \tag{3.9}$$

Tarkastellaan tässä osoittajaa. Derivoimalla yhtälö (3.6) puolittain nähdään, että

$$\frac{1}{\sqrt{n}} l'(\theta_0; \boldsymbol{Y}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_1'(\theta_0; Y_i) = \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} l_1'(\theta_0; Y_i),$$

jossa $l_1(\theta; y) = \log f(y; \theta)$ on yhteen havaintoon perustuva log-uskottavuusfunktio. Sovelletaan tähän esitykseen keskeistä raja-arvolausetta. Satunnaismuuttujat $l_1'(\theta_0; Y_i)$ ovat riippumattomia sekä samoin jakautuneita, ja kohdan 2.5.3 apulauseen nojalla kunkin odotusarvo on 0 ja varianssi $i_1(\theta_0)$ (yhden havainnon Fisherin informaatio). Siten (3.9):n osoittaja $\xrightarrow{w} N(0, i_1(\theta_0))$.

Tarkastellaan sitten (3.9):n nimittäjää

$$-\frac{1}{n} l''(\theta_0; \boldsymbol{Y}) = \frac{1}{n} \sum_{i=1}^{n} -l_1''(\theta_0; Y_i).$$

Suurten lukujen lain heikon muodon perusteella tämä suppenee stokastisesti kohti lukua $E[-l_1''(\theta_0; Y_1)] = i_1(\theta_0)$. Yhdistämällä saadut tulokset voidaan päätellä, että

$$\sqrt{n}(\hat{\theta}^{(n)} - \theta_0) \xrightarrow{w} N\left(0, \frac{1}{i_1(\theta_0)}\right)$$

eli

$$\hat{\theta}^{(n)} \underset{\text{as}}{\sim} N\left(\theta_0, \frac{1}{n i_1(\theta_0)}\right),$$

jossa $n i_1(\theta_0) = i(\theta_0)$ on koko mallin Fisherin informaatio. □

3.6.6 Esimerkki: eksponenttimalli. Olkoot $Y_1, \ldots, Y_n \sim Exp(\lambda)$ ⫫. Tehtävässä 2.3 on todettu, että $\hat{\lambda} = 1/\overline{Y}$, jossa $\overline{Y}$ on otoskeskiarvo. Lisäksi tehtävässä 2.11 on laskettu $i(\lambda) = n/\lambda^2$. Siten $\hat{\lambda} \underset{\text{as}}{\sim} N(\lambda, \lambda^2/n)$.

3.6.7 Pistemäärän asymptoottinen jakauma. Edellisen lauseen todistuksen yhteydessä nähtiin, että $l'(\theta_0; \boldsymbol{Y})/\sqrt{n} \xrightarrow{w} N(0, i_1(\theta_0))$, kun $n \to \infty$. Näin ollen riittävän säännöllisen mallin tapauksessa on voimassa

$$l'(\theta_0; \boldsymbol{Y}) \underset{\text{as}}{\sim} N(0, i(\theta_0)).$$

(Jälleen käytettiin havaintoa $n i_1(\theta_0) = i(\theta_0)$.) Tälläkin tuloksella on käyttöä myöhemmin testiteorian yhteydessä.

3.6.8 Su-estimaattorin asymptoottinen normaalisuus, kun $d > 1$. Tarkastellaan kohdan 3.6.1 mallia siinä tapauksessa, että parametri $\boldsymbol{\theta}$ samoin kuin estimaattori $\hat{\boldsymbol{\theta}}^{(n)}$ ovat d-ulotteisia vektoreita, $d > 1$. Yleistämällä edellisen lauseen todistus sopivasti (käyttämällä mm. keskeisen raja-arvolauseen vektoriversiota) saadaan seuraava tulos.

Lause. *Riittävien säännöllisyysehtojen vallitessa*

$$\hat{\boldsymbol{\theta}}^{(n)} \underset{\text{as}}{\sim} N_d(\boldsymbol{\theta}, \boldsymbol{i}^{-1}(\boldsymbol{\theta})).$$

Suurilla n:n arvoilla $\hat{\boldsymbol{\theta}}^{(n)}$ noudattaa siis likimain d-ulotteista normaalijakaumaa, jonka odotusarvovektori on parametrin todellinen arvo ja kovarianssimatriisi Fisherin informaation käänteismatriisi.†

Harjoitustehtäviä

3.1. Johda funktion $g(\boldsymbol{\theta})$ estimaattorin T keskineliövirheelle kohdassa 3.4.1 mainittu hajotelma

$$E_{\boldsymbol{\theta}}[(T - g(\boldsymbol{\theta}))^2] = \text{var}_{\boldsymbol{\theta}}(T) + b(\boldsymbol{\theta})^2,$$

jossa $b(\boldsymbol{\theta})$ on estimaattorin harha.

3.2. Tarkastellaan Poisson-mallia $Y_1, \ldots, Y_n \sim P(\mu)$ ⫫. Varmista, että su-estimaattori $\hat{\mu} = \overline{Y} = (Y_1 + \cdots + Y_n)/n$ on harhaton, ja laske sen varianssi. Onko $\hat{\mu}$ täystehokas?

† Palautetaan todennäköisyyslaskennasta mieleen, että *d-ulotteisen normaalijakauman* $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ määrittelevät *odotusarvovektori* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ ja *kovarianssimatriisi* $\boldsymbol{\Sigma} = (\sigma_{ij})$, joka on ei-negatiivisesti definiitti $d \times d$ -matriisi. Jos $(X_1, \ldots, X_d) \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, niin erityisesti $X_i \sim N(\mu_i, \sigma_{ii})$ ja $\text{cov}(X_i, X_j) = \sigma_{ij}$, kun $i, j = 1, \ldots, d$.

3.3. Oletetaan, että havainnot $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat jakaumaa, jolla on odotusarvo μ ja varianssi σ^2. Tarkastellaan parametrin σ^2 estimointia muotoa cV olevilla estimaattoreilla, kun $c > 0$ on vakio ja $V = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$. Näytä, että cV on harhaton jos ja vain jos $c = 1/(n-1)$. *Ehdotus.* Käytä tehtävän 2.2 hajotelmaa valinnalla $a = \mu$.

3.4. Havainnoista $Y_1, \ldots, Y_n$ oletetaan samoin kuin tehtävässä 3.3. Keksi jokin harhaton estimaattori odotusarvon neliölle μ^2.

3.5. Palautetaan todennäköisyyslaskennasta mieleen, että jos $X_1, \ldots, X_k \sim N(0,1)$ $\perp\!\!\!\perp$, niin $X_1^2 + \cdots + X_k^2$ noudattaa jakaumaa χ_k^2. Tarkista tämän esityksen avulla, että χ_k^2-jakauman odotusarvo on k, varianssi $2k$ ja toinen momentti $k^2 + 2k$. Muista, että standardinormaalijakauman neljäs momentti on 3.

Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, ja olkoot c sekä V kuten tehtävässä 3.3. Todennäköisyyslaskennassa on opittu, että $V/\sigma^2 \sim \chi_{n-1}^2$. Näytä tämän tiedon ja edellisen kohdan perusteella, että estimaattorin cV keskineliövirhe on

$$E[(cV - \sigma^2)^2] = [(n^2-1)c^2 - 2(n-1)c + 1]\,\sigma^4$$

ja se saa pienimmän arvonsa (c:n funktiona) täsmälleen pisteessä $c = 1/(n+1)$.

Varianssin σ^2 estimaattoriksi on normaalijakaumamallissa siis ainakin kolme muotoa cV olevaa hyvää kandidaattia: $V/(n-1)$ (harhaton), V/n (su) ja $V/(n+1)$ (pienin keskineliövirhe).

3.6. Olkoot $x_1, x_2, \ldots$ annettuja nollasta eroavia reaalilukuja. Tarkastellaan regressiomallia $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$, $Y_i \sim N(\beta x_i, \sigma^2)$, jossa β ja σ^2 ovat tuntemattomia parametreja.

a) Varmista, että parametrin β su-estimaattori on

$$\hat{\beta} = \sum_{i=1}^{n} x_i Y_i \Big/ \sum_{i=1}^{n} x_i^2,$$

totea, että se on harhaton, ja laske sen varianssi.

b) Totea, että myös estimaattori

$$T = \sum_{i=1}^{n} Y_i \Big/ \sum_{i=1}^{n} x_i$$

on harhaton, ja laske sen varianssi.

c) Kumpi estimaattori on tehokkaampi?

Ohje. Epäyhtälö $\left(\sum_{i=1}^{n} x_i\right)^2 \le n \sum_{i=1}^{n} x_i^2$ lienee hyödyksi. Siinä yhtäsuuruus pätee vain jos kaikki x_i:t ovat samoja.

3.7. Jatkoa tehtävään 3.6. Oletetaan, että $x_i \ge c$ kaikilla i, jossa $c > 0$ on vakio. Näytä, että $\hat{\beta}$ ja T ovat tarkentuvia.

3.8. Tarkastellaan mallia $K \sim Bin(n, \theta)$, jossa $0 < \theta < 1$. Estimoitavana on θ:n käänteisluku $1/\theta$ (oleellisesti tällainen tarve esiintyi esimerkissä 2.3.3). Osoita, ettei harhattomia estimaattoreita ole olemassa. *Ohje.* Vastaoletus: on olemassa $T = t(K)$, jolle $E(T) = 1/\theta$. Johda ristiriita, kun $\theta \to 0+$.

3.9. Erään elektronisen komponentin kestoikä noudattaa eksponenttijakaumaa, jonka odotusarvo on θ/t, jossa $t > 0$ on komponentin käyttölämpötila ja $\theta > 0$ on tuntematon parametri. Parametrin θ estimoimiseksi testataan n komponenttia toisistaan riippumattomasti lämpötiloissa $t_1, \ldots, t_n$ ja mitataan niiden kestoiät $Y_1, \ldots, Y_n$. Osoita, että

$$T = \sum_{i=1}^{n} Y_i \Big/ \sum_{i=1}^{n} \frac{1}{t_i}$$

on θ:n harhaton mutta ei täystehokas estimaattori.

3.10. Mallissa $Y_1, \ldots, Y_n \sim Tas(0, \theta)$ $\perp\!\!\!\perp$ on su-estimaattoriksi saatu $\hat{\theta} = \max(Y_1, \ldots, Y_n)$ (ks. 2.2.8).

a) Muodosta $\hat{\theta}$:n kertymäfunktio F lähtien havainnosta

$$P\{\hat{\theta} \leq t\} = P\{Y_1 \leq t\} \cdots P\{Y_n \leq t\}$$

ja derivoi siitä tiheysfunktio $f = F'$.

b) Laske $\hat{\theta}$:n odotusarvo ja totea, että $\hat{\theta}$ on harhainen mutta asymptoottisesti harhaton.

c) Laske $\hat{\theta}$:n varianssi ja keskineliövirhe $E[(\hat{\theta} - \theta)^2]$ ja vertaa jälkimmäistä momenttimenetelmän antaman harhattoman estimaattorin $\tilde{\theta} = 2\overline{Y}$ (ks. 3.3.3) varianssiin. Kumpi estimaattori on parempi?

d) Olisiko $\check{\theta} = [(n+1)/n]\hat{\theta}$ hyvä estimaattori?

3.11. Tarkastellaan mallia $Y_1, \ldots, Y_n \sim G(\alpha, 1/\beta)$ $\perp\!\!\!\perp$ (ks. teht. 2.8). Johda momenttimenetelmän antamien estimaattoreiden $\tilde{\alpha}$ ja $\tilde{\beta}$ lausekkeet

$$\tilde{\alpha} = \frac{\overline{Y}^2}{\tilde{\sigma}^2}, \qquad \tilde{\beta} = \frac{\tilde{\sigma}^2}{\overline{Y}},$$

jossa $\tilde{\sigma}^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2/n$.

3.12. Oletetaan, että malli $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ on säännöllinen siinä mielessä kuin informaatioepäyhtälössä oletettiin (ks. 3.4.3 ja 2.5.2). Olkoon T parametrin θ harhaton estimaattori.

a) Osoita, että T on täystehokas jos ja vain jos on olemassa vain θ:sta riippuva luku $a(\theta)$ siten, että $T = \theta + a(\theta) l'(\theta; \boldsymbol{Y})$. *Ohje.* Tarkastele informaatioepäyhtälön todistusta ja palauta todennäköisyyslaskennasta mieleen, että Schwarzin epäyhtälössä yhtäsuuruus pätee jos ja vain jos muuttujat riippuvat toisistaan lineaarisesti.

b) Oletetaan, että su-estimaatti $\hat{\theta}(\boldsymbol{y})$ saadaan uskottavuusyhtälön $l'(\theta; \boldsymbol{y}) = 0$ ratkaisuna. Perustele, että jos T on täystehokas, niin $T = \hat{\theta}(\boldsymbol{Y})$.

3.13. Olkoot $Y_1, \ldots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$. Mitä normaalijakaumaa su-estimaattori $\hat{\mu} = \overline{Y}$ approksimatiivisesti noudattaa, kun n on suuri?

3.14. Olkoon Y jatkuvasti jakautunut satunnaismuuttuja, jonka tiheysfunktio on f, ja olkoon g jonkin toisen jakauman tiheysfunktio. Todista, että

$$E[\log f(Y)] > E[\log g(Y)].$$

Tätä tulosta käytettiin kohdassa 3.6.2. Miten?

Ohje. Todennäköisyyslaskennan Jensenin epäyhtälö kertoo, että jos u on aidosti ylöspäin kupera (eli aidosti konkaavi) funktio ja X on satunnaismuuttuja, niin $E(u(X)) \leq u(E(X))$ ja yhtäsuuruus pätee vain jos X on vakio. Sovella tätä valinnoilla $u(x) = \log x$ ja $X = g(Y)/f(Y)$. Muista myös, että $\log a/b = \log a - \log b$.

3.15. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Parametrin (μ, σ^2) su-estimaattori $(\hat{\mu}, \hat{\sigma}^2)$ on laskettu kohdassa 2.2.6.

a) Palauta todennäköisyyslaskennasta mieleen, mikä on satunnaismuuttujaparin $(\hat{\mu}, \hat{\sigma}^2)$ yhteisjakauma (reunajakaumat ja komponenttien välinen riippuvuus/riippumattomuus).

b) Mitkä ovat sen kaksiulotteisen normaalijakauman parametrit (odotusarvovektori ja kovarianssimatriisi), jota $(\hat{\mu}, \hat{\sigma}^2)$ approksimatiivisesti noudattaa, kun n on suuri?

4 Tyhjentävyys

4.1 Tyhjentävä tunnusluku

4.1.1 Tunnusluku. Olkoon $\boldsymbol{y} = (y_1, \ldots, y_n)$ aineisto, joka tulee analysoida. Kuten jo aikaisemmin on ollut puhetta, mitä tahansa aineistosta $\boldsymbol{y}$ laskettua suuretta eli aineiston funktiota kutsutaan *tunnusluvuksi*. Se voi olla reaaliarvoinen $t = t(\boldsymbol{y})$ tai vektoriarvoinen $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y}) = \big(t_1(\boldsymbol{y}), \ldots, t_k(\boldsymbol{y})\big)$. Esimerkkejä tunnusluvuista ovat

a) otoskeskiarvo $\overline{y} = (y_1 + \cdots + y_n)/n$
b) otosvarianssi $s^2 = \sum_{i=1}^n (y_i - \overline{y})^2/(n-1)$
c) pari $(\overline{y}, s^2)$
d) pienin havainto $y_{(1)} = \min(y_1, \ldots, y_n)$
e) suurin havainto $y_{(n)} = \max(y_1, \ldots, y_n)$
f) järjestystunnusluku $\big(y_{(1)}, \ldots, y_{(n)}\big)$ eli havainnot y_i suuruusjärjestyksessä.

Yleensä tunnusluvun dimensio on pienempi kuin n; edellä näin on lukuunottamatta f-esimerkkiä. Tunnuslukujen avulla voidaan siis tiivistää aineistoa tai toisaalta tuoda esiin sen keskeisiä piirteitä. Lisäksi jokainen tunnusluku $\boldsymbol{t}$ osittaa kaikkien mahdollisten aineistojen avaruuden erillisiin osiin: aineistot $\boldsymbol{y}$ ja $\boldsymbol{y}'$ kuuluvat samaan osaan jos ja vain jos $\boldsymbol{t}(\boldsymbol{y}) = \boldsymbol{t}(\boldsymbol{y}')$ – aineistot $\boldsymbol{y}$ ja $\boldsymbol{y}'$ siis ikään kuin samastetaan, jos niitä katsellaan kyseisen tunnusluvun ”läpi”. Esimerkiksi $\boldsymbol{y} = (1,1,1)$ ja $\boldsymbol{y}' = (1,2,0)$ ovat eri aineistoja, mutta otoskeskiarvon mielessä ne ovat samanarvoisia, koska $\overline{y} = \overline{y}' = 1$.

4.1.2 Tyhjentävän tunnusluvun määritelmä ja tulkinta. Siirrytään nyt parametrisen tilastollisen päättelyn asetelmaan. Ajatellaan, että aineisto $\boldsymbol{y}$ on satunnaisvektorin $\boldsymbol{Y}$ toteutunut arvo, ja spesifioidaan tilastollinen malli eli yptf/ytf $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$. Tällöin jokaista tunnuslukua $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$ vastaa satunnaismuuttuja tai -vektori $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$, jota myös nimitetään tunnusluvuksi.

Kuten ennenkin on todettu, tilastollisen päättelyn tavoitteena on tehdä päätelmiä parametrista $\boldsymbol{\theta}$ aineiston perusteella. Herää kysymys: Voidaanko päätelmät perustaa johonkin tunnuslukuun $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$ koko aineiston $\boldsymbol{y}$ sijasta? Mitä tunnusluvulta tulee vaatia, jotta näin meneteltäessä ei hukata parametrin $\boldsymbol{\theta}$ kannalta relevanttia informaatiota? Tyhjentävän tunnusluvun käsite on syntynyt näistä pohdinnoista. Se on tärkeä ja syvällinen käsite päättelyn teoriassa, mutta tällä kurssilla aihetta käsitellään varsin suppeasti.

Matemaattinen määritelmä on seuraava: Tunnusluku $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on parametrin $\boldsymbol{\theta}$ *tyhjentävä tunnusluku*, jos satunnaisvektorin $\boldsymbol{Y}$ ehdollinen jakauma ehdolla $\boldsymbol{T} = \boldsymbol{t}$ ei koskaan riipu $\boldsymbol{\theta}$:sta eli käytännössä ehdollisen yptf/ytf:n $f_{\boldsymbol{Y}|\boldsymbol{T}}(\boldsymbol{y}|\boldsymbol{t}; \boldsymbol{\theta})$ lauseke ei itse asiassa sisällä $\boldsymbol{\theta}$:aa.

Tulkinnallisesti tämän voi ymmärtää seuraavasti: Jos aineistosta $\boldsymbol{y}$ tiedetään sen verran, että tunnusluvun $\boldsymbol{t}(\boldsymbol{y})$ arvo on $\boldsymbol{t}$, niin aineiston tarkempi tuntemus ei enää tuo mitään lisäinformaatiota parametrista $\boldsymbol{\theta}$, koska vastaava satunnaismekanismi eli $\boldsymbol{Y}$:n ehdollinen jakauma ehdolla $\boldsymbol{t}(\boldsymbol{Y}) = \boldsymbol{t}$ on $\boldsymbol{\theta}$:sta riippumaton.

Sanaa ”informaatio” on edellä käytetty sen yleiskielisessä merkityksessä. Tunnusluvun tyhjentävyys on kuitenkin mahdollista luonnehtia myös eksaktisti Fisherin informaation avulla. Tietyt säännöllisyysehdot olettamalla voidaan nimittäin osoittaa, että mitä tahansa tunnuslukua $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ vastaavan mallin $f_{\boldsymbol{T}}$ Fisherin informaatio on aina pienempi tai yhtä suuri kuin koko aineiston mallin $f_{\boldsymbol{Y}}$ ja että yhtäsuuruus pätee täsmälleen silloin kun $\boldsymbol{T}$ on tyhjentävä. Sivuutamme tämän tuloksen todistuksen (ks. kuitenkin teht. 4.7).

Yleensä pyrkimyksenä on löytää tyhjentävä tunnusluku, jonka dimensio on mahdollisimman pieni, eli tiivistää aineistoa mahdollisimman paljon menettämättä päättelyn kannalta relevanttia informaatiota. Tämän matemaattinen formulointi johtaa ns. *minimaalisen tyhjentävän tunnusluvun* käsitteeseen, jonka käsittely tällä kurssilla sivuutetaan. Tässä yhteydessä on syytä huomata, että aineisto $\boldsymbol{Y}$ itse on aina tyhjentävä tunnusluku (mieti miksi). Siinä tapauksessa, että malli muodostuu riippumattomista samoin jakautuneista havainnoista – kuten tällä kurssilla useimmissa esimerkeissä on –, myös järjestystunnusluku $\big(Y_{(1)}, \ldots, Y_{(n)}\big)$ on tyhjentävä (ks. teht. 4.6). Näitä tunnuslukuja kutsutaan *triviaaleiksi* tyhjentäviksi tunnusluvuiksi.

4.1.3 Esimerkki: toistokoemalli. Palataan kohdan 1.2.1 hehkulamppuesimerkkiin. Siinä aineisto oli $\boldsymbol{y} = (y_1, \ldots, y_n)$, jossa

$$y_i = \begin{cases} 0, & \text{jos } i\text{:s lamppu on ehjä,} \\ 1, & \text{jos } i\text{:s lamppu on rikki.} \end{cases}$$

Vastaavat satunnaismuuttujat olivat $Y_1, \ldots, Y_n \sim B(\theta)$ $\perp\!\!\!\perp$, jossa θ on rikkinäisten suhteellinen osuus perusjoukossa. Tilastollisen mallin spesifioi yptf

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i}.$$

Tarkastellaan tunnuslukua $K = \sum_{i=1}^{n} Y_i$, joka kertoo rikkinäisten lamppujen lukumäärän otoksessa. Osoitetaan, että se on tyhjentävä.

Kuten kohdassa 2.1.5 todettiin, $K \sim Bin(n, \theta)$, jolloin siis K:n ptf on

$$f_K(k; \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Muodostetaan ehdollinen ptf

$$f_{\boldsymbol{Y}|K}(\boldsymbol{y}|k; \theta) = \frac{P\{\boldsymbol{Y} = \boldsymbol{y},\ K = k\}}{P\{K = k\}}.$$

Selvästi $P\{\boldsymbol{Y} = \boldsymbol{y},\ K = k\} = 0$ jos $\sum y_i \neq k$. Siis $f_{\boldsymbol{Y}|K}(\boldsymbol{y}|k; \theta) = 0$ tällöin. Oletetaan, että $\sum y_i = k$. Tällöin tapahtuma $\{\boldsymbol{Y} = \boldsymbol{y}\}$ sisältyy tapahtumaan $\{K = k\}$, joten

$$f_{\boldsymbol{Y}|K}(\boldsymbol{y}|k; \theta) = \frac{P\{\boldsymbol{Y} = \boldsymbol{y}\}}{P\{K = k\}} = \frac{\theta^{\sum y_i}(1-\theta)^{n-\sum y_i}}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

Koska lopputulos ei riipu θ:sta, K on parametrin θ tyhjentävä tunnusluku.

Huomaa, että siirtyminen koko aineistosta $\boldsymbol{y}$ tunnuslukuun $k = \sum y_i$ hukkaa tiedon siitä, missä järjestyksessä ehjät ja rikkinäiset lamput ovat otokseen tulleet. Mutta

intuitiivisestikin on selvää, että tällä ei ole merkitystä parametria θ koskevien päätelmien kannalta vaan rikkinäisten lukumäärä otoksessa sisältää kaiken informaation θ:sta.

4.2 Faktorointikriteeri

4.2.1 Faktorointikriteeri tyhjentävyydelle. Kuten edellinen esimerkki osoittaa, suoraan määritelmän perusteella on varsin hankalaa ja työlästä löytää tyhjentäviä tunnuslukuja yksinkertaisissakaan malleissa. Käytännön tehtävissä on paljon kätevämpää käyttää seuraavaa, *faktorointikriteeriksi* kutsuttua lausetta. Tarkastellaan jälleen mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, jonka parametriavaruus on Ω.

Lause. *Tunnusluku $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on parametrin $\boldsymbol{\theta}$ tyhjentävä tunnusluku jos ja vain jos yptf/ytf $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ voidaan kirjoittaa muodossa*

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = h(\boldsymbol{y})g(\boldsymbol{t}(\boldsymbol{y});\boldsymbol{\theta})$$

kaikilla $\boldsymbol{y}$ ja $\boldsymbol{\theta} \in \Omega$.

Oleellista tässä on se, että tekijä $h(\boldsymbol{y})$ ei riipu lainkaan mallin parametristä ja että tekijä $g(\boldsymbol{t}(\boldsymbol{y});\boldsymbol{\theta})$ riippuu aineistosta $\boldsymbol{y}$ vain tunnusluvun $\boldsymbol{t}(\boldsymbol{y})$ välityksellä. Koska mallin uskottavuusfunktion saattoi kertoa tai jakaa millä tahansa (positiivisella) vain aineistosta riippuvalla vakiolla, voidaan faktorointikriteeri lausua myös näin: tunnusluku $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on tyhjentävä jos ja vain jos mallin uskottavuusfunktio voidaan valita siten, että se riippuu aineistosta vain $\boldsymbol{t}(\boldsymbol{y})$:n välityksellä. Sama pätee tietysti myös log-uskottavuusfunktion suhteen.

Faktorointikriteerin todistusta tarkastellaan vasta esimerkkien jälkeen.

4.2.2 Esimerkki: toistokoemalli. Esimerkissä 4.1.3 tunnusluvun $k = \sum y_i$ tyhjentävyys nähdään faktorointikriteerin avulla suoraan siitä, että mallin yptf $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i}$ riippuu aineistosta vain sen välityksellä.

4.2.3 Esimerkki: normaalimalli. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Kohdassa 2.1.4 todettiin, että tämän mallin uskottavuusfunktio voidaan saattaa muotoon

$$L(\mu, \sigma^2) = (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\overline{y}-\mu)^2}{2\sigma^2}\right\},$$

jossa $\overline{y}$ on otoskeskiarvo ja s^2 otosvarianssi. Siten $(\overline{y}, s^2)$ on parametrin (μ, σ^2) tyhjentävä tunnusluku. Toinen vaihtoehto olisi pari $\left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2\right)$.

Siinä tapauksessa, että varianssi on tunnettu luku $\sigma^2 = \sigma_0^2 > 0$, saatiin uskottavuusfunktiolle lauseke

$$L(\mu) = \exp\left\{-\frac{n(\overline{y}-\mu)^2}{2\sigma_0^2}\right\},$$

joka riippuu aineistosta vain otoskeskiarvon $\overline{y}$ kautta. Tällöin siis $\overline{y}$ on parametrin μ tyhjentävä tunnusluku. Mielenkiintoista on, että normaalimallin tapauksessa voidaan löytää tyhjentävä tunnusluku, jonka dimensio on sama kuin mallin parametrin. Tämä ilmiö toteutuu myös tavallisen normaalijakaumaoletukseen perustuvan regressiomallin kohdalla (ks. teht. 4.4). Yleensä malleilla ei tällaista ominaisuutta suinkaan ole.

4.2.4 Esimerkki: Cauchy-jakauma. Cauchy-jakaumaksi kutsutaan jatkuvaa jakaumaa, jonka tiheysfunktio on

$$f(y;\theta) = \frac{1}{\pi} \cdot \frac{1}{1+(y-\theta)^2}, \qquad y \in \mathbb{R},$$

ja jossa θ on reaalinen parametri. Muodoltaan se muistuttaa varsin paljon $N(\theta,1)$-jakaumaa. Jos $Y_1,\ldots,Y_n$ on riippumaton otos tästä jakaumasta, voidaan osoittaa, että näin muodostuvalla mallilla ei ole mitään ei-triviaalia tyhjentävää tunnuslukua vaan järjestystunnusluku on ”tiivistetyin” mahdollinen (eli minimaalinen) tyhjentävä tunnusluku.

4.2.5 Eksponenttiperheen mallit. Malli $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ (eli yptf/ytf) kuuluu d-ulotteiseen *eksponenttiperheeseen*, mikäli se on muotoa

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\boldsymbol{y})\exp\Big\{\sum_{j=1}^{d}\phi_j(\boldsymbol{\theta})t_j(\boldsymbol{y})\Big\}, \tag{4.1}$$

jossa parametri $\boldsymbol{\theta}$ on d-ulotteinen. Mallin voi uudelleenparametroida myös vektorin $\big(\phi_1(\boldsymbol{\theta}),\ldots,\phi_d(\boldsymbol{\theta})\big)$ avulla (ks. teht. 2.20); se on mallin *luonnollinen parametri*. Faktorointikriteerin perusteella tällä mallilla on tyhjentävä tunnusluku $\big(t_1(\boldsymbol{y}),\ldots,t_d(\boldsymbol{y})\big)$.

Jos $Y_1,\ldots,Y_n$ on riippumaton otos jakaumasta, jonka ptf/tf on muotoa (4.1), eli kullakin Y_i on ptf/tf

$$f(y;\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(y)\exp\Big\{\sum_{j=1}^{d}\phi_j(\boldsymbol{\theta})t_j(y)\Big\},$$

niin syntyvän mallin yptf/ytf on

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}) = c(\boldsymbol{\theta})^n\Big(\prod_{i=1}^{n}h(y_i)\Big)\exp\Big\{\sum_{j=1}^{d}\phi_j(\boldsymbol{\theta})\Big(\sum_{i=1}^{n}t_j(y_i)\Big)\Big\},$$

joka edelleen on tyyppiä (4.1). Eksponenttiperheeseen kuuluvalla mallilla on siis se miellyttävä ominaisuus, että havaintojen lukumäärästä n riippumatta sillä on tyhjentävä tunnusluku, jonka dimensio on sama kuin mallin parametrin. Voidaan kääntäen osoittaa, että eräin poikkeuksin tällainen ominaisuus on vain eksponenttiperheeseen kuuluvilla malleilla. Eksponenttiperhe sisältää mm. Bernoulli-, binomi-, Poisson-, normaali-, gamma- ja eksponenttijakaumat.

4.2.6 Faktorointikriteerin perustelu. Tässä käsitellään vain diskreetti tapaus; jatkuva tapaus on vaikeampi ja sivuutetaan.

Oletetaan, että tunnusluku $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on tyhjentävä. Kirjoitetaan

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) &= P_{\boldsymbol{\theta}}\{\boldsymbol{Y}=\boldsymbol{y}\} \\ &= P_{\boldsymbol{\theta}}\{\boldsymbol{Y}=\boldsymbol{y},\, \boldsymbol{T}=\boldsymbol{t}(\boldsymbol{y})\} \\ &= P_{\boldsymbol{\theta}}\{\boldsymbol{T}=\boldsymbol{t}(\boldsymbol{y})\}P_{\boldsymbol{\theta}}\{\boldsymbol{Y}=\boldsymbol{y}\,|\,\boldsymbol{T}=\boldsymbol{t}(\boldsymbol{y})\}. \end{aligned}$$

Koska $\boldsymbol{T}$ on tyhjentävä, niin jälkimmäinen tekijä ei riipu $\boldsymbol{\theta}$:sta. Toisaalta tekijä $P_{\boldsymbol{\theta}}\{\boldsymbol{T}=\boldsymbol{t}(\boldsymbol{y})\}$ riippuu $\boldsymbol{y}$:stä vain $\boldsymbol{t}(\boldsymbol{y})$:n kautta. Siten edellä on saatu vaaditunlainen faktorisaatio.

Oletetaan kääntäen, että faktorisaatio

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = h(\boldsymbol{y})g(\boldsymbol{t}(\boldsymbol{y});\boldsymbol{\theta})$$

pätee, ja osoitetaan, että $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on tyhjentävä. Lähdetään ehdollisen pistetodennäköisyyden määritelmästä

$$f_{\boldsymbol{Y}|\boldsymbol{T}}(\boldsymbol{y}|\boldsymbol{t};\boldsymbol{\theta}) = \frac{P_{\boldsymbol{\theta}}\{\boldsymbol{Y} = \boldsymbol{y},\, \boldsymbol{T} = \boldsymbol{t}\}}{P_{\boldsymbol{\theta}}\{\boldsymbol{T} = \boldsymbol{t}\}}.$$

Jos $\boldsymbol{t} \neq \boldsymbol{t}(\boldsymbol{y})$, niin osoittajassa oleva tapahtuma on mahdoton ja siten $f_{\boldsymbol{Y}|\boldsymbol{T}}(\boldsymbol{y}|\boldsymbol{t};\boldsymbol{\theta}) = 0$. Oletetaan, että $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{y})$, ja merkitään $A_{\boldsymbol{t}} = \{\boldsymbol{y}' : \boldsymbol{t}(\boldsymbol{y}') = \boldsymbol{t}\}$. Nyt $\{\boldsymbol{Y} = \boldsymbol{y}\} \subset \{\boldsymbol{T} = \boldsymbol{t}\}$ ja $\{\boldsymbol{T} = \boldsymbol{t}\} = \bigcup_{\boldsymbol{y}' \in A_{\boldsymbol{t}}} \{\boldsymbol{Y} = \boldsymbol{y}'\}$, joten

$$\begin{aligned} f_{\boldsymbol{Y}|\boldsymbol{T}}(\boldsymbol{y}|\boldsymbol{t};\boldsymbol{\theta}) &= \frac{P_{\boldsymbol{\theta}}\{\boldsymbol{Y} = \boldsymbol{y}\}}{P_{\boldsymbol{\theta}}\{\boldsymbol{T} = \boldsymbol{t}\}} \\ &= \frac{f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}{\sum_{\boldsymbol{y}' \in A_{\boldsymbol{t}}} f_{\boldsymbol{Y}}(\boldsymbol{y}';\boldsymbol{\theta})} \\ &= \frac{h(\boldsymbol{y})g(\boldsymbol{t};\boldsymbol{\theta})}{\sum_{\boldsymbol{y}' \in A_{\boldsymbol{t}}} h(\boldsymbol{y}')g(\boldsymbol{t};\boldsymbol{\theta})} \\ &= \frac{h(\boldsymbol{y})}{\sum_{\boldsymbol{y}' \in A_{\boldsymbol{t}}} h(\boldsymbol{y}')}. \end{aligned}$$

Koska tämä ei riipu $\boldsymbol{\theta}$:sta, niin $\boldsymbol{T}$ on tyhjentävä.

Harjoitustehtäviä

4.1. Olkoot $Y_1 \ldots, Y_n \sim \mathit{Exp}(\lambda)$ $\perp\!\!\!\perp$. Osoita, että otoskeskiarvo $\overline{Y}$ on parametrin λ tyhjentävä tunnusluku.

4.2. Olkoot $Y_1, \ldots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$. Johda parametrille μ reaalinen tyhjentävä tunnusluku.

4.3. Olkoot $Y_1, \ldots, Y_n \sim \mathit{Tas}(0, \theta)$ $\perp\!\!\!\perp$. Osoita, että suurin havainto $Y_{(n)}$ on parametrin θ tyhjentävä tunnusluku.

4.4. Tarkastellaan yhden selittäjän regressiomallia $Y_1, \ldots, Y_n$ $\perp\!\!\!\perp$, $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, jossa parametri on $(\alpha, \beta, \sigma^2)$. Näytä, että tunnusluku $\left(\sum_{i=1}^n y_i^2, \sum_{i=1}^n y_i, \sum_{i=1}^n y_i x_i\right)$ on tyhjentävä.

4.5. Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ tilastollinen malli, jonka parametrilla on yksikäsitteinen suurimman uskottavuuden estimaatti $\hat{\boldsymbol{\theta}}$. Oletetaan, että $\boldsymbol{T} = \boldsymbol{t}(\boldsymbol{Y})$ on tyhjentävä tunnusluku. Järkeile, että $\hat{\boldsymbol{\theta}}$ riippuu aineistosta $\boldsymbol{y}$ vain tunnusluvun $\boldsymbol{t}(\boldsymbol{y})$ välityksellä.

4.6. Olkoon $Y_1, \ldots, Y_n$ riippumaton otos jakaumasta, jonka ptf/tf on $f(y;\theta)$. Päättele esim. faktorointikriteerin avulla, että järjestystunnusluku $\left(Y_{(1)}, \ldots, Y_{(n)}\right)$ on aina tyhjentävä.

4.7. Olkoot $Y_1, \ldots, Y_n \sim \mathit{Exp}(\lambda)$ $\perp\!\!\!\perp$. Tehtävän 4.1 perusteella $T = Y_1 + \cdots + Y_n$ on mallin parametrin λ tyhjentävä tunnusluku. Ilmoita T:n tiheysfunktio f_T ja laske siitä Fisherin informaatio $i_T(\lambda)$. Vertaa sitä koko aineistoa vastaavasta mallista $f_{\boldsymbol{Y}}$ laskettuun Fisherin informaatioon $i_{\boldsymbol{Y}}(\lambda)$. Mitä huomaat?

Vihje. Todennäköisyyslaskennassa on kerrottu, että $T \sim G(n, \lambda)$.

5 Hypoteesien testaaminen

5.1 Johdanto

Tilastolliset testit ovat käytetyimpiä tilastollisen päättelyn menetelmiä. Niiden avulla pyritään ottamaan kantaa väitteisiin eli hypoteeseihin, jotka koskevat tutkittavaa satunnaisilmiötä. Testit tulevat käyttöön esimerkiksi silloin, kun tilastollisten aineistojen valossa etsitään vastausta seuraavanlaisiin kysymyksiin: Onko tutkittava lantti harhaton? Onko tietyntyyppisen sähkölaitteen keskimääräinen kestoikä yli 100 vuorokautta? Ehkäiseekö C-vitamiini flunssaan sairastumista?

5.1.1 Esimerkki. Hehkulamppujen valmistaja väittää, että tuotetuista lampuista korkeintaan 1 % on rikki. Väitteen todenperäisyyden selvittämiseksi poimitaan sadan lampun satunnaisotos. Havaitaan, että tässä otoksessa on $k = 3$ rikkinäistä lamppua. Miten valmistajan väitteeseen tulisi tämän havainnon valossa suhtautua? Kumpi on luontevampi selitys havainnolle:

a) Valmistaja puhuu totta, mutta sattuma teki kiusaa ja tuotti tarkasteltuun otokseen keskimääräistä enemmän rikkinäisiä lamppuja.

b) Valmistajan väite ei pidä paikkaansa, vaan rikkinäisiä on tuotannossa enemmän kuin 1 %.

Jotta kysymystä voitaisiin tarkastella tilastollisesti, formuloidaan koeasetelmaa kuvaava malli. Jos satunnaismuuttuja K ilmoittaa rikkinäisten lamppujen lukumäärän sadan lampun otoksessa, niin $K \sim Bin(100, \theta)$, jossa $\theta \in (0, 1)$ on rikkinäisten suhteellinen osuus koko tuotannossa. Siis K:n pistetodennäköisyydet ovat

$$P_\theta\{K = k\} = \binom{100}{k} \theta^k (1-\theta)^{100-k}, \qquad k = 0, 1, \ldots, 100.$$

Valmistajan väite on, että $\theta \leq 0.01$. Tämän kannalta kriittisiä ovat ilmeisestikin ne tapaukset, joissa K:n saama arvo on ”suuri” (selvästi suurempi kuin 1).

Kuinka kriittinen nyt saatu havainto $k = 3$ on valmistajan väitteen kannalta? Arvioidaan tätä kysymystä laskemalla havaittuun tapahtumaan liittyviä todennäköisyyksiä a-selityksen vallitessa. Jos valmistajan väite pätee, todennäköisyys sille, että otoksessa on ainakin kolme rikkinäistä lamppua, on $P_\theta\{K \geq 3\}$, jossa $\theta \leq 0.01$. Tämä on ilmeisesti suurimmillaan silloin, kun $\theta = 0.01$, jolloin

$$\begin{aligned} P_{0.01}\{K \geq 3\} &= 1 - P_{0.01}\{K \leq 2\} \\ &= 1 - 0.99^{100} - 100 \cdot 0.01 \cdot 0.99^{99} - 4950 \cdot 0.01^2 \cdot 0.99^{98} \\ &\approx 0.08. \end{aligned}$$

Toistetun aineistonkeruun (otannan) kannalta ajateltuna tämä merkitsee sitä, että valmistajan väitteen pätiessä noin 8 %:ssa otoksista on yhtä paljon tai enemmän rikkinäisiä lamppuja kuin siinä otoksessa, joka nyt on saatu. Vaihtoehdon a mahdollisuus ei siis ole suuri mutta kylläkin otettava huomioon.

Lopputulokseksi jäänee, että otoksen perusteella valmistajan väitettä on syytä hieman epäillä mutta sitä ei kuitenkaan voi täysin tyrmätä. Jos sen sijaan olisi havaittu otoksessa peräti $k = 5$ rikkinäistä lamppua, olisi voitu laskea $P_{0.01}\{K \geq 5\} \approx 0.003$ ja järkeillä, että jos valmistajan väite pätee, näin paljon tai enemmän rikkinäisiä lamppuja esiintyisi otoksessa vain noin 0.3 %:n todennäköisyydellä. Väite vaikuttaisi siis varsin epäuskottavalta ja voitaisiin hylätä.

5.1.2 Testin vaiheet yleisesti. Oletetaan tuttuun tapaan, että tarkasteltavaa satunnaisilmiötä on päädytty kuvaamaan jollakin parametrisella tilastollisella mallilla. Tällöin testausasetelman muotoilu ja testin suorittaminen käytännössä käsittää karkeasti ottaen seuraavat neljä vaihetta, joiden olemukseen tässä luvussa tutustutaan:

1. Asetetaan nollahypoteesi H_0 ja mahdollisesti vastahypoteesi H_1.
2. Valitaan käytettävä testisuure.
3. Lasketaan havaittua aineistoa vastaava testisuureen arvo sekä havaittu merkitsevyystaso eli p-arvo.
4. Tehdään johtopäätökset.

Esimerkissä 5.1.1 oli nollahypoteesina $H_0 \colon \theta \leq 0.01$ ja vastahypoteesina $H_1 \colon \theta > 0.01$. Testisuureena käytettiin rikkinäisten lamppujen lukumäärää k, jonka havaittu arvo oli 3 (tai vastaavasti 5). Havaituksi merkitsevyystasoksi saatiin likimain 0.08 (tai vastaavasti 0.003).

5.2 Peruskäsitteet ja testin suorittaminen

Tarkastellaan yleistä parametrista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jonka parametriavaruus on Ω.

5.2.1 Hypoteesit ja kysymyksenasettelu. *Hypoteesilla* tarkoitetaan mitä tahansa väitettä, joka koskee mallin parametria ja voidaan kirjoittaa muotoon $\boldsymbol{\theta} \in \Omega'$, jossa Ω' on jokin Ω:n epätyhjä osajoukko. Se siis lausuu, että todellinen parametriarvo $\boldsymbol{\theta}$ kuuluu tiettyyn parametriavaruuden osajoukkoon Ω'. Hypoteesi on *yksinkertainen*, jos Ω' sisältää vain yhden pisteen; muuten hypoteesi on *yhdistetty*.

Testausasetelman määrittelyn lähtökohtana on *nollahypoteesi* H_0, jonka mukaan $\boldsymbol{\theta}$ kuuluu Ω:n osajoukkoon Ω_0. Tätä merkitään lyhyesti

$$H_0 \colon \ \boldsymbol{\theta} \in \Omega_0.$$

Testaamisen ensisijaisena tavoitteena on arvioida havaitun aineiston $\boldsymbol{y}$ valossa nollahypoteesin paikkansapitävyyttä: Onko aineisto sopusoinnussa nollahypoteesin kanssa vai saattaako se sen epäilyksenalaiseksi? Jos saattaa, niin kuinka voimakkaasti? Joissakin sovellustilanteissa on suorastaan tehtävä kategorinen *päätös*: nollahypoteesi joko *hyväksytään* tai *hylätään*. Tähän päätöksentekoon liittyy aina erehtymisen riski, ja riskin suuruuden arviointi on olennainen osa testausmenettelyn valintaa sekä testin tulosten raportointia. Yleensä nollahypoteesi edustaa (nimensä mukaisesti) jossain mielessä neutraalia tai oletusarvoista asiaintilaa, joten erityisesti sen virheellistä hylkäämistä halutaan varoa.

Monesti muotoillaan nollahypoteesin H_0 rinnalle myös *vaihtoehtoinen hypoteesi* eli *vastahypoteesi* H_1, jonka mukaan $\boldsymbol{\theta}$ kuuluu Ω:n osajoukkoon Ω_1, siis lyhyesti

$$H_1\colon\ \boldsymbol{\theta} \in \Omega_1.$$

Joukosta Ω_1 oletetaan aina, että se ei leikkaa Ω_0:aa, ja usein pätee myös $\Omega_0 \cup \Omega_1 = \Omega$. Jos vastahypoteesi on asetettu, testauksen tavoitteena on arvioida erityisesti sitä, tukisiko aineisto H_0:n sijasta pikemminkin hypoteesia H_1. Vastahypoteesin asettaminen merkitsee siis jonkinlaista kannanottoa siihen, mitä parametrin $\boldsymbol{\theta}$ ajatellaan tai toivotaan toteuttavan siinä tapauksessa, että nollahypoteesi osoittautuu epäilyksenalaiseksi. Tällöin nollahypoteesin hylkääminen merkitsee aina samalla vastahypoteesin H_1 hyväksymistä.

5.2.2 Testisuure. Testaaminen eli nollahypoteesin paikkansapitävyyden arviointi perustuu siis mallista havaitun aineiston tarkasteluun. On tavalla tai toisella kyettävä mittaamaan nollahypoteesin ja aineiston välistä yhteensopivuutta. Tähän tarkoitukseen käytetään tavallisesti jotakin testisuuretta. *Testisuureella* tarkoitetaan aineistosta laskettavaa reaaliarvoista tunnuslukua $t = t(\boldsymbol{y})$, jolla on seuraavanlainen *monotonisuusominaisuus*: pienet t:n arvot viittaavat siihen, että aineisto on sopusoinnussa H_0:n kanssa, kun taas suuret arvot merkitsevät, että aineisto todistaa H_0:aa vastaan ja tukee H_1:tä, jos tämä on asetettu.

Monotonisuusominaisuudella voi olla eri tilanteissa vaihtoehtoisia muotoja. Joskus t:n pienet arvot viittaavat ristiriitaan ja suuret yhteensopivuuteen H_0:n kanssa. Ns. kaksisuuntaisissa (tai -puolisissa) testausasetelmissa kiinnitetään yleensä huomiota t:n poikkeamaan $|t - t_0|$ jostakin "vertailuarvosta" t_0: suuri poikkeama merkitsee, että aineisto on ristiriidassa nollahypoteesin kanssa. Nämä tilanteet voidaan tietysti haluttaessa palauttaa yllä esiteltyyn standardiasetelmaan korvaamalla t suureella $-t$ tai vastaavasti suureella $|t - t_0|$. Testisuureen arvojen oikea tulkinta selviää aina asiayhteydestä.

Käytettävän testisuureen valinta ei useinkaan ole suoraviivaista, ja siihen liittyykin suuri osa testauksen problematiikasta sekä teoriasta. Valintaan vaikuttavat ainakin malli, nollahypoteesi H_0 ja mahdollinen vastahypoteesi H_1 eli se, millaisia poikkeamia nollahypoteesista halutaan erityisesti paljastaa. Lukuisiin tavanomaisiin testausasetelmiin liittyviä vakiintuneita testisuureita esitellään tilastollisia malleja ja menetelmiä käsittelevässä kirjallisuudessa. On myös olemassa tiettyjä kriteerejä, joilla verrata eri testisuureiden hyvyyttä annetussa testausasetelmassa. Näitä käsitellään pykälässä 5.5. Mallin uskottavuusfunktion avulla ja asymptoottiseen teoriaan turvautumalla on lisäksi mahdollista konstruoida joitakin varsin yleispäteviä testisuureita, kuten pykälissä 5.6 ja 5.7 tullaan näkemään.

5.2.3 Havaittu merkitsevyystaso. Oletetaan, että testaaminen on päätetty perustaa testisuureeseen t, jonka suuret arvot ovat nollahypoteesille H_0 kriittisiä. Olkoon $T = t(\boldsymbol{Y})$ sitä vastaava satunnaismuuttuja, ja olkoon $\boldsymbol{y}$ havaittu aineisto. Tällöin voidaan laskea testin *p-arvo* eli *havaittu merkitsevyystaso* p, johon testin seurauksena tehtävät johtopäätökset perustuvat. Sillä tarkoitetaan todennäköisyyttä

$$p = p(\boldsymbol{y}) = P_{\boldsymbol{\theta}}\{T \geq t(\boldsymbol{y})\}, \qquad \boldsymbol{\theta} \in \Omega_0,$$

jos H_0 on yksinkertainen tai jos tämä todennäköisyys on sama kaikilla $\boldsymbol{\theta} \in \Omega_0$, ja yleisemmin lukua

$$p = p(\boldsymbol{y}) = \sup_{\boldsymbol{\theta} \in \Omega_0} P_{\boldsymbol{\theta}}\{T \geq t(\boldsymbol{y})\}.$$

Kyseessä on siis nollahypoteesin pätiessä laskettu yläraja sille todennäköisyydelle, tai yksinkertaisessa tapauksessa itse todennäköisyys, että satunnaismuuttuja T saa arvon, joka on yhtä suuri tai suurempi kuin nyt havaittu arvo $t(\boldsymbol{y})$.

Pieni p-arvo merkitsee, että nollahypoteesin pätiessä on epätodennäköistä saada aineistoja, jotka olisivat yhtä huonosti tai vielä huonommin sopusoinnussa H_0:n kanssa kuin se aineisto, joka nyt on havaittu. Tällöin siis aineisto todistaa H_0:aa vastaan ja H_0 voidaan asettaa epäilyksenalaiseksi, jopa hylätä. Mikäli p-arvo ei ole kovin pieni, aineisto näyttää olevan sopusoinnussa H_0:n kanssa, joten H_0 hyväksytään. Mitä "pieni" tässä yhteydessä tarkoittaa ja miten p-arvot yleensä on tulkittava, pohditaan enemmän seuraavassa pykälässä.

Kuten määritelmästä huomataan, p-arvon laskeminen edellyttää sitä, että testisuuretta vastaavan satunnaismuuttujan T jakauma hallitaan ainakin kaikilla nollahypoteesiarvoilla $\boldsymbol{\theta} \in \Omega_0$. Tähänkin seikkaan täytyy kiinnittää huomiota testisuuretta valittaessa: on tietysti eduksi, jos testisuureen jakauma on hyvin tunnettu ja kenties jopa sama kaikilla $\boldsymbol{\theta} \in \Omega_0$. Käytännössä p-arvon laskennassa joudutaan kuitenkin usein turvautumaan approksimatiivisista jakaumatuloksista saataviin likiarvoihin.

5.2.4 Esimerkki: lantin harhattomuus. Halutaan tutkia, onko lantti harhaton. Olkoon kruunun todennäköisyys tuntematon θ. Nollahypoteesi väittää, että lantti on harhaton, ts. $H_0\colon \theta = \frac{1}{2}$. Tämä on yksinkertainen hypoteesi. Tutkitaan sen todenperäisyyttä heittämällä tuhat heittoa. Jos K on kruunujen lukumäärä satunnaismuuttujana, niin $K \sim Bin(1000, \theta)$ ja erityisesti H_0:n pätiessä $K \sim Bin(1000, \frac{1}{2})$.

Oletetaan, että suoritetussa heittosarjassa havaitaan k kruunua. Koska $E_{H_0}(K) = 500$, on ilmeistä, että nollahypoteesin kannalta kriittisiä ovat havainnot, joissa poikkeama $|k-500|$ on suuri eli joissa k on selvästi alle 500 tai selvästi yli 500. Kyseessä on siis *kaksisuuntainen* testausasetelma. Havaitun merkitsevyystason eli p-arvon määritelmä on tällöin

$$\begin{aligned} p &= P_{H_0}\{|K - 500| \geq |k - 500|\} \\ &= P_{H_0}\{K \leq 500 - |k - 500|\} + P_{H_0}\{K \geq 500 + |k - 500|\}. \end{aligned}$$

Alaindeksi H_0 viittaa siihen, että todennäköisyydet on laskettava nollahypoteesin pätiessä eli parametriarvolla $\theta = \frac{1}{2}$. Numeerinen esimerkki tästä tilanteesta on tehtävässä 5.2.

5.3 Havaitun merkitsevyystason tulkinnasta

5.3.1 P-arvo on tunnusluku. Määritelmänsä perusteella testin p-arvo on aineistosta laskettu tunnusluku $p = p(\boldsymbol{y})$. Se mittaa aineiston ja nollahypoteesin välistä yhteensopivuutta asteikolla $[0, 1]$. Useimmissa sovellustilanteissa on järkevintä raportoida testin lopputuloksena saatu p-arvo likimääräisesti ja tulkita sanallisesti sen merkitystä. Tulkintaan voidaan antaa esimerkiksi seuraavat nyrkkisäännöt:

p	Tulkinta
> 0.1	aineisto on (kohtuullisessa) sopusoinnussa H_0:n kanssa
≈ 0.05	aineisto todistaa lievästi H_0:aa vastaan
≤ 0.01	aineisto todistaa voimakkaasti H_0:aa vastaan

5.3.2 Päätösteoreettinen lähestymistapa. Joissakin sovelluksissa on tarpeen tehdä havaitun aineiston perusteella selkeä päätös: joko H_0 hyväksytään tai se hylätään ja mahdollinen vastahypoteesi H_1 hyväksytään. Tällöin menetellään seuraavasti: Kiinnitetään jo ennalta jokin luku $\alpha \in (0,1)$, jota kutsutaan *merkitsevyystasoksi*. Lasketaan aineistosta p-arvo $p = p(\boldsymbol{y})$ ja verrataan sitä valittuun merkitsevyystasoon: jos $p > \alpha$, niin H_0 hyväksytään, ja jos taas $p \leq \alpha$, niin H_0 hylätään ja mahdollinen H_1 hyväksytään.

Perinteisesti merkitsevyystasona α on käytetty jotakin ns. tavanomaisista merkitsevyystasoista 0.05, 0.01 ja 0.001. Jos H_0 hylätään jollakin näistä, on tapana sanoa, että aineiston osoittama poikkeama nollahypoteesista on vastaavasti tilastollisesti *melkein merkitsevä*, *merkitsevä* tai *erittäin merkitsevä*.

Tilastollisen testin lopputuloksena tehtävä päätös nollahypoteesin hyväksymisestä tai hylkäämisestä ei juuri koskaan merkitse sitä, että tämä hypoteesi olisi "todistettu" oikeaksi tai vääräksi jossain loogisesti sitovassa tai matemaattisessa mielessä. Kyseessä on vain hypoteesin tilastollinen punninta, johon aina liittyy virheen mahdollisuus. Virheiden esiintymistä ja niistä käytettäviä nimityksiä voi kuvata seuraavan taulukon avulla:

Todellisuus	Päätös	
	Hyväksytään H_0	Hylätään H_0
H_0 tosi	oikea päätös	*hylkäämisvirhe*†
H_0 epätosi	*hyväksymisvirhe*	oikea päätös

Käytettävä merkitsevyystaso α on aina yläraja hylkäämisvirheen riskille eli todennäköisyydelle. P-arvon määritelmästä nimittäin johtuu, että H_0:n pätiessä tapahtuman $\{p(\boldsymbol{Y}) \leq \alpha\}$ todennäköisyys on korkeintaan α.‡ Testauksen luonteeseen kuuluu, että tämä halutaan pitää varsin pienenä (esim. jonakin tavanomaisista merkitsevyystasoista); H_0 siis hylätään vain jos aineisto todistaa kyllin voimakkaasti sitä vastaan. Siten H_0:n hyväksyminen merkitsee tosiasiallisesti yleensä vain sitä, että riittävää näyttöä sen hylkäämiseksi ei ole saatu. Tästä syystä jotkut tilastotieteilijät pitävät parempana sanoa, että H_0 "jää voimaan" kuin että se "hyväksytään".

Koska merkitsevyystaso eli hylkäämisvirheen riski on etukäteen kiinnitetty, on luonnollista pyrkiä valitsemaan käytettävä testausmenetelmä eli testisuure siten, että hyväksymisvirheen riski olisi mahdollisimman pieni. Tätä kysymystä tarkastellaan pykälässä 5.5 alla.

5.3.3 P-arvo ei ole todennäköisyys sille, että H_0 pätee. Tilastollisten menetelmien soveltajat ajattelevat joskus virheellisesti, että p-arvo olisi todennäköisyys sille, että nollahypoteesi on tosi eli $\boldsymbol{\theta} \in \Omega_0$. Frekventistisessä päättelyssä (ks. 1.4) ei koskaan voida liittää todennäköisyyslausumia hypoteesiin tai muutenkaan mallin parametriin

† Kirjallisuudessa hylkäämisvirhettä on perinteisesti kutsuttu "I lajin virheeksi" ja hyväksymisvirhettä "II lajin virheeksi".

‡ Tarkka perustelu: Oletetaan, että $T = t(\boldsymbol{Y})$ on käytettävä testisuure. Kiinnitetään $\boldsymbol{\theta} \in \Omega_0$. Olkoon A niiden $t \in \mathbb{R}$ joukko, joille $P_{\boldsymbol{\theta}}\{T \geq t\} \leq \alpha$. Koska tässä esiintyvä todennäköisyys on t:n funktiona vähenevä, niin A on oikealta rajoittamaton väli, ts. a) $A = [t_0, \infty)$ tai b) $A = (t_0, \infty)$, jossa $t_0 \in \mathbb{R}$. Kummassakin tapauksessa on $P_{\boldsymbol{\theta}}\{T \in A\} \leq \alpha$. Tapauksessa a tämä johtuu siitä, että $t_0 \in A$. Tapauksessa b A:n määritelmästä seuraa, että $P_{\boldsymbol{\theta}}\{T \geq t\} \leq \alpha$ kaikilla $t > t_0$, jolloin rajankäynnillä $t \to t_0+$ saadaan $P_{\boldsymbol{\theta}}\{T \in A\} = P_{\boldsymbol{\theta}}\{T > t_0\} \leq \alpha$. Lopuksi todetaan, että jos $p(\boldsymbol{y}) \leq \alpha$, niin p-arvon määritelmän nojalla $t(\boldsymbol{y}) \in A$. Siten $P_{\boldsymbol{\theta}}\{p(\boldsymbol{Y}) \leq \alpha\} \leq P_{\boldsymbol{\theta}}\{T \in A\} \leq \alpha$.

suoraan. P-arvo on kyllä määritelmänsä mukaan eräs todennäköisyys, tai yleisemmin eräiden todennäköisyyksien pienin yläraja, mutta tämä todennäköisyys viittaa testisuureeseen ja sitä kautta aineistoon, ja se on ymmärrettävä toistetun aineistonkeruun mielessä.

Kiteyttäen: pieni p-arvo ei tarkoita, että nollahypoteesi olisi ”epätodennäköinen” vaan että on saatu sellainen aineisto, joka on nollahypoteesin pätiessä epätodennäköinen.

5.3.4 Valintakorjaus. Tilastollisissa tutkimuksissa tapaa toisinaan menettelyä, jossa samaa tai samantapaisia nollahypoteeseja testataan usealla eri testillä ja raportoidaan saaduista p-arvoista pienin eli tilastollisesti merkitsevin. Jos se on pienempi kuin esimerkiksi jokin tavanomaisista merkitsevyystasoista, päätellään sitten, että tutkittavaan ilmiöön liittyy jokin mielenkiintoinen riippuvuus tai muu piirre. Samasta menettelystä on implisiittisesti kyse silloin, kun aineistosta etsimällä etsitään, vaikkapa graafisten menetelmien avulla, epätavalliselta näyttäviä, ”ei-satunnaisia” piirteitä ja sitten varta vasten valituilla testeillä ”osoitetaan”, että ne ovat tilastollisesti merkitseviä.

Meneteltäessä yllä kuvatulla tavalla unohdetaan, että jokainen aineisto on omalla tavallaan ainutkertainen ja varmasti sisältää jotain ”poikkeuksellista” jo pelkästään sattuman oikusta. Esimerkiksi on yleensä täysin mahdollista löytää aineistosta jokin sellainen osa-aineisto, johon sovellettuna tietty testi tuottaa tilastollisesti merkitsevän p-arvon, vaikka koko aineiston testi ei sellaista tuotakaan.

Tilastollisesti pätevä menettelytapa on, että ennen aineistoon tutustumista tai jopa ennen sen keruuta päätetään, millaisia hypoteeseja halutaan tutkia ja mitä testejä tähän tarkoitukseen käytetään. Mikäli samaa nollahypoteesia testataan useammalla testillä ja valitaan saaduista p-arvoista pienin, tähän on tehtävä ns. valintakorjaus, jota kuvataan seuraavassa.

Oletetaan, että nollahypoteesin $H_0\colon \boldsymbol{\theta} \in \Omega_0$ testaamiseksi suoritetaan k eri testiä, joiden p-arvot ovat $p_j(\boldsymbol{y})$, $j = 1, \ldots, k$. Ajatellaan näitä toistetun aineistonkeruun kannalta eli siirrytään vastaaviin satunnaismuuttujiin $P_j = p_j(\boldsymbol{Y})$. Olkoon lisäksi $Q = \min(P_1, \ldots, P_k)$ pienin p-arvo satunnaismuuttujana, ja tutkitaan sen jakaumaa H_0:n pätiessä, erityisesti kuinka suurella todennäköisyydellä se saa pieniä arvoja.

Oletetaan, että H_0 pätee eli $\boldsymbol{\theta} \in \Omega_0$. Kohdassa 5.3.2 todettiin, että $P_{\boldsymbol{\theta}}\{P_j \leq p\} \leq p$ kaikilla $p \in (0, 1)$. Jos testit ovat riippumattomia siinä mielessä, että $P_1, \ldots, P_k \perp\!\!\!\perp$, saadaan kaikilla $q \in (0, 1)$ arvio

$$\begin{aligned} P_{\boldsymbol{\theta}}\{Q \leq q\} &= 1 - P_{\boldsymbol{\theta}}\{Q > q\} \\ &= 1 - P_{\boldsymbol{\theta}}\{P_1 > q, \ldots, P_k > q\} \\ &= 1 - P_{\boldsymbol{\theta}}\{P_1 > q\} \cdots P_{\boldsymbol{\theta}}\{P_k > q\} \\ &\leq 1 - (1 - q)^k. \end{aligned}$$

Yhtäsuuruus pätee, jos $P_{\boldsymbol{\theta}}\{P_j \leq q\} = q$ kaikilla j. Jos siis pienin p-arvo on q, voidaan arvioida, että todellinen merkitsevyystaso on jopa $1 - (1 - q)^k$. Tämä luku on aina suurempi kuin q. Esimerkki: On suoritettu $k = 3$ riippumatonta testiä, ja pienin p-arvo on $q = 0.05$ (melkein merkitsevä). Tällöin valintakorjattu p-arvo on $1 - 0.95^3 \approx 0.14$, joten H_0:n pätiessä näin pieni tai vielä pienempi Q:n arvo saadaan mahdollisesti jopa todennäköisyydellä 0.14. Nollahypoteesia tuskin voi tällä perusteella hylätä.

Yleensä testien riippumattomuudesta ei voi olettaa mitään. Tällöin voidaan kuitenkin arvioida vielä karkeammin

$$P_{\boldsymbol{\theta}}\{Q \le q\} = P_{\boldsymbol{\theta}}\Big(\bigcup_{j=1}^{k}\{P_j \le q\}\Big) \le \sum_{j=1}^{k} P_{\boldsymbol{\theta}}\{P_j \le q\} \le kq.$$

Luku kq on siis yleispätevä yläraja todelliselle merkitsevyystasolle silloin, kun on suoritettu k testiä ja pienin p-arvo on ollut q. Sitä kutsutaan *Bonferroni-korjatuksi* p-arvoksi. Yo. esimerkin tapauksessa se olisi ollut $3 \cdot 0.05 = 0.15$. Bonferroni-korjaus on helpoin ja yleisimmin käytetty tapa suorittaa valintakorjaus eli ottaa huomioon usean testin käytöstä ja pienimmän p-arvon valinnasta aiheutuva vääristymä merkitsevyystasoon.

5.4 Normaalimallin perustestit

Tässä pykälässä palautetaan mieleen jo tilastotieteen johdantokursseilta tutut testit odotusarvolle μ ja varianssille σ^2 mallissa

$$Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp . \tag{5.1}$$

5.4.1 Odotusarvon testi, kun varianssi tunnettu. Oletetaan, että $\sigma^2 = \sigma_0^2 > 0$ on tunnettu luku. Tarkastellaan kaksisuuntaista testausasetelmaa

$$H_0\colon\ \mu = \mu_0, \qquad H_1\colon\ \mu \ne \mu_0, \tag{5.2}$$

jossa μ_0 on annettu kiinteä reaaliluku. Testaus on luontevaa perustaa otoskeskiarvon $\overline{y}$ poikkeamaan μ_0:sta. Käytännön kannalta mukavin muoto testisuureesta on $z = \sqrt{n}(\overline{y} - \mu_0)/\sigma_0$, sillä sitä vastaava satunnaismuuttuja Z noudattaa standardinormaalijakaumaa H_0:n pätiessä.

Nollahypoteesin H_0 kannalta kriittisiä ja vastahypoteesia H_1 tukevia ovat tapaukset, joissa $|z|$ on suuri. Testin p-arvo lasketaan siis kaavalla

$$p = P_{\mu_0}\{|Z| \ge |z|\} = 2[1 - \Phi(|z|)],$$

jossa Φ on standardinormaalijakauman kertymäfunktio. Koska $\Phi(1.96) \approx 0.975$, nähdään esimerkiksi, että H_0 voidaan hylätä ja H_1 hyväksyä 5 %:n merkitsevyystasolla silloin, kun $|z| \ge 1.96$.

Yksisuuntaisessa asetelmassa

$$H_0\colon\ \mu = \mu_0, \qquad H_1\colon\ \mu > \mu_0$$

(jossa H_0 voisi olla myös $\mu \le \mu_0$) kiinnitetään huomiota ainoastaan $\overline{y}$:n poikkeamiin μ_0:sta ylöspäin. Siten p-arvo saadaan normaalijakauman oikeanpuoleisena häntätodennäköisyytenä

$$p = P_{\mu_0}\{Z \ge z\} = 1 - \Phi(z).$$

5.4.2 Odotusarvon testi, kun varianssi kiusaparametri. Tarkastellaan mallia (5.1) ja hypoteeseja (5.2) ilman epärealistista oletusta, että varianssi olisi tunnettu. Nyt myös H_0 on yhdistetty: vastaava joukko Ω_0 on puolisuora $\{(\mu_0, \sigma^2) : 0 < \sigma^2 < \infty\}$.

Tässä asetelmassa muuttuja $Z = \sqrt{n}(\overline{Y} - \mu_0)/\sigma$ riippuu tuntemattomasta parametrista σ^2, joten se ei ole tunnusluku eikä siis kelpaa testisuureeksi. Kaikeksi onneksi σ^2 voidaan korvata harhattomalla estimaattorillaan

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

jolloin päädytään tärkeään t-testisuureeseen

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}.$$

Nollahypoteesin pätiessä tämä noudattaa Studentin t-jakaumaa, jonka vapausasteluku on $n-1$ ja joka on hyvin lähellä standardinormaalijakaumaa, kun n on suuri. Sitä käytetään kuten z-testisuuretta edellä.

5.4.3 Varianssin testi. Tarkastellaan mallin (5.1) puitteissa esim. hypoteeseja

$$H_0\colon\ \sigma^2 = \sigma_0^2, \qquad H_1\colon\ \sigma^2 > \sigma_0^2.$$

Nytkin H_0 on itse asiassa yhdistetty, koska se ei määrittele μ:tä lainkaan.

On luontevaa perustaa testaus σ^2:n estimaattiin eli otosvarianssiin s^2, jonka suuret arvot puhuvat H_0:aa vastaan ja H_1:n puolesta. Koska todennäköisyyslaskennasta tiedetään, että $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ (χ^2-jakauma, jonka vapausasteluku on $n-1$), niin lopulliseksi testisuureeksi on parasta valita $(n-1)s^2/\sigma_0^2$ ja testin p-arvoksi saadaan vastaava oikeanpuoleinen häntätodennäköisyys χ^2_{n-1}-jakaumasta.

5.5 Testin voima ja Neyman–Pearson-teoria

5.5.1 Kriittiset alueet ja voimafunktio. Tarkastellaan mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, jonka parametriavaruus on Ω ja johon liittyen on asetettu nolla- ja vastahypoteesi

$$H_0\colon\ \boldsymbol{\theta} \in \Omega_0, \qquad H_1\colon\ \boldsymbol{\theta} \in \Omega_1. \tag{5.3}$$

Lähestytään testausta jyrkän päätösteoreettisesti: on joko hyväksyttävä H_0 tai hylättävä se ja hyväksyttävä H_1. Edellä kohdassa 5.3.2 selostettiin, kuinka tällöin tehdään oikea päätös tai hylkäämis- tai hyväksymisvirhe sen mukaan, päteekö H_0 todellisuudessa vai ei.

Olkoon $\alpha \in (0,1)$ valittu merkitsevyystaso, esimerkiksi $\alpha = 0.05$. Olkoon lisäksi t käytettävä testisuure ja p siitä laskettu p-arvo. Merkitään C_α:lla niiden aineistojen $\boldsymbol{y}$ joukkoa, joiden havaitseminen johtaa H_0:n hylkäämiseen, ts.

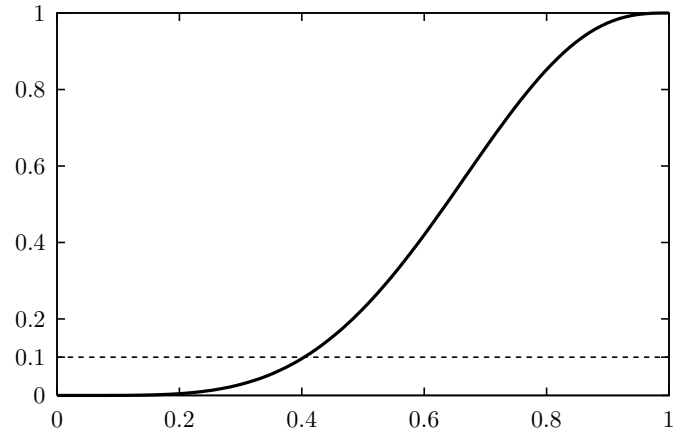
$$C_\alpha = \{\boldsymbol{y} : p(\boldsymbol{y}) \le \alpha\}.$$

Tätä joukkoa kutsutaan testisuureen t indusoimaksi α-tasoiseksi *kriittiseksi alueeksi*. Jos H_0 pätee, on H_0:n hylkääminen tietysti hylkäämisvirhe, ja kuten kohdassa 5.3.2 todettiin, merkitsevyystaso α on yläraja tämän todennäköisyydelle:

$$P_{\boldsymbol{\theta}}\{\boldsymbol{Y} \in C_\alpha\} = P_{\boldsymbol{\theta}}(H_0 \text{ hylätään}) \le \alpha, \qquad \boldsymbol{\theta} \in \Omega_0.$$

Toista virhelajia eli hyväksymisvirhettä on mukavinta lähestyä testin voiman käsitteen avulla. Edellä kuvatussa asetelmassa määritellään, että testisuureen t α-tasoinen *voima* tai *voimafunktio* on

$$\pi_\alpha(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}\{\boldsymbol{Y} \in C_\alpha\} = P_{\boldsymbol{\theta}}(H_0 \text{ hylätään}), \qquad \boldsymbol{\theta} \in \Omega.$$



Kuva 5.1. Testisuureen $K \sim Bin(7, \theta)$ 0.1-tasoinen voima, kun $H_0 \colon \theta \leq 0.4$ ja $H_1 \colon \theta > 0.4$.

Tämä merkitsee, että

$$\pi_\alpha(\boldsymbol{\theta}) = \begin{cases} \text{hylkäämisvirheen todennäköisyys, kun } \boldsymbol{\theta} \in \Omega_0, \\ 1 - \text{hyväksymisvirheen todennäköisyys, kun } \boldsymbol{\theta} \notin \Omega_0. \end{cases}$$

Toimittaessa valitulla merkitsevyystasolla α on siis aina $\pi_\alpha(\boldsymbol{\theta}) \leq \alpha$ joukossa Ω_0. Voimafunktioon onkin aiheellista kiinnittää huomiota lähinnä vastahypoteesijoukossa Ω_1. Perustavoitteena on, että käytettävän testisuureen voima olisi siellä mahdollisimman suuri eli lähellä ykköstä. Testin tulisi siis hylätä H_0 mahdollisimman suurella todennäköisyydellä silloin kun H_1 pätee. Tätä tavoitetta tarkastellaan lähemmin seuraavien esimerkkien jälkeen.

5.5.2 Esimerkki: toistokoemalli. Tarkastellaan toistokoemallia $K \sim Bin(7, \theta)$. Testattavana on $H_0 \colon \theta \leq 0.4$ vastaan $H_1 \colon \theta > 0.4$, ja päätetään toimia merkitsevyystasolla $\alpha = 0.1$. Luonteva testisuure on k itse; sen suuret arvot ovat nollahypoteesille kriittisiä, ja sen pistetodennäköisyydet ovat

$$P_\theta\{K = k\} = \binom{7}{k} \theta^k (1-\theta)^{7-k}, \qquad k = 0, 1, \ldots, 7.$$

Kriittisen alueen selvittämiseksi lasketaan oikeanpuoleisia häntätodennäköisyyksiä $P_\theta\{K \geq k\}$ nollahypoteesivälillä $\theta \leq 0.4$. Kiinteällä k tällainen todennäköisyys on ilmeisesti suurimmillaan silloin, kun $\theta = 0.4$. Koska

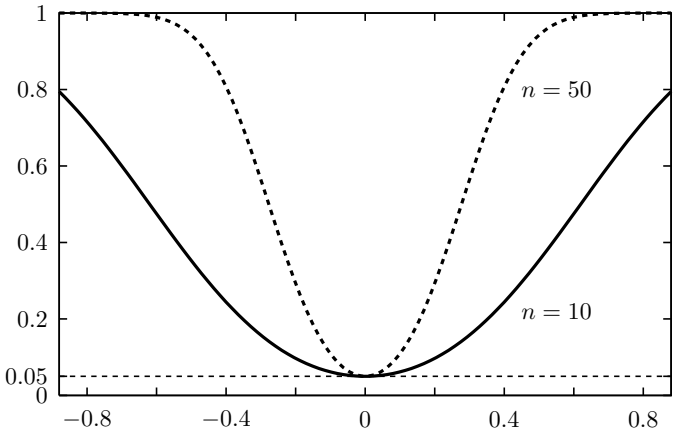
$$\begin{aligned} P_{0.4}\{K \geq 5\} &= 0.0774 + 0.0172 + 0.0016 = 0.0962 \leq 0.1, \\ P_{0.4}\{K \geq 4\} &= 0.1935 + 0.0962 = 0.2897 > 0.1, \end{aligned}$$

nähdään, että kriittinen alue on $\{k : k \geq 5\} = \{5, 6, 7\}$. Huomaa, että testisuureen jakauman diskreettisyydestä johtuen tämän todellinen ”koko” eli hylkäämisvirheen todennäköisyys on suurimmillaankin aidosti pienempi kuin valittu merkitsevyystaso.

Testin 0.1-tasoinen voimafunktio on nyt määritelmän mukaan

$$\pi_{0.1}(\theta) = P_\theta\{K \geq 5\} = 21 \cdot \theta^5 (1-\theta)^2 + 7 \cdot \theta^6 (1-\theta) + \theta^7.$$

Tämän kuvaaja on piirretty kuvaan 5.1.



Kuva 5.2. Mallin $Y_1, \ldots, Y_n \sim N(\mu, 1) \perp\!\!\!\perp$ z-testisuureen 0.05-tasoinen voimafunktio kaksisuuntaisessa testissä $H_0\colon \mu = 0$ vastaan $H_1\colon \mu \neq 0$, kun $n = 10$ ja $n = 50$.

5.5.3 Esimerkki: normaalimallin odotusarvo. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2) \perp\!\!\!\perp$, jossa $\sigma_0^2 > 0$ on tunnettu. Testausasetelma olkoon kaksisuuntainen

$$H_0\colon\ \mu = 0, \qquad H_1\colon\ \mu \neq 0.$$

Kohdassa 5.4.1 opittiin, että testisuureena on luontevaa käyttää $z = z(\boldsymbol{y}) = \sqrt{n}\overline{y}/\sigma_0$ ja testin p-arvo on $p = 2[1 - \Phi(|z|)]$, jossa Φ on standardinormaalijakauman kertymäfunktio.

Kiinnitetään merkitsevyystaso $\alpha \in (0, 1)$. Tällöin H_0 hylätään ja H_1 hyväksytään täsmälleen silloin kun $p \leq \alpha$ eli $\Phi(|z|) \geq 1 - \alpha/2$. Jos $z_{\alpha/2}$ on se piste, jolle $\Phi(z_{\alpha/2}) = 1 - \alpha/2$, niin tämä toteutuu jos ja vain jos $|z| \geq z_{\alpha/2}$ eli yhtäpitävästi $|\overline{y}| \geq z_{\alpha/2}\sigma_0/\sqrt{n}$. Testin kriittinen alue on siis

$$C_\alpha = \{\boldsymbol{y} : |z(\boldsymbol{y})| \geq z_{\alpha/2}\} = \left\{\boldsymbol{y} : |\overline{y}| \geq \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right\}.$$

Lasketaan vielä testin voimafunktio. Tätä varten on selvitettävä z-testisuuretta vastaavan satunnaismuuttujan $Z = \sqrt{n}\overline{Y}/\sigma_0$ jakauma myös muulloin kuin H_0:n pätiessä. Koska $\overline{Y} \sim N(\mu, \sigma_0^2/n)$, nähdään, että $Z \sim N(\sqrt{n}\mu/\sigma_0, 1)$. Siten α-tasoinen voimafunktio on

$$\begin{aligned}\pi_\alpha(\mu) &= P_\mu\{|Z| \geq z_{\alpha/2}\} = P_\mu\{Z \geq z_{\alpha/2} \text{ tai } Z \leq -z_{\alpha/2}\} \\ &= 1 - \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}\mu}{\sigma_0}\right) + \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}\mu}{\sigma_0}\right).\end{aligned}$$

Esimerkiksi jos $\alpha = 0.05$, niin $z_{0.025} \approx 1.96$ ja

$$\pi_{0.05}(\mu) = 1 - \Phi\left(1.96 - \frac{\sqrt{n}\mu}{\sigma_0}\right) + \Phi\left(-1.96 - \frac{\sqrt{n}\mu}{\sigma_0}\right).$$

Kuvaan 5.2 on piirretty voimafunktion $\pi_{0.05}$ kuvaaja tapauksessa $\sigma_0^2 = 1$, kun $n = 10$ tai $n = 50$. Kiinnitä huomiota siihen, miten havaintojen lukumäärä vaikuttaa voimafunktion kulkuun. Mitä enemmän havaintoja, sitä todennäköisemmin testi havaitsee pienetkin poikkeamat nollahypoteesista $\mu = 0$.

5.5.4 Testien voiman vertailu ja Neyman–Pearson-apulause. Oletetaan, että t ja t' ovat kaksi testisuuretta, joita voidaan käyttää samojen hypoteesien (5.3) testaukseen. Olkoot niiden α-tasoiset voimafunktiot $\pi_\alpha(\boldsymbol{\theta}; t)$ ja $\pi_\alpha(\boldsymbol{\theta}; t')$.

Asetetaan seuraavat luonnolliset määritelmät:

a) t on *voimakkaampi* kuin t' pisteessä $\boldsymbol{\theta} \in \Omega_1$, jos

$$\pi_\alpha(\boldsymbol{\theta}; t) \geq \pi_\alpha(\boldsymbol{\theta}; t'). \tag{5.4}$$

b) t on *tasaisesti voimakkaampi* kuin t', jos (5.4) pätee kaikilla $\boldsymbol{\theta} \in \Omega_1$.

c) t on *voimakkain* testisuure pisteessä $\boldsymbol{\theta}$, jos (5.4) pätee kaikilla testisuureilla t'.

d) t on *tasaisesti voimakkain* testisuure, jos (5.4) pätee kaikilla $\boldsymbol{\theta} \in \Omega_1$ ja kaikilla testisuureilla t'.

Parasta tietenkin olisi, jos jokaiseen testausasetelmaan olisi löydettävissä tasaisesti voimakkain testi. Näin ei kuitenkaan yleensä ole aivan yksinkertaisimpia tilanteita lukuunottamatta. Erästä tällaista tilannetta käsittelee seuraava kuuluisa tulos:

Neyman–Pearson-apulause. *Tarkastellaan tilastollista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ ja yksinkertaisia hypoteeseja $H_0 \colon \boldsymbol{\theta} = \boldsymbol{\theta}_0$ ja $H_1 \colon \boldsymbol{\theta} = \boldsymbol{\theta}_1$. Merkitään*

$$v(\boldsymbol{y}) = \frac{L(\boldsymbol{\theta}_1; \boldsymbol{y})}{L(\boldsymbol{\theta}_0; \boldsymbol{y})} = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}_0)}.$$

Tällöin v on voimakkain testisuure em. hypoteeseille jokaisella sellaisella merkitsevyystasolla $\alpha \in (0, 1)$, jolle pätee $P_{\boldsymbol{\theta}_0}\{v(\boldsymbol{Y}) \geq v_\alpha\} = \alpha$ jollakin v_α.†

Tässä $L(\boldsymbol{\theta}; \boldsymbol{y}) = c(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ on mallin uskottavuusfunktio. Testisuuretta $v(\boldsymbol{y})$ kutsutaan *uskottavuusosamäärän testisuureeksi*. Ekvivalentti testi eli samat p-arvot ja kriittiset alueet saadaan myös mistä tahansa sen aidosti kasvavasta muunnoksesta. Myöhemmin osoittautuu käyttökelpoiseksi erityisesti muunnos

$$2 \log v(\boldsymbol{y}) = 2 \left[l(\boldsymbol{\theta}_1; \boldsymbol{y}) - l(\boldsymbol{\theta}_0; \boldsymbol{y}) \right],$$

jossa $l(\boldsymbol{\theta}; \boldsymbol{y})$ on log-uskottavuusfunktio.

Käytännön sovellusten kannalta Neyman–Pearson-apulauseen asetelma on tietenkin epärealistinen, koska yleensä ainakin vastahypoteesit ovat yhdistettyjä. Eräissä tilanteissa kuitenkin osoittautuu, että saatava testi on tasaisesti voimakkain yksisuuntaiselle yhdistetylle vastahypoteesille.

5.5.5 Esimerkki: uskottavuusosamäärän testi normaalimallin odotusarvolle. Tarkastellaan jälleen mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$. Tämän uskottavuusfunktioksi on esimerkissä 2.1.4 saatu

$$L(\mu; \boldsymbol{y}) = \exp\left\{ -\frac{n(\overline{y} - \mu)^2}{2\sigma_0^2} \right\}.$$

† Tässä on oletettava, että käytettävä merkitsevyystaso α ”saavutetaan” siinä mielessä, että v:n avulla voidaan rajata kriittinen alue, jonka todennäköisyys on tasan α eikä aidosti sen alle (vrt. 5.5.1). Jatkuvassa tapauksessa näin on mille tahansa α mutta diskreetissä tapauksessa yleensä ei (vrt. 5.5.2). Tämä rajoitus voitaisiin poistaa ottamalla käyttöön ns. satunnaistetut testit, joiden käsittely tällä kurssilla kuitenkin sivuutetaan.

Etsitään Neyman–Pearson-apulauseen avulla voimakkain testisuure nollahypoteesille $H_0\colon \mu = \mu_0$, kun vastahypoteesi on $H_1\colon \mu = \mu_1$, jossa $\mu_1 > \mu_0$. Uskottavuusosamäärä saadaan muotoon

$$\begin{aligned}\frac{L(\mu_1; \boldsymbol{y})}{L(\mu_0; \boldsymbol{y})} &= \exp\biggl\{-\frac{n}{2\sigma_0^2}[(\overline{y}-\mu_1)^2 - (\overline{y}-\mu_0)^2]\biggr\} \\ &= \exp\biggl\{\frac{n}{2\sigma_0^2}[2\overline{y}(\mu_1-\mu_0) - (\mu_1^2-\mu_0^2)]\biggr\}.\end{aligned}$$

Koska $\mu_1 - \mu_0 > 0$, havaitaan, että tämä on otoskeskiarvon $\overline{y}$ suhteen aidosti kasvava funktio. Siispä uskottavuusosamäärän testin kriittinen alue $L(\mu_1; \boldsymbol{y})/L(\mu_0; \boldsymbol{y}) \geq c$ on yhtäpitävästi muotoa $\overline{y} \geq c'$ tai muotoa $z \geq c''$, jossa $z = \sqrt{n}(\overline{y}-\mu_0)/\sigma_0^2$. Johtopäätös onkin, että tuttu z-testisuure (ks. 5.4.1) on itse asiassa voimakkain testisuure yo. asetelmassa.

Lopuksi huomattakoon, että saatavan testin kriittinen alue ei mitenkään riipu kiinnitetystä vastahypoteesipisteestä μ_1, kunhan vain $\mu_1 > \mu_0$. Siten voidaan päätellä, että kyseessä on tasaisesti voimakkain testisuure hypoteeseille

$$H_0\colon\ \mu = \mu_0, \qquad H_1\colon\ \mu > \mu_0.$$

5.5.6 Monotoninen uskottavuusosamäärä. Edellisen esimerkin lopussa havaittu ilmiö on sen verran hyödyllinen, että se ansaitsee oman nimityksen.

Tarkastellaan yleisesti mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$, jonka parametri on yksiulotteinen. Siis Ω on reaaliakselin osajoukko, yleensä väli. Sanotaan, että tällä mallilla on *monotoninen uskottavuusosamäärä*, mikäli on olemassa tunnusluku $t(\boldsymbol{y})$ siten, että

$$v(\boldsymbol{y}) = \frac{L(\theta'; \boldsymbol{y})}{L(\theta; \boldsymbol{y})} = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta')}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)}$$

riippuu aineistosta vain $t(\boldsymbol{y})$:n välityksellä ja on lisäksi tämän aidosti kasvava funktio kaikilla $\theta, \theta' \in \Omega$, joille $\theta < \theta'$.

Tässä tilanteessa $t(\boldsymbol{y})$ on tasaisesti voimakkain testisuure mille tahansa yksinkertaiselle nollahypoteesille $H_0\colon \theta = \theta_0$ ja yksisuuntaiselle yhdistetylle vastahypoteesille $H_1\colon \theta > \theta_0$ (vast. $\theta < \theta_0$). Nollahypoteesille kriittisiä ovat $t(\boldsymbol{y})$:n suuret arvot (vast. pienet arvot).

5.5.7 Esimerkki: eksponenttimalli. Olkoot $Y_1, \ldots, Y_n \sim \mathit{Exp}(\lambda)$ $\perp\!\!\!\perp$. Tällöin ytf on $f_{\boldsymbol{Y}}(\boldsymbol{y}; \lambda) = \lambda^n \exp\{-\lambda \sum_{i=1}^n y_i\}$. Jos $0 < \lambda < \lambda'$, on siis

$$\frac{f_{\boldsymbol{Y}}(\boldsymbol{y}; \lambda')}{f_{\boldsymbol{Y}}(\boldsymbol{y}; \lambda)} = \left(\frac{\lambda'}{\lambda}\right)^n \exp\biggl\{-(\lambda'-\lambda)\sum_{i=1}^n y_i\biggr\}.$$

Koska $\lambda' - \lambda > 0$, tämä on tunnusluvun $-\sum_{i=1}^n y_i$ aidosti kasvava funktio. Mallilla on siis monotoninen uskottavuusosamäärä.

Testattaessa hypoteesia $H_0\colon \lambda = \lambda_0$ vastaan $H_1\colon \lambda > \lambda_0$ tasaisesti voimakkaimman testin antaa näin ollen testisuure $t(\boldsymbol{y}) = \sum_{i=1}^n y_i$, jonka pienet arvot ovat kriittisiä (huomaa etumerkin vaihto, joka kääntää kriittisten arvojen suunnan). Testin p-arvo on vasemmanpuoleinen häntätodennäköisyys

$$p = P_{\lambda_0}\{T \leq t(\boldsymbol{y})\},$$

jossa $T = \sum_{i=1}^n Y_i$. Todennäköisyyslaskennan kurssilla on opittu, että riippumattomien $\mathit{Exp}(\lambda_0)$-muuttujien summana T noudattaa gammajakaumaa $G(n, \lambda_0)$ H_0:n pätiessä.

Jos käytössä on vain normaalit todennäköisyystaulukot eikä esim. tietokoneohjelmaa, joka suoraan laskisi gammajakauman kertymäfunktion arvoja, on vielä tarpeen siirtyä muuttujaan $2\lambda_0 T$, joka noudattaa χ^2_{2n}-jakaumaa.

5.5.8 Neyman–Pearson-apulauseen todistus. Käsitellään diskreetti tapaus. Jatkuva tapaus on samanlainen; summat vain on korvattava integraaleilla.

Merkitään $C = \{\boldsymbol{y} : v(\boldsymbol{y}) \geq v_\alpha\}$. Tällöin

$$\begin{cases} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1) \geq v_\alpha f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_0), & \boldsymbol{y} \in C, \\ f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1) < v_\alpha f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_0), & \boldsymbol{y} \notin C. \end{cases} \tag{5.5}$$

Olkoon $t(\boldsymbol{y})$ jokin toinen testisuure ja D sen indusoima α-tasoinen kriittinen alue. Merkitään $C^* = C \setminus D$ ja $D^* = D \setminus C$ (piirrä kuva!). Tällöin pätee

$$\begin{cases} \alpha = P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C\} = P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C^*\} + P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C \cap D\}, \\ \alpha \geq P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in D\} = P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in D^*\} + P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C \cap D\}, \end{cases}$$

josta seuraa, että $P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C^*\} \geq P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in D^*\}$ eli

$$\sum_{\boldsymbol{y} \in C^*} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_0) \geq \sum_{\boldsymbol{y} \in D^*} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_0).$$

Käyttämällä epäyhtälöitä (5.5) ja havaitsemalla, että $C^* \subset C$ mutta $D^* \cap C = \emptyset$, saadaan

$$\sum_{\boldsymbol{y} \in C^*} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1) \geq \sum_{\boldsymbol{y} \in D^*} f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}_1)$$

eli $P_{\boldsymbol{\theta}_1}\{\boldsymbol{Y} \in C^*\} \geq P_{\boldsymbol{\theta}_1}\{\boldsymbol{Y} \in D^*\}$. Lisäämällä tämän kummallekin puolelle todennäköisyys $P_{\boldsymbol{\theta}_1}\{\boldsymbol{Y} \in C \cap D\}$ päädytään epäyhtälöön $P_{\boldsymbol{\theta}_1}\{\boldsymbol{Y} \in C\} \geq P_{\boldsymbol{\theta}_1}\{\boldsymbol{Y} \in D\}$, mikä merkitsee, että v:n voima on ainakin yhtä suuri kuin t:n voima vastahypoteesipisteessä $\boldsymbol{\theta}_1$. Apulause on siis todistettu.

5.6 Uskottavuusfunktioon perustuvia testejä I

5.6.1 Johdanto. Realistisissa testausasetelmissa ainakin vastahypoteesi on yhdistetty eikä yleensä ole mahdollista löytää tasaisesti voimakkainta testiä. Poikkeuksen muodostavat lähinnä ne mallit, joiden parametri on yksiulotteinen ja joilla on monotoninen uskottavuusosamäärä. Mikäli mallin parametri on useampiulotteinen, on normaalisti myös nollahypoteesi yhdistetty, mikä entisestään mutkistaa testisuureen valintaa. Herääkin kysymys, miten tällaisessa tilanteessa voidaan löytää ylipäätään mitään järkeviä ja käyttökelpoisia testisuureita.

Osoittautuu, että uskottavuusfunktion pohjalta voidaan muodostaa ainakin kolme erilaista testisuuretta, jotka soveltuvat varsin yleisten testien suorittamiseen. Näillä testisuureilla on suuri merkitys käytännön tilastoanalyyseissa. Ominaista tarkasteltavalle tilanteelle on se, että testisuureiden nollahypoteesijakaumia harvoin osataan eksaktisti johtaa. Siten joudutaan turvautumaan likimääräisiin, asymptoottisiin jakaumatuloksiin ja asettamaan tarkasteltavalle mallille tiettyjä säännöllisyysvaatimuksia aivan niin kuin pykälässä 3.6, jossa puhuttiin suurimman uskottavuuden estimaattorien asymptotiikasta.

Tässä pykälässä käsitellään seuraavaa yksinkertaistettua tilannetta:

a) Havainnot $Y_1, \ldots, Y_n$ ovat riippumattomia ja samoin jakautuneita.

b) Parametri θ on yksiulotteinen eli parametriavaruus Ω on tyypillisesti reaalilukuväli.

c) Nollahypoteesi on yksinkertainen $H_0\colon \theta = \theta_0$. Vastahypoteesi (mikäli se halutaan spesifioida) on pääsääntöisesti kaksisuuntainen $H_1\colon \theta \neq \theta_0$, mutta myös yksisuuntaista asetelmaa voidaan tarkastella.

d) Malli täyttää riittävät säännöllisyysvaatimukset (vrt. 3.6).

Oletusta a voidaan lieventää huomattavastikin, mutta tällä kurssilla emme perehdy asiaan tarkemmin. Useampiulotteisen parametrin tapausta käsitellään seuraavassa pykälässä.

5.6.2 Uskottavuusosamäärän testisuure. Ensimmäinen esiteltävistä testisuureista on muunnelma Neyman–Pearson-apulauseen testisuureesta. Nyt kun vastahypoteesi ei ole yksinkertainen vaan yhdistetty, verrataan uskottavuusfunktion arvoa nollahypoteesipisteessä θ_0 sen globaaliin maksimiarvoon eli arvoon suurimman uskottavuuden pisteessä $\hat{\theta} = \hat{\theta}(\boldsymbol{y})$. Intuitiivisesti on selvää, että jos uskottavuus pisteessä θ_0 on huomattavasti pienempi kuin uskottavuus pisteessä $\hat{\theta}$, niin aineisto todistaa H_0:aa vastaan.

Jakaumatarkastelujen kannalta mukavimmaksi testisuureen muodoksi osoittautuu

$$r(\boldsymbol{y}) = 2 \log \frac{L(\hat{\theta}; \boldsymbol{y})}{L(\theta_0; \boldsymbol{y})} = 2\,[l(\hat{\theta}; \boldsymbol{y}) - l(\theta_0; \boldsymbol{y})],$$

jossa $l(\theta; \boldsymbol{y})$ on log-uskottavuusfunktio. Tätä kutsutaan *uskottavuusosamäärän testisuureeksi*. Huomaa, että se on aina ei-negatiivinen ja että sen suuret arvot ovat H_0:lle kriittisiä.

5.6.3 Esimerkki: normaalimalli, kun varianssi tunnettu. Joskus muuttujan $r(\boldsymbol{Y})$ eksakti nollahypoteesijakauma osataan helposti johtaa. Tarkastellaan esimerkkinä mallia $Y_1, \ldots, Y_n \sim N(\mu, 1)$ $\perp\!\!\!\perp$. Tällöin (ks. 2.1.7) $l(\mu; \boldsymbol{y}) = -n(\overline{y} - \mu)^2/2$ ja $\hat{\mu} = \overline{y}$, joten jos testattavana on nollahypoteesi $H_0\colon \mu = \mu_0$, niin $r(\boldsymbol{y}) = n(\overline{y} - \mu_0)^2$.

Jos H_0 pätee, on tunnetusti $\overline{Y} \sim N(\mu_0, 1/n)$, joten $\sqrt{n}(\overline{Y} - \mu_0) \sim N(0, 1)$. Todennäköisyyslaskennassa määritellään, että standardinormaalijakautuneen satunnaismuuttujan neliö on χ^2-jakautunut yhdellä vapausasteella, joten siis $r(\boldsymbol{Y}) \sim \chi^2_1$.

5.6.4 Uskottavuusosamäärän testisuureen asymptoottinen jakauma. Palataan nyt kohdan 5.6.1 yleiseen asetelmaan. Tällöin osoittautuu, että edellisen esimerkin nollahypoteesijakauma on voimassa asymptoottisesti (eli raja-arvomielessä kun $n \to \infty$): säännöllisyysehtojen vallitessa

$$r(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_1, \qquad \text{kun } H_0 \text{ pätee.} \tag{5.6}$$

Tämä merkitsee käytännössä sitä, että kun havaintojen lukumäärä n on suuri, uskottavuusosamäärän testin p-arvo saadaan approksimatiivisesti χ^2_1-jakauman oikeana häntätodennäköisyytenä:

$$p = P_{H_0}\{r(\boldsymbol{Y}) \geq r(\boldsymbol{y})\} \approx P\{\chi^2_1 \geq r(\boldsymbol{y})\}.$$

Siitä, kuinka paljon havaintoja on oltava, jotta approksimaatio olisi riittävän tarkka, on mahdotonta antaa mitään nyrkkisääntöä. Joissakin malleissa jo muutama kymmenen voi olla riittävä määrä, joissakin toisissa havaintoja tarvitaan satoja tai tuhansia.

**Tuloksen* (5.6) *todistuksen idea.* Lähtökohdaksi otetaan Taylorin kaavasta saatava log-uskottavuusfunktion "normaaliapproksimaatio"

$$l(\theta;\boldsymbol{y}) - l(\hat{\theta};\boldsymbol{y}) \approx -\tfrac{1}{2} j(\hat{\theta};\boldsymbol{y})(\theta-\hat{\theta})^2,$$

joka johdettiin kohdan 2.4.2 kaavassa (2.7). Kun H_0 pätee ja n on suuri, seuraa tästä $r(\boldsymbol{Y}) \approx j(\hat{\theta};\boldsymbol{Y})(\hat{\theta}-\theta_0)^2$, sillä su-estimaattorin tarkentuvuuden (ks. 3.6.2) nojalla $\hat{\theta}$ on lähellä θ_0:aa. Käyttämällä edelleen $\hat{\theta}$:n tarkentuvuutta ja lisäksi suurten lukujen lakia (vrt. lauseen 3.6.5 todistuksen loppu), nähdään, että $j(\hat{\theta};\boldsymbol{Y}) \approx i(\theta_0)$, jossa $i(\theta_0)$ on odotettu eli Fisherin informaatio pisteessä θ_0. Siten $r(\boldsymbol{Y}) \approx i(\theta_0)(\hat{\theta}-\theta_0)^2$. Kohdassa 3.6.5 osoitettiin, että $\hat{\theta} \underset{\text{as}}{\sim} N(\theta_0, 1/i(\theta_0))$, joten

$$\sqrt{i(\theta_0)}(\hat{\theta}-\theta_0) \underset{\text{as}}{\sim} N(0,1).$$

Koska χ^2_1-jakauma on määritelmänsä mukaan standardinormaalijakauman neliö, saadaan $r(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_1$. □

5.6.5 Waldin testisuure. Toinen tärkeä testisuure perustuu erotukseen $\hat{\theta}-\theta_0$. Jos tämä erotus on itseisarvoltaan suuri eli aineiston valossa uskottavin parametriarvo on kaukana nollahypoteesipisteestä, on selvästikin syytä asettaa H_0 kyseenalaiseksi.

Käyttökelpoinen testisuure, jonka asymptoottinen jakauma tunnetaan, on yo. erotuksesta varsin helposti muodostettavissa ja se perustuu suoraan su-estimaattorien asymptoottiseen normaalisuuteen. Määritellään

$$w^{1/2}(\boldsymbol{y}) = \sqrt{i(\theta_0)}(\hat{\theta}-\theta_0), \qquad w(\boldsymbol{y}) = i(\theta_0)(\hat{\theta}-\theta_0)^2.$$

Edellisen todistuksen lopussa todettiin näitä vastaavista satunnaismuuttujista, että

$$w^{1/2}(\boldsymbol{Y}) \underset{\text{as}}{\sim} N(0,1), \qquad w(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_1, \qquad \text{kun } H_0 \text{ pätee.}$$

Muuttujia $w^{1/2}$ ja w kutsutaan *Waldin testisuureiksi*. Vaihtoehtoisia versioita niistä saadaan korvaamalla luku $i(\theta_0)$ jollakin luvuista $i(\hat{\theta})$, $j(\theta_0;\boldsymbol{y})$ ja $j(\hat{\theta};\boldsymbol{y})$. Seuraavassa pykälässä tarkasteltavan yleistyksen ja myöhemmän luottamusvälitulkinnan kannalta $i(\hat{\theta})$ ja $j(\hat{\theta};\boldsymbol{y})$ ovat tavallaan luontevimmat valinnat.

Testisuure $w^{1/2}$ soveltuu yksisuuntaisen testin suorittamiseen. Jos vastahypoteesi on $H_1\colon \theta > \theta_0$, muodostuu kriittinen alue $w^{1/2}$:n suurista positiivisista arvoista ja siten approksimatiivinen p-arvo saadaan standardinormaalijakauman oikeanpuoleisena häntätodennäköisyytenä. Jos taas vastahypoteesi on $H_1\colon \theta < \theta_0$, kriittisiä ovat $w^{1/2}$:n voimakkaasti negatiiviset arvot. Kaksisuuntaisen vastahypoteesin $H_1\colon \theta \neq \theta_0$ testauksessa voidaan käyttää joko testisuuretta $w^{1/2}$ ja kaksipuolista kriittistä aluetta tai yhtäpitävästi testisuureen neliöityä versiota w, jolloin approksimatiivinen p-arvo saadaan χ^2_1-jakaumasta aivan kuten uskottavuusosamäärän testissä edellä.

5.6.6 Raon testisuure. Kolmas uskottavuusfunktioon pohjautuva yleinen testi on *Raon pistemäärätesti*. Siinä testataan, poikkeaako nollahypoteesin mukainen pistemäärä $l'(\theta_0;\boldsymbol{y})$ liian paljon nollasta. Jos nimittäin H_0 pätee, on säännöllisessä mallissa pistemäärän odotusarvo θ_0:ssa nolla: $E[l'(\theta_0;\boldsymbol{Y})] = 0$ (ks. 2.5.3). Kohdassa 3.6.7 vieläpä todettiin, että $l'(\theta_0;\boldsymbol{Y}) \underset{\text{as}}{\sim} N(0, i(\theta_0))$ H_0:n pätiessä. Siispä jos määritellään

$$u^{1/2}(\boldsymbol{y}) = \frac{l'(\theta_0;\boldsymbol{y})}{\sqrt{i(\theta_0)}}, \qquad u(\boldsymbol{y}) = \frac{l'(\theta_0;\boldsymbol{y})^2}{i(\theta_0)},$$

nähdään, että

$$u^{1/2}(\boldsymbol{Y}) \underset{\text{as}}{\sim} N(0,1), \qquad u(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_1, \qquad \text{kun } H_0 \text{ pätee.}$$

Testisuureita $u^{1/2}$ ja u käytetään aivan samalla tavalla kuin Waldin testisuureita. Niiden määritelmässä esiintyvä Fisherin informaatio $i(\theta_0)$ voidaan korvata havaitulla informaatiolla $j(\theta_0; \boldsymbol{y})$. Periaatteessa myös lukuja $i(\hat{\theta})$ ja $j(\hat{\theta}; \boldsymbol{y})$ voitaisiin käyttää, mutta näin harvoin tehdään. Raon testisuureen etu onkin siinä, että sen muodostaminen ei edellytä su-estimaatin $\hat{\theta}$ laskemista.

5.6.7 Esimerkki: eksponenttimalli. Muodostetaan edellä mainitut kolme testisuuretta riippumattomien eksponenttihavaintojen

$$Y_1, \ldots, Y_n \sim Exp(1/\mu) \perp\!\!\!\perp$$

mallissa. Nollahypoteesi on $H_0 \colon \mu = \mu_0$, jossa $\mu_0 > 0$.

Mallia vastaava ytf on $f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu) = \mu^{-n} e^{-n\overline{y}/\mu}$, joten log-uskottavuusfunktio on

$$l(\mu; \boldsymbol{y}) = -n \log \mu - n\overline{y}/\mu.$$

Lisäksi tiedetään, että $\hat{\mu} = \overline{y}$ (ks. 2.3.2). Näin ollen uskottavuusosamäärän testisuure saa muodon

$$\begin{aligned} r(\boldsymbol{y}) &= 2\left[l(\hat{\mu}; \boldsymbol{y}) - l(\mu_0; \boldsymbol{y})\right] \\ &= 2n\left[\overline{y}/\mu_0 - \log(\overline{y}/\mu_0) - 1\right] \end{aligned}$$

Derivoimalla saadaan

$$\begin{aligned} l'(\mu; \boldsymbol{y}) &= -n/\mu + n\overline{y}/\mu^2 = n(\overline{y} - \mu)/\mu^2, \\ l''(\mu; \boldsymbol{y}) &= n/\mu^2 - 2n\overline{y}/\mu^3 = n(\mu - 2\overline{y})/\mu^3, \end{aligned}$$

joten mallin odotettu eli Fisherin informaatio on $i(\mu) = E[-l''(\mu; \boldsymbol{Y})] = n/\mu^2$. Muodostetaan Waldin testisuure $w^{1/2}$ käyttämällä siinä Fisherin informaation arvoa suurimman uskottavuuden pisteessä:

$$w^{1/2}(\boldsymbol{y}) = \sqrt{i(\hat{\mu})}(\hat{\mu} - \mu_0) = \frac{\sqrt{n}(\overline{y} - \mu_0)}{\overline{y}}.$$

Raon pistemäärätestisuureeksi puolestaan saadaan

$$u^{1/2}(\boldsymbol{y}) = \frac{l'(\mu_0; \boldsymbol{y})}{\sqrt{i(\mu_0)}} = \frac{\sqrt{n}(\overline{y} - \mu_0)}{\mu_0}.$$

Sovelletaan testisuureita $w^{1/2}$ ja $u^{1/2}$ numeeriseen esimerkkiin, jossa μ on sähkölaitteen keskimääräinen kestoikä tunneissa (vrt. 1.2.2). Valmistaja väittää, että keskimääräinen kestoikä on (ainakin) tuhat tuntia. Nollahypoteesi on siis $H_0 \colon \mu = 1000$ tai yhtä hyvin $\mu \geq 1000$, ja vastahypoteesiksi on luontevaa asettaa $H_1 \colon \mu < 1000$. Oletetaan, että otoksen koko on $n = 50$ ja siinä kestoikien keskiarvoksi havaitaan $\overline{y} = 800$. Testisuureiden arvot ovat siis

$$\begin{aligned} w^{1/2}(\boldsymbol{y}) &= \frac{\sqrt{50}(800 - 1000)}{800} \approx -1.77, \\ u^{1/2}(\boldsymbol{y}) &= \frac{\sqrt{50}(800 - 1000)}{1000} \approx -1.41. \end{aligned}$$

Koska valmistajan väitteen kannalta kriittisiä ovat pienet testisuureen arvot, vastaavat approksimatiiviset p-arvot lasketaan standardinormaalijakauman vasemmanpuoleisina häntätodennäköisyyksinä. Ne ovat $\Phi(-1.77) \approx 0.038$ ja $\Phi(-1.41) \approx 0.079$.

Tässä esimerkissä voidaan itse asiassa laskea myös tarkka p-arvo lähtien siitä tiedosta, että H_0:n pätiessä $50\overline{Y} = Y_1 + \cdots + Y_{50} \sim G(50, 1/1000)$ tai $\overline{Y}/10 \sim \chi^2_{100}$ (vrt. 5.5.7). Tarkka p-arvo on siis

$$P_{H_0}\{\overline{Y} \leq 800\} = P\{\chi^2_{100} \leq 80\} \approx 0.070.$$

Huomaa, että erityisesti Waldin testistä saatava approksimatiivinen p-arvo eroaa merkittävästi tarkasta arvosta. Jos vaikkapa olisi tehtävä päätös H_0:n hyväksymisestä tai hylkäämisestä merkitsevyystasolla 0.05, tulos olisi eri kuin tarkkaan p-arvoon perustuva. Näyttää siis siltä, että eksponenttimallin tapauksessa otoskoko $n = 50$ ei ole vielä lainkaan riittävä, jotta asymptoottisiin jakaumatuloksiin voisi täysin luottaa. Tämä ei liene yllättävää, kun otetaan huomioon, että eksponenttijakauma on sangen vino jakauma ja siis muodoltaan kaukana normaalijakaumasta.

5.7 Uskottavuusfunktioon perustuvia testejä II

5.7.1 Johdanto. Tässä pykälässä pohditaan edellä tarkasteltujen kolmen testisuureen yleistystä malleihin, joiden parametri on useampiulotteinen, ja testausasetelmiin, joissa nollahypoteesi on mahdollisesti yhdistetty. Tyypillisesti nollahypoteesi ottaa kantaa vain joihinkin parametrivektorin komponentteihin.

Yleisesti muotoiltuna asetelma on seuraava: On annettu malli $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jonka parametriavaruus on $\Omega \subset \mathbb{R}^d$, ja nollahypoteesi $H_0\colon \boldsymbol{\theta} \in \Omega_0$, jossa Ω_0 on Ω:n osajoukko. Tehtävänä on siis testata, onko havaittu aineisto sopusoinnussa sen hypoteesin kanssa, että todellinen parametriarvo kuuluisi joukkoon Ω_0.

Kysymystä voidaan lähestyä ajattelemalla, että tarkasteltavana on kaksi mallia:

$$\begin{aligned} \textit{vapaa}\text{ malli:} &\quad f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Omega, \\ \textit{rajoitettu}\text{ malli:} &\quad f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Omega_0. \end{aligned}$$

Niiden kummankin puitteissa voidaan aineistosta $\boldsymbol{y}$ laskea suurimman uskottavuuden estimaatit: vapaa su-estimaatti $\hat{\boldsymbol{\theta}} \in \Omega$ ja rajoitettu su-estimaatti $\hat{\boldsymbol{\theta}}_0 \in \Omega_0$, jotka määräytyvät ehdoista

$$\begin{aligned} L(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) &= \max_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta}; \boldsymbol{y}), \\ L(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y}) &= \max_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\theta}; \boldsymbol{y}). \end{aligned}$$

Tässä $L(\boldsymbol{\theta}; \boldsymbol{y}) = c(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ on tuttuun tapaan uskottavuusfunktio.

Uskottavuusosamäärän testisuureeksi on nyt luontevaa valita

$$r(\boldsymbol{y}) = 2 \log \frac{L(\hat{\boldsymbol{\theta}}; \boldsymbol{y})}{L(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})} = 2\,[l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) - l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})].$$

Ilmeisesti tämän suuret arvot ovat nollahypoteesin kannalta kriittisiä, sillä nehän merkitsevät, että rajoitetun mallin antama selitys havaitulle aineistolle on parhaimmillaankin paljon epäuskottavampi kuin vapaan mallin antama.

Jatkossa nähdään, että myös Waldin ja Raon testisuureet voidaan yleistää. Ensin mainittu perustuu erotusvektoriin $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0$ ja jälkimmäinen log-uskottavuusfunktion gradientin arvoon $\nabla l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})$.

5.7.2 Oletukset. Merkintöjen kiinnittämiseksi ja käytettävien testisuureiden asymptoottisten nollahypoteesijakaumien hallitsemiseksi on spesifioitava tarkasteltava testausasetelma hieman tarkemmin kuin edellä. Oletetaan jatkossa, että

a) mallin parametri voidaan osittaa muotoon $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$, jossa $\boldsymbol{\psi} = (\theta_1, \ldots, \theta_q)$ ja $\boldsymbol{\lambda} = (\theta_{q+1}, \ldots, \theta_d)$,

b) parametriavaruus Ω voidaan kirjoittaa vastaavasti tulona $\Omega = \Omega' \times \Omega''$, jossa $\Omega' \subset \mathbb{R}^q$ ja $\Omega'' \subset \mathbb{R}^{d-q}$, ja

c) nollahypoteesi on $H_0 \colon \boldsymbol{\psi} = \boldsymbol{\psi}_0$, jossa $\boldsymbol{\psi}_0 \in \Omega'$ on tunnettu kiinteä vektori, ts.

$$\Omega_0 = \{\boldsymbol{\psi}_0\} \times \Omega'' = \{(\boldsymbol{\psi}_0, \boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Omega''\}.$$

Lisäksi on vaadittava, että malli toteuttaa tietyt säännöllisyysehdot aivan kuten edellisessäkin pykälässä.

Nollahypoteesi siis kiinnittää symbolilla $\boldsymbol{\psi}$ merkityn parametrivektorin osan mutta ei ota mitään kantaa osaan $\boldsymbol{\lambda}$. Koska tutkijan mielenkiinto on tässä testausasetelmassa kohdistunut osaan $\boldsymbol{\psi}$, sitä voidaan kutsua *kiinnostavaksi parametriksi*. Osa $\boldsymbol{\lambda}$ puolestaan on *kiusaparametri*.

Huomaa, että nollahypoteesin rakenteesta johtuu, että rajoitettu su-estimaatti on nyt muotoa $\hat{\boldsymbol{\theta}}_0 = (\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0)$, jossa $\hat{\boldsymbol{\lambda}}_0$ saadaan maksimointitehtävän

$$L(\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0; \boldsymbol{y}) = \max_{\boldsymbol{\lambda} \in \Omega''} L(\boldsymbol{\psi}_0, \boldsymbol{\lambda}; \boldsymbol{y})$$

ratkaisuna eli estimoimalla malli $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\psi}_0, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \Omega''$, suurimman uskottavuuden menetelmällä.

5.7.3 Esimerkkejä. a) Mallissa $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$ parametri on (μ, σ^2) ja parametriavaruus $\mathbb{R} \times (0, \infty)$. Jos $H_0 \colon \mu = \mu_0$, niin μ on kiinnostava parametri ja σ^2 on kiusaparametri.

b) Yhden selittäjän regressiomallissa $Y_1, \ldots, Y_n \perp\!\!\!\perp$, $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ (ks. 1.2.4) parametri on $(\alpha, \beta, \sigma^2)$ ja parametriavaruus $\mathbb{R} \times \mathbb{R} \times (0, \infty)$. Tavallisesti halutaan testata hypoteesia $H_0 \colon \beta = 0$, jolloin β on kiinnostavan parametrin asemassa ja (α, σ^2) on kiusaparametri. Jos taas testattavana on $H_0 \colon \alpha = 0$, niin α on kiinnostava parametri ja (β, σ^2) kiusaparametri.

c) Oletusten 5.7.2a–c mukaisen testausasetelman sovellusaluetta voi usein laajentaa mallin sopivan uudelleenparametroinnin avulla. Tarkastellaan esimerkkinä tilannetta, jossa havaintoja vastaavat satunnaismuuttujat ovat $X_1, \ldots, X_m, Y_1, \ldots, Y_n \perp\!\!\!\perp$ ja

$$X_1, \ldots, X_m \sim N(\mu, \sigma^2), \qquad Y_1, \ldots, Y_n \sim N(\nu, \tau^2).$$

Tämän mallin parametri on $(\mu, \nu, \sigma^2, \tau^2)$.

Halutaan testata hypoteesia $H_0 \colon \mu = \nu$ eli tutkia, voisivatko x-havainnot ja y-havainnot olla peräisin normaalijakaumista, joilla on sama odotusarvo. Tämä hypoteesi ei suoraan ole c-oletuksen mukainen. Siksi tehdään malliin uudelleenparametrointi $(\mu, \nu, \sigma^2, \tau^2) \mapsto (\mu, \delta, \sigma^2, \tau^2)$, jossa $\delta = \nu - \mu$ eli kääntäen $\nu = \mu + \delta$. Nyt H_0 on yhtäpitävä hypoteesin $H_0' \colon \delta = 0$ kanssa, joka on c-oletuksen tyyppiä.

Tämän esimerkin testausasetelmaa kutsutaan tilastollisen päättelyn kirjoissa perinteisesti *Behrensin ja Fisherin ongelmaksi*.

5.7.4 Uskottavuusosamäärän testisuure. Tarkastellaan edellä kohdissa 5.7.1 ja 5.7.2 kuvattua asetelmaa. Olkoon $\hat{\boldsymbol{\theta}}$ vapaa su-estimaatti ja $\hat{\boldsymbol{\theta}}_0$ rajoitettu eli H_0:n puitteissa muodostettu su-estimaatti. Kuten todettua, *uskottavuusosamäärän testisuure* määritellään tällöin kaavalla

$$r(\boldsymbol{y}) = 2\,[l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) - l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})].$$

Yleistämällä yksiulotteisen parametrin tapauksessa suoritettua päättelyä (ks. 5.6.4) ja olettamalla riittävät säännöllisyysehdot voidaan osoittaa, että

$$r(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_q, \qquad \text{kun } H_0 \text{ pätee.}$$

Uskottavuusosamäärän testin approksimatiivinen p-arvo saadaan siis χ^2_q-jakaumasta: $p \approx P\{\chi^2_q \geq r(\boldsymbol{y})\}$. Huomaa, että vapausasteiden lukumäärä q on sama kuin kiinnostavan parametrin dimensio eli H_0:n asettamien (skalaaristen) side-ehtojen lukumäärä.

5.7.5 Waldin testisuure. Palautetaan pykälästä 2.6 mieleen, että malliin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ liittyvä Fisherin informaatiomatriisi on symmetrinen $d \times d$ -matriisi

$$\boldsymbol{i}(\boldsymbol{\theta}) = \begin{bmatrix} i_{1,1}(\boldsymbol{\theta}) & \cdots & i_{1,d}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ i_{d,1}(\boldsymbol{\theta}) & \cdots & i_{d,d}(\boldsymbol{\theta}) \end{bmatrix},$$

jossa

$$i_{a,b}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2}{\partial\theta_a \partial\theta_b} l(\boldsymbol{\theta}; \boldsymbol{Y})\right]$$

kun $a, b = 1, \ldots, d$.

Kohdassa 3.6.8 on todettu, että

$$\hat{\boldsymbol{\theta}} \underset{\text{as}}{\sim} N_d(\boldsymbol{\theta}, \boldsymbol{i}^{-1}(\boldsymbol{\theta})),$$

jossa $\boldsymbol{i}^{-1}(\boldsymbol{\theta})$ on informaatiomatriisin käänteismatriisi. Lohkotaan tämä neljään lohkoon kirjoittamalla

$$\boldsymbol{i}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{i}^{\boldsymbol{\psi},\boldsymbol{\psi}}(\boldsymbol{\theta}) & \boldsymbol{i}^{\boldsymbol{\psi},\boldsymbol{\lambda}}(\boldsymbol{\theta}) \\ \boldsymbol{i}^{\boldsymbol{\lambda},\boldsymbol{\psi}}(\boldsymbol{\theta}) & \boldsymbol{i}^{\boldsymbol{\lambda},\boldsymbol{\lambda}}(\boldsymbol{\theta}) \end{bmatrix},$$

jossa $\boldsymbol{i}^{\boldsymbol{\psi},\boldsymbol{\psi}}(\boldsymbol{\theta})$ on $q \times q$ -matriisi. Kun jaetaan vastaavasti su-estimaattori $\hat{\boldsymbol{\theta}}$ kahteen osaan kirjoittamalla $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}})$, niin pätee

$$\hat{\boldsymbol{\psi}} \underset{\text{as}}{\sim} N_q(\boldsymbol{\psi}, \boldsymbol{i}^{\boldsymbol{\psi},\boldsymbol{\psi}}(\boldsymbol{\theta})).$$

Tästä seuraa, että†

$$(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})'\, \boldsymbol{i}^{\boldsymbol{\psi},\boldsymbol{\psi}}(\boldsymbol{\theta})^{-1}\, (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \underset{\text{as}}{\sim} \chi^2_q.$$

(Muista, että matriisilaskuissa vektorit ajatellaan pystyvektoreiksi, jolloin niiden transpoosit ovat tietysti vaakavektoreita.)

† Pätee seuraava yleinen tulos: jos $\boldsymbol{X} \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$ (q-ulotteinen normaalijakauma, jonka odotusarvovektori on $\boldsymbol{0}$ ja kovarianssimatriisi $\boldsymbol{\Sigma}$) ja $\boldsymbol{\Sigma}$ on ei-singulaarinen, niin $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \sim \chi^2_q$. Tässä $\boldsymbol{X}'$ on vektorin $\boldsymbol{X}$ transpoosi. Tämä tulos todistetaan lineaaristen mallien kurssilla.

Näiden tarkastelujen pohjalta *Waldin testisuureen* määritelmäksi kohdan 5.7.2 tilanteessa otetaan

$$w(\boldsymbol{y}) = (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)' \, \boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}})^{-1} \, (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

jossa $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}})$ on vapaa su-estimaatti. Matriisin $\boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}})$ sijasta voidaan käyttää myös vastaavaa havaitusta informaatiosta saatavaa matriisia. Säännöllisyysoletusten vallitessa on voimassa asymptoottinen jakaumatulos

$$w(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_q, \qquad \text{kun } H_0 \text{ pätee.}$$

Testisuureen $w(\boldsymbol{y})$ suuret arvot asettavat nollahypoteesin kyseenalaiseksi, joten approksimatiivinen p-arvo lasketaan samoin kuin uskottavuusosamäärän testissä.

5.7.6 Raon testisuure. Palautetaan pykälästä 2.6 mieleen, että vektoriparametrisen mallin pistemääräfunktiolla tarkoitetaan log-uskottavuusfunktion gradienttia

$$\nabla l(\boldsymbol{\theta}; \boldsymbol{y}) = \left(\frac{\partial}{\partial \theta_1} l(\boldsymbol{\theta}; \boldsymbol{y}), \ldots, \frac{\partial}{\partial \theta_d} l(\boldsymbol{\theta}; \boldsymbol{y}) \right)$$

Ositetaan tämä kahteen osaan

$$\nabla l(\boldsymbol{\theta}; \boldsymbol{y}) = (\nabla_{\boldsymbol{\psi}} \, l(\boldsymbol{\theta}; \boldsymbol{y}), \nabla_{\boldsymbol{\lambda}} \, l(\boldsymbol{\theta}; \boldsymbol{y}))$$

siten, että $\nabla_{\boldsymbol{\psi}} \, l(\boldsymbol{\theta}; \boldsymbol{y})$ koostuu osittaisderivaatoista muuttujien $\theta_1, \ldots, \theta_q$ suhteen ja $\nabla_{\boldsymbol{\lambda}} \, l(\boldsymbol{\theta}; \boldsymbol{y})$ osittaisderivaatoista muuttujien $\theta_{q+1}, \ldots, \theta_d$ suhteen.

Raon pistemäärätestisuure määritellään nyt kaavalla

$$u(\boldsymbol{y}) = [\nabla_{\boldsymbol{\psi}} \, l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})]' \, \boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}}_0) \, [\nabla_{\boldsymbol{\psi}} \, l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})],$$

jossa $\hat{\boldsymbol{\theta}}_0 = (\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0)$ on rajoitettu su-estimaatti. Määritelmässä voidaan $\boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}}_0)$ korvata vastaavalla havaitusta informaatiosta saatavalla matriisilla. Riittävien säännöllisyysehtojen vallitessa voidaan osoittaa, että

$$u(\boldsymbol{Y}) \underset{\text{as}}{\sim} \chi^2_q, \qquad \text{kun } H_0 \text{ pätee,}$$

ja approksimatiivinen p-arvo lasketaan kuten uskottavuusosamäärän ja Waldin testeissä.

5.7.7 Testisuureiden vertailua. Kaikki kolme uskottavuusfunktioon perustuvaa testisuuretta noudattavat nollahypoteesin pätiessä asymptoottisesti samaa jakaumaa χ^2_q. Ne kuitenkin eroavat toisistaan vaadittavan suurimman uskottavuuden estimoinnin suhteen. Uskottavuusosamäärän testisuureen $r(\boldsymbol{y})$ muodostamiseksi on estimoitava sekä vapaa että rajoitettu malli. Waldin testisuure $w(\boldsymbol{y})$ puolestaan perustuu pelkästään vapaaseen su-estimaattiin ja Raon testisuure $u(\boldsymbol{y})$ rajoitettuun su-estimaattiin. Näillä seikoilla on oma merkityksensä, kun tarkastellaan monimutkaisia malleja, joissa estimointi edellyttää raskasta numeerista laskentaa. Laskennalliset ongelmat ovat tosin viime vuosina paljolti poistuneet tietokoneiden laskentakapasiteetin kehityksen myötä.

5.7.8 Esimerkki: Raon testi normaalimallille. Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Muodostetaan Raon testisuure hypoteesille $H_0 \colon \mu = \mu_0$. Nyt siis σ^2 on kiusaparametri, johon H_0 ei ota mitään kantaa.

Raon testisuureen muodostamiseksi on etsittävä rajoitetun mallin su-estimaatti $(\mu_0, \hat{\sigma}_0^2)$, jossa $\hat{\sigma}_0^2$ saadaan maksimoimalla log-uskottavuusfunktio

$$l(\mu_0, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_0)^2$$

muuttujan σ^2 suhteen. Yhtälön

$$\frac{\partial}{\partial(\sigma^2)}l(\mu_0, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu_0)^2 = 0$$

ratkaisuksi nähdään helposti piste

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2,$$

ja esim. toisen derivaatan arvo laskemalla nähdään, että kyseessä on todellakin globaali maksimikohta.

Lisäksi tarvitaan pistemääräfunktion kiinnostavaa parametria eli μ:tä vastaava komponentti

$$\frac{\partial}{\partial\mu}l(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = \frac{n(\overline{y} - \mu)}{\sigma^2},$$

jonka arvo rajoitetun su-estimaatin pisteessä on $n(\overline{y} - \mu_0)/\hat{\sigma}_0^2$.

Kohdassa 2.6.3 on saatu tarkasteltavan mallin Fisherin informaatioksi

$$\boldsymbol{i}(\mu, \sigma^2) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix},$$

jonka käänteismatriisi on

$$\boldsymbol{i}^{-1}(\mu, \sigma^2) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

Raon testisuureen lausekkeessa esiintyy tämän ensimmäinen (eli μ:tä vastaava) diagonaalialkio rajoitetun su-estimaatin pisteessä laskettuna, siis luku $\hat{\sigma}_0^2/n$.

Kaiken kaikkiaan saadaan Raon testisuureen lausekkeeksi

$$u(\boldsymbol{y}) = \frac{n(\overline{y} - \mu_0)}{\hat{\sigma}_0^2} \cdot \frac{\hat{\sigma}_0^2}{n} \cdot \frac{n(\overline{y} - \mu_0)}{\hat{\sigma}_0^2} = \frac{n(\overline{y} - \mu_0)^2}{\hat{\sigma}_0^2}.$$

Osoittautuu, että tästä saatava testi on itse asiassa yhtäpitävä tavallisen kaksisuuntaisen t-testin kanssa. Voidaan nimittäin kirjoittaa (ks. teht. 2.2)

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2 = \hat{\sigma}^2 + (\overline{y} - \mu_0)^2,$$

kun $\hat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2/n$ on varianssin vapaa su-estimaatti, ja siten

$$u(\boldsymbol{y}) = \frac{n(\overline{y} - \mu_0)^2}{\hat{\sigma}^2 + (\overline{y} - \mu_0)^2} = n\left(1 - \frac{1}{1 + t(\boldsymbol{y})^2/(n-1)}\right),$$

jossa

$$t(\boldsymbol{y}) = \frac{\overline{y} - \mu_0}{s/\sqrt{n}}$$

on t-testisuure (ks. 5.4.2). Nähdään siis, että $u(\boldsymbol{y})$ on aidosti kasvava muunnos neliöstä $t(\boldsymbol{y})^2$ tai yhtä hyvin itseisarvosta $|t(\boldsymbol{y})|$. Tällöin tapahtuma $\{u(\boldsymbol{Y}) \geq u(\boldsymbol{y})\}$ on aina sama kuin tapahtuma $\{|t(\boldsymbol{Y})| \geq |t(\boldsymbol{y})|\}$, joten Raon pistemäärätesti on yhtäpitävä kaksisuuntaisen t-testin kanssa siinä mielessä, että niistä laskettavat p-arvot yhtyvät ja siten myös kriittiset alueet ovat samat.

Tehtävässä 5.15 todetaan vastaava asia myös Waldin testin ja uskottavuusosamäärän testin osalta. Nämä tulokset ovat omiaan motivoimaan sitä, että t-testi on "oikea" testi hypoteesille $H_0\colon \mu = \mu_0$ mallissa $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$.

Harjoitustehtäviä

5.1. Tilastollinen malli muodostuu kahdesta riippumattomasta havainnosta $Y_1 \sim N(\mu_1, 1)$ ja $Y_2 \sim N(\mu_2, 1)$. Parametri on (μ_1, μ_2). Ilmoita, mitkä seuraavista hypoteeseista ovat yksinkertaisia ja mitkä yhdistettyjä: a) $\mu_1 = 1$, b) Y_1 ja Y_2 ovat samoin jakautuneet, c) kummankin mediaani on 1, d) $P(Y_1 > Y_2) > \frac{1}{2}$. Piirrä myös vastaavat Ω_0-joukot (μ_1, μ_2)-tasossa.

5.2. a) Esimerkin 5.2.4 koeasetelmassa saadaan 560 kruunua. Laske vastaava p-arvo ja pohdi, voidaanko lanttia pitää harhattomana. (Käytä normaaliapproksimaatiota binomijakauman todennäköisyyksien laskentaan.)

b) Ovatko johtopäätökset toisenlaiset, jos heittoja onkin sata ja saadaan 56 kruunua?

5.3. Kemiantehtaassa kone annostelee erästä kemikaalia kanistereihin. Oletetaan, että kerralla annostellun kemikaalin määrä (litroina) noudattaa normaalijakaumaa. Pyrkimyksenä on säätää kone siten, että keskimääräinen annos μ on 10 ja keskihajonta σ korkeintaan 0.2. Tutkittiin 20 kanisteria ja havaittiin, että niissä oli kemikaalia keskimäärin $\overline{y} = 9.86$ (litraa), keskihajonnan ollessa $s = 0.25$. Testaa kaksisuuntaisella t-testillä ja yksisuuntaisella χ^2-testillä, onko kone säädön tarpeessa. Käytä 5 %:n merkitsevyystasoa.

5.4. Kauppias myy männynsiemeniä, joiden itävyyden väitetään olevan ainakin 80 %. Neljä asiakasta ostaa kukin pussillisen eli 10 siementä. He havaitsevat, että itäviä siemeniä oli 9, 5, 6 ja 8.

a) Kukin asiakas testaa itävyysväitettä oman havaintonsa valossa tavallista binomijakaumamallia käyttäen. Mitkä ovat asiakkaiden saamat p-arvot? Onko jollakin heistä aihetta hylätä väite 5 %:n merkitsevyystasolla?

b) Toinen asiakas tulee valittamaan siementen laadusta kauppiaalle ja kertoo oman p-arvonsa. Miten kauppias voi arvioida itävyysväitettä, kun hän otaksuu, että muut kolme ovat olleet laatuun tyytyväisiä? *Vihje.* Valintakorjaus.

c) Testaa itävyysväitettä kokonaisaineiston (40 siemenestä 28 iti) valossa. (Voit käyttää normaaliapproksimaatiota.)

5.5. Olkoot Y_1 ja Y_2 kaksi riippumatonta havaintoa Poisson-jakaumasta $P(\mu)$. Halutaan testata hypoteesia $H_0\colon \mu = 2$ vastaan $H_1\colon \mu < 2$. Testisuureena on $T = Y_1 + Y_2 \sim P(2\mu)$.

a) Millaiset testisuureen arvot t todistavat mielestäsi H_0:aa vastaan ja H_1:n puolesta: pienet vai suuret?

b) Mitkä t johtavat H_0:n hylkäämiseen ja H_1:n hyväksymiseen merkitsevyystasolla 0.1? Mitkä havaintoparit (y_1, y_2) kuuluvat vastaavaan kriittiseen alueeseen?

c) Hahmottele voimafunktion $\pi_{0.1}(\mu)$ kuvaajaa välillä $(0, 2]$ laskemalla sen arvot ainakin muutamassa eri pisteessä.

5.6. Malli on $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jossa $\sigma_0^2 > 0$ on tunnettu luku. Testataan $H_0\colon \mu = 0$ vastaan $H_1\colon \mu > 0$ käyttämällä yksisuuntaista z-testiä (ks. 5.4.1). Muodosta 0.05-tasoinen voimafunktio ja hahmottele sen kuvaajaa. Vertaa sitä kaksisuuntaisen z-testin voimaan (kuva 5.2). Miten näiden ero pitäisi ymmärtää, erityisesti joukossa $\mu > 0$?

5.7. Oletetaan, että $Y_1, \ldots, Y_n \sim N(\mu, 1)$ $\perp\!\!\!\perp$, ja tarkastellaan yksisuuntaista testausasetelmaa $H_0\colon \mu = 0$, $H_1\colon \mu > 0$. Testisuureena on tavalliseen tapaan $z = \sqrt{n}\overline{y}$. Kuinka suuri on havaintojen lukumäärän n oltava, jotta testin 0.01-tasoinen voima pisteessä $\mu = 0.5$ olisi ≥ 0.6 (eli hyväksymisvirheen riski ko. pisteessä ≤ 0.4)?

5.8. Olkoot $Y_1, \ldots, Y_{10} \sim Tas(0, \theta)$ $\perp\!\!\!\perp$, jossa $\theta > 0$. Testataan hypoteesia $H_0\colon \theta = 2$ vastaan $H_1\colon \theta < 2$. Testisuureena käytetään suurinta havaintoa $t = y_{(10)} = \max(y_1, \ldots, y_{10})$ ja H_0 hylätään, jos $t \leq 1.5$. Laske testin merkitsevyystaso (eli hylkäämisvirheen todennäköisyys) ja voimafunktio. *Apu.* T:n jakauma on selvitetty tehtävässä 3.10.

5.9. Diskreetin satunnaismuuttujan Y jakauma riippuu parametrista θ, jolla on kaksi mahdollista arvoa: 0 ja 1. Vastaavat Y:n pistetodennäköisyydet on esitetty taulukossa alla.

y	1	2	3	4	5	6	7
$f_Y(y;0)$	.01	.01	.01	.01	.01	.01	.94
$f_Y(y;1)$	.06	.05	.04	.03	.02	.01	.79

Halutaan testata $H_0\colon \theta = 0$ vastaan $H_1\colon \theta = 1$. Laske suhteet $f_Y(y;1)/f_Y(y;0)$ ja määritä Neyman–Pearson-apulauseeseen vetoamalla voimakkain testi (esim. ilmoittamalla kriittinen alue), kun merkitsevyystasoksi valitaan 0.04. Kuinka suuri on hyväksymisvirheen todennäköisyys?

5.10. Toistokoemallin $K \sim Bin(n, \theta)$ parametrina on $\theta \in (0, 1)$. Tarkastellaan hypoteeseja $H_0\colon \theta = \theta_0$ ja $H_1\colon \theta = \theta_1$, jossa $\theta_0 < \theta_1$. Osoita uskottavuusosamäärää tutkimalla eli Neyman–Pearson-apulauseen avulla, että voimakkain testi saadaan testisuureesta k. Järkeile myös, että kyseessä on tasaisesti voimakkain testi yhdistetylle vastahypoteesille $H_1\colon \theta > \theta_0$.

Vihje. Muokkaa uskottavuusosamäärää niin, että saat näkyviin suhteet $\theta_0/(1-\theta_0)$ ja $\theta_1/(1-\theta_1)$.

5.11. Olkoon $Y_1, \ldots, Y_n$ riippumaton otos eksponenttiperheen jakaumasta, jonka ptf/tf on muotoa

$$f(y;\theta) = c(\theta)h(y)e^{\phi(\theta)t(y)},$$

jossa c ja h ovat ei-negatiivisia funktioita ja $\phi(\theta)$ on aidosti kasvava funktio reaalisesta parametrista θ (vrt. 4.2.5 ja teht. 2.20). Näytä, että syntyvällä mallilla $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ on monotoninen uskottavuusosamäärä. Mitä muotoa ovat kriittiset alueet tasaisesti voimakkaimmassa yksisuuntaisessa testissä?

5.12. a) Olkoot $Y_1, \ldots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$. Johda uskottavuusosamäärän testisuureen lauseke, kun testattavana on $H_0\colon \mu = \mu_0$.

b) Eräässä tienristeyksessä on pitkällä aikavälillä sattunut keskimäärin 7.2 onnettomuutta kuukaudessa. Risteykseen asennetaan liikennevalot. Sitä seuraavan vuoden aikana sattuu yhteensä 60 onnettomuutta. Testaa uskottavuusosamäärän testiä ja χ^2-approksimaatiota käyttämällä, voidaanko valojen asentamisen katsoa vaikuttaneen onnettomuuksien määrään. Oletetaan, että onnettomuuksien lukumäärä kuukaudessa on Poisson-jakautunut.

5.13. Olkoot $Y_1, \ldots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$. Muodosta Waldin testisuureen $w^{1/2}(\boldsymbol{y})$ ja Raon pistemäärätestisuureen $u^{1/2}(\boldsymbol{y})$ lausekkeet.

5.14. Esimerkin 2.4.6 a-asetelmassa saadaan otokseen ($n = 50$) genotyyppejä rr, rR ja RR vastaavasti 4, 20 ja 26 yksilöä. Laske parametrin θ suurimman uskottavuuden estimaatti ja testaa kaksisuuntaisella Waldin testillä nollahypoteesia $H_0\colon \theta = 0.2$.

5.15. Oletetaan, että $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa parametri on (μ, σ^2). Testattavana on $H_0\colon \mu = \mu_0$. Johda uskottavuusosamäärän testisuure ja Waldin testisuure ja totea, että saatavat testit ovat yhtäpitäviä tavallisen kaksisuuntaisen t-testin kanssa.

Huom. Sama asia Raon testisuureen osalta on todettu esimerkissä 5.7.8.

5.16. Eräiden kuulalaakereiden kestoa (miljoonaa kierrosta) on totuttu kuvaamaan Weibull-jakaumalla, jonka tiheysfunktio on

$$f(y;\beta,\lambda) = \lambda\beta y^{\beta-1}\exp(-\lambda y^{\beta}), \qquad y > 0,$$

ja jossa β ja λ ovat positiivisia parametreja. Halutaan testata hypoteesia $H_0\colon \beta = 1$ eli tutkia, voisiko kestoa kuvata eksponenttijakaumalla. Poimitaan 23 laakerin otos ja mitataan otosyksiköiden kestot:

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84
51.96	54.12	55.56	67.80	68.64	68.64	68.88	84.12
93.12	98.64	105.12	105.84	127.92	128.04	173.40	

Suorita testi käyttämällä uskottavuusosamäärän testisuuretta ja χ^2-approksimaatiota.

Apu. Vapaan mallin su-estimaatteja ei voi lausua suljetussa muodossa, mutta uskottavuusyhtälöt numeerisesti ratkaisemalla nähdään, että $\hat{\beta} = 2.1021$ ja $\hat{\lambda} = 9.515 \cdot 10^{-5}$.

6 Luottamusjoukot

6.1 Määritelmä ja tulkinta

6.1.1 Johdanto. Luvuissa 2 ja 3 tarkasteltiin piste-estimointia, jossa tavoitteena oli spesifioida mallin parametriavaruudesta piste, joka olisi ”hyvä” arvio mallin parametrille. Pelkän piste-estimaatin esittäminen on kuitenkin harvoin riittävä vastaus annettuun estimointitehtävään. Onhan nimittäin yleensä täysin epärealistista ajatella, että voitaisiin juuri tarkalleen löytää se oikea parametriarvo, joka on aineiston tuottanut. Siksi onkin tarpeellista pyrkiä jollakin tavalla arvioimaan esitettävien piste-estimaattien tarkkuutta. Pelkistetymmin ja hieman yleisemmin voidaan tarkastella ”joukkoestimointitehtävää”, jossa aineiston perusteella on rajattava parametriavaruudesta osajoukko – mielellään mahdollisimman pieni –, joka varsin suurella ja etukäteen annetulla varmuudella sisältäisi todellisen parametriarvon.

Tämä kysymyksenasettelu johtaa luottamusjoukkojen teoriaan. Jos estimoitava parametri on yksiulotteinen, kyseiset joukot ovat tavallisesti välejä, joten niitä kutsutaan luottamusväleiksi ja niiden muodostamista väliestimoinniksi. Osoittautuu, että luottamusjoukkojen teoria on läheisessä yhteydessä testiteorian kanssa, ja siten luvun 5 lopussa käsitellyt, uskottavuusfunktioon ja asymptotiikkaan perustuvat menetelmät mahdollistavat melko yleisten luottamusjoukkojen konstruoimisen.

6.1.2 Luottamusjoukon määritelmä. Tarkastellaan mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, jonka parametriavaruus on $\Omega \subset \mathbb{R}^d$. Olkoon lisäksi $0 < \alpha < 1$. Aineistosta riippuva Ω:n osajoukko $A(\boldsymbol{y})$ on parametrin $\boldsymbol{\theta}$ *luottamusjoukko luottamustasolla* $1-\alpha$, jos

$$P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} \in A(\boldsymbol{Y})\} \geq 1-\alpha \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega. \tag{6.1}$$

Jos $d = 1$ ja $A(\boldsymbol{y})$ on väli, sitä kutsutaan myös *luottamusväliksi*. Esimerkiksi $A(\boldsymbol{y})$ voisi olla avoin väli $A(\boldsymbol{y}) = (a(\boldsymbol{y}), b(\boldsymbol{y}))$, jossa

$$P_\theta\{a(\boldsymbol{Y}) < \theta < b(\boldsymbol{Y})\} \geq 1-\alpha \qquad \text{kaikilla } \theta \in \Omega.$$

Monesti tapauksessa $d > 1$ on mielenkiinnon kohteena koko parametrivektorin $\boldsymbol{\theta}$ sijasta vain jokin sen komponentti tai osavektori $\boldsymbol{\psi}$ (vrt. 5.7.2). Joukko $A(\boldsymbol{y})$ on tämän *luottamusjoukko luottamustasolla* $1-\alpha$, jos

$$P_{\boldsymbol{\theta}}\{\boldsymbol{\psi} \in A(\boldsymbol{Y})\} \geq 1-\alpha \qquad \text{kaikilla } \boldsymbol{\theta} \in \Omega.$$

Ensisijaisena pyrkimyksenä on etsiä luottamusjoukkoja, joilla em. todennäköisyydet olisivat tasan $1-\alpha$. Tämä voi kuitenkin olla vaikeaa tai mahdotonta johtuen mallin diskreettisyydestä tai siitä, että todennäköisyydet riippuvat parametrista $\boldsymbol{\theta}$. Siksi on mukavaa sallia myös epäyhtälö ”$>$”. Tavanomaisia luottamustasoja ovat 95 % ja 99 %, jotka vastaavat lukuja $\alpha = 0.05$ ja $\alpha = 0.01$.

6.1.3 Esimerkki: normaalimalli. Tarkastellaan jälleen tuttua normaalijakaumamallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma_0^2)$ $\perp\!\!\!\perp$, jonka parametri on μ ja jossa $\sigma_0^2 > 0$ on tunnettu luku. Tällöin tunnetusti (vrt. 5.4.1)

$$Z = \frac{\overline{Y} - \mu}{\sigma_0/\sqrt{n}} \sim N(0,1).$$

Olkoon $0 < \alpha < 1$ ja olkoon $z_{\alpha/2}$ se luku, jonka oikealla puolella on osuus $\alpha/2$ standardinormaalijakauman todennäköisyysmassasta. Tällöin

$$P\{|Z| < z_{\alpha/2}\} = P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha,$$

mikä voidaan kirjoittaa myös muodossa

$$P\left\{\overline{Y} - \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}} < \mu < \overline{Y} + \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right\} = 1 - \alpha.$$

Tämä merkitsee, että väli

$$\left(\overline{y} - \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}},\ \overline{y} + \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right) \tag{6.2}$$

on μ:n luottamusväli luottamustasolla $1-\alpha$. Esimerkiksi toimittaessa luottamustasolla 95 % on $\alpha = 0.05$ ja $z_{0.025} \approx 1.96 \approx 2$, joten kyseinen luottamusväli on likimain $(\overline{y} - 2\sigma_0/\sqrt{n},\ \overline{y} + 2\sigma_0/\sqrt{n})$.

Luottamusväli ei koskaan ole mitenkään yksikäsitteinen. Päättelemällä samaan tapaan kuin yllä nähdään, että mikä tahansa väli

$$\left(\overline{y} - \frac{z_{\alpha_1}\sigma_0}{\sqrt{n}},\ \overline{y} + \frac{z_{\alpha_2}\sigma_0}{\sqrt{n}}\right),$$

jossa $\alpha_1 + \alpha_2 = \alpha$, on μ:n luottamusväli luottamustasolla $1 - \alpha$. Rajatapauksina saadaan myös rajoittamattomat luottamusvälit

$$\left(-\infty,\ \overline{y} + \frac{z_\alpha\sigma_0}{\sqrt{n}}\right) \qquad \text{ja} \qquad \left(\overline{y} - \frac{z_\alpha\sigma_0}{\sqrt{n}},\ \infty\right),$$

joiden äärellisiä päätepisteitä kutsutaan toisinaan μ:n *ylemmäksi* ja vastaavasti *alemmaksi luottamusrajaksi*. Voidaan osoittaa, että otoskeskiarvon suhteen symmetrinen väli (6.2) on lyhin mainituista luottamusväleistä, joten eräässä mielessä se on siis tarkin. Se onkin useimmin käytetty luottamusväli nyt tarkasteltavassa mallissa. Joissakin sovellustilanteissa saattaa kuitenkin pelkän toispuolisen luottamusrajan esittäminen olla järkevämpää.

Tarkastellaan lopuksi mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, jossa myös σ^2 on tuntematon parametri mutta luottamusväli halutaan muodostaa edelleen vain μ:lle. Tällöin otetaan lähtökohdaksi jakaumatulos

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

(vrt. 5.4.2), jossa $S^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)$. Välin (6.2) sijasta saadaan nyt luottamusväliksi

$$\left(\overline{y} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}},\ \overline{y} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right),$$

jossa luku $t_{n-1}(\alpha/2)$ on valittu siten, että sen oikealla puolella on t_{n-1}-jakauman massasta osuus $\alpha/2$. Koska t-jakaumissa on paksummat hännät kuin standardinormaalijakaumassa, tämä väli on hieman leveämpi kuin (6.2). Ero on tosin merkityksettömän pieni, jos havaintoja on enemmän kuin muutama kymmenen.

6.1.4 Luottamusjoukon tulkinta. Oletetaan konkreettisuuden vuoksi, että $\alpha = 0.05$ eli on päätetty toimia luottamustasolla 95 %. Luottamusjoukon määrittelevä ehto (6.1) merkitsee tällöin sanallisesti ilmaistuna, että satunnainen joukko $A(\boldsymbol{Y})$ peittää vähintään todennäköisyydellä 0.95 todellisen parametriarvon $\boldsymbol{\theta}$, joka on kiinteä eisatunnainen piste parametriavaruudessa. Todennäköisyyden käsitteen käyttö viittaa siis hypoteettiseen toistettuun aineistonkeruuseen: jos aineistonkeruu tarkasteltavasta satunnaisilmiöstä voitaisiin toistaa yhä uudelleen ja uudelleen ja jokaisesta saadusta aineistosta $\boldsymbol{y}$ laskettaisiin luottamusjoukko $A(\boldsymbol{y})$, niin saaduista joukoista keskimäärin 95 % tai useampi sisältäisi todellisen parametriarvon.

Todellisuudessa tutkijalla on tietenkin analysoitavanaan vain yksi aineisto $\boldsymbol{y}$ ja sitä vastaava luottamusjoukko $A(\boldsymbol{y})$. Tästä yksittäisestä luottamusjoukon realisaatiosta ei voi sanoa, että se "todennäköisyydellä 0.95" sisältäisi todellisen parametriarvon!

6.1.5 Esimerkki. Olkoot $Y_1, Y_2 \sim Tas(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ $\perp\!\!\!\perp$. Toteutuneet havainnot ovat y_1 ja y_2. Olkoon $y_{(1)}$ niistä pienempi ja $y_{(2)}$ suurempi, ts.

$$y_{(1)} = \min(y_1, y_2), \qquad y_{(2)} = \max(y_1, y_2).$$

Tarkastellaan vastaavia satunnaismuuttujia $Y_{(1)}$ ja $Y_{(2)}$. Tapahtuma $\{Y_{(1)} < \theta < Y_{(2)}\}$ muodostuu toisensa poissulkevista tapahtumista $\{Y_1 < \theta < Y_2\}$ ja $\{Y_2 < \theta < Y_1\}$, joiden kummankin todennäköisyys on $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Siten

$$P\{Y_{(1)} < \theta < Y_{(2)}\} = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}.$$

Tämä merkitsee, että väli $(y_{(1)}, y_{(2)})$ on 50 %:n luottamusväli parametrille θ. Kuitenkin siinä tapauksessa, että ko. välin pituus $y_{(2)} - y_{(1)}$ on suurempi kuin $\frac{1}{2}$, on täysin varmaa, että θ kuuluu tähän väliin (piirrä tilanne lukusuoralle ja mieti).

6.2 Yhteys testeihin ja saranasuureet

6.2.1 Testien ja luottamusjoukkojen dualiteetti. Esimerkissä 6.1.3 muodostettiin luottamusvälejä normaalimallin odotusarvolle z- ja t-testisuureen avulla. Aivan yleisestikin pätee, että testien avulla voidaan muodostaa luottamusjoukkoja (ja päinvastoin). Siten testiteoria ja luottamusjoukkojen teoria ovat tietyssä mielessä ekvivalentteja.

Tarkastellaan mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jonka parametriavaruus on Ω, ja olkoon $0 < \alpha < 1$. Jos $\boldsymbol{\theta}_0 \in \Omega$, olkoon $C_\alpha(\boldsymbol{\theta}_0)$ jonkin hypoteesia $H_0 \colon \boldsymbol{\theta} = \boldsymbol{\theta}_0$ koskevan testin α-tasoinen kriittinen alue. Kun $\boldsymbol{y}$ on havaittu aineisto, hypoteesi H_0 tulee siis hylätyksi, jos $\boldsymbol{y} \in C_\alpha(\boldsymbol{\theta}_0)$, ja hyväksytyksi, jos $\boldsymbol{y} \notin C_\alpha(\boldsymbol{\theta}_0)$. Lisäksi

$$P_{\boldsymbol{\theta}_0}\{\boldsymbol{Y} \in C_\alpha(\boldsymbol{\theta}_0)\} \leq \alpha$$

kuten kohdissa 5.3.2 ja 5.5.1 on todettu. Merkitään $A(\boldsymbol{y})$:llä kaikkien niiden nollahypoteesiarvojen $\boldsymbol{\theta}_0$ joukkoa, jotka tulevat hyväksytyiksi, ts.

$$A(\boldsymbol{y}) = \{\boldsymbol{\theta}_0 \in \Omega : \boldsymbol{y} \notin C_\alpha(\boldsymbol{\theta}_0)\}.$$

Nyt tämä joukko on parametrin $\boldsymbol{\theta}$ luottamusjoukko luottamustasolla $1 - \alpha$: nimittäin jokaisella $\boldsymbol{\theta} \in \Omega$ pätee

$$P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} \in A(\boldsymbol{Y})\} = P_{\boldsymbol{\theta}}\{\boldsymbol{Y} \notin C_\alpha(\boldsymbol{\theta})\} = 1 - P_{\boldsymbol{\theta}}\{\boldsymbol{Y} \in C_\alpha(\boldsymbol{\theta})\} \geq 1 - \alpha.$$

Tässä yhteydessä on paikallaan huomauttaa, että erityisesti diskreeteissä malleissa ei ole yleensä mahdollista löytää joukkoja $C_\alpha(\boldsymbol{\theta}_0)$ siten, että tapahtuman $\{\boldsymbol{Y} \in C_\alpha(\boldsymbol{\theta}_0)\}$

todennäköisyys olisi tasan ennalta annetun α:n suuruinen jokaisella $\boldsymbol{\theta}_0$. Tämä ilmiö kohdattiin vaikkapa esimerkissä 5.5.2. Siksi myös luottamusjoukon määritelmässä on syytä sallia aito epäyhtälö. Seuraavissa pykälissä tarkastellaan uskottavuusfunktioon perustuvista testeistä saatavia approksimatiivisia luottamusvälejä, ja silloin tämä ilmiö häviää näkyvistä, koska testisuureiden asymptoottiset nollahypoteesijakaumat ovat jatkuvia.

6.2.2 Saranasuureet. Joissakin kyllin yksinkertaisissa malleissa voi luottamusjoukkoja muodostaa varsin kätevästi saranasuureiden avulla. Satunnaismuuttujaa $Q(\boldsymbol{Y};\boldsymbol{\theta})$, joka on siis aineistosta ja tarkasteltavan mallin parametrista riippuva suure, kutsutaan *saranasuureeksi*, jos sen jakauma on sama kaikilla $\boldsymbol{\theta}$. Tällöin muotoa $\{Q(\boldsymbol{Y};\boldsymbol{\theta}) \in B\}$ olevan tapahtuman todennäköisyys ei riipu $\boldsymbol{\theta}$:sta, joten voidaan päätellä, että jos tämä todennäköisyys on $\geq 1-\alpha$, niin joukko $A(\boldsymbol{y}) = \{\boldsymbol{\theta} : Q(\boldsymbol{y};\boldsymbol{\theta}) \in B\}$ on $\boldsymbol{\theta}$:n luottamusjoukko luottamustasolla $1-\alpha$.

6.2.3 Esimerkkejä. a) Mallissa $Y_1,\ldots,Y_n \sim N(\mu,\sigma^2)$ $\perp\!\!\!\perp$ on $T = \sqrt{n}(\overline{Y}-\mu)/S$ saranasuure. Jos $\sigma^2 = \sigma_0^2 > 0$ on tunnettu, myös $Z = \sqrt{n}(\overline{Y}-\mu)/\sigma_0$ on saranasuure. (Ks. 6.1.3.)

b) Olkoot $Y_1,\ldots,Y_n \sim \mathit{Tas}(0,\theta)$ $\perp\!\!\!\perp$, jossa $\theta > 0$, ja $Y_{(n)} = \max(Y_1,\ldots,Y_n)$. Kohdassa 2.2.8 todettiin, että $Y_{(n)}$ on parametrin θ su-estimaattori. Sen kertymäfunktio on (vrt. teht. 3.10)

$$P\{Y_{(n)} \leq t\} = P\{Y_1 \leq t\}\cdots P\{Y_n \leq t\} = (t/\theta)^n, \qquad 0 < t < \theta.$$

Tarkastellaan muuttujaa $Y_{(n)}/\theta$. Sen arvojoukko on $(0,1)$ ja kertymäfunktio

$$P\{Y_{(n)}/\theta \leq q\} = P\{Y_{(n)} \leq q\theta\} = (q\theta/\theta)^n = q^n, \qquad 0 < q < 1.$$

Koska tämä ei riipu θ:sta, niin $Y_{(n)}/\theta$ on saranasuure.

Saranasuureen $Y_{(n)}/\theta$ avulla voidaan muodostaa esimerkiksi 95 %:n luottamusväli θ:lle seuraavasti: Valitaan luku a, jolle $P\{a < Y_{(n)}/\theta\} = P\{a < Y_{(n)}/\theta < 1\} = 0.95$ kaikilla $\theta > 0$; itse asiassa $a = \sqrt[n]{0.05}$. Ratkaisemalla tässä esiintyvä epäyhtälöpari θ:n suhteen voidaan yhtäpitävästi kirjoittaa $P\{Y_{(n)} < \theta < Y_{(n)}/a\} = 0.95$ kaikilla $\theta > 0$. Tämä merkitsee, että $(y_{(n)}, y_{(n)}/a)$ on 95 %:n luottamusväli.

6.3 Uskottavuusosamäärään perustuvat luottamusjoukot

Edellä kuvattua testien ja luottamusjoukkojen yhteyttä voidaan soveltaa luvun 5 lopussa tarkasteltuihin uskottavuusfunktioon perustuviin testeihin. Tällä tavalla saadaan asymptotiikkaan perustuvia yleisiä menetelmiä approksimatiivisten luottamusjoukkojen muodostamiseen. Tässä pykälässä tarkastellaan uskottavuusosamäärän testiin perustuvia luottamusjoukkoja.

6.3.1 Uskottavuusosamäärään perustuva luottamusväli kun $d = 1$. Tarkastellaan kyllin säännöllistä reaaliparametrista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$. Kohdassa 5.6.4 opittiin, että uskottavuusosamäärän testisuuretta $r(\boldsymbol{y}) = 2\left[l(\hat{\theta};\boldsymbol{y}) - l(\theta_0;\boldsymbol{y})\right]$ vastaava satunnaismuuttuja noudattaa suurissa otoksissa approksimatiivisesti χ^2_1-jakaumaa, kun $\theta = \theta_0$. Olkoon $\chi^2_1(\alpha)$ se piste, jonka oikealla puolella on osuus α tämän jakauman todennäköisyysmassasta. Tällöin $P_{\theta_0}\{r(\boldsymbol{Y}) < \chi^2_1(\alpha)\} \approx 1-\alpha$, joten voidaan todeta, että

joukko

$$
\begin{aligned}
A(\boldsymbol{y}) &= \{\theta : 2\,[l(\hat{\theta};\boldsymbol{y}) - l(\theta;\boldsymbol{y})] < \chi^2_1(\alpha)\} \\
&= \{\theta : l(\theta;\boldsymbol{y}) - l(\hat{\theta};\boldsymbol{y}) > -\tfrac{1}{2}\chi^2_1(\alpha)\}
\end{aligned} \tag{6.3}
$$

on θ:n *approksimatiivinen luottamusjoukko* luottamustasolla $1-\alpha$. Yleensä tämä joukko on *väli*.

Muodostetulla luottamusvälillä on myös elegantti uskottavuuspohjainen tulkinta. Kun L on tarkasteltavan mallin uskottavuusfunktio, määritellään *normitettu* (eli suhteellinen) *uskottavuusfunktio*

$$L_0(\theta;\boldsymbol{y}) = \frac{L(\theta;\boldsymbol{y})}{L(\hat{\theta};\boldsymbol{y})}.$$

Tällöin aina $0 \le L_0 \le 1$ ja $L_0(\hat{\theta};\boldsymbol{y}) = 1$. Tämän funktion logaritmi

$$l_0(\theta;\boldsymbol{y}) = l(\theta;\boldsymbol{y}) - l(\hat{\theta};\boldsymbol{y})$$

on *normitettu log-uskottavuusfunktio*, ja se esiintyi jo kohdassa 2.4.2. Sille pätee $l_0 \le 0$ sekä $l_0(\hat{\theta};\boldsymbol{y}) = 0$. Jos nyt $0 < c < 1$, niin joukkoa

$$\{\theta : L_0(\theta;\boldsymbol{y}) > c\} = \{\theta : l_0(\theta;\boldsymbol{y}) > \log c\},$$

joka yleensä on väli, sanotaan $100c\,\%$:n *uskottavuusväliksi* parametrille θ. Siihen kuuluvat siis ne parametriarvot, joiden uskottavuus on enemmän kuin $100c\,\%$ uskottavuuden maksimiarvosta eli arvosta pisteessä $\hat{\theta}$.

Kaavassa (6.3) johdettu approksimatiivinen luottamusväli luottamustasolla $1-\alpha$ on siis itse asiassa uskottavuusväli, jonka "uskottavuustaso" c määräytyy yhtälöstä $\log c = -\frac{1}{2}\chi^2_1(\alpha)$. Tarkastellaan erityisesti tapausta $\alpha = 0.05$ eli 95 %:n luottamustasoa. Tällöin $\chi^2_1(0.05) \approx 3.84$, joten $\log c \approx -1.92$ ja $c \approx e^{-1.92} \approx 0.147$. Näin ollen kyseessä on 14.7 %:n uskottavuusväli

$$A(\boldsymbol{y}) = \{\theta : l_0(\theta;\boldsymbol{y}) > -1.92\}.$$

Graafisesti tämä väli muodostetaan seuraavasti: Piirretään log-uskottavuusfunktion kuvaaja siten, että maksimikohdassa $\theta = \hat{\theta}$ sen arvo on nolla. Sitten piirretään vaakasuora tasolle -1.92 ja etsitään ne pisteet θ, joissa log-uskottavuusfunktion kuvaaja on tämän vaakasuoran yläpuolella.

6.3.2 Esimerkki: eksponenttimalli. Malliin $Y_1, \ldots, Y_n \sim Exp(1/\mu)$ $\perp\!\!\!\perp$ liittyvä uskottavuusosamäärän testisuure laskettiin kohdassa 5.6.7:

$$r(\boldsymbol{y}) = 2n\,[\overline{y}/\mu_0 - \log(\overline{y}/\mu_0) - 1].$$

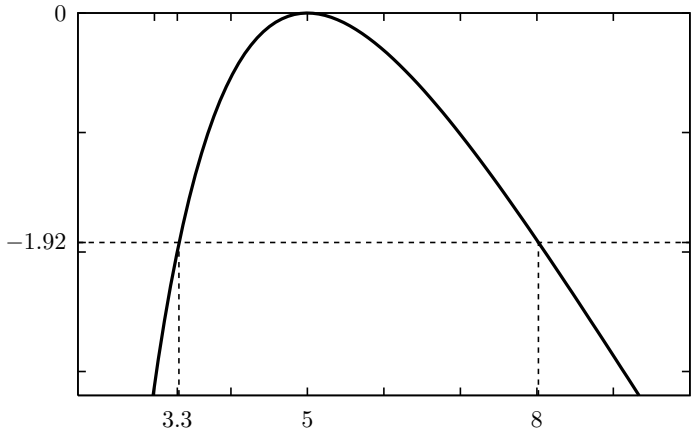
Siten normitettu log-uskottavuusfunktio on

$$l_0(\mu;\boldsymbol{y}) = n\,[\log(\overline{y}/\mu) - \overline{y}/\mu + 1]$$

ja 14.7 %:n uskottavuusväli (eli approksimatiivinen 95 %:n luottamusväli) on

$$\{\mu > 0 : n\,[\log(\overline{y}/\mu) - \overline{y}/\mu + 1] > -1.92\}.$$

Esimerkiksi jos havaintoja on $n = 20$ ja niiden keskiarvo on $\overline{y} = 5$, saadaan 14.7 %:n uskottavuusväliksi likimain $(3.3, 8.0)$ (ks. kuva 6.1).



Kuva 6.1. Normitettu log-uskottavuusfunktio ja 14.7 %:n uskottavuusvälin muodostaminen mallissa $Y_1, \ldots, Y_n \sim Exp(1/\mu)$ ⫫, kun $n = 20$ ja $\overline{y} = 5$.

6.3.3 Uskottavuusosamäärään perustuva luottamusjoukko kun $d > 1$. Oletetaan, että mallin $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$ parametri on d-ulotteinen vektori. Tällöin tavoitteena on useimmiten muodostaa luottamusväli tai -joukko jollekin parametrin komponentille tai osavektorille, ei koko parametrille. Niinpä oletetaankin, aivan kuten pykälässä 5.7, että malli on riittävän säännöllinen ja sen parametri voidaan osittaa muotoon $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$, jossa osavektorin $\boldsymbol{\psi}$ dimensio on q. Tehtävänä on muodostaa luottamusjoukko vektorille $\boldsymbol{\psi}$.

Kohdassa 5.7.4 johdettiin nollahypoteesin $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ testaamiseksi uskottavuusosamäärän testisuure $r(\boldsymbol{y}) = 2\,[l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) - l(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y})]$ ja todettiin, että nollahypoteesin pätiessä sitä vastaava satunnaismuuttuja noudattaa asymptoottisesti χ^2_q-jakaumaa. Tässä $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}})$ on parametrin $\boldsymbol{\theta}$ tavanomainen su-estimaatti ja $\hat{\boldsymbol{\theta}}_0$ puolestaan on rajoitettu su-estimaatti muotoa $\hat{\boldsymbol{\theta}}_0 = (\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}_0))$, jossa piste $\hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}_0)$ määräytyy maksimointitehtävästä

$$L(\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}_0); \boldsymbol{y}) = \max_{\boldsymbol{\lambda}} L(\boldsymbol{\psi}_0, \boldsymbol{\lambda}; \boldsymbol{y}).$$

Maksimikohta tavallisesti riippuu pisteestä $\boldsymbol{\psi}_0$; siksi tämä on merkitty sulkuihin symbolin $\hat{\boldsymbol{\lambda}}_0$ perään. Huomaa erityisesti, että $\hat{\boldsymbol{\lambda}}_0(\hat{\boldsymbol{\psi}}) = \hat{\boldsymbol{\lambda}}$. Päättelemällä samaan tapaan kuin yllä kohdassa 6.3.1 saadaan nyt vektorille $\boldsymbol{\psi}$ approksimatiivinen luottamusjoukko luottamustasolla $1 - \alpha$:

$$A(\boldsymbol{y}) = \{\boldsymbol{\psi} : 2\,[l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) - l(\boldsymbol{\psi}, \hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}); \boldsymbol{y})] < \chi^2_q(\alpha)\}.$$

Kyseessä on q-ulotteisen euklidisen avaruuden osajoukko.

Aivan kuten reaaliparametrisessa tilanteessa edellä joukolle $A(\boldsymbol{y})$ voidaan nytkin antaa yksinkertainen uskottavuustulkinta. Se perustuu *profiiliuskottavuuden* käsitteeseen. Määritellään vektorin $\boldsymbol{\psi}$ *profiiliuskottavuusfunktio* ja *logaritminen profiiliuskottavuusfunktio* asettamalla

$$L_{\mathrm{P}}(\boldsymbol{\psi}; \boldsymbol{y}) = L(\boldsymbol{\psi}, \hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}); \boldsymbol{y}),$$
$$l_{\mathrm{P}}(\boldsymbol{\psi}; \boldsymbol{y}) = l(\boldsymbol{\psi}, \hat{\boldsymbol{\lambda}}_0(\boldsymbol{\psi}); \boldsymbol{y}).$$

Nämä funktiot siis saadaan tavallisesta uskottavuus- ja log-uskottavuusfunktiosta suorittamalla jokaisella kiinnostavan parametrin $\boldsymbol{\psi}$ arvolla maksimointi kiusaparametrin $\boldsymbol{\lambda}$ suhteen. Edelleen niistä voidaan muodostaa normitetut eli suhteelliset versiot ja

määritellä niihin liittyvä *uskottavuusjoukon* käsite samalla periaatteella kuin kohdassa 6.3.1. Kirjoittamalla joukko $A(\boldsymbol{y})$ muodossa

$$A(\boldsymbol{y}) = \{\boldsymbol{\psi} : l_{\mathrm{P}}(\boldsymbol{\psi}; \boldsymbol{y}) - l_{\mathrm{P}}(\hat{\boldsymbol{\psi}}; \boldsymbol{y}) > -\tfrac{1}{2}\chi^2_q(\alpha)\}$$

nähdäänkin, että se on $\boldsymbol{\psi}$:n profiiliuskottavuuteen liittyvä uskottavuusjoukko.

6.4 Waldin testiin perustuvat luottamusjoukot

Waldin testi perustui suoraan suurimman uskottavuuden estimaattorien asymptoottiseen normaalisuuteen säännöllisissä malleissa. Tässä pykälässä opitaan, millaisiin approksimatiivisiin luottamusjoukkoihin tätä kautta päädytään. Erityisesti silloin kun kiinnostava parametri on yksiulotteinen, näin saatavat luottamusvälit ovat hyvin yksinkertaisia ja laajalti käytettyjä eri sovelluksissa.

6.4.1 Waldin testiin perustuva luottamusväli kun $d = 1$. Oletetaan, että malli on $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$, jonka parametri on reaalinen. Palautetaan mieleen kohdasta 5.6.5 yksisuuntainen Waldin testisuure $w^{1/2}(\boldsymbol{Y}) = i(\hat{\theta})^{1/2}(\hat{\theta} - \theta_0)$, joka noudattaa asymptoottisesti standardinormaalijakaumaa, kun $\theta = \theta_0$. Jos luku $z_{\alpha/2}$ on valittu siten, että sen oikealla puolella on standardinormaalijakauman todennäköisyysmassasta osuus $\alpha/2$, niin

$$P_{\theta_0}\{|w^{1/2}(\boldsymbol{Y})| < z_{\alpha/2}\} = P_{\theta_0}\left\{|\hat{\theta} - \theta_0| < \frac{z_{\alpha/2}}{i(\hat{\theta})^{1/2}}\right\} \approx 1 - \alpha.$$

Tämä merkitsee, että $\hat{\theta}$-keskinen väli

$$\left(\hat{\theta} - \frac{z_{\alpha/2}}{i(\hat{\theta})^{1/2}},\ \hat{\theta} + \frac{z_{\alpha/2}}{i(\hat{\theta})^{1/2}}\right)$$

on θ:n approksimatiivinen luottamusväli luottamustasolla $1 - \alpha$.

6.4.2 Keskivirheen käsite. Suurimman uskottavuuden estimaattorien asymptoottisen teorian (ks. 3.6.5) mukaan estimaattorin $\hat{\theta}$ asymptoottinen keskihajonta eli varianssin neliöjuuri on $1/\sqrt{i(\theta)}$, jossa θ on todellinen parametriarvo ja siis tuntematon. Luku $1/i(\hat{\theta})^{1/2}$ yllä on tämän arvio eli estimaatti, ja sitä sanotaan su-estimaattorin $\hat{\theta}$ *keskivirheeksi* (engl. *standard error*) ja merkitään symbolilla $\mathrm{s.e.}(\hat{\theta})$.

Keskivirhe voidaan laskea Fisherin informaation sijasta myös havaitusta informaatiosta: $\mathrm{s.e.}(\hat{\theta}) = 1/j(\hat{\theta}; \boldsymbol{y})^{1/2}$. Siinä tapauksessa, että estimaattorin $\hat{\theta}$ varianssi $v(\theta) = \mathrm{var}_\theta(\hat{\theta})$ osataan muodostaa (θ:sta riippuvana lausekkeena), voidaan käyttää myös suoraan tästä saatavaa arviota: $\mathrm{s.e.}(\hat{\theta}) = v(\hat{\theta})^{1/2}$.

Kun tilastollisen mallin estimoinnin tuloksia raportoidaan, on tavallista ja erittäin hyödyllistä ilmoittaa estimaattien arvojen yhteydessä niiden tavalla tai toisella lasketut keskivirheet. Monet tietokoneohjelmat tekevätkin tämän automaattisesti. Näin lukija saa jonkinlaisen käsityksen estimaattien tarkkuudesta: hän voi esimerkiksi valita mieleisensä luottamustason $1 - \alpha$ ja todeta, että väli

$$(\hat{\theta} - z_{\alpha/2}\,\mathrm{s.e.}(\hat{\theta}),\ \hat{\theta} + z_{\alpha/2}\,\mathrm{s.e.}(\hat{\theta}))$$

on θ:n approksimatiivinen luottamusväli tällä tasolla. Koska $z_{0.025} \approx 1.96 \approx 2$, on nyrkkisääntönä hyvä muistaa, että erityisesti $(\hat{\theta} - 2\,\mathrm{s.e.}(\hat{\theta}),\ \hat{\theta} + 2\,\mathrm{s.e.}(\hat{\theta}))$ on likimääräinen 95 %:n luottamusväli.

6.4.3 Esimerkki: eksponenttimalli. Mallissa $Y_1, \ldots, Y_n \sim Exp(1/\mu)$ $\perp\!\!\!\perp$ on $\hat{\mu} = \overline{y}$ ja $i(\mu) = n/\mu^2$ (ks. 5.6.7). Keskivirheeksi saadaan siten $\text{s.e.}(\hat{\mu}) = 1/i(\hat{\mu})^{1/2} = \overline{y}/\sqrt{n}$. Samaan tulokseen päädytään myös laskemalla $\text{var}(\hat{\mu}) = \mu^2/n$ ja ottamalla tämän neliöjuuri sekä korvaamalla μ estimaatillaan $\overline{y}$.

Esimerkiksi jos $n = 20$ ja $\hat{\mu} = \overline{y} = 5$, niin $\text{s.e.}(\hat{\mu}) = 5/\sqrt{20} \approx 1.12$ ja approksimatiiviseksi 95 %:n luottamusväliksi saadaan $(3.8, 7.2)$ (vrt. 6.3.2).

6.4.4 Waldin testiin perustuva luottamusjoukko kun $d > 1$. Tarkastellaan jälleen vektoriparametrista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta})$, jossa $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$. Tehtävänä on muodostaa luottamusjoukko q-ulotteiselle kiinnostavalle parametrille $\boldsymbol{\psi} = (\theta_1, \ldots, \theta_q)$.

Kohdassa 5.7.5 esiteltiin Waldin testisuureen vektoriversio

$$w(\boldsymbol{Y}) = (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)' \, \boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}})^{-1} \, (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

joka noudattaa asymptoottisesti χ^2_q-jakaumaa, kun $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Tässä $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}})$ on tavallinen su-estimaattori ja $\boldsymbol{i}^{\psi,\psi}$ viittaa Fisherin informaatiomatriisin käänteismatriisin vasempaan ylälohkoon $(q \times q)$. Näin ollen vektorille $\boldsymbol{\psi}$ saadaan luottamustasolla $1 - \alpha$ approksimatiivinen luottamusjoukko

$$\{\boldsymbol{\psi} : (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})' \, \boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}})^{-1} \, (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) < \chi^2_q(\alpha)\}, \tag{6.4}$$

kun $\chi^2_q(\alpha)$ on se piste, jonka oikealla puolella on osuus α χ^2_q-jakauman todennäköisyysmassasta.

Joukko (6.4) on $\hat{\boldsymbol{\psi}}$-keskisen ellipsoidin rajoittama alue q-ulotteisessa avaruudessa. Ellipsoidin muodon eli akselit määrää symmetrinen matriisi $\boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}})$. Seuraavassa tarkastellaan hieman lähemmin tapauksia $q = 1$ ja $q = 2$.

6.4.5 $q = 1$: luottamusväli parametrin yhdelle komponentille. Jos $q = 1$, joukko (6.4) on parametrin ensimmäisen komponentin θ_1 luottamusjoukko

$$\{\theta_1 : (\hat{\theta}_1 - \theta_1)^2 / i^{1,1}(\hat{\boldsymbol{\theta}}) < \chi^2_1(\alpha)\}.$$

Koska χ^2_1 on standardinormaalijakauman neliö, tämä on itse asiassa väli

$$(\hat{\theta}_1 - z_{\alpha/2} \, i^{1,1}(\hat{\boldsymbol{\theta}})^{1/2}, \ \hat{\theta}_1 + z_{\alpha/2} \, i^{1,1}(\hat{\boldsymbol{\theta}})^{1/2}).$$

Nyt siis su-estimaattorin $\hat{\theta}_1$ keskivirhe eli keskihajonnan estimaatti on $\text{s.e.}(\hat{\theta}_1) = i^{1,1}(\hat{\boldsymbol{\theta}})^{1/2}$. Kaikki, mitä kohdassa 6.4.2 keskivirheestä todettiin, on relevanttia tässäkin tapauksessa.

Huomattakoon lopuksi, että tässä esitetty luottamusvälin konstruktio on luonnollisestikin sovellettavissa mihin tahansa $\boldsymbol{\theta}$:n komponenttiin θ_j, $j = 1, \ldots, d$, kunhan vain korvataan indeksi 1 indeksillä j. Estimaattorin $\hat{\theta}_j$ keskivirhe saadaan ottamalla neliöjuuri vastaavasta informaatiomatriisin käänteismatriisin alkiosta: $\text{s.e.}(\hat{\theta}_j) = i^{j,j}(\hat{\boldsymbol{\theta}})^{1/2}$. (Ks. merkinnät kohdassa 3.4.6.)

6.4.6 $q = 2$: luottamusellipsi parametrin komponenttien parille. Jos $q = 2$, joukko (6.4) on parin (θ_1, θ_2) luottamusjoukko, joka on ellipsin rajoittama alue tasossa. Tämän *luottamusellipsin* keskipiste on $(\hat{\theta}_1, \hat{\theta}_2)$, ja sen akselit määrää symmetrinen kerroinmatriisi

$$\boldsymbol{i}^{\psi,\psi}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} i^{1,1}(\hat{\boldsymbol{\theta}}) & i^{1,2}(\hat{\boldsymbol{\theta}}) \\ i^{2,1}(\hat{\boldsymbol{\theta}}) & i^{2,2}(\hat{\boldsymbol{\theta}}) \end{bmatrix}.$$

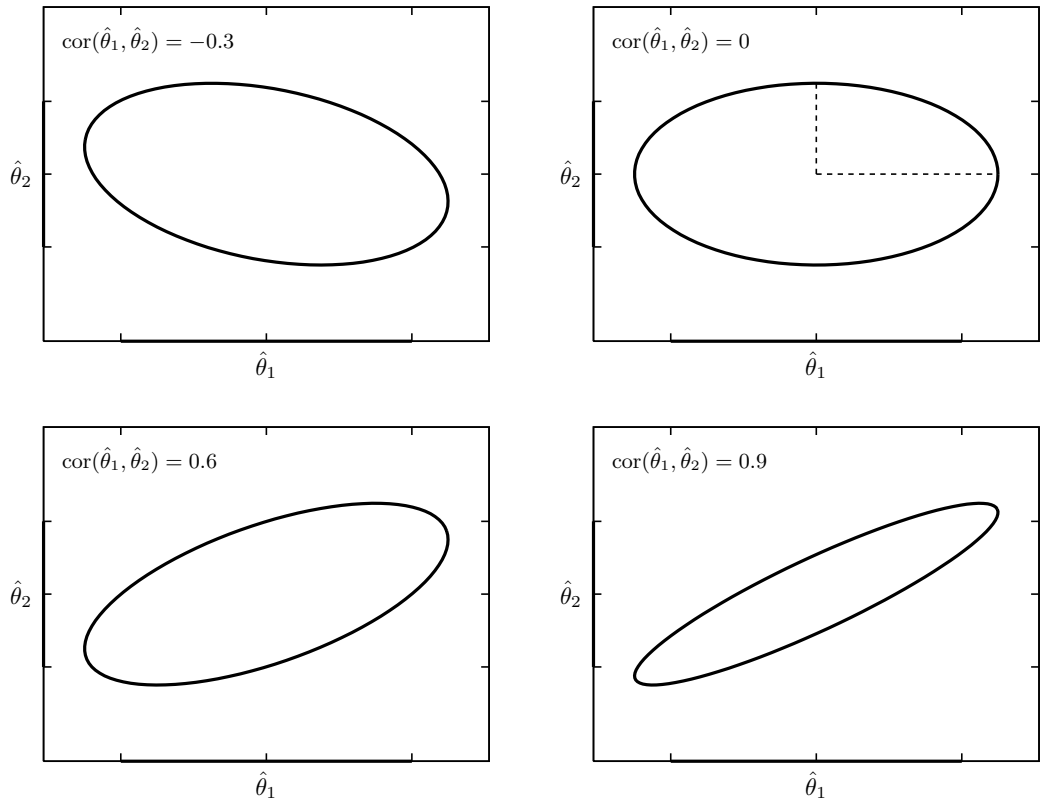
Diagonaalialkiot ovat keskivirheiden neliöt eli estimaatit vastaavien su-estimaattorien $\hat{\theta}_1$ ja $\hat{\theta}_2$ variansseille. Alkio $i^{1,2}(\hat{\boldsymbol{\theta}}) = i^{2,1}(\hat{\boldsymbol{\theta}})$ puolestaan on estimaatti kovarianssille. Niinpä voidaan sanoa, että luku

$$\operatorname{cor}(\hat{\theta}_1, \hat{\theta}_2) = \frac{i^{1,2}(\hat{\boldsymbol{\theta}})}{i^{1,1}(\hat{\boldsymbol{\theta}})^{1/2} \cdot i^{2,2}(\hat{\boldsymbol{\theta}})^{1/2}}$$

on arvio estimaattorien $\hat{\theta}_1$ ja $\hat{\theta}_2$ korrelaatiolle.

Muodostetun luottamusellipsin akselit ovat koordinaattiakselien suuntaiset jos ja vain jos $i^{1,2}(\hat{\boldsymbol{\theta}}) = 0$ eli $\operatorname{cor}(\hat{\theta}_1, \hat{\theta}_2) = 0$. Näin käy silloin kun θ_1 ja θ_2 kuuluvat toisiinsa nähden ortogonaalisiin parametrin osiin (ks. 2.6.2). Puoliakselien pituudet ovat tällöin $[\chi^2_2(\alpha) i^{j,j}(\hat{\boldsymbol{\theta}})]^{1/2}$, $j = 1, 2$.

Jos $\operatorname{cor}(\hat{\theta}_1, \hat{\theta}_2)$ on itseisarvoltaan suuri (lähellä yhtä), luottamusellipsi on koordinaattiakseleihin nähden hyvin kalteva ja eksentrinen. Tällainen tilanne viittaa epäonnistumiseen mallin ja sen parametroinnin valinnassa, sillä estimaattorit $\hat{\theta}_1$ ja $\hat{\theta}_2$ riippuvat voimakkaasti toisistaan ja saattavat olla epästabiileja. Erityisesti yksiulotteiset luottamusvälit antavat tällöin virheellisen kuvan parin (θ_1, θ_2) yhteisestä luottamusjoukosta. Katso kuvaa 6.2.



Kuva 6.2. Waldin testisuureeseen perustuvia 95 %:n luottamusellipsejä parille (θ_1, θ_2), kun $\text{s. e.}(\hat{\theta}_1) = i^{1,1}(\hat{\boldsymbol{\theta}})^{1/2} = 2$ ja $\text{s. e.}(\hat{\theta}_2) = i^{2,2}(\hat{\boldsymbol{\theta}})^{1/2} = 1$. Tapauksessa $\operatorname{cor}(\hat{\theta}_1, \hat{\theta}_2) = 0$ puoliakselien (katkoviiva) pituudet ovat likimain 4.90 ja 2.45, koska $\chi^2_2(0.05)^{1/2} \approx 2.45$. Kuviin on merkitty myös 95 %:n luottamusvälit $\big(\hat{\theta}_j - 1.96\,\text{s. e.}(\hat{\theta}_j),\ \hat{\theta}_j + 1.96\,\text{s. e.}(\hat{\theta}_j)\big)$.

Harjoitustehtäviä

6.1. Olkoot $Y_1, \ldots, Y_{25} \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, ja merkitään $S^2 = \sum_{i=1}^{25}(Y_i - \overline{Y})^2/24$. Osoita, että S/σ on saranasuure eli sen jakauma ei riipu μ:stä eikä σ^2:sta. Etsi keskihajonnalle σ ylempi 95 %:n luottamusraja b eli 95 %:n luottamusväli muotoa $(0, b)$, kun on havaittu $s = 10$.

6.2. Olkoot $Y_1 \perp\!\!\!\perp Y_2$ ja $Y_1 \sim N(\mu_1, 1)$ sekä $Y_2 \sim N(\mu_2, 1)$. Etsi sellaiset luvut $a, b > 0$, että

$$P\{|Y_1 - \mu_1| \le a,\ |Y_2 - \mu_2| \le a\} = 0.95,$$
$$P\{(Y_1 - \mu_1)^2 + (Y_2 - \mu_2)^2 \le b^2\} = 0.95.$$

Havaittu aineisto on $(y_1, y_2) = (1, 0.5)$. Mitkä kaksi 95 %:n luottamusjoukkoa saadaan yo. yhtälöiden perusteella parametriparille (μ_1, μ_2)? Piirrä kuva. Kumpi luottamusjoukoista on mielestäsi parempi? *Ohje.* Tarvitset jakaumien $N(0, 1)$ ja χ^2_2 taulukoita.

6.3. Olkoot $Y_1, \ldots, Y_{10} \sim N(\mu, 1)$ $\perp\!\!\!\perp$. Tunnetusti parametrilla μ on 95 %:n luottamusväli $(\overline{y} - 1.96/\sqrt{10},\ \overline{y} + 1.96/\sqrt{10})$, kun $\overline{y} = (y_1 + \cdots + y_{10})/10$ on havaintojen keskiarvo. Oletetaan, että 11. havainto Y_{11} noudattaa myös samaa jakaumaa $N(\mu, 1)$ ja on riippumaton muuttujista $Y_1, \ldots, Y_{10}$. Olkoon p todennäköisyys sille, että Y_{11} kuuluu em. luottamusväliä vastaavaan satunnaiseen väliin $(\overline{Y} - 1.96/\sqrt{10},\ \overline{Y} + 1.96/\sqrt{10})$. Arvaa, onko p suurempi, pienempi vai yhtäsuuri kuin 0.95. Tarkista arvauksesi laskemalla.

6.4. Muodosta tehtävän 2.10 tilanteessa odotusarvolle μ approksimatiivinen 99 %:n luottamusväli a) uskottavuusosamäärän testisuureeseen, b) Waldin testisuureeseen perustuen.

6.5. Oletetaan, että $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat kukin jatkuvaa jakaumaa, jonka tiheysfunktio on

$$f(y; \theta) = 2\theta^{-1} y \exp(-y^2/\theta), \qquad y > 0,$$

ja jossa θ on positiivinen parametri (ks. teht. 2.5). Laske keskivirhe $\mathrm{s.e.}(\hat{\theta})$ ja muodosta sen avulla θ:lle approksimatiivinen 95 %:n luottamusväli, kun aineisto on $\boldsymbol{y} = (y_1, \ldots, y_n)$.

Liite: jakaumia

Taulukossa on lueteltu useimmat näissä muistiinpanoissa esiintyvät jakaumat, niiden tunnukset, pistetodennäköisyys- tai tiheysfunktiot ja odotusarvot sekä varianssit (multinormaalijakauman osalta odotusarvovektori ja kovarianssimatriisi).

Symbolilla Γ on merkitty Eulerin gammafunktiota:

$$\Gamma(\kappa) = \int_0^\infty t^{\kappa-1} e^{-t}\, dt, \qquad \kappa > 0.$$

Nimi	Tunnus	Parametrit	Arvojoukko	Pistetodennäköisyys- tai tiheysfunktio	Odotus-arvo	Varianssi
Bernoulli	$B(\theta)$	$0 < \theta < 1$	$\{0, 1\}$	$\theta^y (1-\theta)^{1-y}$	θ	$\theta(1-\theta)$
binomi	$Bin(n, \theta)$	$n = 1, 2, \ldots,$ $0 < \theta < 1$	$\{0, 1, \ldots, n\}$	$\binom{n}{y} \theta^y (1-\theta)^{n-y}$	$n\theta$	$n\theta(1-\theta)$
geometrinen	$Geom(\theta)$	$0 < \theta < 1$	$\{1, 2, 3, \ldots\}$	$\theta(1-\theta)^{y-1}$	$\frac{1}{\theta}$	$\frac{1-\theta}{\theta^2}$
Poisson	$P(\mu)$	$\mu > 0$	$\{0, 1, 2, \ldots\}$	$e^{-\mu} \frac{\mu^y}{y!}$	μ	μ
eksponentti	$Exp(\lambda)$	$\lambda > 0$	$(0, \infty)$	$\lambda e^{-\lambda y}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
normaali	$N(\mu, \sigma^2)$	$\mu \in \mathbb{R},$ $\sigma^2 > 0$	$\mathbb{R}$	$\frac{\exp\{-(y-\mu)^2/2\sigma^2\}}{\sqrt{2\pi\sigma^2}}$	μ	σ^2
tasainen	$Tas(\alpha, \beta)$	$\alpha < \beta$	(α, β)	$\frac{1}{\beta - \alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$
gamma	$G(\kappa, \lambda)$	$\kappa > 0,$ $\lambda > 0$	$(0, \infty)$	$\frac{\lambda^\kappa}{\Gamma(\kappa)} y^{\kappa-1} e^{-\lambda y}$	$\frac{\kappa}{\lambda}$	$\frac{\kappa}{\lambda^2}$
khii-toiseen	χ^2_n	$n = 1, 2, \ldots$	$(0, \infty)$	$\frac{1}{2^{n/2}\Gamma(\frac{1}{2}n)} y^{n/2-1} e^{-y/2}$	n	$2n$
t	t_n	$n = 1, 2, \ldots$	$\mathbb{R}$	$\frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)\sqrt{n\pi}} \left(1 + \frac{y^2}{n}\right)^{-(n+1)/2}$	0	$\frac{n}{n-2}$
multi-normaali	$N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$d = 1, 2, \ldots,$ $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$	$\mathbb{R}^d$	$\frac{\exp\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu})\}}{(2\pi)^{d/2}\sqrt{\det(\boldsymbol{\Sigma})}}$	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$