Tilastollisen päättelyn jatkokurssi

Pentti Saikkonen

Syksy 2010

Sisältö

1 Asymptoottista teoriaa[1]

1.1 Konvergenssikäsitteitä ja tuloksia

Stokastinen konvergenssi. Palautetaan mieleen stokastisen konvergenssin määritelmä.[2] Jono satunnaismuuttujia (sm) $X_1, X_2, ...$ konvergoi stokastisesti kohti sm:aa Z, jos

$$\mathsf{P}\left\{|X_n - Z| > \varepsilon\right\} \underset{n\to\infty}{\to} 0 \quad \text{kaikilla } \varepsilon > 0. \tag{1.1}$$

Tästä käytetään merkintää $X_n \xrightarrow{p} Z$ tai $plim_{n\to\infty} X_n = Z$. Huomaa, että määritelmässä voi vaihtoehtoisesti olla tapahtuma $\{|X_n - Z| \geq \varepsilon\}$. Jos $X_1, X_2, ...$ ja Z ovat $k \times 1$ satunnaisvektoreita (sv), määritellään stokastinen konvergenssi komponenteittain eli

$$X_n \xrightarrow{p} Z \quad \Leftrightarrow \quad X_{a,n} \xrightarrow{p} Z_a \quad \text{kaikilla } a = 1, ..., k,$$

kun esimerkiksi $X_{a,n}$ on X_n:n a. komponentti. On helppo nähdä, että tämä on yhtäpitävää sen kanssa, että konvergenssissa (1.1) itseisarvo korvataan vektorin normilla $\|\cdot\|$.[3] Satunnaismatriisiin tapauksessa stokastisen konvergenssin määritelmä voidaan palauttaa satunnaisvektorin määritelmään suorittamalla "vektorointi".

Tilastollisissa sovelluksissa on usein tarpeen tietää miten stokastinen konvergenssi käyttäytyy (jatkuvissa) funktiomuunnoksissa. Tätä varten todetaan seuraava tulos. Tässä kuten myöhemminkin tarkoitetaan vakiolla ei-satunnaista (äärellistä) suuretta.

Lause 1.1. Jos jono sv:ta $X_1, X_2, ...$ $(k \times 1)$ konvergoi stokastisesti kohti vakiovektoria c eli $X_n \xrightarrow{p} c$ ja $g : \mathbb{R}^k \to \mathbb{R}^l$ jatkuva pisteessä c, niin $g(X_n) \xrightarrow{p} g(c)$.

Todistus: Oletetan yksinkertaisuuden vuoksi, että $k = l = 1$. Koska g on jatkuva pisteessä c, niin jatkuvuuden määritelmän mukaan jokaista $\varepsilon > 0$ kohti on olemassa $\delta > 0$ siten, että $|g(x) - g(c)| < \varepsilon$, kun $|x - c| < \delta$. Tästä seuraa tapahtumarelaatio $\{|X_n - c| < \delta\} \subseteq \{|g(X_n) - g(c)| < \varepsilon\}$, joten

$$\mathsf{P}\left\{|g(X_n) - g(c)| < \varepsilon\right\} \geq \mathsf{P}\left\{|X_n - c| < \delta\right\} \underset{n\to\infty}{\to} 1.$$

Tässä konvergenssi seuraa oletuksesta $X_n \xrightarrow{p} c$, kun stokastisen konvergenssin määrittelyehdon (1.1) tapahtuma korvataan komplementtitapahtumallaan. $\square$

Lauseen tulos pitää paikkansa myös, kun raja-arvo on satunnainen, mutta todistus ei onnistu kopioimalla edellä esitettyä todistusta aivan suoraan. Jatkossa tätä yleistystä ei kuitenkaan tarvita. Lausetta voidaan soveltaa esimerkiksi tapauksissa $g(x_1, x_2) = x_1 + x_2$ ja $g(x_1, x_2) = x_1 x_2$ (olettaen summa ja tulo määritellyiksi).

[1] Tämän jakson oheislukemistona voi käyttää C.R. Raon teoksen "Linear Statistical Inference and Its Applications. Second Edition" (Wiley, 1973) lukua 2.c tai R.J. Serflingin teoksen "Approximation Theorems of Mathematical Statistics" (Wiley, 1980) lukua 1.

[2] Stokastista konvergenssia käytetään tilastotieteessä mm. estimaattorin (heikon) tarkentuvuuden määrittelyyn. Tarkentuvuus voidaan määritellä myös käyttäen melkein varmaa konvergenssia, jota ei kurssilla kuitenkaan käytetä tarkentuvuuden eikä muidenkaan tarkastelujen yhteydessä.

[3] Jos $x \in \mathbb{R}^k$, niin $\|x\| = \sqrt{x_1^2 + \cdots + x_k^2}$.

Jakaumakonvergenssi. Tarkastellan seuraavaksi jakaumakonvergenssia. Olkoon $F_{X_1}, F_{X_2}, , ...$ jono sm:ien $X_1, X_2, ...$ kertymäfunktioita ja F_Z sm:n Z kertymäfunktio (esim. $F_Z(x) = \mathsf{P}\{Z \leq x\}$). Jonon $X_1, X_2, ...$ sanotaan konvergoivan jakaumaltaan kohti sm:aa Z, jos

$$\lim_{n\to\infty} F_{X_n}(x) = F_Z(x)$$

kaikissa F_Z:n jatkuvuuspisteissä. Yleistys sv:lle sujuu käyttämällä sv:ien kertymäfunktioita (esim. $F_Z(x) = \mathsf{P}\{Z_1 \leq x_1, ..., Z_k \leq x_k\}$, kun $Z = (Z_1, ..., Z_k)$). Jakaumakonvergenssista käytetään merkintää $X_n \xrightarrow{d} Z$.

Tilastotieteessä rajajakauma on useimmiten jatkuva, joten edellä $F_Z(x)$ on jatkuva kaikilla x. Epäjatkuvuuspisteitä voi yleisestikin olla korkeintaan numeroituva määrä. (Tämä seuraa kertymäfunktion monotonisuudesta ja tunnetusta monotonisia funktioita koskevasta tuloksesta.) Kun F_Z on jatkuva, niin "suurilla" n:n arvoilla ja kaikilla x pätee $\mathsf{P}\{X_n \leq x\} \approx \mathsf{P}\{Z \leq x\}$, joten kaikkien sm:aa X_n koskevien tapahtumien todennäköisyyksiä voidaan approksimoida käyttäen sm:n Z jakaumaa. Tyypillisessä tilastollisessa sovelluksessa X_n on testisuure tai (muunnettu) estimaattori.

Edellä esitetyn jakaumakonvergenssin tai toiselta nimeltään heikon konvergenssin määritelmä on havainnollinen, mutta se voi olla etenkin vektoritapauksessa matemaattisesti hankala. Seuraavassa lauseessa esitettävät tulokset ovat siten teoreettisesti hyödyllisiä. Tässä lauseessa käytetään sv:n karakteristista funktiota, jonka määritelmä on sv:n $Z = (Z_1, ..., Z_k)$ tapauksessa

$$\varphi_Z(t) = \mathsf{E}\left[\exp\left(it'Z\right)\right] = \mathsf{E}\left[\cos\left(t'Z\right)\right] + i\mathsf{E}\left[\sin\left(t'Z\right)\right],$$

jossa $i = \sqrt{-1}$, $t = (t_1, ..., t_k)$ ja $t'X = \sum_{j=1}^{k} t_j Z_j$.[4]

Lause 1.2. Olkoon $X_1, X_2, ...$ ja Z sv:ta $(k \times 1)$. Seuraavat kolme konvergenssia pätevät yhtäpitävästi.

(i) $X_n \xrightarrow{d} Z$

(ii) $\varphi_{X_n}(t) \underset{n\to\infty}{\rightarrow} \varphi_Z(t)$ jokaisella $t \in \mathbb{R}^k$.

(iii) $\mathsf{E}[h(X_n)] \underset{n\to\infty}{\rightarrow} \mathsf{E}[h(Z)]$ kaikilla reaaliarvoisilla rajoitetuilla ja jatkuvilla funktioilla $h : \mathbb{R}^k \to \mathbb{R}$.

Lauseen todistus ei ole aivan yksinkertainen ja sivuutetaan. Ekvivalenssit (i) ja (ii) osoittavat, että jakaumakonvergenssi voidaan määritellä karakterististen funktioiden konvergenssin avulla. Jos kohdan (ii) konvergenssi voidaan todeta ja raja-arvo $\varphi_Z(t)$ tunnistetaan esimerkiksi $\mathsf{N}\left(0, \sigma^2\right)$–jakauman karakteristiseksi funktioksi, voidaan päätellä, että kohdan (i) jakaumakonvergenssi pätee ja $Z \sim \mathsf{N}\left(0, \sigma^2\right)$. Viimeksi mainittu seikka perustuu todennäköisyyslaskennan tunnettuun tulokseen, jonka

[4] Vektorit tulkitaan matriisilaskuissa aina pystyvektoreiksi ja niistä voidaan käyttää myös matriislaskennan merkintää eli esimerkiksi $x = [x_1 \ \cdots \ x_k]'$, jossa pilkku osoittaa vektorin (samoin kuin myöhemmin matriisin) transponointia.

mukaan karakteristinen funktio määrää jakauman yksikäsitteisesti. Tällä tavoin voidaan todistaa verraten helposti keskeinen raja-arvolause riippumattomien ja samoin jakautuneiden muuttujien eli ns. iid–*muuttujien* tapauksessa.

Kohdan (iii) ehdon yhteyttä jakaumakonvergenssiin voi olla ensi silmäyksellä hankala hahmottaa. Se tulee havainnollisemmaksi, kun tarkastellaan tapausta $k = 1$ ja asetetaan funktion h paikalle (epäjatkuva) välin $(-\infty, x]$ indikaattorifunktio $\mathsf{I}(X_n \leq x)$, joka saa arvon 1, kun $X_n \leq x$ ja arvon 0, kun $X_n > x$. Tällöin kohdan (iii) ensimmäisen odotusarvon paikalla on $\mathsf{E}[\mathsf{I}(X_n \leq x)] = \mathsf{P}\{X_n \leq x\} = F_{X_n}(x)$. Jos unohdetaan funktion h jatkuvuusvaatimus, saadaan yhteys jakaumakonvergenssiin. Epäjatkuvuus voidaan tässä kuitenkin "hoitaa" huomaamalla, että välin $(-\infty, x]$ indikaattorifunktiota voidaan approksimoida mielivaltaisen tarkasti jatkuvalla funktiolla (ajattele esim. $\mathsf{N}(x, \sigma^2)$–jakauman kertymäfunktiota, kun $\sigma^2 \to 0$).

Jakaumakonvergenssin "jatkuvan kuvauksen" lause voidaan johtaa helposti Lauseen 1.2 kohtien (i) ja (iii) ekvivalenssista. Jos $g : \mathbb{R}^k \to \mathbb{R}^l$ on jatkuva ja $h : \mathbb{R}^l \to \mathbb{R}$ on rajoitettu ja jatkuva, niin yhdistetty kuvaus $(h \circ g) : \mathbb{R}^k \to \mathbb{R}$ on rajoitettu ja jatkuva. Jakaumakonvergenssista $X_n \xrightarrow{d} Z$ seuraa siten (Lause 1.2; (i) $\Rightarrow$ (iii)) $\mathsf{E}[(h \circ g)(X_n)] \to \mathsf{E}[(h \circ g)(Z)]$ eli $\mathsf{E}[h(g(X_n))] \to \mathsf{E}[h(g(Z))]$, kun $n \to \infty$. Tästä puolestaan seuraa (Lause 1.2; (iii) $\Rightarrow$ (i)) $g(X_n) \xrightarrow{d} g(Z)$. Esitetään tämä vielä lauseena.

Lause 1.3. Jos jono sv:ta $X_1, X_2, \ldots$ $(k \times 1)$ konvergoi jakaumaltaan kohti sv:ia Z eli $X_n \xrightarrow{d} Z$ ja $g : \mathbb{R}^k \to \mathbb{R}^l$ on jatkuva, niin $g(X_n) \xrightarrow{d} g(Z)$.

Mainittakoon, että tämän lauseen jatkuvuusoletusta voidaan lieventää. Funktion g jatkuvuus riittää vaatia vain joukossa C_g, jolle pätee $\mathsf{P}\{Z \in C_g\} = 1$.

Stokastista ja jakaumakonvergenssia koskevia tuloksia. Parametrien estimointia ja hypoteesien testausta koskevissa asymptoottisissa tarkasteluissa joudutaan usein käyttämään sekä jakaumakonvergenssia että stokastista konvergenssia. Seuraava lause koskee tällaista tilannetta. Sen todistuksessa käytetään surkastunutta sm:aa Y, jolle pätee $\mathsf{P}\{Y = c\} = 1$ jollain vakiolla c. Tällöin Y:n kertymäfunktio on $F_Y(y) = \mathsf{I}(y \geq c)$ eli

$$F_Y(y) = \begin{cases} 0, & \text{kun } y < c \\ 1, & \text{kun } y \geq c. \end{cases}$$

Vektoritapauksessa määritelmä on samanlainen (kertymäfunktiossa epäyhtälö tulkitaan komponenteittain). Surkastunut satunnaismuuttuja on tunnetusti riippumaton mistä tahansa toisesta sm:sta.[5]

Lause 1.4. Oletetaan, että sv–jonoille $X_1, X_2, \ldots$ $(k \times 1)$ ja $Y_1, Y_2, \ldots$ $(l \times 1)$ pätee $X_n \xrightarrow{d} Z$ ja $Y_n \xrightarrow{p} c$, jossa c on vakiovektori. Tällöin $(X_n, Y_n) \xrightarrow{d} (Z, c)$.

[5] Jos Y on surkastunut ja X mikä tahansa toinen satunnaismuuttuja, niin $\mathsf{P}\{Y \leq y\}$ on nolla tai yksi, sillä esimerkiksi edellisessä tapauksessa $0 \leq \mathsf{P}\{X \leq x, Y \leq y\} \leq \mathsf{P}\{Y \leq y\}\,\mathsf{P}\{X \leq x, Y \leq y\} = 0 = \mathsf{P}\{X \leq x\}\,\mathsf{P}\{Y \leq y\}$.

Todistus: Oletetaan $k = l = 1$. Olkoon x mikä tahansa Z:n kertymäfunktion F_Z jatkuvuuspiste. Jos $y < c$, niin

$$0 \leq \mathsf{P}\{X_n \leq x, Y_n < y\} \leq \mathsf{P}\{Y_n < y\} \underset{n\to\infty}{\to} 0. \tag{a}$$

Tässä konvergenssi seuraa oletuksesta $Y_n \xrightarrow{p} c$, koska millä tahansa $\varepsilon < c - y$ pätee $\mathsf{P}\{Y_n < y\} \leq \mathsf{P}\{|Y_n - c| > \varepsilon\}$ (piirrä kuva!). Vastaavalla tavalla nähdään, että jokaisella $y > c$ pätee

$$\mathsf{P}\{Y_n \geq y\} \underset{n\to\infty}{\to} 0. \tag{b}$$

Koska $\mathsf{P}\{A\} - \mathsf{P}\{B^c\} \leq \mathsf{P}\{A \cap B\}$ mille tahansa tapahtumille A ja B,[6] pätee tapauksessa $y > c$ lisäksi

$$\mathsf{P}\{X_n \leq x\} - \mathsf{P}\{Y_n \geq y\} \leq \mathsf{P}\{X_n \leq x, Y_n < y\} \leq \mathsf{P}\{X_n \leq x\}. \tag{c}$$

Koska $X_n \xrightarrow{d} Z$, seuraa tuloksista (a), (b) ja (c)

$$\mathsf{P}\{X_n \leq x, Y_n < y\} \underset{n\to\infty}{\to} \begin{cases} 0, & \text{kun } y < c \\ F_Z(x), & \text{kun } y > c. \end{cases}$$

Jos sm:lle Y pätee $\mathsf{P}\{Y = c\} = 1$, niin raja-arvo voidaan ilmaista $F_Z(x)\,\mathsf{I}(c < y) = \mathsf{P}\{Z \leq x\}\,\mathsf{P}\{Y < y\} = \mathsf{P}\{Z \leq x, Y < y\}$, jossa jälkimmäinen yhtälö seuraa, koska $Z \perp\!\!\!\perp Y$. Tapauksessa $y = c$, raja-arvo ei ole määritelty. Jos se määritellään tässä tapauksessa kaavan $F_Z(x)\,\mathsf{I}(y \geq c)$ mukaisesti, pätee konvergenssi kaikissa sv:n (Z, Y) kertymäfunktion jatkuvuuspisteissä, sillä (x, c) on epäjatkuvuuspiste aina, kun $F_Z(x) > 0$. Väite seuraa jakaumaidentiteetistä $(Z, Y) \stackrel{d}{=} (Z, c)$. $\square$

Lauseiden 1.3 ja 1.4 suorina seurauksia saadaan mm. seuraavat tulokset.

Seuraus 1.1. Oletetaan Lauseen 1.4 tilanne eli $X_n \xrightarrow{d} Z$ ja $Y_n \xrightarrow{p} c$ (vakio) ja lisäksi että $k = l = 1$. Tällöin,

(i) $X_n + Y_n \xrightarrow{d} Z + c$

(ii) $X_n Y_n \xrightarrow{d} Zc$

(iii) $X_n / Y_n \xrightarrow{d} Z/c$, kun $c \neq 0$.

Tässä ensimmäinen tulos pätee tietysti myös vektoritapauksessa, jota tarkastellaan seuraavassa, jossa Y_n on $k \times l$ matriisi.

Seuraus 1.2. Olkoon $X_1, X_2, ...$ ja Z $(k \times 1)$ kuten Lauseessa 1.4 eli $X_n \xrightarrow{d} Z$ ja $Y_1, Y_2, ...$ jono $l \times k$ matriiseja, jolle pätee $Y_n \xrightarrow{p} A$, jossa A $(l \times k)$ on vakiomatriisi. Tällöin,

[6] Tästä kannattaa jälleen piirtää kuva. Formaalisti väite perustuu inkluusioon (B^c on joukon B komplementti) $A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) \subseteq (A \cap B) \cup B^c$.

(i) $Y_n X_n \xrightarrow{d} AZ$

Jos $k = l$, pätee myös

(ii) $X_n' Y_n X_n \xrightarrow{d} Z'AZ$

(iii) $X_n' Y_n^{-1} X_n \xrightarrow{d} Z'A^{-1}Z$, kun A on epäsingulaarinen (eli kääntyvä).

Nämä tulokset seuraavat funktioiden $(A, z) \mapsto Az$, $(A, z) \mapsto z'Az$ ja $A \mapsto A^{-1}$ (A epäsingulaarinen) jatkuvuudesta ja Lauseista 1.3 ja 1.4. Käänteismatriisin jatkuvuus voidaan perustella seuraavasti. Todetaan ensin, että determinanttifunktio $A \mapsto \det(A)$ on jatkuva, koska determinantti voidaan muodostaa summaamalla matriisin alkioista laskettuja tuloja (katso 2×2 tapausta). Koska käänteismatriisin A^{-1} alkiot voidaan lausua A:n ja sen alimatriisien determinanttien osamäärien avulla (jakajana $\det(A)$), on funktio $A \mapsto A^{-1}$ jatkuvien funktioiden yhdisteenä jatkuva.

Seurauksessa 1.1(iii) on implisiittisesti oletettu, että $Y_n \neq 0$ kaikilla n (vastaavasti Seurauksessa 1.2(iii) on oletettu Y_n:n epäsingulaarisuus kaikilla n). Vaikka oletuksista $Y_n \xrightarrow{p} c$ ja $c \neq 0$ seuraa, että tapahtuman $\{Y_n \neq 0\}$ todennäköisyys lähestyy ykköstä, ei ehdon $Y_n \neq 0$ toteutuminen äärellisillä n:n arvoilla ei ole kuitenkaan taattu ilman lisäoletuksia (kuten Y_n:n jakauman jatkuvuutta). Tämä teoreettinen hankaluus voitaisiin ottaa huomioon määrittelemällä $1/Y_n$ jollain (mielivaltaisella) tavalla ilman, että esitetty jakaumakonvergenssi muuttuisi mitenkään. Kuten kirjallisuudessa yleensäkin, ei tätä seikkaa oteta yksinkertaisuuden vuoksi eksplisiittisesti huomioon, vaan Seurauksen 1.1(iii) kaltaisissa tilanteissa $Y_n \neq 0$ oletetaan, jos Y_n:n stokastinen raja-arvo on nollasta poikkeava. Vastaava tulkinta tehdään ilman eri mainintaa Seurauksen 1.2(iii) kaltaisissa tilanteissa.

Valitsemalla Seurauksessa 1.1(i) $X_n = Z$ kaikilla n, $Y_n = U_n - Z$ ja $c = 0$ nähdään, että stokastisesta konvergenssista $U_n \xrightarrow{p} Z$ seuraa jakaumakonvergenssi $U_n \xrightarrow{d} Z$. Käyttäen samantapaisia argumentteja kuin Lauseen 1.4 todistuksessa voidaan osoittaa, että käänteinen implikaatio pätee siinä erikoistapauksessa, että jakaumakonvergenssi tapahtuu kohti vakiota.

Sovellus deltamenetelmään. Edellä esitettyjen tulosten avulla voidaan perustella ns. *deltamenetelmä*, joka reaaliarvoisessa tapauksessa on seuraava. Olkoon $\theta \in \Theta \subset \mathbb{R}$ estimoitava parametri ja θ_0 sen "todellinen" arvo. Oletetaan (yksinkertaisuuden vuoksi), että parametrivaruus Θ on avoin väli ja tarkastellaan θ_0:n asymptoottisesti normaalia estimaattoria $\hat{\theta}_n$ eli[7]

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathsf{N}\left(0, \sigma^2(\theta_0)\right) \qquad \left(\sigma^2(\theta_0) > 0\right).$$

[7] Tässä kuten usein myöhemminkin ilmaistaan jakaumakonvergenssin raja-arvo käyttäen jakauman symbolia, joten esimerkiksi $X_n \xrightarrow{d} \mathsf{N}(0, \sigma^2)$ tarkoittaa samaa kuin $X_n \xrightarrow{d} Z$, jossa $Z \sim \mathsf{N}(0, \sigma^2)$. Toinen vastaava merkintä on $X_n \xrightarrow{d} \chi^2_k$.

Jos $h : \Theta \to \mathbb{R}$ on jatkuvasti derivoituva funktio ja $h'(\theta_0) \neq 0$, niin muunnoksen $h(\theta_0)$ estimaattorille $h(\hat{\theta}_n)$ pätee

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \mathsf{N}\left(0, [h'(\theta_0)]^2 \sigma^2(\theta_0)\right).$$

Väliarvolauseeseen ja Lauseisiin 1.1, 1.3 ja 1.4 (tai jälkimmäisten seuraukseen) perustuva todistus jätetään tehtäväksi.

Tasainen stokastinen konvergenssi. Tilastollisessa päättelyssä tarkastellaan usein parametrista riippuvia sm:ia ja niistä muodostettuja jonoja. Esimerkkinä voidaan mainita uskottavuusfunktio ja sen johdannaiset. Olkoon $X_1(\theta), X_2(\theta), ..., \theta \in \Theta$, mahdollisesti vektori- tai matriisiarvoinen jono tällaisia satunnaisfunktioita (myös parametri θ voi olla vektori). Kun n viittaa havaintojen lukumäärään, riippuu $X_n(\theta)$ yleensä myös havaintovektorista $Y_n = (Y_1, ..., Y_n)$, jota ei kuitenkaan ole tässä merkitty näkyviin. Tyypillinen esimerkki on otoskeskiarvo $X_n(\theta) = n^{-1}\sum_{i=1}^{n} q(Y_i;\theta)$, jossa $q(\cdot;\cdot)$ on tunnettu funktio kuten esimerkiksi yksittäisen havainnon logaritmoitu uskottavuusfunktio riippumattomien samoin jakautuneiden havaintojen mallissa.

Määritellään tasainen stokastinen konvergenssi rajoittuen tilanteeseeen, jossa raja-arvo on ei-satunnainen funktio $c(\theta)$.[8] Määrittelyehto on reaaliarvoisessa tapauksessa

$$\sup_{\theta\in\Theta} |X_n(\theta) - c(\theta)| \xrightarrow{p} 0. \tag{1.2}$$

Koska tässä tarkastellaan poikkeaman $|X_n(\theta) - c(\theta)|$ maksimaalista arvoa, on $X_n(\theta)$ "suurilla" n ja "lähes ykkösen" todennäköisyydellä "lähellä" funktiota $c(\theta)$ eikä "läheisyys" riipu siitä mikä parametriarvo on kysymyksessä. Tällaista samanaikaista läheisyyttä ei vaadita lievemmässä pisteittäisessä stokastisessa konvergenssissa, jonka määrittelee ehto $X_n(\theta) \xrightarrow{p} c(\theta)$ jokaisella $\theta \in \Theta$ tai yhtäpitävästi

$$|X_n(\theta) - c(\theta)| \xrightarrow{p} 0 \quad \text{jokaisella } \theta \in \Theta. \tag{1.3}$$

Huomaa erityisesti, ettei ehto (1.3) takaa tulosta $|X_n(\theta_n) - c(\theta_n)| \xrightarrow{p} 0$, jossa $\{\theta_n\}$ on joukon Θ jono. Ehto (1.2) sen sijaan takaa tämän, sillä $|X_n(\theta_n) - c(\theta_n)|$ on pienempi tai yhtä kuin (1.2):n vasen puoli ja ehdoista $0 \leq Y_n \leq Z_n$ ja $Z_n \xrightarrow{p} 0$ seuraa $Y_n \xrightarrow{p} 0$ (perustelu jätetään tehtäväksi).[9]

Jos $X_n(\theta)$ on vektoriarvoinen, määritellään tasainen stokastinen konvergenssi komponenteittain tai korvaamalla ehdossa (1.2) itseisarvo normilla. Matriisitapaus voidaan palauttaa vektoritapaukseen matriisin "vektoroinnilla".

Tasaista stokastista konvergenssia voidaan käyttää todistettaessa suurimman uskottavuuden (SU) estimaattorin (ja muiden saman tyyppisten estimaattoreiden) tarkentuvuus. Toinen sovellus on seuraava. Olkoon $\hat{\theta}_n$ parametrin θ_0 tarkentuva estimaattori ($\hat{\theta}_n$, $\theta_0 \in \Theta$). Vaikka pisteittäinen stokastinen konvergenssi $X_n(\theta_0) \xrightarrow{p} c(\theta_0)$

[8] Tapaus, jossa raja-arvo on satunnainen $X(\theta)$, saadaan asettamalla $c(\theta) = 0$ ja tarkastelemalla jonoa $\{X_n(\theta) - X(\theta)\}_{n=1}^{\infty}$.

[9] Jos esimerkiksi $\mathsf{P}\{Z = 1\} = \mathsf{P}\{Z = -1\} = 1/2$, niin (keinotekoiselle) jonolle $X_n(\theta) = Zn\theta \mathsf{I}\left(0 \leq \theta \leq n^{-1}\right)$, $n \geq 1$, pätee $X_n(\theta) \xrightarrow{p} 0$ jokaisella $\theta \geq 0$, mutta $X_n\left(n^{-1}\right) = Z$, joten $\sup_{\theta\geq 0} |X_n(\theta)| \geq 1$ ja tasainen stokastinen konvergenssi nollaan ei päde.

tai yleisemmin $X_n(\theta) \xrightarrow{p} c(\theta)$ jokaisella $\theta \in \Theta$ pätisi, ei tästä ja tarkentuvuudesta $\hat{\theta}_n \xrightarrow{p} \theta_0$ voida päätellä tulosta $X_n(\hat{\theta}_n) \xrightarrow{p} c(\theta_0)$. Tasainen stokastinen konvergenssi ja funktion $c(\theta)$ jatkuvuus pisteessä θ_0 riittävät kuitenkin tähän. Tätä tulosta, joka esitetään täsmällisemmin seuraavassa lauseessa, käytetään myöhemmin tilanteessa, jossa $X_n(\theta)$ on n:llä jaettu havaittu informaatiomatriisi ja siten tyyppiä $X_n(\theta) = n^{-1}\sum_{i=1}^n q(Y_i;\theta)$ tai yleisemmin tyyppiä $X_n(\theta) = n^{-1}\sum_{i=1}^n q_i(Y_i;\theta)$.

Lause 1.5. Oletetaan, että jonolle satunnaisfunktioita $X_1(\theta), X_2(\theta), ...,\ \theta \in \Theta$, pätee $X_n(\theta) \xrightarrow{p} c(\theta)$ tasaisesti joukossa Θ. Jos $\hat{\theta}_n \xrightarrow{p} \theta_0$ ja funktio $c(\theta)$ on jatkuva pisteessä θ_0, niin $X_n(\hat{\theta}_n) \xrightarrow{p} c(\theta_0)$.

Todistus: Oletetaan reaaliarvoinen tapaus. Käyttäen kolmioepäyhtälöä saadaan

$$\begin{aligned}\left|X_n(\hat{\theta}_n) - c(\theta_0)\right| &\le \left|X_n(\hat{\theta}_n) - c(\hat{\theta}_n)\right| + \left|c(\hat{\theta}_n) - c(\theta_0)\right| \\ &\le \sup_{\theta\in\Theta}|X_n(\theta) - c(\theta)| + \left|c(\hat{\theta}_n) - c(\theta_0)\right|.\end{aligned}$$

Lauseen väite seuraa, koska viimeisen summan molemmat yhteenlaskettavat konvergoivat stokastisesti nollaan. Jälkimmäisen kohdalla tämä nähdään Lauseesta 1.1 ja edellisen kohdalla määrittelyehdosta (1.2). □

1.2 Raja-arvolauseita

1.2.1 Suurten lukujen laki

Todennäköisyyslaskennan kurssilla on esitetty iid–jonojen suurten lukujen laki (SLL). Jos $X_1, X_2, ...$ on iid–jono, jolla $\mathsf{E}(X_i) = \mu$ ja $\mathsf{Var}(X_i) = \sigma^2 < \infty$, niin

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Todistus seuraa helposti Tšebyševin (tai Markovin) epäyhtälöstä. Koska tilastollisen päättelyn sovelluksissa joudutaan usein tilanteisiin, joissa riippumattomuus ja samoin jakautuneisuus eivät päde, tarkastellaan seuraavassa yleistyksiä, joissa näistä vaatimuksista luovutaan. Riippumattomuutta ei yleensä voida perustella aikasarja-aineistoissa, joista esimerkkinä vaikkapa päivittäinen lämpötila Kaisaniemessä kesäkuussa 2010. Sama pätee lääketieteellisiin kokeisiin, joissa havaintoina on mittaustuloksia samoista koehenkilöistä eri ajankohtina. Tällöin eri koehenkilöiden mittaustuloksia voidaan kuitenkin pitää yleensä riippumattomina.

Oletetaan ensin, että jono $X_1, X_2, ...$ on riippumaton, mutta odotusarvot $\mathsf{E}(X_i) = \mu_i$ ja varianssit $\mathsf{Var}(X_i) = \sigma_i^2 < \infty$ riippuvat i:stä. Tällöin $\mathsf{E}\left(\bar{X}_n\right) = n^{-1}\sum_{i=1}^n \mu_i$ ja, koska riippumattomuuden nojalla $\mathsf{Var}\left(\bar{X}_n\right) = n^{-2}\sum_{i=1}^n \sigma_i^2$, saadaan Tšebyševin epäyhtälöstä

$$\mathsf{P}\left\{\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^n \mu_i\right| > \varepsilon\right\} \le \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^n \sigma_i^2.$$

Jos epäyhtälön oikea puoli konvergoi nollaan n:n kasvaessa rajatta, pätee SLL muodossa

$$\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i \xrightarrow{p} 0. \tag{1.4}$$

Yksinkertainen riittävä ehto tälle on, että varianssit ovat rajoitettuja eli $\sigma_i^2 \leq C < \infty$ kaikilla i. Jos tämän lisäksi oletetaan konvergenssi $n^{-1}\sum_{i=1}^{n}\mu_i \rightarrow \mu$ (äärellinen), saadaan tulos (ks. Lause 1.1)

$$\bar{X}_n \xrightarrow{p} \mu. \tag{1.5}$$

Tulkitsemalla tässä $\bar{X}_n$ satunnaisvektorin komponentista muodostetuksi keskiarvoksi saadaan perustelu vastaavalle moniulotteiselle yleistykselle.

Yksinkertaisena esimerkkinä riippuvuudesta tarkastellaan ns. m–riippuvuutta. Jono sv:ta $X_1, X_2, ...$ on *m–riippuva*, jos kaikilla $k \geq 1$ pätee

$$(X_1, ..., X_k) \perp\!\!\!\perp (X_{k+l}, X_{k+l+1}, ...), \quad \text{kun} \quad l > m.$$

Jos tässä l viittaa havaintoyksikköön, niin havainnot, joiden välinen "etäisyys" on suurempi kuin m ovat riippumattomia. Tämä on luonteva oletus, kun tarkastellaan riippumattomista koehenkilöistä saatuja mittauksia, kun kustakin koehenkilöstä on korkeintaan m mittausta (ks. jakso 2.1). Aikasarjatapauksessa riippumattomuus puolestaan pätee kaikille tapahtumille, joiden välinen ajallinen etäisyys on vähintään m aikayksikköä (ei välttämättä luontevaa edellä mainitussa lämpötilaesimerkissä).

SSL voidaan todeta helposti m–riippuville jonoille. Otoskeskiarvon varianssi on

$$\begin{aligned} \mathsf{Var}\left(\bar{X}_n\right) &= \frac{1}{n^2}\mathsf{Var}\left(\sum_{i=1}^{n}X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{Var}\left(X_i\right) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathsf{Cov}\left(X_i, X_j\right). \end{aligned}$$

Oletetaan jälleen yksinkertaisuuden vuoksi, että $\mathsf{Var}\left(X_i\right) \leq C < \infty$ pätee kaikilla i. Tällöin Cauchy-Schwarz'n epäyhtälön perusteella $|\mathsf{Cov}\left(X_i, X_j\right)| \leq [\mathsf{Var}\left(X_i\right)\mathsf{Var}\left(X_j\right)]^{1/2}$ $\leq C$ ja käyttäen m–riippuvuudesta seuraavaa tulosta $\mathsf{Cov}\left(X_i, X_j\right) = 0$, $|i-j| > m$, nähdään, että

$$\mathsf{Var}\left(\bar{X}_n\right) \leq \frac{C}{n} + \frac{2C(n-1)}{n^2} + \cdots + \frac{2C(n-m)}{n^2} \underset{n\rightarrow\infty}{\rightarrow} 0.$$

Siten (1.4) pätee tässäkin tapauksessa ja, kun $n^{-1}\sum_{i=1}^{n}\mu_i \rightarrow \mu$, pätee myös (1.5). Vektoritapaukset saadaan ilmeisinä yleistyksinä kuten edelläkin. Huomaa, että edellä m–riippuvuuden asemesta riittää lievempi oletus $\mathsf{Cov}\left(X_i, X_j\right) = 0$, kun $|i-j| > m$, joten SLL pätee erityisesti korreloimattomalle jonolle, kun varianssit ovat rajoitettuja.

SLL voidaan osoittaa myös tapauksissa, joissa $X_i \perp\!\!\!\perp X_j$ ei päde olipa $|i-j|$ kuinka suuri tahansa. Vaatimuksena on kuitenkin, että X_i:n ja X_j:n välinen riippuvuus (tai korrelaatio) "häviää", kun $|i-j| \rightarrow \infty$. Tälle intuitiiviselle luonnehdinnalle on useita matemaattisia muotoilija, joita käyttäen SLL:ja on todistettu sallimalla riippuvuuden lisäksi mahdollisesti myös jakaumien heterogeenisuus. Näihin kysymyksiin ei syvennytä tällä kurssilla tämän enempää. On kuitenkin syytä muistaa, ettei SLL päde, jos riippuvuus ja/tai jakaumien heterogeenisuus on "liian" voimakasta.

SLL:lle on olemassa myös useita tasaisia versioita, joissa riippuvuus ja jakaumien heterogeenisuus voidaan sallia. Näitä tasaisia SLL:ja voidaan käyttää mm. Lauseen 1.5 tyyppisissä tilanteissa ja todistettaessa SU–estimaattorin tarkentuvuus. Jos jonolle satunnaisfunktioita $X_1(\theta), X_2(\theta), ..., \theta \in \Theta$, pätee

$$\bar{X}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i(\theta) \xrightarrow{p} \mu(\theta) \quad \text{tasaisesti joukossa } \Theta,$$

sanotaan tasaisen SLL:n pätevän. Tässä (äärelliseksi oletettu) $\mu(\theta)$ tulkitaan funktiojonon $n^{-1}\sum_{i=1}^{n} \mathsf{E}[X_i(\theta)]$ raja-arvoksi.

Vaikka pisteittäinen SLL $\bar{X}_n(\theta) \xrightarrow{p} \mu(\theta)$ (jokaisella $\theta \in \Theta$) voidaan usein todeta kohtuullisen helposti, on tasaisten SSL:ien todistaminen tai edes riittävien ehtojen esittäminen matemaattisesti paljon hankalampaa. Mainitakoon, että iid–tapauksessa ja kun $X_i(\theta) = q(Y_i;\theta)$ riittää, että (i) funktio $\theta \mapsto q(y;\theta)$ on jatkuva kaikilla Y_i:n mahdollisilla arvoilla y, (ii) joukko Θ on kompakti (eli suljettu ja rajoitettu) ja (iii) $\mathsf{E}(\sup_{\theta\in\Theta}|q(Y_1;\theta)|) < \infty$.

1.2.2 Keskeinen raja-arvolause

Jos $X_1, X_2, ...$ on reaaliarvoinen iid–jono, jolla $\mathsf{E}(X_i) = \mu$ ja $\mathsf{Var}(X_i) = \sigma^2$ $(0 < \sigma^2 < \infty)$, niin todennäköisyyslaskennan kurssilla esitetyn keskeisen raja-arvolauseen (KRL) mukaan

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathsf{N}\left(0, \sigma^2\right)$$

tai vaihtoehtoisesti esitettynä (ks. Lause 1.3)

$$\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma \xrightarrow{d} \mathsf{N}(0,1).$$

Tuloksen käytännön merkitys on, että $\bar{X}_n$:n (tuntematonta) jakaumaa voidaan "suurilla" n:n arvoilla approksimoida $\mathsf{N}\left(\mu, \sigma^2/n\right)$–jakaumalla. Tähän liittyen käytetään usein merkintää $\bar{X}_n \underset{as}{\sim} \mathsf{N}\left(\mu, \sigma^2/n\right)$, jonka täsmällinen merkitys ilmenee edellä esitetystä jakaumakonvergenssista.

Vastaava tulos pätee, jos $X_1, X_2, ...$ on iid–jono satunnaisvektoreita $(k \times 1)$, joilla $\mathsf{E}(X_i) = \mu$ $(k \times 1)$ ja $\mathsf{Cov}(X_i) = \Sigma$ $(k \times k)$ on äärellinen ja nollasta poikkeava. Tällöin[10]

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathsf{N}(0, \Sigma).$$

Koska joissakin sovelluksissa päädytään singulaariseen raja-jakaumaan, ei kovarianssimatriisia Σ ole oletettu tässä (toisin kuin yleensä myöhemmin) epäsingulaariseksi.

KRL:sta on olemassa lukuisia versioita, joissa sallitaan jakaumien heterogeenisuus ja muuttujien riippuvuus (esim. m–riippuvuus). Tällaiset KRL:t esitetään (reaaliarvoisessa tapauksessa) usein muodossa $\left(\bar{X}_n - \mu_n\right)/\sigma_n \xrightarrow{d} \mathsf{N}(0,1)$, jossa $\mu_n = \mathsf{E}\left(\bar{X}_n\right)$ ja

[10] Jos tarkasteltavan multimormaalijakauman dimensio on aiheellista merkitä näkyviin, se osoitetaan alaindeksillä eli esimerkiksi tässä $\mathsf{N}_k(0,\Sigma)$.

$\sigma_n^2 = \mathsf{Var}\left(\bar{X}_n\right)$. Tällöin $\bar{X}_n$:n jakaumalle saadaan approksimaatio $\bar{X}_n \underset{as}{\sim} \mathsf{N}\left(\mu_n, \sigma_n^2\right)$. Kuten SLL:ien tapauksessa, ei KRL kuitenkaan päde, jos riippuvuus ja/tai jakaumien heterogeenisuus on "liiallista". Seuraavassa esitettävä ns. martingaalien keskeinen raja-arvolause ja sen eri versiot ovat käyttökelppoisia (erityisesti uskottavuusfunktioon perustuvassa) tilastollisessa päättelyssä.

Martingaalien keskeinen raja-arvolause. Koska tarvittava martingaalin määritelmä perustuu ehdolliseen odotusarvoon, palautetaan ensin mieleen joitakin ehdollista odotusarvoa koskevia tuloksia.

Jos Y ja X ovat satunnaisvektoreita, joilla on jatkuva yhteisjakauma ja Y:n odotusarvo on äärellisenä olemassa, on Y:n ehdollinen odotusarvo ehdolla $X = x$ määritelmän mukaan

$$\mathsf{E}\left(Y \mid X = x\right) = \int_{-\infty}^{\infty} y \frac{f_{Y,X}(y,x)}{f_X(x)} dy = \int_{-\infty}^{\infty} y f_{Y|X}(y;x)\, dy,$$

jossa $f_{Y,X}(y,x)$ on sv:n (Y,X) yhteistiheysfunktio, $f_X(x) = \int_{-\infty}^{\infty} f_{Y,X}(y,x)\, dy$ on X:n reunajakauman tiheysfunktio ja $f_{Y|X}(y|x) = f_{Y,X}(y,x)/f_X(x)$ on Y:n ehdollinen tiheysfunktio ehdolla $X = x$. Kun x:n annetaan vaihdella yli sv:n X mahdollisten arvojen, määrittelee x:n funktio $\mathsf{E}\left(Y \mid X = x\right)$ satunnaisvektorin, josta on luonteva käyttää merkintää $\mathsf{E}\left(Y \mid X\right)$. Tätä merkintää käyttäen todetaan seuraavat ehdollisen odotusarvon ominaisuudet, jotka pätevät yleisesti ilman oletusta yhteisjakauman jatkuvuudesta.

EO1 $\mathsf{E}\left(aY_1 + bY_2 \mid X\right) = a\mathsf{E}\left(Y_1 \mid X\right) + b\mathsf{E}\left(Y_2 \mid X\right)$, kun a ja b ovat vakioita.

EO2 $\mathsf{E}\left(Y \mid X\right) = \mathsf{E}(Y)$, kun $Y \perp\!\!\!\perp X$.

EO3 $\mathsf{E}(Y) = \mathsf{E}\left[\mathsf{E}\left(Y \mid X\right)\right]$ (ns. iteroidun odotusarvon laki)

EO4 $\mathsf{E}\left[g(X)\, Y \mid X\right] = g(X)\, \mathsf{E}\left(Y \mid X\right)$ mille tahansa funktiolle g (olettaen, että tulo $g(X)\, Y$ on määritelty ja sen odostusarvo on äärellisenä olemassa).

Esitetään nyt seuraava yksinkertaistettu versio martingaalin määritelmästä.

Määritelmä 2.1. Olkoon $Z_1, Z_2, ...$ jono sm:ia ja $\mathcal{F}_i = (Z_1, ..., Z_i)$. Satunnaismuuttujajonoa $M_1, M_2, ...$ sanotaan martingaaliksi informaatiojoukon $\mathcal{F}_i$ suhteen, jos

(i) $M_i = g_i(\mathcal{F}_i)$ jollain funktiolla g_i ja kaikilla $i = 1, 2, ...$

(ii) $\mathsf{E}\left(|M_i|\right) < \infty$ kaikilla $i = 1, 2, ...$

(iii) $\mathsf{E}\left(M_i \mid \mathcal{F}_{i-1}\right) = M_{i-1}$ kaikilla $i = 2, 3, ...$

Yleistys tapaukseen, jossa Z_i ja M_i ovat vektoreita on samanlainen, kun kohdassa (ii) itseisarvo korvataan normilla $\|M_i\|$ (tai odotusarvon äärellisyys vaaditaan kaikilta sv:n M_i komponenteilta).

Nimitys informaatiojoukko liittyy määritelmän ehtoon (iii), jossa ehdollinen odotusarvo voidaan tunnetusti tulkita (ei-havaittavan) muuttujan M_i keskineliövirheen mielessä optimaaliseksi ennusteeksi, kun ennuste perustetaan (havaittavaan) sv:iin $\mathcal{F}_{i-1}$. Ehdon mukaan M_i:n optimaalinen ennuste on jonon edellinen arvo M_{i-1}. Syistä, jotka jätetään pohdittaviksi, voidaan martingaali tulkita reilun pelin malliksi. Koska informaatiojoukko on usein selvä asiayhteydestä, ei sitä aina mainita.

Tyypillinen erikoistapaus määritelmän ehdosta (i) on $M_i = g_i(\mathcal{F}_i) = Z_i$ kaikilla i. Yleisempi muotoilu on tarpeen esimerkiksi, kun $Z_i = (Y_i, X_i)$ on vektori ja $M_i = Y_i$ (mahdollisesti myös vektori). Tällöin M_i sisältää vain osan Z_i:n komponenteista muiden komponenttien ollessa M_i:tä ennustavia "ulkopuolisia" muuttujia (vrt. selittävät muuttujat regressiomallissa). Määritelmän ehto (ii) tarvitaan, jotta ehdossa (iii) oleva ehdollinen odotusarvo olisi määritelty.

Tarkastellaan nyt erotusta $X_i = M_i - M_{i-1}$, $i \geq 2$, jolle pätee (ks. EO1 ja EO4)

$$\mathsf{E}(X_i | \mathcal{F}_{i-1}) = \mathsf{E}(M_i | \mathcal{F}_{i-1}) - \mathsf{E}(M_{i-1} | \mathcal{F}_{i-1}) = M_{i-1} - M_{i-1} = 0 \quad \text{kaikilla } i \geq 2.$$

Tästä seuraa (ks. EO3) $\mathsf{E}(X_i) = \mathsf{E}[\mathsf{E}(X_i | \mathcal{F}_{i-1})] = 0$ kaikilla $i \geq 2$. Määrittelemällä $X_1 = M_1$ saadaan jono $X_1, X_2, ...$, jota sanotaan *martingaalidifferenssiksi (MD) informaatiojoukon $\mathcal{F}_i$ suhteen.*

Kun $i,\ j \geq 1$, pätee reaaliarvoisessa tapauksessa ja olettaen äärelliset toiset momentit

$$\mathsf{E}(X_i X_{i+j}) \overset{\text{(EO3)}}{=} \mathsf{E}[\mathsf{E}(X_i X_{i+j} | \mathcal{F}_{i+j-1})] \overset{\text{(EO4)}}{=} \mathsf{E}[X_i \mathsf{E}(X_{i+j} | \mathcal{F}_{i+j-1})] = 0,$$

mistä yhdessä tuloksen $\mathsf{E}(X_i) = 0$ $(i \geq 2)$ kanssa seuraa $\mathsf{Cov}(X_i, X_j) = 0$, $i \neq j$, eli MD–jono on korreloimaton. Laskelmissa X_i oletettiin reaaliseksi, mutta tulos pätee ilmeisellä tavalla myös vektoritapauksessa. On syytä huomata, että korreloimattomuudesta huolimatta MD–jonon riippuvuus saattaa olla hyvinkin voimakasta.

Martingaalin asemesta lähtökohtana on usein martingaalidifferenssi eli jono $X_1, X_2, ...$, jolle pätee

(i) $X_i = h_i(\mathcal{F}_i)$ jollain funktiolla h_i ja kaikilla $i = 1, 2, ...$

(ii) $\mathsf{E}(|X_i|) < \infty$ (tai vektoritapauksessa $\mathsf{E}(\|X_i\|) < \infty$) kaikilla $i = 1, 2, ...$ ja

(iii) $\mathsf{E}(X_i | \mathcal{F}_{i-1}) = 0$ kaikilla $i = 2, 3, ...$

Tällöin $M_i = X_1 + \cdots + X_i = M_{i-1} + X_i$ $(M_0 = 0)$ on selvästikin martingaali informaatiojoukon $\mathcal{F}_i$ suhteen. Tässä yhteydessä oletetaan usein lisäksi $\mathsf{E}(X_1) = 0$, jolloin $\mathsf{E}(X_i) = 0$ pätee kaikilla $i \geq 1$.

MD–jonon korreloimattomuudesta seuraa, että reaalisessa tapauksessa $\mathsf{Var}(M_n) = \sum_{i=1}^{n} \mathsf{Var}(X_i)$ ja vastaavasti vektoritapauksessa

$$\mathsf{Cov}(M_n) = \sum_{i=1}^{n} \mathsf{Cov}(X_i) \quad \text{ja} \quad \mathsf{Cov}(M_n) = \sum_{i=1}^{n} \mathsf{E}\left(X_i X_i'\right), \text{ kun } \mathsf{E}(X_1) = 0. \quad (1.6)$$

Seuraavassa esimerkkejä MD–jonoista. Yksinkertaisuuden vuoksi jonot oletetaan reaalisiksi.

Esimerkki 1.1. (i) Reaalinen iid–jono $\{X_i\}_{i=1}^{\infty}$ on MD, kun $\mathsf{E}(X_1) = 0$ ja $\mathcal{F}_i = (X_1, ..., X_i)$.

(ii) Jos sm–jonot $\{X_i\}_{i=1}^{\infty}$ ja $\{Z_i\}_{i=1}^{\infty}$ ovat riippumattomia ja edellinen on kuten kohdassa (i), niin jono $\{Z_i X_i\}_{i=1}^{\infty}$ on MD, kun $\mathsf{E}(|Z_i|) < \infty$ kaikilla $i \geq 1$ ja $\mathcal{F}_i = (X_1, ..., X_i, Z_1, ..., Z_{i+1})$.

(iii) Jos $\{X_i\}_{i=1}^{\infty}$ on kuten kohdassa (i), $Z_1 = c$ (vakio) ja $Z_i = g_i(X_1, ..., X_{i-1})$, $i \geq 2$, niin jono $\{Z_i X_i\}_{i=1}^{\infty}$ on MD, kun $\mathsf{E}(|Z_i|) < \infty$ kaikilla $i \geq 2$ ja $\mathcal{F}_i = (X_1, ..., X_i)$.

Kohta (i) on ilmeinen ominaisuuden EO2 perusteella. Kohtien (ii) ja (iii) perustelu jätetään tehtäväksi (perustelu ei muutu olennaisesti, vaikka Z_i olisi vektori).

Koska MD–jono on korreloimaton, pätee SLL, jos varianssit ovat esimerkiksi rajoitettuja kuten tuloksessa (1.4) oletettiin. Myös KRL voidaan osoittaa usein eri oletuksin. Seuraavassa eräs versio, joka esitetään suoraan moniulotteisena. Tämän lauseen todistus sivuutetaan.[11]

Lause 1.6. Olkoon $X_1, X_2, ...$ $(k \times 1)$ MD–jono jonkin informaatiojoukon $\mathcal{F}_i$ suhteen ja $X_i = (X_{1,i}, ..., X_{k,i})$, $i = 1, 2, ...$. Jos

(i) $\mathsf{E}(|X_{a,i}|^{2+\delta}) \leq C < \infty$ jollain $\delta > 0$, kaikilla $i \geq 1$ ja kaikilla $a = 1, ..., k$ ja

(ii) $n^{-1} \sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \Sigma$ (nollasta poikkeava ja äärellinen),

niin, $\mathsf{Cov}\left(n^{-1/2} \sum_{i=1}^{n} X_i\right) \underset{n\to\infty}{\to} \Sigma$ tai yhtäpitävästi $n^{-1} \sum_{i=1}^{n} \mathsf{E}(X_i X_i') \underset{n\to\infty}{\to} \Sigma$ (ks. (1.6)) ja

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathsf{N}(0, \Sigma).$$

Lauseen oletus (i) vaatii, että sv:ien X_i komponenttien hieman toista kertalukua korkeammat momentit ovat rajoitettuja. Tämä (tai jotain vastaavaa) tarvitaan rajoittamaan havaintojen jakaumien heterogeenisuutta, mitä selittää osaltaan se, että sv:ien $X_1, X_2, ...$ ollessa samoin jakautuneita, voidaan oletuksessa (i) asettaa $\delta = 0$. Vaatimus $\delta > 0$ tuntuu käytännön kannalta siten melko harmittomalta. Oletus (ii) vaatii toisten otosmomenttien SLL:n, joka voidaan perustella käyttäen jaksossa 1.2.1 esitettyjä tuloksia tai niiden laajennuksia. Käytännön kannalta tämä ei ole sikäli rajoittavaa, että tilastollisissa sovelluksissa kovarianssimatriisille Σ tarvitaan joka tapauksessa tarkentuva estimaattori. Lauseen oletukset eivät ole lievimmät mahdolliset, mutta ne ovat joitakin lievempiä vaihtoehtoja ehkä intuitiivisemmat ja helpommat esittää.

[11] Lause on modifioitu versio lauseesta, joka löytyy yksiulotteisena J. Davidsonin teoksesta "Stochastic Limit Theory" (Oxford University Press, 1994, s. 383). Modifiointi vahvistaa hieman tämän lauseen ehtoa (b) (Lauseessa 1.6 ehto (i)) ja lisää kaksi ensin mainittua konvergenssia. Modifioinnit ja yleistys moniulotteiseen tapaukseen voidaan perustella käyttäen alaviitteessä 1 mainituissa teoksissa esitettyjä tuloksia.

2 Uskottavuuspäättelyä

2.1 Tilastollinen malli

Kuten tilastollisen päättelyn kurssilla on todettu, on tilastollisen tutkimuksen lähtökohtana aineisto $\boldsymbol{y} = (y_1, ..., y_n)$, jossa y_i on havaintoyksikköön i liittyvä havainto.[12] Havainnot voivat olla joko reaaliarvoisia tai vektoriarvoisia ja ne tulkitaan sm:ien tai sv:ien $Y_1, ..., Y_n$ havaituiksi arvoiksi. Esimerkki vektoriarvoisesta tapauksesta on $Y_1, ..., Y_n \sim \mathsf{N}_k(\mu, \Sigma) \ \underline{\|}$.

Tilastollinen malli on sv:n $\boldsymbol{Y} = (Y_1, ..., Y_n)$ yhteistiheysfunktio (ytf) tai yhteispistetodennäköisyysfunktio (yptf), joka rajataan käytettävissä olevan taustatiedon perusteella. Kurssilla tarkasteltavassa parametrisessa tilastollisessa päättelyssä oletetaan, että ytf/yptf $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ riippuu tuntemattomasta parametrista $\theta = (\theta_1, .., \theta_d) \in \Theta \subseteq \mathbb{R}^d$. Mahdollisten parametriarvojen joukon Θ määrittely on osa mallin spesifiointia. Tavoitteena on tehdä päätelmiä tuntemattomasta parametrista θ aineiston $\boldsymbol{y}$ perusteella. Joissakin yhteyksissä on hyödyllistä osoittaa havaintomäärä vektoreissa $\boldsymbol{y}$ ja $\boldsymbol{Y}$. Tällöin merkitään $\boldsymbol{y}_n$ ja $\boldsymbol{Y}_n$.

Tilastollisen päättelyn kurssilla havainnot oletetaan tyypillisesti riippumattomiksi ja samoin jakautuneiksi. Seuraavassa kaksi tästä poikkeavaa esimerkkiä, joista ensimmäinen mainitaan myös tilastollisen päättelyn kurssilla.

Autoregressiivinen aikasarjamalli. Autoregressiivisen aikasarjamallin määrittely perustuu yhtälöön

$$Y_i = \phi Y_{i-1} + \varepsilon_i, \quad i = 1, ..., n, \tag{2.1}$$

jossa $Y_0 = y_0$ on tunnettu vakio ja $\varepsilon_1, ..., \varepsilon_n \sim \mathsf{N}(0, \sigma^2) \ \underline{\|}$. Mallin mukaan havainto Y_i riippuu lineaarisesti edellisestä havainnosta Y_{i-1} ja ei-havaittavasta virhetermistä ε_i aivan kuten lineaarisessa mallissa. Tässä tapauksessa havaintojen riippuvuus on ilmeistä (ellei $\phi = 0$). Havainnot eivät myöskään ole samoin jakautuneita, mikä nähdään esimerkiksi laskemalla muuttujien $Y_1 = \phi y_0 + \varepsilon_1$ ja $Y_2 = \phi^2 y_0 + \phi\varepsilon_1 + \varepsilon_2$ varianssit. Usein parametrille ϕ asetetaan rajoite $|\phi| < 1$, kun taas varianssi σ^2 on luonteva rajoittaa positiiviseksi. On verraten helppo todeta, että tapauksessa $|\phi| \geq 1$ poikkeaa parametrien ϕ ja σ^2 asymptoottinen estimointi- ja testiteoria tavanomaisesta, sillä edellisessä jaksossa esitetyn tyyppiset SLL:t ja KRL:t eivät päde. Tähän esimerkkiin liittyvään havaintojen yhteistiheysfunktioon $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ $\left(\theta = (\phi, \sigma^2)\right)$ palataan seuraavassa jaksossa.

Mallin (2.1) ilmeinen yleistys saadaan lisäämällä yhtälön oikealle puolelle Y_i:n useita viivästettyjä arvoja. Toinen yleistys saadaan tulkitsemalla Y_i ja ε_i vektoreiksi $(k \times 1)$ ja ϕ matriisiksi $(k \times k)$. Yleistys usean Y_i:n viivästetyn arvon tapaukseen on tällöinkin mahdollinen. Erilaiset autoregressiiviset mallit ovat olleet suosittuja taloudellisten aikasarjojen analysoinnissa, mutta niitä käytetään yleisesti muillakin tilastotieteen sovellusalueilla.

[12] Tässä ja vastaavissa myöhemmissä viittauksissa tarkoitetaan P. Niemisen & P. Saikkosen monisteessa "Tilastollisen päättelyn kurssi" esitettyä materiaalia.

Lineaarisen mallin yleistyksiä. Tarkastellaan tilannetta, jossa havainnot on saatu riippumattomasti N:stä "yksilöstä" (esimerkiksi koehenkilö tai kotitalous) ja yksilölle j spesifioidaan (aluksi) tavanomainen yhden selittäjän lineaarinen regressiomalli

$$\begin{bmatrix} Y_{j,1} \\ \vdots \\ Y_{j,n_j} \end{bmatrix} = \alpha_j \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} x_{j,1} \\ \vdots \\ x_{j,n_j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{j,1} \\ \vdots \\ \varepsilon_{j,n_j} \end{bmatrix}$$

tai ilmeisin vektorimerkinnöin

$$Y_j = \alpha_j 1_{n_j} + \gamma x_j + \varepsilon_j, \quad j = 1, ..., N. \tag{2.2}$$

Kiinteäksi oletetulla selittävällä muuttujalla ajatellaan olevan sama tulkinta kaikilla yksilöillä. Mallin virhetermeistä oletetaan

$$\varepsilon_1, ..., \varepsilon_N \perp\!\!\!\perp, \quad \varepsilon_j \sim \mathsf{N}\left(0, \sigma^2 I_{n_j}\right),$$

jolloin kaikki havainnot ovat riippumattomia.[13] Yksilöiden välisten havaintojen riippumattomuutta voidaan usein pitää perusteltuna, mutta yksilökohtaisten havaintojen riippumattomuus on yleensä vähintäänkin kyseenalainen. Palataan tähän kuitenkin myöhemmin.

Mallissa (2.2) sallitaan regressiovakion tai tasoparametrin vaihtelu yksilöiden välillä, mutta regressiokerroin oletetaan samaksi kaikilla yksilöillä. Jälkimmäinen oletus on pelkästään yksinkertaisuuden vuoksi, sillä seuraavat tarkastelut voidaan yleistää suoraviivaisesti (tehtäväksi jätettävään) tapaukseen, jossa regressiokerroin γ riippuu yksilöindeksistä j. Usein tämän tyyppistä mallia sovellettaessa yksilöiden lukumäärä N on suuri, mutta yksittäisistä yksilöistä on vain vähän havaintoja (esim. $n_j < 10$). Tällöin tasoparametrien $\alpha_1, ..., \alpha_N$ estimointi on epäluotettavaa. Eräs ratkaisu tälle hankaluudelle on olettaa tasoparametri α_j ei-havaittavan satunnaismuuttujan saamaksi arvoksi. Taustalla olevassa (suuressa) populaatiossa tasoparametrin voidaan tällöin ajatella vaihtelevan yksilöstä toiseen ja yksittäisen yksilön kohdalla tasoparametrin arvo "määräytyy satunnaisesti".

Tasoparametrien vaihtelua yli koko populaation kuvataan spesifioimalla tasoparametrien yhteisjakauma. Tätä varten oletetaan

$$\alpha_1, ..., \alpha_N \sim \mathsf{N}\left(\alpha, \omega^2\right) \perp\!\!\!\perp,$$

jossa oletetaan riippumattomuus, jotta eri yksilöiden välisten havaintojen riippumattomuus toteutuisi. Mallin määrittelyä varten oletetaan lisäksi riippumattomuus

$$(\alpha_1, ..., \alpha_N) \perp\!\!\!\perp (\varepsilon_1, ..., \varepsilon_N).$$

Yhtälö (2.2) voidaan kirjoittaa

$$Y_j = Z_j \beta + \eta_j, \quad j = 1, ..., N, \tag{2.3}$$

[13] Merkintä I_k tarkoittaa $k \times k$ yksikkömatriisia.

jossa $Z_j = \left[1_{n_j} : x_j\right]$ $(n_j \times 2)$, $\beta = [\alpha \quad \gamma]'$ ja $\eta_j = (\alpha_j - \alpha)\, 1_{n_j} + \varepsilon_j$. Tehtyjen oletusten ja multinormaalijakauman tunnettujen ominaisuuksien nojalla pätee

$$\eta_j \sim \mathsf{N}\left(0, \omega^2 J_{n_j} + \sigma^2 I_{n_j}\right) \quad \underline{\|} \ , \tag{2.4}$$

jossa matriisi $J_{n_j} = 1_{n_j} 1'_{n_j}$ on tunnettu (kaikki alkiot ykkösiä). Näin määritellyn mallin parametriavaruus määräytyy ehdoista $\beta \in \mathbb{R}^2$, $\sigma^2 > 0$ ja $\omega^2 \geq 0$. Jos $\omega^2 = 0$, voivat (satunnaiset) $\alpha_1, ..., \alpha_N$ saada vain vakioarvon α, jolloin päädytään tavanomaiseen lineaariseen regressiomalliin. Koska matriisi J_{n_j} ei ole diagonaalinen, ovat yksilökohtaiset havainnot mallin mukaan riippuvia, kun $\omega^2 > 0$. Jos oletetaan $n_j \leq m$ kaikilla $j \geq 1$, ovat kaikki $n_1 + \cdots + n_N$ havaintoa m–riippuvia.

Käyttäen jakaumaoletusta (2.4) saadaan tilastolliseksi malliksi

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \prod_{j=1}^{N} (2\pi)^{-n_j/2} \det\left(V_j\left(\omega^2, \sigma^2\right)\right)^{-1/2} \exp\left\{-\tfrac{1}{2}(y_j - Z_j\beta)' V_j\left(\omega^2, \sigma^2\right)^{-1}(y_j - Z_j\beta)\right\}, \tag{2.5}$$

jossa $V_j\left(\omega^2, \sigma^2\right) = \omega^2 J_{n_j} + \sigma^2 I_{n_j}$ ja $\theta = \left(\beta, \sigma^2, \omega^2\right) \in \mathbb{R}^2 \times (0, \infty) \times [0, \infty)$. Tätä tavanomaisen lineaarisen mallin yleistystä kutsutaan kovarianssirakennemalleiksi tai virhekomponenttimalliksi (johtuen virhetermin η_j kahdesta komponentista). Huomaa, että tässä havaintovektorit $y_1, ..., y_N$ eivät ole välttämättä samaa dimensiota.

Vaikka edellä esitetyssä mallissa yksilökohtaiset havainnot ovatkin riippuvia, on (satunnaisesta tasoparametrista aiheutuva) riippuvuus varsin erityistä eikä riitä aina kuvaamaan havaintojen riippuvuutta. Tällaisissa tilanteissa mallia voidaan yleistää lisäämällä virhetermiin η_j kolmas komponentti eli olettamalla $\eta_j = (\alpha_j - \alpha)\, 1_{n_j} + \xi_j + \varepsilon_j$, jossa uusi virhekomponentti $\xi_j = (\xi_{j,1}, ..., \xi_{j,n_j})$ oletetaan riippumattomaksi η_j:n kahdesta muusta komponentista ja riippumattomuus yli yksilöiden (eli yli j:n arvojen) oletetaan myös. Usein yksilöistä saadut havainnot muodostavat aikasarjan, jolloin riippuvuuden mallintamiseksi voidaan komponentille ξ_j spesifioida esimerkiksi autoregressiivinen malli (vrt. (2.1)) $\xi_{j,i} = \phi\xi_{j,i-1} + \epsilon_{j,i}$, $i = 1, ..., n_j$, jossa $\epsilon_{j,i} \sim \mathsf{N}\left(0, \lambda^2\right)$, $\lambda \geq 0$, ja $\xi_{j,0} = 0$ (jotta $\mathsf{E}\left(\xi_{j,i}\right) = 0$). Muitakin vaihtoehtoja kuitenkin käytetään erityisesti, kun yksilökohtaiset havainnot on havaittu epätasaisesti ajassa.

Kuten aiemmin mainittiin, voitaisiin yhtälöä (2.2) yleistää sallimalla regressiokertoimen riippuvuus indeksistä j. Menemättä yksityiskohtiin todetaan vain, että näin saatava malli samoin kuin edellä tarkasteltu kolmen virhekomponentin malli sekä mainittujen mallien erilaiset yhdistelmät ovat erikoistapauksia lineaarisen mallin yleistyksestä, jossa havaintojen yhteistiheysfunktio on

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \prod_{j=1}^{N} (2\pi)^{-n_j/2} \det\left(V_j(\psi)\right)^{-1/2} \exp\left\{-\tfrac{1}{2}(y_j - Z_j\beta)' V_j(\psi)^{-1}(y_j - Z_j\beta)\right\}, \tag{2.6}$$

jossa $\mathsf{Cov}(Y_j) = V_j(\psi)$ on parametrivektorin ψ (tunnettua muotoa oleva) funktio, $\theta = (\beta, \psi)$ ja matriisi Z_j voi sisältää useita selittäviä muuttujia. Tällaisia malleja sovelletaan paljon tilastotieteen eri sovellusalueilla.

2.2 Uskottavuusfunktio ja SU–estimaattori

Uskottavuusfunktio $L(\theta; \boldsymbol{y})$ on määritelmän mukaan edellisessä jaksossa määritelty tilastollinen malli tulkittuna tuntemattoman parametrin θ funktiona ja kerrottuna θ:sta riippumattomalla, mutta aineistosta mahdollisesti riippuvalla positiivisella suureella $c(\boldsymbol{y})$. Toisin sanoen,

$$L(\theta; \boldsymbol{y}) = c(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta), \quad \theta \in \Theta \subseteq \mathbb{R}^d,$$

jossa $c(\boldsymbol{y}) > 0$ (myöhemmin usein $c(\boldsymbol{y}) = 1$). Monissa uskottavuuspäättelyn tarkasteluissa käytetään logaritmoitua uskottavuusfunktiota

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

Kun on aiheellista korostaa uskottavuusfunktion ja sen logaritmin riippuvuutta havaintojen lukumäärästä, merkitään $L_n(\theta; \boldsymbol{y}_n)$ ja $l_n(\theta; \boldsymbol{y}_n)$.

Mallin tuloesitys. Jos havainnot $Y_1, ..., Y_n$ ovat *riippumattomia*, voidaan tunnetusti kirjoittaa

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

jossa $f_{Y_i}(y_i; \theta)$ on sv:n $\boldsymbol{Y}$ komponentin Y_i tf tai ptf. Näin käy edellisessä jaksossa mainitussa yleisessä mallissa (2.6) ja sen erikoistapauksissa (ks. esim. (2.5)). Myös autoregressiivisen mallin (2.1) tapauksessa saadaan yhteistiheysfunktiolle ja siten uskottavuusfunktiolle tuloesitys käyttäen seuraavaa yleistä menettelyä.

Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ sv:n $\boldsymbol{Y} = (Y_1, ..., Y_n)$ ytf tai yptf ($Y_1, ..., Y_n$ voivat olla vektoreita). Merkitään $\boldsymbol{Y}_i = (Y_1, ..., Y_i)$ ja vastaavasti $\boldsymbol{y}_i = (y_1, ..., y_i)$, $i = 1, ..., n$. Seuraavassa näitä i:n havainnon aineistoon liittyviä muuttujia käytetään vain apuna johdettaessa mallille jatkon kannalta kätevä tuloesitys. Jättäen argumentit yksinkertaisuuden vuoksi pois merkitään sv:n $\boldsymbol{Y}_i$ ytf:ta/yptf:ta symbolilla $f_{\boldsymbol{Y}_i}$. Tarkastelun kohteena olevan n:n havainnon aineistoa vastaavan sv:n $\boldsymbol{Y}_n$ ytf/yptf on siten $f_{\boldsymbol{Y}_n}$. Koska $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i-1}, Y_i)$, saadaan ehdollisen tf:n/ptf:n määritelmää käyttäen

$$f_{\boldsymbol{Y}_n} = f_{Y_n|\boldsymbol{Y}_{n-1}} \cdot f_{\boldsymbol{Y}_{n-1}} = f_{Y_n|\boldsymbol{Y}_{n-1}} \cdot f_{Y_{n-1}|\boldsymbol{Y}_{n-2}} \cdot f_{\boldsymbol{Y}_{n-2}} = \cdots = f_{Y_1} \cdot \prod_{i=2}^{n} f_{Y_i|\boldsymbol{Y}_{i-1}}.$$

Kun lisätään poisjätetyt argumentit tavanomaiseen tapaan, voidaan havaintojen ytf/yptf kirjoittaa (merkitsemättä alaindeksiä n vektoreihin $\boldsymbol{Y}$ ja $\boldsymbol{y}$)

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = f_{Y_1}(y_1; \theta) \cdot \prod_{i=2}^{n} f_{Y_i|\boldsymbol{Y}_{i-1}}\left(y_i|\boldsymbol{y}_{i-1}; \theta\right)$$

tai, käyttäen mukavuussyistä lyhennysmerkintää $f_{i-1}(y_i; \theta) = f_{Y_i|\boldsymbol{Y}_{i-1}}\left(y_i|\boldsymbol{y}_{i-1}; \theta\right)$,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = f_{Y_1}(y_1; \theta) \cdot \prod_{i=2}^{n} f_{i-1}(y_i; \theta). \tag{2.7}$$

Merkintöjen ja ehdollisen tf:n/ptf:n määritelmän mukaan on

$$f_{i-1}(y_i;\theta) = \frac{f_{\mathbf{Y}_{i-1},Y_i}\left(\boldsymbol{y}_{i-1}, y_i;\theta\right)}{f_{\mathbf{Y}_{i-1}}\left(\boldsymbol{y}_{i-1};\theta\right)} = \frac{f_{Y_1,\dots,Y_i}(y_1,\dots,y_i;\theta)}{f_{Y_1,\dots,Y_{i-1}}(y_1,\dots,y_{i-1};\theta)} \quad i = 2,\dots,n. \tag{2.8}$$

Tämä yleistää edellä mainitun riippumattomien havaintojen tuloesityksen, joka saadaan erikoistapauksena, koska riippumattomuudesta seuraa $f_{i-1}(y_i;\theta) = f_{Y_i}(y_i;\theta)$, $i = 2,\dots,n$.

Uskottavuusfunktiolle saadaan edellä käytetyin merkinnöin tuloesitys

$$L(\theta;\boldsymbol{y}) = c(\boldsymbol{y}) \cdot f_{Y_1}(y_1;\theta) \cdot \prod_{i=2}^{n} f_{i-1}(y_i;\theta) \tag{2.9}$$

ja sen logaritmille vastaavasti summaesitys

$$l(\theta;\boldsymbol{y}) = \log c(\boldsymbol{y}) + \log f_{Y_1}(y_1;\theta) + \sum_{i=2}^{n} \log f_{i-1}(y_i;\theta). \tag{2.10}$$

Edellä käytetty ehdollistamiseen perustuvan tilastollisen mallin tuloesitys (2.7) on yleinen ja toimii aina. Jotta siinä olevilla ehdollisilla jakaumilla tai niiden johdannaisilla kuten ehdollisilla odotusarvoilla olisi järkevä tulkinta, on havaintojen oltava jossain mielessä luontevassa järjestyksessä. Aikasarja-aineistossa tällainen järjestys on luonnostaan. Mallin tuloesitys olettaa tietoa myös havaintojen välisestä riippuvuudesta, joka on kuitenkin (usein jopa keskeinen) osa mallinnusta.

Koska mallin tuloesitys (2.7) toimii aina, antaa se kätevän yleisen kehikon tarkastella uskottavuuspäättelyä riippuvien havaintojen tapauksessa. Joissakin tapauksissa uskottavuusfunktio on kuitenkin helppo johtaa havaintojen riippuvuudesta huolimatta suoraan nojautumatta edellä käytettyyn ehdollistamismenettelyyn. Edellisessä jaksossa tarkastellut kovarianssirakennemallit ovat esimerkkejä tästä (tosin niissäkin voitaisiin riippumattomien sv:ien $Y_1,\dots,Y_N$ yhteisjakaumiin soveltaa ehdollistamista).

Autoregressiivinen aikasarjamalli. Edellä esitettyä voidaan soveltaa autoregressiiviseen aikasarjamalliin (2.1). Peräkkäisillä sijoituksilla nähdään helposti, että $Y_i = \phi^i y_0 + \varepsilon_i + \phi\varepsilon_{i-1} + \cdots + \phi^{i-1}\varepsilon_1$, joten $\mathbf{Y}_{i-1} \perp\!\!\!\perp \varepsilon_i$. Tämän ja yhtälön (2.1) avulla voidaan helposti todeta (yksityiskohdat jätetään tehtäväksi) ehdollinen jakaumaidenttisyys $Y_i|\mathbf{Y}_{i-1} \overset{d}{=} Y_i|Y_{i-1}$ samoin kuin, että Y_i:n ehdollinen jakauma ehdolla $Y_{i-1} = y_{i-1}$ on $\mathsf{N}\left(\phi y_{i-1}, \sigma^2\right)$ $(i = 2,\dots,n)$.[14] Koska lisäksi $Y_1 \sim \mathsf{N}\left(\phi y_0, \sigma^2\right)$, saadaan malliksi

$$\begin{aligned} f_{\mathbf{Y}}(\boldsymbol{y};\theta) &= \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \phi y_{i-1})^2\right\} \\ &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \phi y_{i-1})^2\right\}, \quad \theta = \left(\phi, \sigma^2\right). \end{aligned}$$

[14] Merkintä $X \overset{d}{=} Y$ tarkoittaa, että sm:lla tai sv:lla X ja Y on sama jakauma.

Vastaava periaate yleistyy suoraviivaisesti myös edellisessä jaksossa mainituille mallin (2.1) yleistyksille. Jos malliyhtälön oikealla puolella on p kappaletta viivästettyjä havaintoja, riippuu ehdollinen tf $f_{Y_i|\boldsymbol{Y}_{i-1}}\left(y_i|\boldsymbol{y}_{i-1};\theta\right)$ p:stä edellisestä havainnosta $y_{i-1},...,y_{i-p}$ (tällöin $y_{-p-1},..,y_0$ tulkitaan tunnetuiksi vakioiksi). Yleisesti kutsutaan mallia, jolla on tällainen ominaisuus astetta p olevaksi Markov-malliksi, jollaisia monet riippuvien havaintojen mallinnuksessa käytettävät mallit ovat.

SU–estimaattori. *SU–estimaatti* $\hat{\theta} = \hat{\theta}(\boldsymbol{y})$ toteuttaa määritelmän mukaan ehdon

$$L(\hat{\theta};\boldsymbol{y}) \geq L(\theta;\boldsymbol{y}) \quad \text{kaikilla } \theta \in \Theta \subseteq \mathbb{R}^d.$$

Koska logaritmifunktio on kasvava, voidaan SU–estimaatti määritellä samalla tavalla myös log-uskottavuusfunktiota $l(\theta;\boldsymbol{y})$ käyttäen. Tulkittuna satunnaisena eli $\hat{\theta} = \hat{\theta}(\boldsymbol{Y})$ on $\hat{\theta}$ *SU–estimaattori.* Jatkossa käytetään yleensä merkintää $\hat{\theta}$ tai $\hat{\theta}_n$, jos riippuvuutta havaintojen lukumäärästä halutaan korostaa. Asiayhteydestä selviää onko kysymys estimaatista vai estimaattorista.

Edellä oletettiin, että SU–estimaatti on olemassa eli että uskottavuusfunktio todella saavuttaa maksimiarvonsa pisteessä $\hat{\theta}$ ja lisäksi että $\hat{\theta}(\boldsymbol{Y})$ on satunnaismuuttuja tai satunnaisvektori. Tämä ei ole itsestään selvää. Menemättä (mitta)teoreettisiin yksityiskohtiin mainitaan vain, että (eräät) riittävät ehdot ovat, että uskottavuusfunktio $\theta \mapsto L(\theta;\boldsymbol{y})$ on jatkuva kaikilla mahdollisilla aineistoilla $\boldsymbol{y}$ ja että parametriavaruus on kompakti (vrt. analyysin tulos, jonka mukaan jatkuva funktio saavuttaa maksimiarvonsa kompaktissa joukossa).

Jatkossa SU–estimaattorin olemassaolo tullaan olettamaan. Huomaa kuitenkin, ettei olemassaolo takaa SU–estimaatin yksikäsitteisyyttä. Useimmissa monimutkaisemmissa malleissa SU–estimaattia ei voida myöskään ratkaista suljetussa muodossa, vaan estimointi joudutaan suorittamaan numeerisia menetelmiä käyttäen.

2.3 Ehdollinen malli ja margnaalimalli

Useissa tilanteissa aineiston havainnot on luontevaa jakaa kahteen komponenttiin, jolloin aineisto on $\boldsymbol{w} = (w_1,...,w_n)$, jossa $w_i = (y_i, x_i)$, $i = 1,...,n$. Jako komponentteihin perustuu taustatietoon tai -teoriaan ja ideana on, että y–muuttujat ovat regressioanalyysin tapaan selitettäviä muuttujia ja x–muuttujat niiden vaihtelua selittäviä "syymuuttujia". Kumpaankin muuttujaan ajatellaan liittyvän satunnaisvaihtelua (molempien havainnot on esimerkiksi saatu satunnaisotantaa käyttäen). Merkitään jälleen vastaavia satunnaismuuttujia suurilla kirjaimilla $\boldsymbol{W}$, W_i, Y_i ja X_i.

Edellä kuvatussa tilanteessa pääasiallinen mielenkiinto kohdistuu usein y–muuttujien riippuvuuteen x–muuttujista, jolloin y–muuttujia tuntuu luontevalta tarkastella ehdollistamalla x–muuttujien suhteen. Ennen tähän liittyvää yleistä tilastollista mallia tarkastellaan yksinkertaista regressioesimerkkiä.

Multinormaalijakaumaan perustuva esimerkki. Olkoon $W_1,...,W_n$ riippumaton otos $\mathsf{N}_k(\mu,\Sigma)$–jakaumasta. Ositetaan $W_i = (Y_i, X_i)$, jossa Y_i on reaalinen ja X_i

$((k-1)\times 1)$. Ositetaan odotusarvo μ ja (positiivisesti definiitti) kovarianssimatriisi Σ vastaavalla tavalla[15]

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{ja} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Tällöin sv:n W_i yhteistiheysfunktio voidaan kirjoittaa ilmeisin lyhennysmerkinnöin $f_{W_i} = f_{Y_i|X_i} \cdot f_{X_i}$ ja multinormaalijakauman tunnettujen ominaisuuksien nojalla

$$Y_i| \left(X_i = x_i\right) \sim \mathsf{N}\left(\mu_1 - \sigma_{12}\Sigma_{22}^{-1}\left(x_i - \mu_2\right), \sigma^2\right), \quad \sigma^2 = \sigma_1^2 - \sigma_{12}\Sigma_{22}^{-1}\sigma_{21},$$

ja

$$X_i \sim \mathsf{N}_{k-1}\left(\mu_2, \Sigma_{22}\right).$$

Kun mielenkiinto kohdistuu edellä johdettuun ehdolliseen jakaumaan, on luontevaa ottaa käyttöön parametrointi

$$\alpha = \mu_1 + \sigma_{12}\Sigma_{22}^{-1}\mu_2 \quad \text{ja} \quad \gamma' = -\sigma_{12}\Sigma_{22}^{-1},$$

jolloin

$$Y_i| \left(X_i = x_i\right) \sim \mathsf{N}\left(z_i'\beta, \sigma^2\right),$$

jossa $z_i = [1 \;\; x_i']'$ ja $\beta = [\alpha \;\; \gamma']'$ ovat $k \times 1$ vektoreita. Merkitsemällä $\psi = \left(\beta, \sigma^2\right)$ ja kokoamalla vektorissa μ_2 ja kovarianssimatriisissa Σ_{22} olevat parametrit vektoriin λ saadaan tilastolliseksi malliksi

$$\begin{aligned} f_{\boldsymbol{W}}\left(\boldsymbol{w};\theta\right) &= \prod_{i=1}^{n} f_{Y_i|X_i}\left(y_i|x_i;\psi\right)\cdot f_{X_i}\left(x_i;\lambda\right) \qquad (2.11) \\ &= \left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - z_i'\beta\right)^2\right\}\cdot\prod_{i=1}^{n} f_{X_i}\left(x_i;\lambda\right), \end{aligned}$$

jossa $\theta = (\psi, \lambda)$ ja parametrit ψ ja λ vaihtelevat vapaasti ($\sigma^2 > 0$ ja λ:ssa oleva Σ_{22} positiivisesti definiitti).

Yhtälön (2.11) oikealla puolella olevan sv:n $\boldsymbol{X} = (X_1, ..., X_n)$ ytf:ta ei ole kirjoitettu eksplisiittisesti näkyviin, koska sen ei ajatella olevan kiinnostuksen kohteena. Siitä käytetään nimitystä *marginaalimalli*, kun taas mallin toista komponenttia nimitetään *ehdolliseksi malliksi*. Kun mielenkiinto kohdistuu ehdolliseen malliin ja sen parametreihin, voidaan marginaalimalli sivuuttaa parametrivektorista ψ riippumattomana tekijänä ja ψ:tä koskeva tilastollinen päättely voidaan perustaa pelkästään ehdolliseen malliin. Tällöin sv:n $\boldsymbol{X}$ havainnot voidaan tulkita kiinteiksi ei–satunnaisiksi selittäviksi muuttujiksi perinteisen lineaarisen mallin tapaan.

On syytä korostaa, että havaintojen jako komponentteihin perustuu ei–tilastollisiin argumentteihin ja siihen, ettei y–muuttujaa ole relevanttia tulkita mitään x:n komponenttia selittäväksi "syymuuttujaksi". Jos näin on, johtaa käytetty ehdollistaminen virheellisiin johtopäätöksiin. Tällaisessa tilanteessa y–muuttujaan tulisi sisällyttää ne x:n komponentit, jotka on aiheellista tulkita "selitettäviksi" muuttujiksi.

[15] Symmetrinen matriisi A $(k \times k)$ on positiivisesti definiitti, jos $x'Ax > 0$ kaikilla nollasta poikkeavilla vektoreilla x $(k \times 1)$. Jos $x'Ax \geq 0$ kaikilla x, sanotaan (symmetristä) matriisia A positiivisesti semidefiniitiksi. Positiivisesti definiitti matriisi on tunnetusti epäsingulaarinen.

Yleinen tapaus. Sovelletaan nyt edellä esitettyä edellisen jakson yleisessä kehikossa. Olkoon aineistoa $\boldsymbol{w} = (w_1, ..., w_n)$ vastaava satunnaisvektori $\boldsymbol{W} = (W_1, ..., W_n)$ ja $f_{\boldsymbol{W}}(\boldsymbol{w}; \theta)$ sen ytf/yptf, joka riippuu tuntemattomasta parametrivektorista θ. Kuten jaksossa 2.2, voidaan kirjoittaa

$$f_{\boldsymbol{W}}(\boldsymbol{w}; \theta) = f_{W_1}(w_1; \theta) \prod_{i=2}^{n} f_{W_i|\boldsymbol{W}_{i-1}}(w_i|\boldsymbol{w}_{i-1}; \theta),$$

jossa $\boldsymbol{W}_i = (W_1, ..., W_i)$ ja $\boldsymbol{w}_i$ määritellään vastaavasti. Ositetaan $W_i = (Y_i, X_i)$, jossa komponenttien Y_i ja X_i rooli on kuten edellisessä esimerkissäkin. Jätetään argumentit lyhyyden vuoksi merkitsemättä ja todetaan, että

$$f_{W_i|\boldsymbol{W}_{i-1}} = \frac{f_{\boldsymbol{W}_{i-1}, Y_i, X_i}}{f_{\boldsymbol{W}_{i-1}}} = \frac{f_{Y_i|(\boldsymbol{W}_{i-1}, X_i)} \cdot f_{\boldsymbol{W}_{i-1}, X_i}}{f_{\boldsymbol{W}_{i-1}}} = f_{Y_i|(\boldsymbol{W}_{i-1}, X_i)} \cdot f_{X_i|\boldsymbol{W}_{i-1}}.$$

Tästä ja identiteetistä $f_{W_1} = f_{Y_1|X_1} f_{X_1}$ saadaan argumentit lisäten

$$\begin{aligned} f_{\boldsymbol{W}}(\boldsymbol{w}; \theta) &= f_{Y_1|X_1}(y_1|x_1; \theta) \prod_{i=2}^{n} f_{Y_i|(\boldsymbol{W}_{i-1}, X_i)}(y_i|\boldsymbol{w}_{i-1}, x_i; \theta) \\ &\quad \times f_{X_1}(x_1; \theta) \prod_{i=2}^{n} f_{X_i|\boldsymbol{W}_{i-1}}(x_i|\boldsymbol{w}_{i-1}; \theta). \end{aligned} \tag{2.12}$$

Yhtälön (2.12) oikealla puolella kertomerkkiä edeltävä tekijä vastaa esimerkin (2.11) *ehdollista mallia* ja kertomerkin jälkeinen tekijä *marginaalimallia*. Koska parametrivektori θ esiintyy molemmissa, ei tilastollista päättelyä voida yleisesti ottaen perustaa pelkästään ehdolliseen malliin menettämättä informaatiota. Oletetaan näin ollen esimerkin (2.11) mukainen tilanne eli että parametri θ voidaan osittaa $\theta = (\psi, \lambda)$, jossa λ esiintyy vain marginaalimallissa ja ψ ehdollisessa mallissa. Kuten edellisessä esimerkissäkin, parametrin ψ määritelmän yhteydessä on voitu käyttää uudelleen parametrointia. Olennaista kuitenkin on, että ψ ei riipu parametrista λ. Parametriavaruus on siten $\Psi \times \Lambda$ ja malliksi saadaan

$$\begin{aligned} f_{\boldsymbol{W}}(\boldsymbol{w}; \theta) &= f_{Y_1|X_1}(y_1|x_1; \psi) \prod_{i=2}^{n} f_{Y_i|(\boldsymbol{W}_{i-1}, X_i)}(y_i|\boldsymbol{w}_{i-1}, x_i; \psi) \\ &\quad \times f_{X_1}(x_1; \lambda) \prod_{i=2}^{n} f_{X_i|\boldsymbol{W}_{i-1}}(x_i|\boldsymbol{w}_{i-1}; \lambda), \quad (\psi, \lambda) \in \Psi \times \Lambda. \end{aligned} \tag{2.13}$$

Kun mielenkiinto kohdistuu vain ehdolliseen malliin ja sen parametriin ψ, voidaan marginaalimalli tulkita parametrista riippumatomaksi vakioksi. Tällöin päädytään ehdolliseen uskottavuusfunktioon (ilman vakiotekijää)

$$L^{(c)}(\psi; \boldsymbol{w}) = f_{Y_1|X_1}(y_1|x_1; \psi) \prod_{i=2}^{n} f_{Y_i|(\boldsymbol{W}_{i-1}, X_i)}(y_i|\boldsymbol{w}_{i-1}, x_i; \psi), \tag{2.14}$$

johon parametria ψ koskeva päättely voidaan perustaa menettämättä informaatiota.

Kuten edellä tarkastellussa esimerkkitapauksessakin, oletetaan ehdollista uskottavuusfunktiota sovellettaessa, että muuttujien jako komponentteihin voidaan perustella ei–tilastollisin argumentein eli että x–muuttujat ovat y–muuttujien vaihtelua selittäviä "syymuuttujia", mutta ei päinvastoin. Tässä "syymuuttujan" tulkinta ei kuitenkaan ole yhtä selkeä kuin aikaisemmassa erikoistapauksessa, koska marginaalimallin ehdolliset tf:t/ptf:t $f_{X_i|\boldsymbol{W}_{i-1}}(x_i|\boldsymbol{w}_{i-1};\lambda)$ voivat riippua muuttujista $y_1, ..., y_{i-1}$. Tilanne on selkeämpi, jos tällaista riippuvuutta ei ole eli jos $X_i|\boldsymbol{W}_{i-1} \overset{d}{=} X_i|\boldsymbol{X}_{i-1}$, jossa $\boldsymbol{X}_{i-1} = (X_1, ..., X_{i-1})$ $(i \geq 2)$. Kuten multinormaalijakaumaan perustuvassa esimerkkissä, sv:n $(X_1, ..., X_n)$ havainnot voidaan tällöin tulkita ehdollisessa mallissa kiinteiksi ei–satunnaisiksi selittäviksi muuttujiksi. Erityisesti tällöin jätetään marginaalimallin tarkempi täsmennys usein tekemättä.

Seuraavassa edellä esitettyä sovelletaan lineaariseen malliin. Siinä kuten useissa muissakin erikoistapauksissa ehdolliset ytf:t $f_{Y_i|(\boldsymbol{W}_{i-1},X_i)}(y_i|\boldsymbol{w}_{i-1}, x_i;\psi)$ (tai yptf:t) yksinkertaistuvat, koska ehdollisten jakaumien oletetaan riippuvan vain kiinteästä määrästä vektorin $(\boldsymbol{w}_{i-1}, x_i)$ komponentteja (esimerkiksi vain komponenteista x_i ja y_{i-1}). Esitetyt periaatteet eivät rajoitu lineaariseen malliin, vaan ovat relevantteja myös muissa malleissa, joissa on "selitettäviä" ja "selittäviä" muuttujia.

Sovellus lineaariseen malliin. Oletetaan yhtälön (2.13) tilanne ja keskitytään ehdolliseen malliin ja sen parametriin. Oletetaan lisäksi, että $Y_i|(\boldsymbol{W}_{i-1}, X_i) \overset{d}{=} Y_i|Z_i$ ja

$$Y_i|(Z_i = z_i) \sim \mathsf{N}\left(z_i'\beta, \sigma^2\right), \tag{2.15}$$

jolloin päädytään normaaliseen lineaariseen regressiomalliin. Tässä kiinteädimensioinen vektori Z_i $(p \times 1)$ sisältää relevantteina pidetyt vektorin $(\boldsymbol{W}_{i-1}, X_i)$ komponentit ja lisäksi mahdolliset kiinteät selittäjät kuten vakiotermin ja indikaattorimuuttujat, joilla otetaan huomioon poikkeavia havaintoja. Konkreettisina esimerkkeinä voidaan tarkastella tapauksia $Z_i = (1, X_i)$ ja $Z_i = (1, Y_{i-1}, X_i)$. Edellinen johtaa "tavanomaiseen" regressiomalliin, jossa on vakiotermin lisäksi selittävänä muuttujana vektori X_i. Jälkimmäinen yleistää jaksossa 2.1 esitetyn autoregressiivisen aikasarjamallin tapaukseen, jossa on myös vakiotermi ja ulkopuolisia selittäviä muuttujia. Tapauksessa $i = 1$ oletetaan tällöin, että $Y_0 = y_0$ on tunnettu vakio. Mallin määrittely ja uskottavuusfunktion johto voidaan yleistää suoraviivaisesti tapaukseen, jossa vektorissa Z_i on useita edeltäviä selitettävän muuttujan havaintoja tai edeltäviä selittävien muuttujan havaintoja.

Malli voidaan esittää myös käyttäen yhtälöä

$$Y_i = Z_i'\beta + \varepsilon_i, \quad i = 1, ..., n, \tag{2.16}$$

jossa $\varepsilon_1, ..., \varepsilon_n \sim \mathsf{N}\left(0, \sigma^2\right) \perp\!\!\!\perp$ ja $(\boldsymbol{W}_{i-1}, X_i) \perp\!\!\!\perp \varepsilon_i$ kaikilla $i = 1, ..., n$. Tällöin pätee myös $Z_i \perp\!\!\!\perp \varepsilon_i$ ja menettelemällä kuten autoregressiivisen aikasarjamallin tapauksessa nähdään, että yhtälöstä (2.16) päädytään ehdolliseen normaalijakaumaan (2.15). Yhtälö (2.16) otetaan usein mallin lähtökohdaksi. Mallin johtaminen edellisessä jaksossa esitettyä tapaa käyttäen osoittaa kuitenkin, mitä satunnaisten selittäjien tapauksessa täytyy olettaa, jotta mallia voidaan käyttää ajatellulla tavalla. Esimerkiksi tapauksessa $Z_i = (1, X_i)$ oletetaan selittävät muuttujat X_i usein kiinteiksi perinteisen lineaarisen mallin tapaan. Kuten edellisessä jaksossa todettiin, tämä on loogista, jos $X_i|\boldsymbol{W}_{i-1} \overset{d}{=} X_i|\boldsymbol{X}_{i-1}$, mutta ei, jos X_i riippuu esimerkiksi satunnaiseksi tulkitusta Y_{i-1}:stä (ja siten myös ε_{i-1}:stä).

Oletetusta jakaumarelaatiosta (2.15) ja edellisessä jaksossa esitetystä seuraa, että ehdollinen uskottavuusfunktio (ks. (2.14)) on muodollisesti samanlainen kuin tavanomaisessa lineaarisessa mallissa eli

$$L^{(c)}\left(\beta, \sigma^2; \boldsymbol{w}\right) = \left(\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - z_i'\beta\right)^2\right\}, \quad \beta \in \mathbb{R}^p,\ \sigma^2 > 0, \quad (2.17)$$

kun parametriavaruus määritellään regressiomalleissa tavanomaisella tavalla. Parametrien β ja σ^2 SU–estimaattoreiksi saadaan siten (ks. lineaaristen mallien kurssi)

$$\hat{\beta} = \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \sum_{i=1}^{n} Z_i Y_i \quad \text{ja} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Z_i'\hat{\beta})^2. \qquad (2.18)$$

Kiinnostavan parametrin β estimointi sujuu näin ollen helposti pienimmän neliösumman (PNS) menetelmällä.

2.4 Pistemääräfunktio ja informaatio

Oletetaan, että uskottavuusfunktio $\theta \mapsto L(\theta; \boldsymbol{y})$ on kahdesti jatkuvasti derivoituva jokaisella $\boldsymbol{y}$. Log-uskottavuusfunktion $l(\theta; \boldsymbol{y}) = \log L(\theta; \boldsymbol{y})$ gradienttivektoria

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

sanotaan *pistemääräfunktioksi* tai *pistemääräksi* ja -1 kertaa toisten osittaisderivaattojen matriisia

$$\mathcal{J}(\theta; \boldsymbol{y}) := -\frac{\partial^2}{\partial \theta \partial \theta'} l(\theta; \boldsymbol{y}) = \left[-\frac{\partial^2}{\partial \theta_a \partial \theta_b} l(\theta; \boldsymbol{y})\right]_{a,b=1}^{d}$$

sanotaan havaituksi *informaatio(matriisi)ksi.*[16] *Fisherin informaatio(matriisi)* on määritelmän mukaan

$$\mathcal{I}(\theta) := \mathsf{E}\left[\mathcal{J}(\theta; \mathbf{Y})\right] = \mathsf{E}\left[-\frac{\partial^2}{\partial \theta \partial \theta'} l(\theta; \mathbf{Y})\right],$$

jossa odotusarvon äärellisyys tulkitaan määritelmään sisältyväksi oletukseksi.

Log-uskottavuusfunktion summalausekkeesta (2.10) nähdään, että pistemäärä voidaan kirjoittaa summana

$$s(\theta; \boldsymbol{y}) = \sum_{i=1}^{n} u_i(\theta; y_i), \quad u_i(\theta; y_i) = \frac{\partial}{\partial \theta} \log f_{i-1}(y_i; \theta), \qquad (2.19)$$

[16] Symboli := (vast. =:) määrittelee vasemman (vast. oikean) puolen merkinnän.

jossa $f_0(y_1;\theta) = f_{Y_1}(y_1;\theta)$. Vastaavasti havaitulle informaatiolle ja Fisherin informaatiolle saadaan summaesitykset

$$\mathcal{J}(\theta;\boldsymbol{y}) = -\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'} \log f_{i-1}(y_i;\theta) \tag{2.20}$$

ja

$$\mathcal{I}(\theta) = \sum_{i=1}^{n} \mathsf{E}\left[-\frac{\partial^2}{\partial\theta\partial\theta'} \log f_{i-1}(\mathbf{Y}_i;\theta)\right]. \tag{2.21}$$

Havaintojen ollessa riippumattomia ja samoin jakautuneita ovat viimeksi mainitun summaesityksen yhteenlaskettavat identtiset ja $\mathcal{I}(\theta) = n\mathsf{E}\left[-\partial^2 \log f_{Y_1}(\mathbf{Y}_1;\theta)/\partial\theta\partial\theta'\right]$. Jos yksittäisen havainnon Fisherin informaatio on positiivisesti definiitti, seuraa positiivisesti definiitin matriisin määritelmästä, että jokaisen parametrin informaatio kasvaa havaintojen määrän kasvaessa. Yleisessä esityksessä (2.21) samanlaisen informaation kasvamisen takaa, että jokainen yhteenlaskettava on positiivisesti definiitti. Yleensä mallit toteuttavat tällaisen "säännöllisyysehdon", joskin poikkeustapauksia esiintyy.

Tilastollisen päättelyn kurssilla on käsitelty pistemäärän sekä Fisherin informaation ja havaitun informaation ominaisuuksia ja merkitystä tilastollisen päättelyn teoriassa. Ns. säännöllisille malleille on todettu tulokset

$$\mathsf{E}[s(\theta;\mathbf{Y})] = 0 \tag{2.22}$$

ja

$$\mathsf{Cov}[s(\theta;\mathbf{Y})] = \mathsf{E}\left[s(\theta;\mathbf{Y})\, s(\theta;\mathbf{Y})'\right] = \mathcal{I}(\theta). \tag{2.23}$$

Näiden lisäksi on todettu, että Fisherin informaatiomatriisin avulla voidaan vastata kysymykseen kuinka tarkasti mallin parametria θ tai sen funktiota voidaan estimoida. Eräs muotoilu tälle ns. *informaatioepäyhtälölle* sanoo, että parametrivektorin θ mille tahansa harhattomalle estimaattorille $T = t(\mathbf{Y})$ pätee

$$a'\mathsf{Cov}(T)\, a \geq a'\mathcal{I}(\theta)^{-1} a \quad \text{kaikilla } a \in \mathbb{R}^d$$

tai toisin ilmaistuna

$$\mathsf{Cov}(T) - \mathcal{I}(\theta)^{-1} \geq 0, \tag{2.24}$$

jossa epäyhtälö tarkoittaa, että vasemmalla oleva matriisi on positiivisesti semidefiniitti. Jos $\mathsf{Cov}(T) = \mathcal{I}(\theta)^{-1}$, on estimaattori T *täystehokas*. SU–estimaattori on todettu täystehokkaaksi "suurissa otoksissa" ja "riittävien säännöllisyysehtojen vallitessa", sillä tällöin $\hat{\theta} \underset{as}{\sim} \mathsf{N}\left(\theta, \mathcal{I}(\theta)^{-1}\right)$.

Parametrien ortogonaalisuus. Tarkastellaan tilannetta, jossa mallin $f_{\mathbf{Y}}(\boldsymbol{y};\theta)$ parametrivektori voidaan osittaa kahteen komponenttiin $\theta = (\psi, \lambda)$, jossa sekä ψ

että λ voivat olla vektoreita. Kuten tilastollisen päättelyn kurssilla, ovat parametrit ψ ja λ määritelmän mukaan *ortogonaalisia*, jos Fisherin informaatiomatriisi on lohkodiagonaalinen. Toisin sanoen, jos ositetaan

$$\mathcal{I}(\theta) = \begin{bmatrix} \mathcal{I}_{\psi\psi}(\theta) & \mathcal{I}_{\psi\lambda}(\theta) \\ \mathcal{I}_{\lambda\psi}(\theta) & \mathcal{I}_{\lambda\lambda}(\theta) \end{bmatrix},$$

niin ortogonallisuus pätee, kun

$$\mathcal{I}_{\psi\lambda}(\theta) = \mathcal{I}_{\psi\lambda}(\theta)' = \mathsf{E}\left[-\frac{\partial^2}{\partial\psi\partial\lambda'} l(\theta; \boldsymbol{Y})\right] = 0.$$

Tämä on hyödyllinen ominaisuus, sillä sen voimassa ollessa parametreja ψ ja λ koskevat estimointi- ja testaustarkastelut ovat (suurissa otoksissa) likimain riippumattomia. Ortogonaalisuuden määritelmää käyttäen nähdään helposti, että mallissa (2.13) ehdollisen mallin ja marginaalimallin parametrit ψ ja λ ovat ortogonalliset (perustelu jätetään tehtäväksi).

Esimerkki lineaarisesta mallista. Tarkastellaan edellisen jakson lineaarista mallia käyttäen ehdollista uskottavuusfunktiota (2.17). Merkitsemällä $l^{(c)}(\beta, \sigma^2; \boldsymbol{w}) = \log L^{(c)}(\beta, \sigma^2; \boldsymbol{w})$ nähdään suoraviivaisella derivoinnilla (vrt. lineaaristen mallien kurssi), että

$$\frac{\partial}{\partial\beta} l^{(c)}(\beta, \sigma^2; \boldsymbol{w}) = \frac{1}{\sigma^2}\sum_{i=1}^{n} z_i (y_i - z_i'\beta)$$

ja

$$\frac{\partial}{\partial\sigma^2} l^{(c)}(\beta, \sigma^2; \boldsymbol{w}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} (y_i - z_i'\beta)^2.$$

Käyttäen yhtälön (2.16) jälkeen mainittuja riippumattomuutta $(\boldsymbol{W}_{i-1}, X_i) \perp\!\!\!\perp \varepsilon_i$, $i = 1, ..., n$ (ja sen seurausta $Z_i \perp\!\!\!\perp \varepsilon_i$) voidaan helposti todeta, että

$$\mathsf{E}_\theta\left[\frac{\partial}{\partial\beta} l^{(c)}(\beta, \sigma^2; \boldsymbol{W})\right] = 0 \quad \text{ja} \quad \mathsf{E}_\theta\left[\frac{\partial}{\partial\sigma^2} l^{(c)}(\beta, \sigma^2; \boldsymbol{W})\right] = 0.$$

Samaan tapaan voidaan myös todeta tulos

$$\mathsf{E}_\theta\left[\frac{\partial}{\partial\sigma^2} l^{(c)}(\beta, \sigma^2; \boldsymbol{W}) \frac{\partial}{\partial\beta} l^{(c)}(\beta, \sigma^2; \boldsymbol{W})\right] = 0,$$

josta seuraa, että parametrit β ja σ^2 ovat ortogonaaliset. Parametrien β ja σ^2 havaittu informaatiomatriisi ja edelleen Fisherin informaatiomatriisi (tai sen puuttuvat osat) voidaan myös johtaa helposti. Laskelmat ja lopputulos eivät poikkea muodollisesti lineaaristen mallien kurssilla esitetystä.

Edellä mainitut pistemäärän odotusarvoa koskevat tulokset voidaan perustella myös toteamalla, että $\{Z_i\varepsilon_i\}_{i=1}^{\infty}$ on MD–jono informaatiojoukon $(\boldsymbol{W}_i, X_{i+1})$ suhteen. Tämä nähdään helposti käyttäen riippumattomuutta $(\boldsymbol{W}_{i-1}, X_i) \perp\!\!\!\perp \varepsilon_i$, $i = 1, ..., n$, ja yhtälöstä (2.16) suoraan saatavaa yhtälöä $\varepsilon_i = Y_i - Z_i'\beta$ (yksityiskohdat jätetään tehtäväksi). Pistemäärä on tässä tapauksessa siten martingaali. Vastaava tulos pätee itse asiassa varsin yleisesti, kuten seuraavassa nähdään.

Pistemäärän martingaaliominaisuus. Tilastollisen päättelyn kurssilla SU–estimaattorin asymptoottisen normaalisuuden perustelu hahmotellaan iid–tapauksessa. Jaksossa 2.5.2 vastaavanlainen perustelu esitetään olettamatta riippumattomuutta ja jakaumien identtisyyttä. Käyttäen jaksoissa 1.1 ja 1.2 esitettyjä tuloksia esitetään perustelu lisäksi hieman täsmällisemmin. Näitä tarkasteluja varten todetaan edellisen esimerkin yleistyksenä kuitenkin ensin, että "säännöllisessä" mallissa pistemäärä on martingaali. Tarvittaviksi säännöllisyysehdoiksi kelpaavat tilastollisen päättelyn kurssilla käytetyt ehdot, jotka ovat vektoriparametrin tapauksessa seuraavat.

Säännölliset mallit. Jatkuva malli (eli ytf) $f_{\mathbf{Y}}(\boldsymbol{y};\theta)$, $\theta\in\Theta\subseteq\mathbb{R}^d$, on säännöllinen, jos seuraavat ehdot ovat voimassa (kaikilla havaintomäärillä $n\geq 1$):

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

(b) funktiolla $\theta\mapsto f_{\mathbf{Y}}(\boldsymbol{y};\theta)$ on jatkuvat toiset osittaisderivaatat jokaisella $\boldsymbol{y}$,

(c) aina kun $T=t(\mathbf{Y})$ on tunnusluku, jolle $\mathsf{E}_\theta(T)$ on olemassa kaikilla θ, pätee

$$\frac{\partial}{\partial\theta_a}\int t(\boldsymbol{y})\,f_{\mathbf{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y}=\int t(\boldsymbol{y})\,\frac{\partial}{\partial\theta_a}f_{\mathbf{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y},\quad a=1,...,d,$$

ja lisäksi

(d)
$$\frac{\partial^2}{\partial\theta_a\partial\theta_b}\int f_{\mathbf{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y}=\int\frac{\partial^2}{\partial\theta_a\partial\theta_b}f_{\mathbf{Y}}(\boldsymbol{y};\theta)\,d\boldsymbol{y},\quad a,b=1,...,d,$$

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

Tulosten (2.22) ja (2.23) todistamista varten voidaan ehdossa (c) rajoittua tapaukseen $t(\boldsymbol{y})=1$. Yleisemmässä muodossaan ehtoa (c) tarvitaan informaatioepäyhtälön (2.24) todistamiseen (ks. tilastollisen päättelyn kurssi). Kuten seuraavan lauseen todistuksesta nähdään, riittää tapaus $t(\boldsymbol{y})=1$ pistemäärän martingaaliominaisuuden toteamiseen. Tässä lauseessa ja sen todistuksessa on aiheellista osoittaa merkinnöissä havaintojen lukumäärä, joten merkintöjen $s(\theta;\boldsymbol{y})$ ja $\mathcal{I}(\theta)$ asemesta käytetään merkintöjä $s_n(\theta;\boldsymbol{y}_n)=\partial l_n(\theta;\boldsymbol{y}_n)/\partial\theta$ ja $\mathcal{I}_n(\theta)$. Kuten säännöllisyysehdossa (c), lisätään odotusarvo- ja kovarianssioperaattoreihin myös selvyyden vuoksi parametri θ osoittamaan laskelmissa käytettävää jakaumaa.

Lause 2.1. Jos malli on säännöllinen, niin (i) $\mathsf{E}_\theta[s_n(\theta;\mathbf{Y}_n)]=0$ kaikilla $n\geq 1$ ja (ii) $\mathsf{E}_\theta[s_n(\theta;\mathbf{Y}_n)\,|\,\mathbf{Y}_{n-1}]=s_{n-1}(\theta;\mathbf{Y}_{n-1})$ kaikilla $n\geq 2$, joten jono $s_n(\theta;\mathbf{Y}_n)$, $n=1,2,...$, on martingaali informaatiojoukon $\mathbf{Y}_n=(Y_1,...,Y_n)$ suhteen. Lisäksi pätee (iii) $\mathsf{Cov}_\theta[s_n(\theta;\mathbf{Y}_n)]=\mathcal{I}_n(\theta)$.

Lausen todistuksesta nähdään, että (θ:n ollessa "todellinen" parametriarvo) $u_i(\theta;Y_i)$, $i\geq 1$ (ks. (2.19)), on MD–jono informaatiojoukon $\mathbf{Y}_i$ suhteen. Lause mahdollistaa martingaalien KRL:den soveltamisen pistemääräfunktioon, mikä on keskeinen osa

SU–estimaattorin asymptoottisen normaalisuuden toteamisessa. Jaksossa 1.2 esitetty KRL (Lause 1.6) on yksi mahdollisuus tässä suhteessa. Mallin (varsin lievän) säännöllisyyden lisäksi vaaditaan MD–jonolta $u_i(\theta; Y_i)$, $i \geq 1$, hieman lisäehtoja, mutta iid–tapauksesta voidaan kuitenkin poiketa huomattavasti.

Lauseen 2.1 todistus: Koska väitteet (i) ja (iii) on todistettu (tapauksessa $d = 1$) tilastollisen päättelyn kurssilla, todistetaan vain väite (ii) olettaen jatkuva malli. Todetaan ensin, että Määritelmän 2.1 vaatimus (i) valinnalla $\mathcal{F}_i = \boldsymbol{Y}_i$ on ilmeinen. Yhtälöstä (2.23) nähdään, että $s_n(\theta; \boldsymbol{Y}_n)$:n komponenteilla on äärelliset toiset momentit, joten Määritelmän 2.1 vaatimus (ii) on myös täytetty. Näin ollen riittää todeta, että $s_n(\theta; \boldsymbol{Y}_n)$ toteuttaa Määritelmän 2.1 vaatimuksen (iii) eli (ks. väite (ii))

$$\mathsf{E}_\theta\left[s_n(\theta; \boldsymbol{Y}_n) \,|\, \boldsymbol{Y}_{n-1} = \boldsymbol{y}_{n-1}\right] = s_{n-1}\left(\theta; \boldsymbol{y}_{n-1}\right) \quad \text{kaikilla } \boldsymbol{y}_{n-1} \text{ ja kaikilla } n \geq 2.$$

Yhtälöstä (2.19) nähdään, että vektorin $s_n(\theta; \boldsymbol{Y}_n)$ a. komponentti on

$$s_{a,n}(\theta; \boldsymbol{Y}_n) = s_{a,n-1}(\theta; \boldsymbol{Y}_{n-1}) + \frac{\partial}{\partial\theta_a} \log f_{n-1}(Y_n; \theta) \qquad (n \geq 2),$$

joten ehdollisen odotusarvon ominaisuuden EO4 perusteella riittää osoittaa, että

$$\Delta_{a,n}\left(\boldsymbol{y}_{n-1}\right) := \mathsf{E}_\theta\left[\frac{\partial}{\partial\theta_a} \log f_{n-1}(Y_n; \theta) \,\middle|\, \boldsymbol{Y}_{n-1} = \boldsymbol{y}_{n-1}\right] = 0, \quad a = 1, ..., d.$$

Koska $f_{Y_n|\boldsymbol{Y}_{n-1}}\left(y_n|\boldsymbol{y}_{n-1}; \theta\right) = f_{n-1}(y_n; \theta)$ ja (ks. yhtälö (2.8))

$$\log f_{n-1}(y_n; \theta) = \log f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta) - \log f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right),$$

niin

$$\begin{aligned}
\Delta_{a,n}\left(\boldsymbol{y}_{n-1}\right) &= \int \frac{\partial}{\partial\theta_a} \log f_{n-1}(y_n; \theta) \cdot f_{n-1}(y_n; \theta)\, dy_n \\
&= \int \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta) \cdot f_{n-1}(y_n; \theta)\, dy_n - \int \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right) \cdot f_{n-1}(y_n; \theta)\, dy_n \\
&= \int \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta) \cdot f_{n-1}(y_n; \theta)\, dy_n - \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_n-1}\left(\boldsymbol{y}_{n-1}; \theta\right) \int f_{n-1}(y_n; \theta)\, dy_n \\
&= \int \frac{\partial f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta) / \partial\theta_a}{f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta)} \cdot f_{n-1}(y_n; \theta)\, dy_n - \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_n-1}\left(\boldsymbol{y}_{n-1}; \theta\right).
\end{aligned}$$

Väite seuraa, koska viimeisellä rivillä oleva integraali voidaan kirjoittaa (ks. (2.8))

$$\begin{aligned}
\frac{1}{f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right)} \int \frac{\partial}{\partial\theta_a} f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta)\, dy_n &= \frac{1}{f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right)} \frac{\partial}{\partial\theta_a} \int f_{\boldsymbol{Y}_{n-1}, Y_n}\left(\boldsymbol{y}_{n-1}, y_n; \theta\right) dy_n \\
&= \frac{1}{f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right)} \frac{\partial}{\partial\theta_a} f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right) \\
&= \frac{\partial}{\partial\theta_a} \log f_{\boldsymbol{Y}_{n-1}}\left(\boldsymbol{y}_{n-1}; \theta\right).
\end{aligned}$$

Tässä ensimmäinen yhtälö perustuu säännöllisyysehdon (c) ensimmäiseen osaan ($t(\boldsymbol{y}) = 1$) ja toinen kaavaan, jolla sv:n $\boldsymbol{Y}_{n-1}$ reunajakauman tf saadaan sv:n $(\boldsymbol{Y}_{n-1}, Y_n)$ yhteisjakauman tf:sta. $\square$

2.5 SU–estimaattorin asymptotiikka[17]

2.5.1 Tarkentuvuus

Kuten tilastollisen päättelyn kurssilla on todettu, täytyy SU–estimaattorin tarkentuvuutta tarkasteltaessa ottaa käyttöön oma merkintä θ_0 "todelliselle" parametriarvolle. Toisin sanoen, $\theta_0 \in \Theta$ on se parametriavaruuden piste, joka vastaa aineiston tuottanutta todennäköisyysmekanismia, kun taas θ esiintyy mallissa "vapaana" muuttujana ja kuvaa mahdollisia vaihtoehtoisia parametriarvoja. Tarve todelliselle parametriarvolle tulee pitkälti siitä, että SU–estimaattoria ei voida yleisesti lausua suljetussa muodossa havaintojen funktiona kuten esimerkiksi otoskeskiarvo eli odotusarvon SU–estimaattori normaalimallissa tai yleisemmin PNS–estimaattori (normaalissa) lineaarisessa mallissa. Tällaisissa tapauksissa tarkentuvuuden osoittaminen vaatii yleensä vain SLL:n soveltamista. Seuraava esimerkki havainnollistaa tätä.

PNS–estimaattorin tarkentuvuus. Jatketaan aikaisempaa lineaarisen mallin esimerkkiä ja tarkastellaan PNS–estimaattoria (ks. jakson (2.3) loppuosa)

$$\hat{\beta} = \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \sum_{i=1}^{n} Z_i Y_i = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} Z_i \varepsilon_i, \tag{2.25}$$

jossa β viittaa siis todelliseen parametriarvoon. Viimeisen summan jälkimmäisessä termissä esiintyy kaksi keskiarvoa, joihin voidaan "sopivin" ehdoin soveltaa SLL:ia. Tällaisia ehtoja tarkasteltiin jaksossa 1.2.1, joten tässä yhteydessä asetetaan vain hyvin yleisluonteiset ehdot. Oletetaan ensinnäkin, että

$$\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \xrightarrow{p} Q, \quad Q \text{ ei–satunnainen ja positiivisesti definiitti.} \tag{2.26}$$

Tässä stokastinen konvergenssi pätee (ks. jakso 1.2.1), kun $\{Z_i\}_{i=1}^{\infty}$ on esimerkiksi iid–jono tai yleisemmin m–riippuva ja $\mathsf{E}(Z_{a,i}^4) \leq C < \infty$ kaikilla a ja i (Cauchy-Schwarzin epäyhtälön avulla nähdään, että tällöin matriisin $Z_i Z_i'$ alkioiden toiset momentit ovat rajoitettuja). Kun tähän lisätään vaatimus raja-arvon positiivisdefiniittisyydestä, saadaan (ks. Lause 1.1 ja Seurauksen 1.2 jälkeinen keskustelu)

$$\left(\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \xrightarrow{p} Q^{-1}.$$

Koska $\{Z_i \varepsilon_i\}_{i=1}^{\infty}$ on MD–jono (ks. edellinen jakso), se on korreloimaton ja lisäksi $\mathsf{E}(Z_{a,i}\varepsilon_i) = 0$ sekä $\mathsf{E}(Z_{a,i}^2 \varepsilon_i^2) = \mathsf{E}(Z_{a,i}^2)\mathsf{E}(\varepsilon_i^2) = \sigma^2 \mathsf{E}(Z_{a,i}^2)$, jossa edellinen yhtälö seuraa riippumattomuudesta. Edellä jo käytetty momenttien rajoittuneisuus takaa siten SLL:n

$$\frac{1}{n} \sum_{i=1}^{n} Z_i \varepsilon_i \xrightarrow{p} 0.$$

[17] Tämän jakson oheislukemistona voi käyttää esimerkiksi T. Amemiyan teoksen "Advanced Econometrics" (Harvard University Press, 1985) lukuja 4.1 ja 4.2.

Yhdistämällä edellä johdetut stokastiset konvergenssit PNS–estimaattorin summaesitykseen saadaan Lauseen 1.1 avulla tulos $\hat{\beta} \xrightarrow{p} \beta$ eli $\hat{\beta}$:n tarkentuvuus. Huomaa, että esitetyt päättelyt eivät vaatineet normaalisuusoletusta missään, joten ne pätevät yleisesti, kun $\{\varepsilon_i\}_{i=1}^{\infty}$ iid–jono odotusarvona nolla ja varianssina $\sigma^2 < \infty$.

SU–estimaattorin tarkentuvuus. Seuraavassa lauseessa esitetään yleisluonteiset riittävät ehdot SU–estimaattorin $\hat{\theta}_n$ tarkentuvuudelle eli tulokselle $\hat{\theta}_n \xrightarrow{p} \theta_0$. Huomaa, että tässä samoin kuin seuraavassa esitettävissä konvergenssituloksissa ja muissa todennäköisyyspäätelmissä oletetaan aineistoa vastaavan sv:n $\boldsymbol{Y}_n$ noudattavan todellisen parametriarvon mukaista mallia $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta_0)$ Havainnollisuuden vuoksi osoitetaan SU–estimaattorin samoin kuin log-uskottavuusfunktion riippuvuus havaintojen lukumäärästä n. Kuten aiemmin mainittiin, oletetaan SU–estimaattorin olemassaolo.

Koska SU–estimaattorille ei yleisessä tapauksessa ole eksplisiittistä lauseketta, joudutaan tarkentuvuuden osoittamisessa käyttämään pelkästään uskottavuusfunktiota tai pikemminkin log–uskottavuusfunktiota. Ennen tarkentuvuuden toteavan lauseen muotoilua esitetään intuitiivinen motivaatio sen kahdelle oletukselle.

Ensimmäisessä oletuksessa vaaditaan, että havaintojen lukumäärällä skaalattu log–uskottavuusfunktio $n^{-1}l_n(\theta; \boldsymbol{Y}_n)$ konvergoi stokastisesti kohti ei-satunnaista funktiota $\bar{l}(\theta)$, joka voidaan tulkita odotusarvon $\mathsf{E}_{\theta_0}\left[n^{-1}l_n(\theta; \boldsymbol{Y}_n)\right]$ raja-arvoksi. Koska tämä stokastinen konvergenssi oletetaan tasaiseksi parametriavaruudessa Θ, on $n^{-1}l_n(\theta; \boldsymbol{Y}_n)$ "suurilla" havaintomäärillä ja todennäköisyydellä, joka on lähes yksi, "lähellä" rajafunktiota $\bar{l}(\theta)$ eikä "läheisyys" riipu siitä mikä parametriarvo on kysymyksessä. Koska lauseen toinen oletus takaa, että rajafunktiolla $\bar{l}(\theta)$ on yksikäsitteinen maksimi pisteessä θ_0, tuntuu intuitiivisesti luonnolliselta, että log-uskottavuusfunktion $n^{-1}l_n(\theta; \boldsymbol{Y}_n)$ maksimipiste $\hat{\theta}_n$ konvergoi stokastisesti kohti $\bar{l}(\theta)$:n maksimipistettä eli todellista parametriarvoa θ_0.

Lause 2.2. Olkoon $\bar{l} : \Theta \to \mathbb{R}$ ei-satunnainen funktio. Jos

(i) $\dfrac{1}{n}l_n(\theta; \boldsymbol{Y}_n) \xrightarrow{p} \bar{l}(\theta)$ tasaisesti joukossa Θ ja

(ii) jokaisella $\varepsilon > 0$, $\sup_{\|\theta-\theta_0\|\geq\varepsilon} \bar{l}(\theta) < \bar{l}(\theta_0)$,

niin SU–estimaattori on tarkentuva eli $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Lauseen todistus ei ole kovin pitkä, mutta tekninen ja esitetään jakson lopuksi. Oletusten yleisluonteisuuden vuoksi lauseen merkitys on pitkälti periaatteellinen. Konkreettisissa tilanteissa oletukset (i) ja (ii) täytyy tarkistaa käyttäen tarkasteltavan mallin erityispiirteitä ja oletuksia. Huomaa, että uskottavuusfunktion derivoituvuutta ja siten edellisessä jaksossa määriteltyä mallin säännöllisyyttä ei tässä vaadita.

Yhtälöstä (2.10) nähdään, että $n^{-1}l_n(\theta; \boldsymbol{Y}_n)$ on otoskeskiarvo, joten konkreettisissa tilanteissa oletus (i) todetaan soveltamalla jotain useista kirjallisuudessa esitetyistä tasaisista SLL:sta. Kuten jakson 1.2.1 lopussa mainittiin, eivät tasaisten SLL:ien takaavat ehdot ole yleensä aivan yksinkertaisia ja joissakin tapauksissa niiden paikkansa pitävyys voi olla hankala varmistaa. On kuitenkin positiivista, että tasaisia SLL:ja on esitetty varsin monin erilaisin oletuksin sallien havaintojen riippuvuus ja jakaumien heterogeenisuus.

Lauseen oletuksessa (ii) $\|\theta - \theta_0\|$ voidaan tulkita parametriarvon θ etäisyydeksi todellisesta parametriarvosta θ_0 ja koko oletus voidaan tulkita parametria θ koskevaksi identifiointiehdoksi. Tämän oletuksen mukaan $\bar{l}\,(\theta)$ eli funktion $\bar{l}$ arvo missä tahansa pisteessä θ, jonka etäisyys todellisesta parametriarvosta θ_0 on vähintään mielivaltaisen (pienen) positiivisen reaaliluvun ε kokoinen, on aina pienempi kuin $\bar{l}\,(\theta_0)$ eli funktion $\bar{l}$ arvo todellisessa parametriarvossa θ_0. Tämän perusteella on selvää, että oletus (ii) takaa, että funktiolla $\bar{l}$ on yksikäsitteinen maksimi pisteessä θ_0 (piirrä kuva tapauksessa $\Theta \subseteq \mathbb{R}^2$!).[18]

Riippumattomien ja samoin jakautuneiden havaintojen tapauksessa on $\bar{l}\,(\theta) = \mathsf{E}_{\theta_0}\left[\log f_{Y_1}(Y_1;\theta)\right]$. Tällöin oletus (ii) vaatii olennaisesti, että $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n;\theta)$ ja $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n;\theta_0)$ ovat eri jakaumien pistetodennäköisyys- tai tiheysfunktioita, kun $\theta \neq \theta_0$ (vrt. tilastollisen päättelyn kurssilla mainittu identifiointiehto).

Edellä esitetyistä konkreettisia tilanteita koskevista varauksista huolimatta on Lause 2.2 sikäli hyödyllinen, että se osoittaa yhden varsin paljon käytetyn periaatteen, jolla SU–estimaattorin (ja muiden saman tyyppisten estimaatorien) tarkentuvuus voidaan todeta yleisesti eikä vain iid–tapauksessa. Jotkut kirjoittajat puhuvat oletusten (i) ja (ii) yhteydessä tarkentuvuustodistuksen perusrakenteesta. Tämän perusrakenteen ja siihen liittyvien yleisten periaatteiden ymmärtäminen on tämän kurssin kannalta olennaisempaa kuin tekniset yksityiskohdat.

Lauseen 2.2 todistus: Olkoon $\varepsilon > 0$ mielivaltainen. Oletuksen (ii) mukaan ehdosta $\|\theta - \theta_0\| \geq \varepsilon$ seuraa, että jollain $\eta > 0$ pätee joukossa $\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}$

$$\bar{l}\,(\theta) - \bar{l}\,(\theta_0) \leq \sup_{\|\theta-\theta_0\|\geq\varepsilon} \bar{l}\,(\theta) - \bar{l}\,(\theta_0) < -\eta.$$

Tästä voidaan päätellä $\{\|\,\hat{\theta}_n - \theta_0\,\| \geq \varepsilon\} \subseteq \{\bar{l}(\hat{\theta}_n) - \bar{l}\,(\theta_0) < -\eta\}$ ja edelleen

$$\mathsf{P}\{\|\,\hat{\theta}_n - \theta_0\,\| \ \geq \varepsilon\} \leq \mathsf{P}\{\bar{l}\,(\theta_0) - \bar{l}(\hat{\theta}_n) > \eta\}. \qquad (*)$$

Riittää siis osoittaa, että epäyhtälön oikea puoli konvergoi nollaan. Oletuksesta (ii) seuraa, että funktio $\bar{l}$ maksimoituu pisteessä θ_0, joten $\bar{l}\,(\theta_0) - \bar{l}(\hat{\theta}_n) \geq 0$ ja riittää osoittaa, että $\bar{l}\,(\theta_0) - \bar{l}(\hat{\theta}_n) \xrightarrow{p} 0$ (ks. stokastisen konvergenssin määrittelyehto (1.1)).

[18] Oletuksen (ii) asemesta Lauseessa 2.2 olisi ollut miellyttävämpää olettaa, että funktiolla $\bar{l}$ on yksikäsitteinen maksimi pisteessä θ_0. Tämä ei kuitenkaan riitä ilman lisäehtoja eikä välttämättä takaa oletuksen (ii) paikkansa pitävyyttä.

Todetaan seuraavaksi, että

$$\begin{aligned}
0 &\leq \bar{l}\,(\theta_0) - \bar{l}(\hat{\theta}_n) \\
&= \bar{l}\,(\theta_0) - \frac{1}{n} l_n\,(\theta_0; \mathbf{Y}_n) + \frac{1}{n} l_n\,(\theta_0; \mathbf{Y}_n) - \frac{1}{n} l_n(\hat{\theta}_n; \mathbf{Y}_n) + \frac{1}{n} l_n(\hat{\theta}_n; \mathbf{Y}_n) - \bar{l}(\hat{\theta}_n) \\
&\leq \bar{l}\,(\theta_0) - \frac{1}{n} l_n\,(\theta_0; \mathbf{Y}_n) + \frac{1}{n} l_n(\hat{\theta}_n; \mathbf{Y}_n) - \bar{l}(\hat{\theta}_n) \\
&\leq \left| \bar{l}\,(\theta_0) - \frac{1}{n} l_n\,(\theta_0; \mathbf{Y}_n) \right| + \left| \frac{1}{n} l_n(\hat{\theta}_n; \mathbf{Y}_n) - \bar{l}(\hat{\theta}_n) \right| \\
&\leq 2 \sup_{\theta \in \Theta} \left| \frac{1}{n} l_n\,(\theta; \mathbf{Y}_n) - \bar{l}(\theta) \right|.
\end{aligned}$$

Tässä toinen epäyhtälö seuraa epäyhtälöstä $l_n\,(\theta_0; \mathbf{Y}_n) - l_n(\hat{\theta}_n; \mathbf{Y}_n) \leq 0$, kolmas itseisarvon ominaisuudesta ja neljäs supremumin ominaisuuksista. Oletuksen (i) ja tasaisen stokastisen konvergenssin määritelmän nojalla

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} l_n\,(\theta; \mathbf{Y}_n) - \bar{l}(\theta) \right| \xrightarrow{p} 0,$$

josta seuraa tulos $\bar{l}\,(\theta_0) - \bar{l}(\hat{\theta}_n) \xrightarrow{p} 0$ ja epäyhtälön $(*)$ perusteella edelleen tarkentuvuus. □

2.5.2 Asymptoottinen normaalisuus

Kun SU–estimaattorin tarkentuvuus on todettu, voidaan asymptoottinen normaalisuus osoittaa käyttäen pistemäärän Taylorin kehitelmää tai väliarvolausetta. Tämä vaatii uskottavuusfunktion toisten derivaattojen olemassaolon, joka seuraa mallin oletetusta säännöllisyydestä (ks. jakso 2.4). Keskeinen elementti SU–estimaattorin asymptoottisen normaalisuuden osoittamisessa on näistä ehdoista seuraava pistemäärän martingaaliominaisuus (ks. Lause 2.1), joka sopivin lisäehdoin mahdollistaa martingaalien KRL:een soveltamisen (ks. Lause 1.6). Seuraavassa esitettävä PNS–estimaattoria koskeva esimerkki havainnollistaa yleisessä tapauksessa vaadittavia tarkasteluja.

PNS–estimaattorin asymptoottinen normaalisuus. Edellisen jakson PNS–estimaattoria koskevan esimerkin yhteydessä esitetystä yhtälöstä (2.25) seuraa

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \varepsilon_i. \tag{2.27}$$

Oletetaan edelleen, että oikealla oleva matriisi toteuttaa ehdon (2.26). Koska $\{Z_i \varepsilon_i\}_{i=1}^{\infty}$ on lisäksi todettu MD–jonoksi, toteuttaa $n^{-1/2} \sum_{i=1}^{n} Z_i \varepsilon_i$ sopivin oletuksin KRL:een. Riittävät ehdot saadaan Lauseesta 1.6, mutta erityistapauksissa muitakin vaihtoehtoja voidaan käyttää. Sivuutetaan tässä yhteydessä tekniset yksityiskohdat ja esitetään vain tarkastelujen pääpiirteet. Todetaan ensiksi tulos

$$\mathsf{Cov}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \varepsilon_i \right) = \sigma^2 \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}\left(Z_i Z_i' \right),$$

jonka perustelu jätetään tehtäväksi. Tämän ja ehdon (2.26) perusteella (vrt. Lause 1.6) ei ole rajoittavaa olettaa

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i\varepsilon_i \xrightarrow{d} \mathsf{N}\left(0,\sigma^2 Q\right),$$

jossa matriisi Q on kuten ehdossa (2.26).[19] Käyttäen Lauseita 1.4 ja 1.3 (ks. myös Seuraus 1.2) voidaan edellä todetusta päätellä PNS–estimaattorin asymptoottinen normaalisuus

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathsf{N}\left(0,\sigma^2 Q^{-1}\right)$$

tai toisin ilmaistuna $\hat{\beta} \underset{as}{\sim} \mathsf{N}\left(\beta,\sigma^2 Q^{-1}\right)$. Kuten tarkentuvuuden tapauksessakin, ei virheiden normaalisuusoletus ole tässä välttämätön, kunhan oletetut raja-arvoväittämät vain pätevät. Huomaa myös, että lineaarisen mallin kurssilla esitetty PNS–estimaattorin eksakti normaalisuus ei päde (edes virheiden ollessa normaalisia), kun selittävien muuttujien joukossa on viivästettyjä selitettävän muuttujan arvoja.

Seuraavassa esitettävä yleisen SU–estimaattorin asymptoottisen normaalisuuden perustelu on periaatteessa tässä tarkastellun esimerkkitapauksen kaltainen, mutta vaatii lisätarkasteluja, koska estimaattorin lauseketta ei voida esittää ekplisiittisessä muodossa.

SU–estimaattorin asymptoottinen normaalisuus. Merkitään jälleen $s_n(\theta;\boldsymbol{y}_n) = \partial l_n(\theta;\boldsymbol{y}_n)/\partial\theta$ ja $s_{a,n}(\theta;\boldsymbol{y}_n) = \partial l_n(\theta;\boldsymbol{y}_n)/\partial\theta_a$ $(a=1,...,d)$. Kuten edellisessä jaksossa, oletetaan seuraavassa esitettävissä todennäköisyyspäätelmissä aineistoa vastaavan sv:n $\boldsymbol{Y}_n$ noudattavan mallia $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n;\theta_0)$.

Liiallisten teknisten tarkastelujen välttämiseksi oletetaan, että parametriavaruus $\Theta \subseteq \mathbb{R}^d$ on avoin ja konveksi. Usean muuttujan funktion väliarvolauseesta saadaan tällöin yhtälö [20]

$$s_{a,n}(\hat{\theta}_n;\mathbf{Y}_n) = s_{a,n}(\theta_0;\mathbf{Y}_n) + \frac{\partial}{\partial\theta'}s_{a,n}\left(\bar{\theta}_{a,n};\mathbf{Y}_n\right)(\hat{\theta}_n-\theta_0), \quad a=1,...,d,$$

jossa välipisteelle $\bar{\theta}_{a,n} = c_a\hat{\theta}_n + (1-c_a)\theta_0$, $0 \le c_a \le 1$, pätee $\|\ \bar{\theta}_{a,n}-\theta_0\ \| \le \|\ \hat{\theta}_n-\theta_0\ \|$ $(a=1,...,d)$. Koska $-\partial s_{a,n}(\theta;\mathbf{Y}_n)/\partial\theta'$ on havaitun informaatiomatriisin $\mathcal{J}_n(\theta;\mathbf{Y}_n) = -\partial^2 l_n(\theta;\mathbf{Y}_n)/\partial\theta\partial\theta'$ a. rivi, voidaan edellä esitetyt yhtälöt koota yhtälöksi

$$s_n(\hat{\theta}_n;\mathbf{Y}_n) = s_n(\theta_0;\mathbf{Y}_n) - \bar{\mathcal{J}}_n\left(\bar{\theta}_n;\mathbf{Y}_n\right)(\hat{\theta}_n-\theta_0), \tag{2.28}$$

jossa matriisin $\bar{\mathcal{J}}_n\left(\bar{\theta}_n;\mathbf{Y}_n\right)$ a. rivi on yhtä kuin matriisin $\mathcal{J}_n\left(\bar{\theta}_{a,n};\mathbf{Y}_n\right)$ a. rivi (eli matriisin $\bar{\mathcal{J}}_n\left(\bar{\theta}_n;\mathbf{Y}_n\right)$ eri riveillä on eri välipisteet). Koska SU–estimaattorin olemassaolo

[19] Jos oletetaan sv:n Z_i komponenttien neljänsien momenttien rajoitettuneisuus eli $E\left(Z_{a,i}^4\right) \le C < \infty$, voidaan tämä oletus perustella Lauseen 1.6 avulla.

[20] Jos d:n muuttujan funktio $f:\Theta\to\mathbb{R}$ on derivoituva, niin merkinnällä $\partial f(\theta)/\partial\theta'$ tarkoitetaan osittaisderivaatoista muodostettua vaakavektoria $\partial f(\theta)/\partial\theta' = [\partial f(\theta)/\partial\theta_1 \ \cdots\ \partial f(\theta)/\partial\theta_d]$ $(1\times d)$.

oletetaan, on $s_n(\hat{\theta}_n; \mathbf{Y}_n) = 0$ ja, jos matriisi $\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)$ oletetaan epäsingulaariseksi, saadaan edelleen yhtälö

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \left(\frac{1}{n}\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)\right)^{-1} \frac{1}{\sqrt{n}} s_n\left(\theta_0; \mathbf{Y}_n\right). \tag{2.29}$$

Edellä tarkastellun (normaalisen) lineaarisen mallin tapauksessa tämä yhtälö supistuu yhtälöksi (2.27) (tällöin σ^2 supistetaan pois yhtälön (2.27) oikealle puolelle tulevasta kahdesta tekijästä). Kuten tässä esimerkkitapauksessakin, on tehtävänä perustella yhtälössä (2.29) suoritettu matriisin kääntäminen ja osoittaa, että sen oikea puoli konvergoi jakaumaltaan kohti multinormaalista satunnaisvektoria.

Tarkastellaan ensin yhtälön (2.29) oikean puolen ensimmäistä tulontekijää. Yhtälöstä (2.20) nähdään, että $n^{-1}\mathcal{J}_n\left(\bar{\theta}_{a,n}; \mathbf{Y}_n\right)$ on otoskeskiarvo, joten sopivin ehdoin siihen voidaan soveltaa SLL:ia. Argumentin $\bar{\theta}_{a,n}$ satunnaisuuden vuoksi joudutaan käyttämään havaittua informaatiota koskevaa tasaista SLL:ia (ks. Lause 1.5), jollaista ei tarvittu lineaarisen mallin tapauksessa (ks. ehto (2.26)). Koska Fisherin informaatio on määritelmän mukaan havaitun informaation odotusarvo, on luontevaa olettaa

$$\frac{1}{n}\mathcal{J}_n\left(\theta; \mathbf{Y}_n\right) \xrightarrow{p} \bar{\mathcal{I}}\left(\theta\right) \quad \text{tasaisesti joukossa } \Theta, \tag{2.30}$$

jossa $\bar{\mathcal{I}}\left(\theta\right)$ määritellään raja-arvona (ks. (2.21))

$$\lim_{n\to\infty} \frac{1}{n}\mathcal{I}_n\left(\theta\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}_{\theta_0}\left[-\frac{\partial^2}{\partial\theta\partial\theta'}\log f_{i-1}\left(\mathbf{Y}_i; \theta\right)\right] = \bar{\mathcal{I}}\left(\theta\right), \quad \theta \in \Theta. \tag{2.31}$$

Tässä raja-arvon olemassaolo oletetaan samoin kuin sen jatkossa tarvittava jatkuvuus ja positiivisdefiniittisyys (ja siten epäsingulaarisuus) pisteessä θ_0.

Kuten Lauseessa 2.1 todetaan, yhtälön (2.29) oikealla puolella oleva pistemäärä $s_n\left(\theta_0; \mathbf{Y}_n\right) = \sum_{i=1}^{n} u_i\left(\theta_0; \mathbf{Y}_i\right)$ on martingaali ja, kun MD–jonon $u_i\left(\theta_0; \mathbf{Y}_i\right)$, $i \geq 1$, oletetaan toteuttavan Lauseen 1.6 ehdot (i) ja (ii), saadaan

$$\frac{1}{\sqrt{n}} s_n\left(\theta_0; \mathbf{Y}_n\right) \xrightarrow{d} \mathsf{N}\left(0, \bar{\mathcal{I}}\left(\theta_0\right)\right). \tag{2.32}$$

Tässä asymptoottisen jakauman kovarianssimatriisi perustuu mallin oletetusta säännöllisyydestä seuraavaan tulokseen $\mathsf{Cov}_{\theta_0}\left[n^{-1/2}s_n\left(\theta_0; \mathbf{Y}_0\right)\right] = n^{-1}\mathcal{I}_n\left(\theta_0\right)$ (ks. (2.23)) ja jo tehtyyn oletukseen (2.31). Toinen perustelu on, että Lauseen 1.6 oletetusten voimassa ollessa $n^{-1}\mathcal{I}_n\left(\theta_0\right)$ konvergoi ja raja-arvo on asymptoottisen jakauman kovarianssimatriisi. Esityksen yksinkertaistamiseksi ja myös koska Lauseessa 1.6 esitetyn KRL:n asemesta voi olla mahdollista soveltaa muitakin KRL:ta, otetaan pistemäärän asymptoottinen normaalisuus (2.32) seuraavassa oletukseksi.

Edellä esitettyjen valmistelujen jälkeen SU–estimaattorin asymptoottisen normaalisuuden toteaminen ei ole hankalaa.

Lause 2.3. Olkoon $f_{\mathbf{Y}_n}\left(\boldsymbol{y}_n; \theta\right)$, $\theta \in \Theta$, säännöllinen malli ja parametriavaruus Θ avoin ja konveksi. Oletetaan, että konvergenssit (2.30), (2.31) ja (2.32) pätevät ja että raja-arvo $\bar{\mathcal{I}}(\theta)$ on positiivisesti definiitti ja jatkuva pisteessä θ_0. Jos SU–estimaattori $\hat{\theta}_n$ on lisäksi tarkentuva, niin

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathsf{N}\left(0, \bar{\mathcal{I}}(\theta_0)^{-1}\right).$$

Lisäksi pätee $n^{-1}\mathcal{J}_n(\hat{\theta}_n; \mathbf{Y}_n) \xrightarrow{p} \bar{\mathcal{I}}(\theta_0)$.

Siirretään todistus jakson loppuun, mutta todetaan tässä pääidea, joka perustuu Seuraukseen 1.2 (tai Lauseisiin 1.3 ja 1.4) ja Lauseeseen 1.5. Viimeksi mainitun lauseen ja tehtyjen oletusten avulla nähdään, että yhtälön (2.29) oikealla puolella olevan tulon asymptoottista jakaumaa johdettaessa matriisi $\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)$ voidaan korvata matriisilla $\mathcal{J}_n(\theta_0; \mathbf{Y}_n)$, minkä jälkeen tulos seuraa oletuksista ja Seurauksesta 1.2.

Lause 2.3 kattaa tilastollisen päättelyn kurssilla esitetyn iid–tapauksen, jossa matriisin $\bar{\mathcal{I}}(\theta_0)$ paikalla on yksittäisen havainnon Fisherin informaatio. Konvergenssin (2.31) perusteella $\bar{\mathcal{I}}(\theta_0)$ voidaan tulkita "keskimääräiseksi" Fisherin informaatioksi ja lauseen tulos voidaan esittää myös "merkinnällisesti"

$$\hat{\theta}_n \underset{as}{\sim} \mathsf{N}\left(\theta_0, \mathcal{I}_n(\theta_0)^{-1}\right).$$

Edellä sanotun perusteella voidaan Lauseen 2.3 tulkita osoittavan SU–estimaattorin asymptoottisen täystehokkuuden.

Lauseen jälkimmäisen tuloksen (ja Lauseen 1.1) mukaan asymptoottisen jakauman kovarianssimatriisin $\bar{\mathcal{I}}(\theta_0)^{-1}$ tarkentuva estimaattori saadaan kääntämällä skaalattu havaittu informaatiomatriisi $n^{-1}\mathcal{J}_n(\hat{\theta}_n; \mathbf{Y}_n)$. Tätä tulosta voidaan käyttää SU–estimaattorin $\hat{\theta}_n$ komponenttien keskivirheiden laskemisessa (ks. tilastollisen päättelyn kurssi) ja seuraavassa jaksossa tarkasteltavissa Waldin testeissä.

Lauseen 2.2 tarkentuvuustuloksen tapaan myös Lauseen 2.3 merkitys on oletusten yleisluonteisuuden vuoksi pitkälti periaatteellinen. Konkreettisissa tilanteissa oletetut konvergenssit täytyy tarkistaa käyttäen tarkasteltavan mallin erityispiirteitä ja oletuksia. Aivan kuten Lauseen 2.2 tapauksessa, on tasaista SLL:ia koskeva konvergenssi (2.30) tässä suhteessa yleensä hankalin. Tämän kurssin kannalta teknisiä yksityiskohtia olennaisempaa on kuitenkin ymmärtää SU–estimaattorin asymptoottisen normaalisuuden todistamisessa käytettävät perusideat ja niiden toimiminen tapauksissa, joissa havainnot eivät ole riippumattomia ja samoin jakautuneita.

Mainittakoon Lauseen 2.3 oletuksista vielä, että parametriavaruuden avoimuutta ja konveksisuutta ei välttämättä tarvita, mutta (toisin kuin tarkentuvuuden tapauksessa) vaatii edellä sovelletun väliarvolauseen toimivuus, että todellinen parametriarvo θ_0 on parametriavaruuden sisäpiste. Tämä ei ole kaikissa tapauksissa aivan harmiton oletus. Esimerkiksi yhtälön (2.3) määrittelemän mallin virhetermien η_j jakaumat riippuvat varianssiparametrista ω^2, josta oletetaan $\omega^2 \geq 0$. Reunapiste $\omega^2 = 0$ on mahdollinen ja määrittelee hypoteesin, jonka testaamisesta ollaan usein kiinnostuneita. Esimerkiksi Waldin testissä tarvitaan tällöin SU–estimaattorin asymptoottinen jakauma nollahypoteesin $\omega^2 = 0$ voimassa ollessa, mutta tätä ei saada

Lauseesta 2.3 eikä muustakaan vastaavasta lauseesta, sillä asymptoottisen jakauman tiedetään olevan ei-normaalinen. On myös testaustilanteita, joissa Fisherin informaatiomatriisi ei ole nollahypoteesin voimassa ollessa positiivisesti definiitti, mikä johtaa tavanomaisesta poikkeavaan (asymptoottiseen) estimointi- ja testiteoriaan.

Ortogonaalisten parametrien tapaus. Jaksossa 2.4 tarkasteltiin parametrivektorin θositusta $\theta = (\psi, \lambda)$ ja määriteltiin sen komponenttien ortogonaalisuus Fisherin informaatiomatriisin $\mathcal{I}_n(\theta)$ lohkodiagonaalisuutena. Tästä ja ehdosta (2.31) seuraa edelleen matriisin $\bar{\mathcal{I}}(\theta)$ lohkodiagonaalisuus eli

$$\bar{\mathcal{I}}(\theta) = \begin{bmatrix} \bar{\mathcal{I}}_{\psi\psi}(\theta) & 0 \\ 0 & \bar{\mathcal{I}}_{\lambda\lambda}(\theta) \end{bmatrix},$$

mitä voidaan asymptoottisissa tarkasteluissa käyttää ortogonaalisuuden vastineena (tai määritelmänä). Tällöin SU–estimaattorin $\hat{\theta}_n = (\hat{\psi}_n, \hat{\lambda}_n)$ asymptoottisen normaalijakauman kovarianssimatriisi on lohkodiagonaalinen, joten estimaattorit $\hat{\psi}_n$ ja $\hat{\lambda}_n$ ovat normaalijakauman tunnetun ominaisuuden nojalla asymptoottisesti riippumattomia ja esimerkiksi edelliselle pätee[21]

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} \mathsf{N}\left(0, \bar{\mathcal{I}}_{\psi\psi}(\theta_0)^{-1}\right).$$

Tarkastellun lineaarisen mallin tapauksessa ortogonaalisuuden todettiin pätevän parametrien β ja σ^2 välillä myös äärellisillä havaintomäärillä.

Lauseen 2.3 todistus. Osoitetaan ensin, että matriisi $n^{-1}\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)$ konvergoi stokastisesti kohti matriisia $\bar{\mathcal{I}}(\theta_0)$. Koska matriisin $\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)$ a. rivi on yhtä kuin matriisin $\mathcal{J}_n\left(\bar{\theta}_{a,n}; \mathbf{Y}_n\right)$ a. rivi, riittää osoittaa, että

$$\frac{1}{n}\mathcal{J}_n\left(\bar{\theta}_{a,n}; \mathbf{Y}_n\right) \xrightarrow{p} \bar{\mathcal{I}}(\theta_0), \quad \text{kaikilla } a = 1, ..., d. \tag{$*$}$$

Välipisteen $\bar{\theta}_{a,n}$ todettiin edellä toteuttavan $\| \bar{\theta}_{a,n} - \theta_0 \| \leq \| \hat{\theta}_n - \theta_0 \|$, joten SU–estimaattorin $\hat{\theta}_n$ oletetusta tarkentuvuudesta seuraa $\| \bar{\theta}_{a,n} - \theta_0 \| \xrightarrow{p} 0$ tai yhtäpitävästi $\bar{\theta}_{a,n} \xrightarrow{p} \theta_0$. Tulos $(*)$ seuraa tästä, oletuksesta (2.30) ja Lauseesta 1.5. On selvää, että korvaamalla $\bar{\theta}_{a,n}$ SU–estimaattorilla $\hat{\theta}_n$ voidaan samalla tavalla perustella lauseen jälkimmäinen väite $n^{-1}\mathcal{J}_n(\hat{\theta}_n; \mathbf{Y}_n) \xrightarrow{p} \bar{\mathcal{I}}(\theta_0)$.

Edellä todetusta tuloksesta $n^{-1}\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right) \xrightarrow{p} \bar{\mathcal{I}}(\theta_0)$ ja matriisin $\bar{\mathcal{I}}(\theta_0)$ epäsingulaarisuudesta seuraa

$$\left(\frac{1}{n}\bar{\mathcal{J}}_n\left(\bar{\theta}_n; \mathbf{Y}_n\right)\right)^{-1} \xrightarrow{p} \bar{\mathcal{I}}(\theta_0)^{-1}$$

(ks. Seuraus 1.2 ja sen jälkeinen keskustelu). Käyttäen tätä ja oletusta (2.32) kehitelmässä (2.29) saadaan Seurauksen 1.2(i) perusteella

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \bar{\mathcal{I}}(\theta_0)^{-1} Z, \quad Z \sim \mathsf{N}\left(0, \bar{\mathcal{I}}(\theta_0)\right).$$

[21] Tätä tulosta ajatellen matriisin $\bar{\mathcal{I}}(\theta)$ lohkodiagonaalisuus riittäisi vaatia vain todellisessa parametriarvossa θ_0.

Lauseen ensimmäinen väite seuraa tästä, sillä $\bar{\mathcal{I}}(\theta_0)^{-1} Z \sim \mathsf{N}\left(0, \bar{\mathcal{I}}(\theta_0)^{-1}\right)$ multinormaalijakauman lineaarisuusominaisuuden nojalla.[22] $\square$

2.6 Uskottavuusfunktioon perustuvia testejä

Tarkastellaan (todelliseen parametriarvoon) liitettyä lineaarista nollahypoteesia

$$\mathsf{H}_0 : A\theta_0 = c, \tag{2.33}$$

jossa matriisi A $(q \times d)$ ja vektori c $(q \times 1)$ ovat tunnettuja ja A:n aste on q eli $\mathsf{r}(A) = q$. Vaihtoehtoinen hypoteesi on $A\theta_0 \neq c$. Tyypillinen erikoistapaus on $A = [I_q : 0]$, jolloin nollahypoteesi kiinnittää θ_0:n ensimmäiset q komponenttia vektorin c vastaaviksi komponenteiksi. Tämä vastaa tilastollisen päättelyn kurssilla tarkasteltua tilannetta.

Seuraavassa tarkastellaan ensin Waldin testiä, jonka asymptoottinen jakauma seuraa suoraviivaisesti SU–estimaattorin asymptoottisesta normaalisuudesta. Koska lineaarisen mallin kurssilla johdettu F–testisuure voidaan tulkita Waldin testiksi, voidaan edellisessä jaksossa esitettyä PNS–estimaattorin asymptoottista normaalisuutta käyttäen perustella vastaavalla tavalla F–testin käyttö tilanteissa, joissa PNS–estimaattorin eksakti normaalisuus ei päde (yksityiskohdat jätetään tehtäväksi).

2.6.1 Waldin testi

Testin johto. Waldin testi perustuu edellisessä jaksossa tarkasteltuun rajoittamattomaan SU–estimaattoriin $\hat{\theta}$, jonka oletetaan toteuttavan Lauseen 2.3 tulos. Ellei tarvetta ole, jätetään havaintojen lukumäärä yksinkertaisuuden vuoksi pois SU–estimaattorista ja muistakin merkinnöistä. Waldin testisuure perustuu erotukseen $A\hat{\theta}-c$. Koska $\hat{\theta}$ on parametrin θ_0 tarkentuva estimaattori riippumatta siitä onko nollahypoteesi tosi vai ei, saa $A\hat{\theta} - c$ tyypillisesti "pieniä" arvoja, kun nollahypoteesi on tosi ja "suuria" arvoja, kun nollahypoteesi ei ole tosi.

Oletetaan, että nollahypoteesi on voimassa, jolloin $\sqrt{n}(A\hat{\theta} - c) = A\sqrt{n}(\hat{\theta} - \theta_0)$ ja Lauseiden 1.3 ja 2.3 sekä multinornaalijakauman lineaarisuusominaisuuden perusteella

$$\sqrt{n}(A\hat{\theta} - c) \xrightarrow{d} Z, \quad Z \sim \mathsf{N}_q\left(0, A\bar{\mathcal{I}}(\theta_0)^{-1} A'\right).$$

Testisuuretta varten tarvitaan asymptoottisen jakauman kovarianssimatriisille tarkentuva estimaattori, joksi kelpaa Lauseiden 2.3 ja 1.1 nojalla $A[n^{-1}\mathcal{J}(\hat{\theta}; \mathbf{Y})]^{-1}A'$. Käyttäen Seurausta 1.2(iii), edellä esitettyä asymptoottista jakaumatulosta ja tunnettua multinormaalijakaumaan ja χ^2–jakaumaan liittyvää tulosta [23] saadaan Waldin testisuure (supistaen $\sqrt{n}$:t ja n^{-1} pois) ja sen asymptoottinen jakauma:

$$\mathsf{W} = (A\hat{\theta} - c)'[A\mathcal{J}(\hat{\theta}; \mathbf{Y})^{-1}A']^{-1}(A\hat{\theta} - c) \underset{\mathsf{H}_0}{\xrightarrow{d}} \chi^2_q. \tag{2.34}$$

[22] Multinormaalijakauman lineaarisuusominaisuudella tarkoitetaan tunnettua tulosta $Z \sim \mathsf{N}(\mu, \Sigma) \Rightarrow AZ + \nu \sim \mathsf{N}(A\mu + \nu, A\Sigma A')$.

[23] Tämän tuloksen mukaan $Z \sim \mathsf{N}_k(\mu, \Sigma) \Rightarrow (Z - \mu)' \Sigma^{-1} (Z - \mu) \sim \chi^2_k$ (olettaen, että Σ on positiivisesti definiitti).

Valitsemalla $A = [I_q : 0]$ saadaan tilastollisen päättelyn kurssilla esitetty erikoistapaus, jossa matriisi $A\mathcal{J}(\hat{\theta}; \boldsymbol{Y})^{-1}A'$ on matriisin $\mathcal{J}(\hat{\theta}; \boldsymbol{Y})^{-1}$ vasemman yläkulman dimensiota $q \times q$ oleva osite.

Testisuure W mittaa luontevasti erotuksen $A\hat{\theta} - c$ suuruutta. Käytännössä testiä sovelletaan laskemalla approksimatiivinen P–arvo

$$\text{P} = \mathsf{P}_{\mathsf{H}_0}\{\mathsf{W} \geq \mathsf{W}(\mathbf{y})\} \approx \mathsf{P}\left\{\chi_q^2 \geq \mathsf{W}(\mathbf{y})\right\},$$

jossa χ_q^2 on χ_q^2–jakaumaa noudattava satunnaismuuttuja ja $\mathsf{W}(\mathbf{y})$ on (satunnaisen) testisuureen W aineistosta laskettu arvo. Jos testisuureen tarkka jakauma tunnetaan, käytetään sitä P–arvon laskemisessa. Yksinkertaisimpia malleja lukuun ottamatta tarkkaa jakaumaa ei kuitenkaan yleensä tunneta.

Vaihtoehtoinen muotoilu. Waldin testisuureen asymptoottinen jakauma ei muutu, jos havaitun informaatiomatriisin paikalla käytetään jotain toista matriisin $\bar{\mathcal{I}}(\theta_0)$ tarkentuvaa estimaattoria. Kuten pistemäärän asymptoottista normaalisuutta (2.32) perusteltaessa todettiin, pätee $s_n(\theta_0; \boldsymbol{Y}_n) = \sum_{i=1}^n u_i(\theta_0; \boldsymbol{Y}_i)$, jossa $u_i(\theta_0; \boldsymbol{Y}_i)$, $i \geq 1$, on MD–jono ja siten korreloimaton. Tästä ja yhtälöistä (2.23) seuraa

$$\frac{1}{n}\mathsf{E}\left[s(\theta_0; \boldsymbol{Y}_n)\, s(\theta_0; \boldsymbol{Y}_n)'\right] = \frac{1}{n}\sum_{i=1}^n \mathsf{E}\left[u_i(\theta_0; Y_i)\, u_i(\theta_0; Y_i)'\right] = \frac{1}{n}\mathcal{I}_n(\theta_0) \underset{n\to\infty}{\to} \bar{\mathcal{I}}(\theta_0),$$

kun konvergenssi (2.31) oletetaan. Tämän perusteella on luontevaa perustaa matriisin $\bar{\mathcal{I}}(\theta_0)$ estimointi ns. ulkotulomatriisiin

$$\mathsf{M}(\theta; \boldsymbol{Y}) = \sum_{i=1}^n u_i(\theta; Y_i) u_i(\theta; Y_i)'.$$

Tarvittava tarkentuvuustulos $n^{-1}\mathsf{M}(\hat{\theta}; \boldsymbol{Y}) \xrightarrow{p} \bar{\mathcal{I}}(\theta_0)$ ei kuitenkaan seuraa Lauseen 2.3 oletuksista. Lauseesta 1.5 nähdään, että riittävä lisäoletus on (vrt. oletus (2.30))

$$\frac{1}{n}\mathsf{M}(\theta; \boldsymbol{Y}) \xrightarrow{p} \bar{\mathcal{I}}(\theta) \quad \text{tasaisesti joukossa } \Theta. \tag{2.35}$$

Jotkut uskottavuusfunktion maksimoinnissa käytettävät numeeriset menetelmät hyödyntävät ulkotulomatriisia log–uskottavuusfunktion Hessen matriisin $-\mathcal{J}(\theta; \boldsymbol{y})$ asemesta. Tällöin ulkotulomatriisia on luonteva käyttää myös Waldin testisuureessa. Toisin kuin Hessen matriisi on ulkotulomatriisi aina positiivisesti semidefiniitti. Monimutkaisissa estimointitehtävissä molemmat matriisit lasketaan yleensä numeerisia derivaattoja käyttäen.

Epälineaarisen hypoteesin tapaus. Waldin testi voidaan yleistää epälineaarisille hypoteeseille

$$h(\theta_0) = 0, \tag{2.36}$$

jossa funktio $h : \Theta \to \mathbb{R}^q$ on jatkuvasti derivoituva ja $q \times d$ derivaattamatriisi $H(\theta) = [\partial h_a(\theta) / \partial \theta_b]$, $a = 1, ..., q$, $b = 1, ..., d$, toteuttaa $\mathsf{r}(H(\theta_0)) = q$. Edellä tarkasteltu lineaarinen hypoteesi saadaan erikoistapauksena valitsemalla $h(\theta) = A\theta - c$. Tässä tapauksessa Waldin testisuure on luentevaa perustaa suureeseen $h(\hat{\theta})$. Käyttäen usean muuttujan funktion väliarvolausetta komponenteittain funktioihin $h_a(\theta)$ $(a = 1, ..., q)$ voidaan jaksossa 1.1 esitetty deltamenetelmä yleistää ja todeta, että nollahypoteesin voimassa ollessa

$$\sqrt{n}h(\hat{\theta}) \overset{as}{=} H(\theta_0)\sqrt{n}(\hat{\theta} - \theta_0),$$

jossa merkintä tarkoittaa, että vasen ja oikea puoli konvergoivat jakaumaltaan kohti samaa rajajakaumaa. Tästä ja edellä lineaariselle hypoteesille esitetystä päättelystä saadaan Waldin testisuureen yleistys (yksityiskohdat jätetään tehtäväksi)

$$\mathsf{W} = h(\hat{\theta})'[H(\hat{\theta})\mathcal{J}(\hat{\theta}; \boldsymbol{Y})^{-1}H(\hat{\theta})']^{-1}h(\hat{\theta}) \xrightarrow[\mathsf{H}_0]{d} \chi^2_q,$$

jossa havaitun informaatiomatriisin $\mathcal{J}(\hat{\theta}; \boldsymbol{Y})$ paikalla voidaan vaihtoehtoisesti käyttää ulkotulomatriisia $\mathsf{M}(\hat{\theta}; \boldsymbol{Y}_n)$ (olettaen ehto (2.35)). Käytännössä testaus sujuu samaan tapaan kuin edellä lineaarisen hypoteesin tapauksessa.

Ortogonaalisten parametrien tapaus. Tarkastellaan nyt Waldin testiä olettaen ositus $\theta = (\psi, \lambda)$, jossa parametrit ψ ja λ ovat ortogonaaliset ja ψ on kiinnostava parametri, jota koskevaa hypoteesia halutaan testata. Rajoitutaan yksinkertaisuuden vuoksi lineaarisen hypoteesin tapaukseen, jolloin hypoteesissa (2.33) matriisilla A on rakenne $A = [A_\psi : 0]$, jossa ositus vastaa parametrin θ ositusta.

Hypoteesina on siis $A_\psi \psi = c$. Tällöin $A\hat{\theta} - c = A_\psi \hat{\psi} - c$ ja, kuten edellä mainittiin, voidaan Waldin testisuureen yleisessä lausekkeessa (2.34) havaittu informaatiomatriisi $\mathcal{J}(\hat{\theta}; \boldsymbol{Y})$ korvata millä tahansa asymptoottisesti yhtäpitävällä vaihtoehdolla. Koska matriisi $\bar{\mathcal{I}}(\theta_0)$ on nyt lohkodiagonaalinen, voidaan $\mathcal{J}(\hat{\theta}; \boldsymbol{Y})$ (tai ulkotulomatriisi $\mathsf{M}(\theta; \boldsymbol{Y})$) rajoittaa siten lohkodiagonaaliseksi eli, ilmeisin merkinnöin, $\mathcal{J}(\hat{\theta}; \boldsymbol{Y})$ voidaan korvata matriisilla $\mathcal{J}_{\mathsf{o}}(\hat{\theta}; \boldsymbol{Y}) = \mathsf{diag}[\mathcal{J}_{\psi\psi}(\hat{\theta}; \boldsymbol{Y}) \ \ \mathcal{J}_{\lambda\lambda}(\hat{\theta}; \boldsymbol{Y})]$ (ks. oletus (2.31)). Koska tällöin

$$\mathcal{J}_{\mathsf{o}}(\hat{\theta}; \boldsymbol{Y})^{-1} = \mathsf{diag}[\mathcal{J}_{\psi\psi}^{-1}(\hat{\theta}; \boldsymbol{Y}) \ \ \mathcal{J}_{\lambda\lambda}^{-1}(\hat{\theta}; \boldsymbol{Y})],$$

saadaan tätä ja oletettua matriisin A rakennetta käyttäen Waldin testisuure (ks. (2.34))

$$\mathsf{W}_\psi = (A_\psi\hat{\psi} - c)'[A_\psi \mathcal{J}_{\psi\psi}^{-1}(\hat{\theta}; \boldsymbol{Y})A_\psi']^{-1}(A_\psi\hat{\psi} - c) \xrightarrow[\mathsf{H}_0]{d} \chi^2_q.$$

Kuten yleisessä tapauksessa voidaan myös testisuureessa W_ψ matriisi $\mathcal{J}_{\psi\psi}(\hat{\theta}; \boldsymbol{Y})$ korvata millä tahansa asymptoottisesti yhtäpitävällä vaihtoehdolla. Jos tällaisen vaihtoehdon paikalle pannaan käytännössä ei-havaittava matriisi $\mathcal{J}_{\psi\psi}(\hat{\psi}, \lambda_0; \boldsymbol{Y})$, voidaan edellä todetusta päätellä, että parametrien ψ ja λ ollessa ortogonaaliset, voidaan Waldin testisuure johtaa olettaen λ_0 ensin tunnetuksi ja korvaamalla saadussa testisuureessa (todellisuudessa tuntematon) λ_0 SU–estimaattorilla $\hat{\lambda}$. Huomaa, että ilman ortogonaalisuutta tällainen menettely ei johda oikeaan tulokseen. Koska matriisi $\mathcal{J}_{\psi\psi}(\theta; \boldsymbol{Y})$ ei yleensä ole riippumaton parametrista λ, joudutaan λ ortogonaalisuudesta huolimatta yleensä estimoimaan (lineaarisen mallin tapauksessa $\lambda = \sigma^2$).

2.6.2 Raon pistemäärätesti

Raon pistemäärätesti perustuu nollahypoteesin huomioon ottavaan rajoitettuun SU–estimaattoriin $\tilde{\theta}$, joka maksimoi uskottavuusfunktion ehdolla $A\theta = c$ ja toteuttaa siten $A\tilde{\theta} = c$. Tarkastellaan ensin hieman tätä estimaattoria sinällään.

Rajoitettu SU–estimointi. Koska matriisi A $(q \times d)$ on astetta q, voidaan nollahypoteesille (2.33) johtaa lineaarialgebraa käyttäen yhtäpitävä esitys (yksityiskohdat jätetään tehtäväksi)

$$\theta_0 = B\delta_0 + e, \tag{2.37}$$

jossa matriisi B $(d \times (d-q))$ ja vektori e $(d \times 1)$ ovat tunnettuja ja δ_0 $((d-q) \times 1)$ on tuntemattoman parametrin δ todellinen arvo. Matriisi B on lisäksi astetta $d-q$ ja toteuttaa $AB = 0$, minkä voi nähdä kertomalla yhtälö (2.37) vasemmalta matriisilla A ja vertaamalla tulosta nollahypoteesiin (2.33). Toisaalta, kertomalla yhtälö (2.37) vasemmalta matriisilla $(B'B)^{-1} B'$ ja ratkaisemalla δ_0 saadaan $\delta_0 = (B'B)^{-1} B' (\theta_0 - e)$. Tästä nähdään, että parametri δ saa arvoja joukossa

$$\Delta = \{\delta : \delta = \left(B'B\right)^{-1} B' (\theta - e), \theta \in \Theta\}.$$

Merkitsemällä $L^{(r)}(\delta; \mathbf{y}) = L(B\delta + e; \mathbf{y})$ voidaan edellä sanotusta päätellä, että rajoitettu SU–estimaatti on $\tilde{\theta} = B\tilde{\delta} + e$, jossa $\tilde{\delta}$ saadaan maksimointitehtävän

$$L^{(r)}(\tilde{\delta}; \mathbf{y}) = \max_{\delta \in \Delta} L^{(r)}(\delta; \mathbf{y})$$

ratkaisuna. Edellisissä jaksoissa esitettyjä tuloksia voidaan soveltaa myös estimaattoriin $\tilde{\delta}$ ja todeta erityisesti sen tarkentuvuus ja asymptoottinen normaalisuus. Rajoitetun SU–estimaattorin $\tilde{\theta}$ vastaavat ominaisuudet voidaan johtaa tästä käyttäen yhtälöä $\tilde{\theta} = B\tilde{\delta} + e$. Tarkentuvuus, jota seuraavassa käytetään, saadaan myös suoraan Lauseesta 2.2, kun parametriavaruus määritellään uudelleen korvaamalla Θ joukolla $\{\theta \in \Theta : A\theta = c\}$.

Testin johto. Raon testin idea on tutkia poikkeaako $s(\tilde{\theta}; \boldsymbol{y}) = \partial l(\tilde{\theta}; \boldsymbol{y})/\partial\theta$ "liikaa" nollasta. Tätä voidaan motivoida seuraavasti. Tarkentuvuuden nojalla voidaan vapaan ja rajoitetun SU–estimaattorin odottaa olevan nollahypoteesin voimassa ollessa lähellä todellista parametriarvoa θ_0 ja siten lähellä toisiaan. Koska vapaa SU–estimaattori $\hat{\theta}$ toteuttaa uskottavuusyhtälöt eli $s(\hat{\theta}; \boldsymbol{y}) = 0$, voidaan nollahypoteesin voimassa ollessa odottaa, että $s(\tilde{\theta}; \boldsymbol{y})$ on myös "lähellä" nollaa. Jos nollahypoteesi ei ole voimassa, ei ole mitään syytä miksi näin kävisi, joten pistemäärän $s(\tilde{\theta}; \boldsymbol{y})$ "suurien" arvojen voidaan tulkita viittaavan nollahypoteesin virheellisyyteen.

Tarkastellaan nyt (satunnaisen) pistemäärän $s(\tilde{\theta}; \boldsymbol{Y})$ asymptoottista jakaumaa olettaen nollahypoteesi. Koska $\tilde{\theta} - \theta_0 = B(\tilde{\delta} - \delta_0)$, voidaan käyttää väliarvolausetta samaan tapaan kuin yhtälöä (2.28) johdettaessa ja osoittaa kehitelmä

$$s(\tilde{\theta}; \boldsymbol{Y}) = s(\theta_0; \boldsymbol{Y}) - \bar{\mathcal{J}}\left(\bar{\theta}; \boldsymbol{Y}\right)(\tilde{\theta} - \theta_0) = s(\theta_0; \boldsymbol{Y}) - \bar{\mathcal{J}}\left(\bar{\theta}; \boldsymbol{Y}\right) B(\tilde{\delta} - \delta_0),$$

jossa $\bar{\mathcal{J}}\left(\bar{\theta};\boldsymbol{Y}\right)$ määritellään kuten yhtälössä (2.28) käyttäen rajoitettua SU–estimaattoria $\tilde{\theta}$ vapaan SU–estimaattorin $\hat{\theta}$ paikalla. Kertomalla edellä esitetty pistemäärän $s(\tilde{\theta};\boldsymbol{Y})$ esitys vasemmalta matriisilla $\sqrt{n}A\bar{\mathcal{J}}\left(\bar{\theta};\boldsymbol{Y}\right)^{-1}$ ja ottamalla huomioon identiteetti $AB=0$ saadaan

$$A\left(\frac{1}{n}\bar{\mathcal{J}}\left(\bar{\theta};\boldsymbol{Y}\right)\right)^{-1}\frac{1}{\sqrt{n}}s(\tilde{\theta};\boldsymbol{Y})=A\left(\frac{1}{n}\bar{\mathcal{J}}\left(\bar{\theta};\boldsymbol{Y}\right)\right)^{-1}\frac{1}{\sqrt{n}}s\left(\theta_0;\boldsymbol{Y}\right).$$

Oikealla puolella $n^{-1/2}s\left(\theta_0;\boldsymbol{Y}\right)\xrightarrow{d}\mathsf{N}(0,\bar{\mathcal{I}}\left(\theta_0\right))$ oletuksen (2.32) nojalla ja aivan kuten Lauseen 2.3 todistuksessa

$$\left[n^{-1}\bar{\mathcal{J}}(\bar{\theta};\boldsymbol{Y})\right]^{-1}\xrightarrow{p}\bar{\mathcal{I}}\left(\theta_0\right)^{-1}\quad\text{ja}\quad[n^{-1}\mathcal{J}(\tilde{\theta};\boldsymbol{Y})]^{-1}\xrightarrow{p}\bar{\mathcal{I}}\left(\theta_0\right)^{-1}.$$

Käyttäen Seurausta 1.2(i) ja multinormaalijakauman lineaarisuusominaisuutta voidaan siten päätellä, että

$$A\left(\frac{1}{n}\mathcal{J}(\tilde{\theta};\boldsymbol{Y})\right)^{-1}\frac{1}{\sqrt{n}}s(\tilde{\theta};\boldsymbol{Y})\xrightarrow{d}Z,\quad Z\sim\mathsf{N}_q\left(0,A\bar{\mathcal{I}}\left(\theta_0\right)^{-1}A'\right).$$

Kuten Waldin testisuureen tapauksessa päädytään tästä edelleen testisuureeseen

$$\mathsf{S}=s(\tilde{\theta};\boldsymbol{Y})'\mathcal{J}(\tilde{\theta};\boldsymbol{Y})^{-1}A'\left[A\mathcal{J}(\tilde{\theta};\boldsymbol{Y})^{-1}A'\right]^{-1}A\mathcal{J}(\tilde{\theta};\boldsymbol{Y})^{-1}s(\tilde{\theta};\boldsymbol{Y})\underset{\mathsf{H}_0}{\xrightarrow{d}}\chi_q^2. \tag{2.38}$$

Raon testisuureella on siten sama asymptoottinen jakauma kuin Waldin testisuurella, joten approksimatiiviset P–arvot lasketaan myös samalla tavalla.

Vaihtoehtoisia muotoiluja. Yhtälössä (2.38) esitetty Raon testisuureen lauseke on monimutkainen, mutta osoittaa miten testisuureen asymptoottiseen jakaumaan voidaan päätyä. Käyttäen matriisilaskentaa voidaan testisuure sieventää yksinkertaiseen muotoon

$$\mathsf{S}=s(\tilde{\theta};\boldsymbol{Y})'\mathcal{J}(\tilde{\theta};\boldsymbol{Y})^{-1}s(\tilde{\theta};\boldsymbol{Y}). \tag{2.39}$$

Sievennyksen vaatimat laskelmat ovat yksinkertaisia tapauksessa $A=[I_q:0]$, jolloin pistemäärän $s(\tilde{\theta};\boldsymbol{Y})$ viimeiset $d-q$ komponenttia ovat nollia ja testisuureen lauseke supistuu tilastollisen päättelyn kurssilla esitettyyn muotoon. Joissakin tapauksissa testisuureen lauseketta voidaan yksinkertaistaa myös käyttämällä havaitun informaatiomatriisin $\mathcal{J}(\tilde{\theta};\boldsymbol{Y})$ paikalla ulkotulomatriisia $\mathsf{M}(\tilde{\theta};\boldsymbol{Y}_n)$ (olettaen ehto (2.35)).

Epälineaarisen hypoteesin tapaus. Raon testi voidaan yleistää tyyppiä (2.36) oleville epälineaarisille hypoteeseille. Menemättä yksityiskohtiin todetaan vain, että samaan tapaan kuin Waldin testissä saadaan testisuure korvaamalla testisuureen (2.38) lausekkeessa matriisi A derivaattamatriisilla $H(\tilde{\theta})$. Erityisesti epälineaaristen rajoitteiden tapauksessa perustetaan rajoitettu SU–estimointi usein Lagrangen kerroinmenettelyyn eli maksimoidaan funktio

$$l\left(\theta,\lambda;\boldsymbol{y}\right)=l\left(\theta;\boldsymbol{y}\right)+\lambda'h\left(\theta\right),$$

jossa vektori $\lambda = [\lambda_1 \ \cdots \ \lambda_q]'$ sisältää Lagrangen kertoimet. Derivoimalla θ:n suhteen ja asettamalla osittaisderivaatat nollaksi saadaan yhtälö

$$\frac{\partial}{\partial\theta} l\,(\theta, \lambda; \boldsymbol{y}) = s(\theta; \boldsymbol{y}) + H\,(\theta)'\,\lambda = 0.$$

Sijoittamalla tässä $\theta = \tilde{\theta}$ ja kertomalla vasemmalta matriisilla

$$[H(\tilde{\theta})\mathcal{J}(\tilde{\theta}; \boldsymbol{Y})^{-1}H(\tilde{\theta})']^{-1}H(\tilde{\theta})\mathcal{J}(\tilde{\theta}; \boldsymbol{Y})^{-1} \quad (q \times d)$$

päädytään λ:n suhteen ratkaisuun

$$\tilde{\lambda} = -[H(\tilde{\theta})\mathcal{J}(\tilde{\theta}; \boldsymbol{y})^{-1}H(\tilde{\theta})']^{-1}H(\tilde{\theta})\mathcal{J}(\tilde{\theta}; \boldsymbol{y})^{-1}s(\tilde{\theta}; \boldsymbol{y}).$$

Korvaamalla testisuureen S lausekeessa (2.38) A matriisilla $H(\tilde{\theta})$ ja käyttämällä vektorin $\tilde{\lambda}$ lauseketta nähdään, että testisuureen S yleiselle lausekkeelle saadaan vaihtoehtoinen esitys

$$\mathsf{S} = \tilde{\lambda}'H(\tilde{\theta})\mathcal{J}(\tilde{\theta}; \boldsymbol{Y})^{-1}H(\tilde{\theta})'\tilde{\lambda} \xrightarrow[\mathsf{H_0}]{d} \chi^2_q.$$

Tämä selittää miksi Raon testiä kutsutaan myös Lagrangen kerrointestiksi.

Ortogonaalisten parametrien tapaus. Ortogonaalisten parametrien tapauksessa voidaan Raon testisuureen asemesta käyttää Waldin testisuureen tapaan modifioitua versiota. Tarkastellaan siis ositetun parametrin tilannetta eli $\theta = (\psi, \lambda)$, jossa parametrit ψ ja λ ovat ortogonaaliset ja ψ on kiinnostava parametri. Testattavassa lineaarisessa hypoteesissa (2.33) matriisilla A on tällöin rakenne $A = [A_\psi : 0]$, jolloin hypoteesi voidaan esittää myös rajoitteena $A_\psi \psi = c$.

Koska myös Raon testisuureessa havaittu informaatiomatriisi $\mathcal{J}(\tilde{\theta}; \boldsymbol{Y})$ voidaan korvata millä tahansa asymptoottisesti yhtäpitävällä vaihtoehdolla, voidaan menetellä kuten Waldin testisuureen tapauksessa ja rajoittaa $\mathcal{J}(\tilde{\theta}; \boldsymbol{Y})$ lohkodiagonaaliseksi. Osittamalla pistemäärävektori $s(\tilde{\theta}; \boldsymbol{Y}) = (s_\psi(\tilde{\theta}; \boldsymbol{Y}), s_\lambda(\tilde{\theta}; \boldsymbol{Y}))$ saadaan tällöin modifioitu Raon testisuure

$$\mathsf{S}_\psi = s_\psi(\tilde{\theta}; \boldsymbol{Y})'\mathcal{J}_{\psi\psi}(\tilde{\theta}; \boldsymbol{Y})^{-1}A'_\psi \left[A_\psi \mathcal{J}_{\psi\psi}(\tilde{\theta}; \boldsymbol{Y})^{-1}A'_\psi\right]^{-1} A_\psi \mathcal{J}_{\psi\psi}(\tilde{\theta}; \boldsymbol{Y})^{-1}s_\psi(\tilde{\theta}; \boldsymbol{Y}) \xrightarrow[\mathsf{H_0}]{d} \chi^2_q.$$

Tämän tuloksen perustelu on suoraviivainen, mutta testisuureen mutkikkaamman johdon vuoksi monivaiheisempi kuin Waldin testin tapauksessa. Kuten yleisessä tapauksessakin (ks. (2.39)), voidaan testisuureelle S_ψ johtaa matriisilaskentaa käyttäen yksinkertaisempi esitys

$$\mathsf{S}_\psi = s_\psi(\tilde{\theta}; \boldsymbol{Y})'\mathcal{J}_{\psi\psi}(\tilde{\theta}; \boldsymbol{Y})^{-1}s_\psi(\tilde{\theta}; \boldsymbol{Y}) \xrightarrow[\mathsf{H_0}]{d} \chi^2_q.$$

Koska myös testisuureessa S_ψ havaittu informaatiomatriisi $\mathcal{J}_{\psi\psi}(\tilde{\theta}; \boldsymbol{Y})$ voidaan korvata asymptoottisesti yhtäpitävällä vaihtoehdolla, nähdään matriisia $\mathcal{J}_{\psi\psi}(\tilde{\psi}, \lambda_0; \boldsymbol{Y})$ käyttäen, että testisuure S_ψ voidaan johtaa olettaen λ_0 ensin tunnetuksi ja korvaamalla saadussa testisuureessa (todellisuudessa tuntematon) λ_0 rajoitetulla SU–estimaattorilla $\tilde{\lambda}$. Jälleen on syytä korostaa, että ilman ortogonaalisuutta tällainen menettely ei johda oikeaan tulokseen.

2.6.3 Uskottavuusosamäärätesti

Uskottavuusosamäärätestissä verrataan mallin uskottavuusfunktion arvoja vapaan SU–estimaatin $\hat{\theta}$ ja rajoitetun SU–estimaatin $\tilde{\theta}$ määrittämissä pisteissä. Testisuure esitetään yleensä muodossa

$$\mathsf{LR} = 2\left[l(\hat{\theta};\mathbf{Y}) - l(\tilde{\theta};\mathbf{Y})\right]. \tag{2.40}$$

Uskottavuusfunktion tulkinta huomioon ottaen on intuitiivisesti selvää, että suuret testisuureen arvot todistavat nollahypoteesia vastaan.

Uskottavuusosamäärätestin asymptoottinen jakauma voidaan johtaa käyttäen toisen asteen Taylorin kehitelmää. Yksityiskohdat ovat algebrallisesti hieman mutkikkaat, joten seuraavassa tarkastellaan erikoistapausta, jossa nollahypoteesi määrää parametriarvon θ_0 täysin eli nollahypoteesin mukaan $\theta_0 = c$. Tällöin

$$\begin{aligned}
\mathsf{LR} &= 2\left[l(\hat{\theta};\mathbf{Y}) - l(c;\mathbf{Y})\right] \\
&= 2s(\hat{\theta};\mathbf{Y})'(\hat{\theta}-c) + (\hat{\theta}-c)'\mathcal{J}(\bar{\theta};\mathbf{Y})(\hat{\theta}-c),
\end{aligned}$$

jossa välipiste $\bar{\theta}$ toteuttaa $\| \bar{\theta} - c \| \leq \| \hat{\theta} - c \|$. Koska $s(\hat{\theta};\mathbf{Y}) = 0$, voidaan vapaan SU–estimaattorin $\hat{\theta}$ tarkentuvuutta käyttäen päätellä kuten Waldin testin tapauksessa (tai Lauseen 2.3 todistuksessa), että

$$\mathsf{LR} \overset{as}{=} (\hat{\theta}-c)'\mathcal{J}(\hat{\theta};\mathbf{Y})(\hat{\theta}-c) \underset{\mathsf{H}_0}{\overset{d}{\to}} \chi^2_d.$$

Yleisen nollahypoteesin (2.33) tai (2.36) tapauksessa edellä esitetty asymptoottinen approksimaatio toimii edelleen, mutta c:n paikalle tulee rajoitettu SU–estimaattori $\tilde{\theta}$. Tällöin jatko sujuu kehittämällä pistemääräfunktiota väliarvolauseen avulla, jolloin saadaan approksimaatio $\hat{\theta} - \tilde{\theta} \overset{as}{=} \mathcal{J}(\tilde{\theta};\mathbf{Y})^{-1}s(\tilde{\theta};\mathbf{Y})$ (ks. (2.28), kun θ_0:n paikalle pannaan $\tilde{\theta}$). Tästä saadaan edelleen yhteys Raon testiin. Kuten edellä viitattiin, mutkistuu asymptoottisen jakauman johtamisessa tarvittava matriisialgebra tämän jälkeen, joten yksityiskohdat sivuutetaan. Lopputulokseksi saadaan kuitenkin, että nollahypotesin voimassa ollessa

$$\mathsf{LR} \underset{\mathsf{H}_0}{\overset{d}{\to}} \chi^2_q.$$

Sama asymptoottinen jakauma pätee siis kaikille kolmelle testisuureelle ja käytännön testaus sujuu uskottavuusosamäärätestillä samalla tavalla kuin Waldin testillä ja Raon testillä.

2.6.4 Sovelluksia lineaariseen malliin

Keskeinen ero Waldin testin ja Raon testin välillä on, että edellinen perustuu rajoittamattomaan SU–estimaattoriin ja jälkimmäinen rajoitettuun SU–estimaattoriin. Waldin testi on siten usein kätevä tai luonteva, kun rajoittamaton malli on yksinkertainen ja sen parametrit on helppo estimoida. Raon testillä on puolestaan vastaava ominaisuus, kun rajoitettu malli on yksinkertainen ja sen parametrien estimointi on helppoa. Toisin kuin Waldin testi ja Raon testi, vaatii uskottavuusosamäärätesti sekä rajoitetun että rajoittamattoman SU–estimoinnin, joten sillä ei ole mainittuja kätevyys- tai luontevuusominaisuukia. Tätä ei kuitenkaan kannata painottaa liikaa, sillä esimerkiksi viime vuosina tapahtunut laskentakapasiteetin kasvu on vähentänyt huomattavasti estimoinnin laskennallisen helppouden merkitystä (mallien monimutkaistumisen myötä sillä voi joissakin tapauksissa olla edelleen merkitystä). Koska SU–estimaattori on lisäksi invariantti vaihtoehtoisille parametroinneille (ks. tilastollisen päättelyn kurssi), on uskottavuusosamäärätestillä vastaava invarianssiominaisuus, jota Waldin testillä ja Raon testillä ei yleisesti ole. Kun testissä joudutaan turvautumaan asymptoottiseen jakaumaan, voi testin tulos näillä testeillä siten riippua käytetystä parametroinnista (tai siitä miten testattava hypoteesi ilmaistaan).

Seuraavassa esitetään kaksi tyypillistä lineaariseen malliin liittyvää testaustilannetta, joista ensimmäisessä Waldin testi on luonteva ja jälkimmäisessä Raon testi on vastaavasti luonteva.

Epälineaaristen rajoitteiden testaus lineaarisessa mallissa. Tarkastellaan jakson 2.3 lopussa esitettyä lineaarista mallia ja sovelletaan ehdollista uskottavuusfunktiota. Oletetaan siis havainnot $Y_1, ..., Y_n$ ja jakson 2.3 merkinnöin, että $Y_i|(\boldsymbol{W}_{i-1}, X_i) \overset{d}{=} Y_i|Z_i$ ja

$$Y_i|(Z_i = z_i) \sim \mathsf{N}\left(z_i'\beta, \sigma^2\right),$$

jossa selittävien muuttujien vektorin z_i $(p \times 1)$ komponentit sisältyvät vektoriin $(\boldsymbol{w}_{i-1}, x_i)$ (vaihtoehtoisesti voidaan käyttää yhtälöön (2.16) perustuvaa mallin esitystapaa). Liitetään parametrivektoriin β $(p \times 1)$ epälineaarinen hypoteesi

$$\mathsf{H}_0 : h(\beta) = 0,$$

jossa funktio $h : \mathbb{R}^p \to \mathbb{R}^q$ on jatkuvasti derivoituva ja $q \times p$ derivaattamatriisi $H(\theta) = [\partial h_a(\theta) / \partial \theta_b]$, $a = 1, ..., q$, $b = 1, ..., p$, toteuttaa $\mathsf{r}(H(\beta)) = q$ (vrt. jakso 6.1 ja huomaa, että aste-ehto riittää olettaa vain todellisessa parametriarvossa). Yksinkertainen konkreettinen esimerkki tällaisesta hypoteesista on $\beta_3 = \beta_1\beta_2$.

Tässä tapauksessa rajoittamaton malli on tavanomainen lineaarinen malli, jonka parametrien estimointi sujuu helposti PNS–menetelmällä, kun käytetään ehdollista uskottavuusfunktiota (ks. jakso 2.3). Rajoittamattoman mallin estimointiin perustuva Waldin testi on siten kätevä. Koska parametrit β ja σ^2 ovat lisäksi ortogonaaliset (ks. jakso 2.3), voidaan testin johtamista yksinkertaistaa.

Olkoon $\hat{\beta}$ ja $\hat{\sigma}^2$ parametrien β ja σ^2 SU–estimaattoreita, jolloin $\hat{\beta}$ on siis PNS–estimaattori (ks. (2.18) jaksossa 2.3). Ortogonaalisuuden nojalla riittää muodostaa parametrin β havaittu informaatiomatriisi, joksi saadaan jaksossa 2.4 esitettyä log–uskottavuusfunktiota derivoimalla (tai lineaarisen mallin kurssilla esitetystä)

$$\mathcal{J}_{\beta\beta}(\beta, \sigma^2; \boldsymbol{W}) := \frac{1}{\sigma^2}\sum_{i=1}^{n} Z_i Z_i' =: \frac{1}{\sigma^2} Q_n.$$

Käyttäen jaksossa 2.6.1 esitettyä Waldin testin yleistystä epälineaaristen hypoteesien tapaukseen saadaan testisuure

$$\mathsf{W}_\beta = \frac{1}{\hat{\sigma}^2} h(\hat{\beta})' \left[H(\hat{\beta}) Q_n^{-1} H(\hat{\beta})' \right]^{-1} h(\hat{\beta}) \underset{\mathsf{H}_0}{\overset{d}{\to}} \chi_q^2.$$

SU–estimaattorin $\hat{\sigma}^2$ asemesta voidaan käyttää lineaarisen mallin harhatonta estimaattoria ilman, että asymptoottinen jakauma muuttuu.

Heteroskedastisuuden testaus lineaarisessa mallissa. Edellisessä esimerkissä tarkasteltu malli olettaa havaintojen $Y_1, ..., Y_n$ ehdolliset varianssit (tai virhevarianssin) vakioksi. Tätä oletusta voidaan kuitenkin epäillä esimerkiksi, kun havaintoyksiköt (kuten kotitalous tai yritys) ovat jossain mielessä eri kokoisia tai muutoin erilaisia (kuten tulojen tai liikevaihdon suhteen). Seuravassa vakiovarianssi- eli homoskedastisuusoletuksen tutkimista varten johdetaan Raon pistemäärätesti, joka on tässä tapauksessa kätevä, koska rajoitettu malli on tavanomainen lineaarinen malli ja rajoitettu SU–estimointi sujuu siten helpposti.

Testiä varten aiemmin tarkasteltua lineaarista mallia laajennetaan olettamalla $Y_i|(\boldsymbol{W}_{i-1}, X_i) \overset{d}{=} Y_i|(Z_i, V_i)$ ja (vrt. edellisen esimerkin vastaava oletus tai (2.15))

$$Y_i|(Z_i = z_i, V_i = v_i) \sim \mathsf{N}\left(z_i'\beta, \sigma_i^2(\delta)\right), \quad \sigma_i^2(\delta) = \exp\left\{v_i'\delta\right\}, \tag{2.41}$$

jossa z_i on kuten edellä ja vektorin v_i $(k \times 1)$ ensimmäinen komponentti on vakio 1 ja muiksi komponenteiksi valitaan sellaiset vektorin $(\boldsymbol{w}_{i-1}, x_i)$ komponentit, joiden uskotaan selittävän ehdollisessa varianssissa mahdollisesti ilmenevä vaihtelu eli heteroskedastisuus. Parametrivektorille $\delta = (\delta_1, \delta_2, ..., \delta_k)$ asetetaan nollahypoteesi

$$\mathsf{H}_0 : \delta_2 = \cdots = \delta_k = 0 \quad \Leftrightarrow \quad \sigma_i^2(\delta) = e^{\delta_1} := \sigma^2,$$

jonka voimassa ollessa homoskedastinen malli on riittävä.

Heteroskedastisuuden kuvaamisessa edellä käytettyä mallia sanotaan usein multiplikatiiviseksi malliksi. Siinä olevan eksponenttifunktion etuina on, että varianssin positiivisuusvaatimus tulee automaattisesti täytetyksi ja testisuureessa tarvittavat osittaisderivaatat tulevat varsin yksinkertaisiksi. Muitakin funktiomuotoja voidaan tietystä käyttää ja on myös käytetty.

Menettelemällä kuten jaksossa 2.4 saadaan ehdolliseksi uskottavuusfunktioksi (vrt. (2.14)) tässä tapauksessa

$$L^{(c)}(\theta; \boldsymbol{w}) = \prod_{i=1}^{n} \sigma_i^2(\delta)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - z_i'\beta)^2}{\sigma_i^2(\delta)} \right\}, \quad \theta = (\beta, \delta) \in \mathbb{R}^p \times \mathbb{R}^k.$$

Samaan uskottavuusfunktioon päädytään myös, kun lähtökohdaksi otetaan vaihtoehtoisesti yhtälön (2.16) yleistys

$$Y_i = Z_i'\beta + \sigma_i(\delta)\eta_i, \quad i = 1, ..., n, \tag{2.42}$$

jossa $\eta_1, ..., \eta_n \perp\!\!\!\perp$, $\eta_i \sim \mathsf{N}(0,1)$ ja $(\boldsymbol{W}_{i-1}, X_i) \perp\!\!\!\perp \eta_i$ kaikilla $i = 1, ..., n$. Nollahypoteesin voimassa ollessa $\sigma_i(\delta)\eta_i = \sigma\eta_i \sim \mathsf{N}(0, \sigma^2)$, joten merkitsemällä $\varepsilon_i = \sigma\eta_i$ palautuu tilanne aikaisempaan homoskedastiseen tapaukseen (2.16).

Käyttämällä määritelmää $\sigma_i^2(\delta) = \exp\{v_i'\delta\}$ ja logaritmoimalla päädytään ehdolliseen log–uskottavuusfunktioon

$$l^{(c)}(\theta; \boldsymbol{w}) = -\frac{1}{2}\sum_{i=1}^{n} v_i'\delta - \frac{1}{2}\sum_{i=1}^{n} \exp\{-v_i'\delta\}\left(y_i - z_i'\beta\right)^2.$$

Lasketaan seuraavaksi parametrien β ja δ pistemäärät. Derivoimalla saadaan

$$\frac{\partial}{\partial\beta} l^{(c)}(\theta; \boldsymbol{w}) = \sum_{i=1}^{n} \exp\{-v_i'\delta\}\left(y_i - z_i'\beta\right) z_i$$

ja

$$\frac{\partial}{\partial\delta} l^{(c)}(\theta; \boldsymbol{w}) = -\frac{1}{2}\sum_{i=1}^{n} v_i + \frac{1}{2}\sum_{i=1}^{n} \exp\{-v_i'\delta\}\left(y_i - z_i'\beta\right)^2 v_i.$$

Käyttäen jakaumarelaatiota (2.41) tai yhtälöä (2.42) ja olettaen tarvittavien momenttien äärellisyys voidaan parametrien β ja δ pistemäärissä olevat yhteenlaskettavat todeta MD–differensseiksi informaatiojoukon $(\boldsymbol{W}_i, X_{i+1})$ suhteen (tämän voi perustella myös mallin säännöllisyydellä). Ne ovat siten erityisesti korreloimattomat, mistä seuraa suoraviivaisella laskulla tulos

$$\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\delta} l^{(c)}(\theta; \boldsymbol{W})\right)\left(\frac{\partial}{\partial\beta'} l^{(c)}(\theta; \boldsymbol{W})\right)\right] = 0$$

eli parametrien δ ja β ortogonaalisuus. Tämän ortogonaalisuuden perusteella Raon testi voidaan perustaa pelkästään parametrin δ pistemäärään ja havaittuun informaatioon. Jälkimmäistä varten lasketaan

$$\frac{\partial^2}{\partial\delta\partial\delta'} l^{(c)}(\theta; \boldsymbol{w}) = -\frac{1}{2}\sum_{i=1}^{n} \exp\{-v_i'\delta\}\left(y_i - z_i'\beta\right)^2 v_i v_i'.$$

Raon testiä varten tarvitaan parametrin θ rajoitettu SU–estimaattori $\tilde{\theta} = (\tilde{\beta}, \tilde{\delta})$. Koska nollahypoteesin voimassa ollessa malli on aiemmin tarkasteltu homoskedastinen lineaarinen malli, on $\tilde{\beta} = \hat{\beta}$ ja $\tilde{\delta} = (\tilde{\delta}_1, 0, ..., 0)$, jossa $\tilde{\delta}_1 = \log\hat{\sigma}^2$, ja $\hat{\beta}$ ja $\hat{\sigma}^2$ on määritelty yhtälöissä (2.18). Parametrin δ pistemäärä, kun argumentiksi valitaan rajoitettu SU–estimaatti, on

$$s_\delta(\tilde{\theta}; \boldsymbol{w}) := \frac{\partial}{\partial\delta} l^{(c)}(\tilde{\theta}; \boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_i - z_i'\tilde{\beta})^2}{\tilde{\sigma}^2} - 1\right) v_i,$$

jossa on loogisuuden vuoksi $\tilde{\sigma}^2 = \hat{\sigma}^2$ ja $\hat{\sigma}^2$ kuten edellä kuvattiin. Havaituksi informaatioksi pisteessä $\tilde{\theta}$ saadaan vastaavin merkinnöin

$$\mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{w}) := -\frac{\partial^2}{\partial\delta\partial\delta'} l^{(c)}(\tilde{\theta}; \boldsymbol{w}) = \frac{1}{2\tilde{\sigma}^2}\sum_{i=1}^{n}(y_i - z_i'\tilde{\beta})^2 v_i v_i'.$$

Käyttäen näitä määritelmiä ja merkintää $A = [0 : I_{k-1}]$ saadaan Raon testisuure

$$\begin{aligned}\mathsf{S}_\delta &= s_\delta(\tilde{\theta}; \boldsymbol{W})' \mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{W})^{-1} A' \left[A \mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{W})^{-1} A' \right]^{-1} A \mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{W})^{-1} s_\delta(\tilde{\theta}; \boldsymbol{W}) \\ &= s_\delta(\tilde{\theta}; \boldsymbol{W})' \mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{W})^{-1} s_\delta(\tilde{\theta}; \boldsymbol{W}) \\ &\xrightarrow[\mathsf{H}_0]{d} \chi^2_{k-1}.\end{aligned}$$

Tässä testisuureen jälkimmäinen lauseke perustuu vastaavaan yleiseen Raon testisuureen lausekkeeseen (2.39) (ks. jakso 2.6.2). Asymptoottisen jakauman perustelu vaatii heteroskedastisuutta selittäviltä muuttujilta "sopivat" ehdot, jotta tarvittavat raja-arvolauseet pätevät.

Mainittakoon, että testisuureen S_δ asemesta käytetään usein yksinkertaisempaa asymptoottisesti yhtäpitävää vaihtoehtoa. Koska $Y_i - Z_i'\beta = \sigma_i(\delta)\, \eta_i$ (olettaen todelliset parametriarvot) ja $\eta_i \perp\!\!\!\perp V_i$, niin nollahypoteesin voimassa ollessa $\mathsf{E}[(Y_i - Z_i'\beta)^2] = \sigma^2 \mathsf{E}(\eta_i^2) = \sigma^2$ ja

$$\mathsf{E}\left[\mathcal{J}_{\delta\delta}(\theta; \boldsymbol{W})\right] = \frac{1}{2\sigma^2} \sum_{i=1}^n \mathsf{E}[\left(Y_i - Z_i'\beta\right)^2] \mathsf{E}\left(V_i V_i'\right) = \frac{1}{2} \sum_{i=1}^n \mathsf{E}\left(V_i V_i'\right).$$

Samaan tapaan kuin perusteltaessa aikaisemmin ulkotulomatriisin $\mathsf{M}(\theta; \boldsymbol{Y})$ käyttöä havaitun informaationmatriisin paikalla yleisessä Waldin testissä ja Raon testissä (ks. jaksot 2.6.1 ja 2.6.2) voidaan edellä esitettyä käyttäen perustella matriisin $\mathcal{J}_{\delta\delta}(\tilde{\theta}; \boldsymbol{W})$ korvaaminen testisuureessa S_δ matriisilla $2^{-1} \sum_{i=1}^n V_i V_i'$. Merkitsemällä $\tilde{U}_i = (Y_i - Z_i'\tilde{\beta})^2/\tilde{\sigma}^2 - 1$ ja harrastaen hieman matriisilaskentaa päädytään testisuureeseen

$$\mathsf{S}_\delta^* = \frac{1}{2} \left(\sum_{i=1}^n \tilde{U}_i V_i' \right) \left(\sum_{i=1}^n V_i V_i' \right)^{-1} \left(\sum_{i=1}^n \tilde{U}_i V_i \right) \xrightarrow[\mathsf{H}_0]{d} \chi^2_{k-1}$$

nollahypoteesin voimassa ollessa.

Edellä esitetyssä testissä käytetty multiplikatiivinen varianssimalli on yleisempi kuin ensi näkemältä voisi ajatella. Valitsemalla esimerkiksi $v_i = (1, \log x_{a,i})$, jossa $x_{a,i}$ on selittävän muuttujan vektorin x_i (positiivinen) a. komponentti saadaan $\sigma_i^2(\delta) = \sigma^2 x_{a,i}^{\delta_2}$. Lisäksi voidaan osoittaa, että samaan testisuureeseen päädytään olettamalla $\sigma_i^2(\delta) = h(v_i'\delta)$, jossa h on mikä tahansa kahdesti jatkuvasti derivoituva positiivinen funktio. Tämä tekee Raon testistä miellyttävän, koska mahdollisen heteroskedastisuuden luonteesta ei ole useinkaan edeltä käsin luotettavaa tietoa.