

Tilastollinen päättely

7. Suurimman uskottavuuden menetelmä ja asymptoottinen teoria

7.1. Suurimman uskottavuuden estimointimenetelmä: Johdanto

Aikasarja, Asymptoottinen normaalisuus, Derivaatta, Ergodisuus, Estimaattori, Estimointi, Harhattomuus, Havainto, Havaintoarvo, Logaritminen uskottavuusfunktio, Maksimointi, Martingaalidifferenssi, Otos, Parametri, Realisaatio, Riippumattomuus, Satunnaisuuttuja, Stationaarisuus, Stokastinen prosessi, Suurimman uskottavuuden estimaattori, Suurimman uskottavuuden menetelmä, Tarkentuvuus, Uskottavuusfunktio, Yhteisjakauma

7.2. Suurimman uskottavuuden estimaattori ja sen asymptoottiset ominaisuudet

Asymptoottinen normaalisuus, Cramérin ja Raon alaraja, Derivaatta, Estimaattori, Estimointi, Fisherin informaatiomatriisi, Gradientti, Harhattomuus, Havainto, Havaintoarvo, Hessen matriisi, Kovarianssimatriisi, Logaritminen uskottavuusfunktio, Luottamustaso, Luottamusväli, Maksimointi, Normaali jakauma, Otos, Parametri, Suurimman uskottavuuden estimaattori, Suurimman uskottavuuden menetelmä, Tarkentuvuus, Tehokas pistemäärä, Tehokkuus, Testi, Tyhjentyvyys, Uskottavuusfunktio, Yhteisjakauma

7.3. Asymptoottiset testit

Asymptoottinen normaalisuus, Derivaatta, Diagnostinen testi, Estimaattori, Estimointi, Havainto, Havaintoarvo, Hessen matriisi, Hylkäysalue, χ^2 -jakauma, Kovarianssimatriisi, Kriittinen arvo, Lagrangen kertojatesti, Logaritminen uskottavuusfunktio, Maksimointi, Nollahypoteesi, Normaali-jakauma, Osamäärätesti, Otos, Parametri, Rajoitus, Side-ehto, Suurimman uskottavuuden estimaattori, Suurimman uskottavuuden menetelmä, Tarkentuvuus, Tehokas pistemäärä, Tehokkuus, Testi, Tyhjentyvyys, Uskottavuusfunktio, Vaihtoehtoinen hypoteesi, Waldin testi, Yhteisjakauma

7.4. Numeerinen optimointi

Apuregressio, Askelpituus, Estimaattori, Fisherin informaatiomatriisi, Gaussin ja Newtonin menetelmä, Gradientti, Iteraatioaskel, Iteratiivinen menetelmä, Maksimointi, Minimointi, Pienimmän neliösumman menetelmä, Suuntavektori, Suurimman uskottavuuden menetelmä, Hessen matriisi, Newtonin ja Raphsonin menetelmä.

7.1. Suurimman uskottavuuden estimointimenetelmä: Johdanto

Otos ja sen jakauma

Olkoot

$$X_1, X_2, \dots, X_n$$

satunnaismuuttujia, joiden yhteisjakauman pistetodennäköisyys- tai tiheysfunktio

$$f(x_1, x_2, \dots, x_n; \theta)$$

riippuu tuntemattomasta parametrista

$$\theta \in \Theta$$

jossa joukko Θ on *parametriavaruus* eli mahdollisten parametrin arvojen joukko.

Oletetaan, että satunnaismuuttujien X_1, X_2, \dots, X_n havaitut arvot ovat

$$x_1, x_2, \dots, x_n$$

Kutsumme satunnaismuuttujia X_1, X_2, \dots, X_n **havainnoiksi** ja niiden havaittuja arvoja x_1, x_2, \dots, x_n **havaintoarvoiksi**.

Satunnaismuuttujat X_1, X_2, \dots, X_n muodostavat **otoksen**, jonka jakauman määrittelee niiden yhteisjakauman pistetodennäköisyys- tai tiheysfunktio

$$f(x_1, x_2, \dots, x_n; \theta)$$

Havaintoarvot

$$x_1, x_2, \dots, x_n$$

ovat kiinteitä eli *ei-satunnaisia* reaalilukuja, mutta ne vaihtelevat satunnaisesti otoksesta toiseen satunnaismuuttujien X_1, X_2, \dots, X_n yhteisjakauman f määräämin todennäköisyyksin.

Uskottavuusfunktio

Otoksen

$$X_1, X_2, \dots, X_n$$

uskottavuusfunktio on satunnaismuuttujien X_1, X_2, \dots, X_n yhteisjakauman pistetodennäköisyys- tai tiheysfunktion

$$f(x_1, x_2, \dots, x_n; \theta)$$

arvo havaintopisteessä $\mathbf{x} = (x_1, x_2, \dots, x_n)$ tulkittuna parametrin θ funktioksi.

Merkitsemme uskottavuusfunktioita seuraavalla tavalla:

$$L(\theta) = L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \theta)$$

Uskottavuusperiaatteen mukaan uskottavuusfunktio tiivistää yhteen kaiken olennaisen parametria θ koskevan informaation otoksesta.

Suurimman uskottavuuden estimaattori

Olkoon

$$L(\theta) = L(\theta; x_1, x_2, \dots, x_n)$$

otoksen X_1, X_2, \dots, X_n uskottavuusfunktio ja olkoon

$$t = g(x_1, x_2, \dots, x_n)$$

sellainen parametrin θ arvo, joka maksimoi uskottavuusfunktion $L(\theta)$ arvon eli

$$L(\mathbf{t}) = \max_{\theta \in \Theta} L(\theta)$$

Uskottavuusfunktion $L(\theta)$ maksimin antava parametrin θ arvo \mathbf{t} on havaintoarvojen x_1, x_2, \dots, x_n funktio.

Valitaan parametrin θ estimaattoriksi satunnaismuuttuja

$$\hat{\theta} = t = g(X_1, X_2, \dots, X_n)$$

Estimaattoria $\hat{\theta}$ kutsutaan parametrin θ suurimman uskottavuuden estimaattoriksi eli **SU-estimaattoriksi**. Suurimman uskottavuuden estimaattori $\hat{\theta}$ tuottaa parametrille θ arvon, joka tekevät saadut havaintoarvot

$$x_1, x_2, \dots, x_n$$

mahdollisimman uskottaviksi (\sim todennäköisiksi).

Parametrin θ suurimman SU-estimaattori määrätään siis maksimoimalla uskottavuusfunktio

$$L(\theta) = L(\theta; x_1, x_2, \dots, x_n)$$

parametrin θ suhteen. Säännöllisissä tapauksissa maksimi löydetään merkitsemällä uskottavuusfunktion $L(\theta)$ derivaatta

$$L'(\theta)$$

nollaksi ja ratkaisemalla θ saadusta normaaliyhtälöstä

$$L'(\theta) = 0$$

Olkoon $t = g(x_1, x_2, \dots, x_n)$ normaaliyhtälön $L'(\theta) = 0$ nollakohta. Piste t vastaa uskottavuusfunktion maksimia, jos

$$L''(t) < 0$$

Uskottavuusfunktion maksimin antava parametrin θ arvo voidaan etsiä myös maksimoimalla **logaritminen uskottavuusfunktio** eli **uskottavuusfunktion logaritmi**

$$l(\theta) = \log L(\theta; x_1, x_2, \dots, x_n)$$

parametrin θ suhteen. Logaritmisen uskottavuusfunktion maksimointi on usein yksinkertaisempaa kuin uskottavuusfunktion itsensä maksimointi.

Huomautus:

Logaritminen uskottavuusfunktio ja uskottavuusfunktio saavuttavat maksiminsa *samassa pisteessä*.

Uskottavuusfunktio ja riippumattomat havainnot

Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertaisen satunnaisotos jakaumasta $f_X(x; \theta)$, jossa θ on jakauman muodon määräävä parametri.

Tällöin satunnaismuuttujat X_1, X_2, \dots, X_n ovat *riippumattomia* ja noudattavat *samaa jakaumaa* $f_X(x; \theta)$:

$$\begin{aligned} X_1, X_2, \dots, X_n &\perp \\ X_i &\sim f_X(x; \theta), i = 1, 2, \dots, n \end{aligned}$$

Koska satunnaismuuttujat X_1, X_2, \dots, X_n oletettiin riippumattomiksi, otoksen X_1, X_2, \dots, X_n uskottavuusfunktio voidaan esittää muodossa

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

jossa

$$f(x_i; \theta), i = 1, 2, \dots, n$$

on havaintoarvoon x_i liittyvä pistetodennäköisyys- tai tiheysfunktio.

Siten vastaava *logaritminen uskottavuusfunktio* voidaan esittää muodossa

$$\begin{aligned} l(\theta) &= \log L(\theta) \\ &= \log(f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)) \\ &= \log f(x_1; \theta) + \log f(x_2; \theta) + \cdots + \log f(x_n; \theta) \\ &= l(\theta; x_1) + l(\theta; x_2) + \cdots + l(\theta; x_n) \end{aligned}$$

jossa

$$l(\theta; x_i) = \log f(\theta; x_i), i = 1, 2, \dots, n$$

on havaintoarvon x_i logaritminen uskottavuusfunktio. Etenkin *riippumattomien* havaintojen tapauksessa *logaritmisen uskottavuusfunktion summaesityksen* maksimointi on tavallisesti helpompaa kuin uskottavuusfunktion itsensä maksimointi.

Suurimman uskottavuuden estimaattorin ominaisuudet

Suurimman uskottavuuden estimaattori ei välttämättä toteuta tavanomaisia *hyvän estimaattorin kriteereitä*, kun otoskoko on *äärellinen* eikä suurimman uskottavuuden estimaattorin *äärellisen otoskoon ominaisuuksia* välttämättä tunneta:

- (i) SU-estimaattori *ei välttämättä ole harhaton*.
- (ii) SU-estimaattorin jakaumaa *ei välttämättä tunneta*.

Onneksi suurimman uskottavuuden estimaattorilla on kuitenkin hyvin yleisin ehdoin *hyvät asymptoottiset ominaisuudet*.

Suurimman uskottavuuden estimaattorin asymptoottiset ominaisuudet

Suurimman uskottavuuden estimaattorilla on hyvin yleisin ehdoin seuraavat *asymptoottiset ominaisuudet* (ks. tarkemmin kappaletta 7.2.).

- (i) SU-estimaattori on **tarkentuva**.
- (ii) SU-estimaattori on **asymptoottisesti normaalin**.

SU-estimaattorin *tarkentuvuus* merkitsee sitä, että SU-estimaattori toteuttaa *suurten lukujen lain*. Siten estimaattorin arvo *konvergoi* (jossakin mielessä) *kohden parametrin ”oikeata” arvoa*, kun havaintojen lukumäärän annetaan kasvaa rajatta.

SU-estimaattorin *asymptoottinen normalisuus* merkitsee sitä, että SU-estimaattori toteuttaa *keskeisen raja-arvolauseen*. Siten SU-estimaattorin jakaumaa voidaan *suurissa otoksissa approksimoida normaalijakaumalla*, jolloin parametrin *luottamusvälit* ja parametreja koskevat *testit* voidaan perustaa *normaalijakaumaan* (tai *t-jakaumaan*).

On suhteellisen yksinkertaista todistaa SU-estimaattorin tarkentuvuus ja asymptoottinen normalisuus, jos otos muodostuu *riippumattomista, samaa jakaumaa noudattavista* havainnoista. Oletuksia havaintojen riippumattomuudesta ja samasta jakaumasta voidaan tietyin edellytyksin *lieventää*, millä on suuri merkitys esimerkiksi *aikasarjamallien* SU-estimoinnissa.

Tilastollinen aikasarja-analyysi perustuu siihen, että havaittu aikasarja tulkitaan jonkin *stokastisen prosessin realisaatioksi*. Stokastiset prosessit ovat tilastollisia malleja havaituille aikasarjoille. Aikasarjamallien parametrin SU-estimaattoreiden *tarkentuvuus* ja *asymptoottinen normalisuus* voidaan *todistaa* esimerkiksi silloin, kun aikasarjan generoinut stokastinen prosessi on jotakin seuraavista tyypeistä:

- *ergodinen*
- *stationaarinen*
- *martingaalidifferenssi*

7.2. Suurimman uskottavuuden estimaattori ja sen asymptoottiset ominaisuudet

1. kertaluvun osittaisderivaattojen vektori ja

2. kertaluvun osittaisderivaattojen matriisi

Olkoon

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$$

p -vektori.

Tällöin

$$\mathbf{D} = \left(\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_p} \right) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \vdots \\ \frac{\partial}{\partial \theta_p} \end{bmatrix}$$

on 1. kertaluvun osittaisderivaattaoperaattoreiden

$$\frac{\partial}{\partial \theta_i}, i = 1, 2, \dots, p$$

muodostama p -vektori ja

$$\mathbf{D}^2 = \mathbf{D}\mathbf{D}' = \begin{bmatrix} \frac{\partial^2}{\partial \theta_1^2} & \dots & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} & \dots & \frac{\partial^2}{\partial \theta_p^2} \end{bmatrix}$$

on 2. kertaluvun osittaisderivaattaoperaattoreiden

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j}, i = 1, 2, \dots, p, j = 1, 2, \dots, p$$

muodostama $p \times p$ -matriisi.

Otos ja sen jakauma

Olkoot

$$X_1, X_2, \dots, X_n$$

satunnaismuuttujia, joiden yhteisjakauman pistetodennäköisyys- tai tiheysfunktio

$$f(x_1, x_2, \dots, x_n; \boldsymbol{\theta})$$

riippuu parametrasta

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$$

Olkoot satunnaismuuttujien X_1, X_2, \dots, X_n havaitut arvot

$$x_1, x_2, \dots, x_n$$

Uskottavuusfunktio

Otoksen

$$X_1, X_2, \dots, X_n$$

uskottavuusfunktio

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \boldsymbol{\theta})$$

on satunnaismuuttujien X_1, X_2, \dots, X_n yhteisjakauman pistetodennäköisyys- tai tiheysfunktion

$$f(x_1, x_2, \dots, x_n; \boldsymbol{\theta})$$

arvo havaintopisteessä $\mathbf{x} = (x_1, x_2, \dots, x_n)$ tulkittuna parametrin $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ funktioksi.

Olkoon

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; x_1, x_2, \dots, x_n)$$

vastaava **logaritminen uskottavuusfunktio**.

Suurimman uskottavuuden estimaattori

Olkoon

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; x_1, x_2, \dots, x_n)$$

otoksen X_1, X_2, \dots, X_n uskottavuusfunktio ja olkoon

$$\mathbf{t} = g(x_1, x_2, \dots, x_n)$$

sellainen parametrin $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ arvo, joka maksimoi uskottavuusfunktion $L(\boldsymbol{\theta})$ arvon.

Uskottavuusfunktion $L(\boldsymbol{\theta})$ maksimin antava parametrin $\boldsymbol{\theta}$ arvo \mathbf{t} on havaintoarvojen x_1, x_2, \dots, x_n funktio.

Valitaan parametrin $\boldsymbol{\theta}$ estimaattoriksi satunnaismuuttuja

$$\hat{\boldsymbol{\theta}} = \mathbf{T} = g(X_1, X_2, \dots, X_n)$$

Estimaattoria $\hat{\boldsymbol{\theta}}$ kutsutaan parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattoriksi eli **SU-estimaattoriksi**.

Koska uskottavuusfunktion $L(\boldsymbol{\theta})$ maksimointi on yhtäpitävää logaritmisestä uskottavuusfunktion

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

maksimoinnin kanssa, niin uskottavuusfunktiolla $L(\boldsymbol{\theta})$ on *lokaali maksimi* pisteessä $\hat{\boldsymbol{\theta}}$, jos

$$\mathbf{D}l(\hat{\boldsymbol{\theta}}) = 0$$

ja

$$-\mathbf{D}^2l(\hat{\boldsymbol{\theta}}) > 0$$

Huomautus:

Merkintä

$$\mathbf{A} > 0$$

tarkoittaa sitä, että matriisi \mathbf{A} on *positiivisesti definiitti* eli että

$$\mathbf{z}'\mathbf{A}\mathbf{z} > 0$$

kaikille $\mathbf{z} \neq \mathbf{0}$. Vastaavasti merkintä

$$\mathbf{A} \geq 0$$

tarkoittaa sitä, että matriisi \mathbf{A} on *ei-negatiivisesti definiitti* eli että

$$\mathbf{z}'\mathbf{A}\mathbf{z} \geq 0$$

kaikille \mathbf{z} .

Tehokas pistemäärä ja Fisherin informaatiomatriisi

Matemaattisessa tilastotieteessä vektoria

$$\mathbf{D}l(\boldsymbol{\theta})$$

kutsutaan **tehokkaaksi pistemääräksi** (engl. *score*) ja matriisia

$$\mathbf{I}(\boldsymbol{\theta}) = -E[\mathbf{D}^2l(\boldsymbol{\theta})] = E[(\mathbf{D}l(\boldsymbol{\theta}))(\mathbf{D}l(\boldsymbol{\theta}))']$$

kutsutaan **Fisherin informaatiomatriisiksi**.

Voidaan osoittaa, että Fisherin informaatiomatriisin $\mathbf{I}(\boldsymbol{\theta})$ käänteismatriisi muodostaa *alarajan* harhattomien estimaattoreiden kovarianssimatriisille seuraavassa mielessä:

Olkoon $\hat{\boldsymbol{\theta}}$ mielivaltainen *harhaton estimaattori* parametrille $\boldsymbol{\theta}$. Tällöin

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq [\mathbf{I}(\boldsymbol{\theta})]^{-1}$$

Matriisi $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$ on *Cramérin ja Raon alaraja* estimaattorin $\hat{\boldsymbol{\theta}}$ kovarianssimatriisille $\text{Cov}(\hat{\boldsymbol{\theta}})$.

Jos *harhattoman* estimaattorin $\hat{\boldsymbol{\theta}}$ kovarianssimatriisi toteuttaa yhtälön

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = [\mathbf{I}(\boldsymbol{\theta})]^{-1}$$

sanomme, että estimaattori $\hat{\boldsymbol{\theta}}$ on (**täys-**) **tehokas**.

Suurimman uskottavuuden estimaattori ja tyhjentyvyys

Olkoon

$$f(\mathbf{x}; \boldsymbol{\theta})$$

otoksen $\mathbf{X} = (X_1, X_2, \dots, X_n)$ yhteisjakauman pistetodennäköisyys- tai tiheysfunktio, joka riippuu parametrista $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$. *Faktorointiteoreeman* mukaan tunnusluku $\mathbf{T}(\mathbf{X})$ on *tyhjentävä* parametrille $\boldsymbol{\theta}$, jos ja vain jos on olemassa funktiot $g(\mathbf{t}; \boldsymbol{\theta})$ ja $h(\mathbf{x})$ siten, että

$$f(\mathbf{x}; \boldsymbol{\theta}) = g(\mathbf{T}(\mathbf{x}); \boldsymbol{\theta})h(\mathbf{x})$$

kaikille havaintopisteille $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ja parametrin $\boldsymbol{\theta}$ mahdollisille arvoille ja funktio g riippuu otoksesta $\mathbf{X} = \mathbf{x}$ vain tunnusluvun $\mathbf{T}(\mathbf{X})$ kautta ja funktio h ei riipu parametrilla $\boldsymbol{\theta}$.

Otoksen \mathbf{X} yhteisjakauman tiheysfunktion faktoroinnista nähdään suoraan, että parametrin $\boldsymbol{\theta}$ *suurimman uskottavuuden estimaattori* $\hat{\boldsymbol{\theta}}$ on tyhjentävän tunnusluvun $\mathbf{T}(\mathbf{X})$ funktio (olettaen siis, että tyhjentävä tunnusluku on olemassa).

Suurimman uskottavuuden estimaattori ja tehokkuus

Olkoon $\hat{\boldsymbol{\theta}}$ *harhaton* ja *tehokas* estimaattori parametrille $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$. Voidaan osoittaa, että tällöin estimaattori $\hat{\boldsymbol{\theta}}$ yhtyy parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattoriin.

Suurimman uskottavuuden estimaattorin tarkentuvuus

Suurimman uskottavuuden estimaattoreiden stokastisia ominaisuuksia ei useinkaan tunneta tai niiden ominaisuudet ovat ainakin vaikeasti hallittavissa, jos otoskoko on *äärellinen*.

Sen sijaan monien estimaattoreiden *asymptoottiset stokastiset ominaisuudet* tunnetaan hyvin. Estimaattoreiden asymptoottisilla stokastisilla ominaisuuksilla tarkoitetaan estimaattoreiden käyttäytymistä satunnaismuuttujina, kun *otoskoon annetaan hypoteettisesti kasvaa rajatta*. Puhumme usein estimaattoreiden *suurten otosten ominaisuuksista* tarkoittaessamme estimaattoreiden asymptoottisia käyttäytymistä.

Perustavaa laatua oleva vaatimus estimaattoreiden asymptoottiselle käyttäytymiselle on se, että estimaattorin arvon pitää *konvergoida stokastisesti* (tai *melkein varmasti*) kohti ”oikeata” parametrin arvoa, kun havaintojen lukumäärän annetaan kasvaa rajatta.

Tunnuslukujen

$$W_n = W_n(X_1, X_2, \dots, X_n), n = 1, 2, 3, \dots$$

jono muodostaa (**heikosti**) tarkentuvan jonon estimaattoreita parametrille θ , jos

$$\lim_{n \rightarrow \infty} \Pr_{\theta}(|W_n - \theta| < \varepsilon) = 1$$

kaikille $\varepsilon > 0$ ja kaikille $\theta \in \Theta$. Sanomme tällöin tavallisesti, että tunnusluku W_n on (**heikosti**) **tarkentuva estimaattori** parametrille θ .

Ekvivalentti ehto estimaattorin W_n tarkentuvuudelle on se, että

$$\lim_{n \rightarrow \infty} \Pr_{\theta}(|W_n - \theta| \geq \varepsilon) = 0$$

kaikille $\varepsilon > 0$ ja kaikille $\theta \in \Theta$.

Tavallisesti tarkentuvuuden todistamisessa ei tarvitse eksplisiittisesti vedota tarkentuvuuden määritelmään, sillä todistamisessa voidaan usein soveltaa *Tshebyshevin epäyhtälöä*.

Tshebyshevin epäyhtälön mukaan

$$\Pr_{\theta}(|W_n - \theta| \geq \varepsilon) \leq \frac{E_{\theta}[(W_n - \theta)^2]}{\varepsilon^2}$$

Siten estimaattori W_n on *tarkentuva* parametrille θ , jos

$$\lim_{n \rightarrow \infty} E_{\theta}[(W_n - \theta)^2] = 0$$

Koska

$$\begin{aligned} E_{\theta}[(W_n - \theta)^2] &= E_{\theta}[(W_n - E_{\theta}(W_n))^2] + [E_{\theta}(W_n) - \theta]^2 \\ &= \text{Var}_{\theta}(W_n) + [\text{Bias}_{\theta}(W_n)]^2 \end{aligned}$$

näemme, että estimaattori W_n on *tarkentuva* parametrille θ , jos

$$\lim_{n \rightarrow \infty} \text{Var}_{\theta}(W_n) = 0$$

ja

$$\lim_{n \rightarrow \infty} \text{Bias}_{\theta}(W_n) = 0$$

Tarkentuvien estimaattoreiden jonot eivät ole yksikäsitteisiä. Tämä nähdään seuraavasta:

Olkoon tunnuslukujen W_n , $n = 1, 2, \dots$ *tarkentuva jono estimaattoreita* parametrille θ ja olkoot a_1, a_2, a_3, \dots ja b_1, b_2, b_3, \dots kaksi jonoa ei-satunnaisia vakioita, joille pätee

$$\lim_{n \rightarrow \infty} a_n = 1$$

$$\lim_{n \rightarrow \infty} b_n = 0$$

Tällöin myös tunnuslukujen jono

$$U_n = a_n W_n + b_n, n = 1, 2, 3, \dots$$

muodostaa *tarkentuvan jonon estimaattoreita* parametrille θ .

Voidaan osoittaa, että *suurimman uskottavuuden estimaattorit* ovat hyvin yleisin ehdoin *tarkentuvia*.

Lause:

Olkoon $\hat{\theta}$ parametrin θ *suurimman uskottavuuden estimaattori* ja olkoon $\tau(\theta)$ parametrin θ jatkuva funktio. Tällöin tunnusluku $\tau(\hat{\theta})$ on hyvin yleisin ehdoin *tarkentuva estimaattori* parametrille $\tau(\theta)$:

$$\lim_{n \rightarrow \infty} \Pr_{\theta} (|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon) = 0$$

kaikille $\varepsilon > 0$ ja kaikille $\theta \in \Theta$.

Olkoon tarkastelun kohteena olevana parametrina *p*-vektori

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$$

ja olkoon parametrin $\boldsymbol{\theta}$ estimaattorina *p*-vektori

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$$

Estimaattorin $\hat{\boldsymbol{\theta}}$ tarkentuvuus voidaan määritellä sen komponenttien

$$\hat{\theta}_i, i = 1, 2, \dots, p$$

tarkentuvuuden kautta.

Suurimman uskottavuuden estimaattorin asymptoottinen tehokkuus ja normalisuus

Tarkentuvuutta voidaan pitää järkevyyksivaatimuksena estimaattoreiden käyttäytymiselle *suurissa otoksissa*. Tilastollisen päättelyn kannalta on kuitenkin tärkeää tuntea myös estimaattorin *varianssi* tai yleisemmin estimaattorin *jakauma suurissa otoksissa*.

Saattaa tuntua houkuttelevalta yrittää määritellä estimaattorin suurten otosten varianssi estimaattorin varianssin raja-arvona, kun otokseen annetaan kasvaa rajatta.

Olkoon tunnusluku T_n estimaattori parametrille θ ja olkoon

$$k_n, n = 1, 2, 3, \dots$$

jono ei-satunnaisia vakioita. Jos

$$\lim_{n \rightarrow \infty} k_n \text{Var}(T_n) = \sigma^2 < \infty$$

niin sanomme, että σ^2 on tunnusluvun T_n **varianssin raja-arvo**. Varianssin raja-arvo ei kuitenkaan ole osoittautunut samalla tavalla hyödylliseksi käsitteeksi kuin seuraavassa määriteltävän *asymptoottisen varianssin* käsite.

Olkoon tunnusluku T_n parametrin θ estimaattori, olkoon $\tau(\theta)$ parametrin θ jatkuva funktio ja olkoon

$$k_n, n = 1, 2, 3, \dots$$

jono ei-satunnaisia vakioita. Jos

$$\lim_{n \rightarrow \infty} k_n [T_n - \tau(\theta)] \rightarrow N(0, \sigma^2)$$

jakaumakonvergenssin mielessä, niin sanomme, että σ^2 on tunnusluvun T_n **asymptoottinen varianssi**. Tämä merkitsee sitä, että tunnusluvun T_n rajajakaumana on otoskoon kasvaessa rajatta normaalijakauma ja σ^2 on tämän rajajakauman varianssi. Monissa tilanteissa tunnusluvun asymptoottinen varianssi on sen varianssin raja-arvo.

Seuraavassa määritelmässä esitetään *Cramérin ja Raon alarajaa* vastaava *asymptoottinen käsite*. Tunnuslukujen $W_n, n = 1, 2, \dots$ jono on **asymptoottisesti tehokas** parametrille $\tau(\theta)$, jos

$$\lim_{n \rightarrow \infty} \sqrt{n} [W_n - \tau(\theta)] \rightarrow N(0, v(\theta))$$

jakaumakonvergenssin mielessä ja

$$v(\theta) = \frac{[\tau'(\theta)]^2}{E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X | \theta) \right)^2 \right]}$$

Tällöin tunnusluku W_n on *asymptoottisesti normaalin* ja sen asymptoottinen varianssi saavuttaa *Cramérin ja Raon alarajan*.

Seuraavan lauseen mukaan *suurten uskottavuuden estimaattori* on *asymptoottisesti normaalin* ja *tehokas*.

Lause:

Olkoon $\hat{\theta}$ parametrin θ suurimman uskottavuuden estimaattori ja olkoon $\tau(\theta)$ parametrin θ jatkuva funktio. Tällöin tunnusluku $\tau(\hat{\theta})$ on hyvin yleisin ehdoin *tarkentuva, asymptoottisesti normaalin* ja *tehokas estimaattori* parametrille $\tau(\theta)$:

$$\lim_{n \rightarrow \infty} \sqrt{n} [\tau(\hat{\theta}) - \tau(\theta)] \rightarrow N(0, v(\theta))$$

jossa $v(\theta)$ on Cramérin ja Raon alaraja.

Merkintä:

$$\tau(\hat{\theta}) \sim_a N(\tau(\hat{\theta}), v(\theta) / n)$$

Olkoon tarkastelun kohteena olevana parametrina nyt *p*-vektori

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$$

ja olkoon parametrin $\boldsymbol{\theta}$ estimaattorina *p*-vektori

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$$

Tällöin parametrin $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ suurimman uskottavuuden estimaattorin $\hat{\boldsymbol{\theta}}$ asymptoottisesta normaalisuutta ja tehokkuutta koskeva tulos voidaan esittää seuraavassa muodossa:

$$\hat{\boldsymbol{\theta}} \sim_a N_p \left(\boldsymbol{\theta}, \frac{1}{n} [\mathbf{IA}(\boldsymbol{\theta})]^{-1} \right)$$

jossa

$$\mathbf{IA}(\boldsymbol{\theta}) = \lim_{n \rightarrow +\infty} \frac{1}{n} \mathbf{I}(\boldsymbol{\theta})$$

on Fisherin informaatiomatriisin $\mathbf{I}(\boldsymbol{\theta})$ asymptoottinen arvo.

Jos parametrin $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ suurimman uskottavuuden estimaattori on asymptoottisesti normaalin, niin parametreja

$$\theta_i, i = 1, 2, \dots, p$$

koskevassa tilastollisessa päättelyssä voidaan nojata suurimman uskottavuuden estimaattorin approksimatiiviseen normalisuuteen suurissa otoksissa:

- (i) Parametreille

$$\theta_i, i = 1, 2, \dots, p$$

voidaan konstruoida *luottamusvälit* samanlaisella tekniikalla kuin normaalijakauman odotusarvolle. Luottamuskertoimet määrätään joko normaalijakaumasta tai *t*-jakaumasta.

- (ii) Parametrien

$$\theta_i, i = 1, 2, \dots, p$$

arvoja koskeville nollahypoteeseille voidaan konstruoida *testit* samaan tapaan kuin normaalijakauman odotusarvolle. Testien hylkäysalueet tai testisuuren arvoja vastaavat *p*-arvot määrätään joko normaalijakaumasta tai *t*-jakaumasta.

On syytä huomata, että SU-estimaattorin normalisuutta koskeva tulos on luonteeltaan *asymptoottinen*, mutta käytännön sovellustilanteissa otokoot ovat aina *äärellisiä*. Siten SU-estimaattorin normalisuuteen perustuvat *luottamusvälit* ja *testit* eivät ole *eksakteja*, vaan ainoastaan *approksimatiivisia*. Siksi luottamuskertoimia ja testien hylkäysalueita tai *p*-arvoja määrättäessä SU-päättelyssä ei tavallisesti käytetä normaalijakaumaa, vaan *t-jakaumaa*. Tämä johtuu siitä, että *t*-jakauman käyttö johtaa *konservatiivisempiin* luottamusväleihin ja testeihin kuin normaali-jakauman käyttö:

- (i) Oletetaan, että konstruoinme luottamusvälit tarkastelun kohteena olevalle parametrille θ *samalla* luottamustasolla $(1-\alpha)$ sekä *t*-jakaumasta että normaalijakaumasta. Tällöin *t*-jakaumaan perustuva luottamusväli on aina *leveämpi* kuin vastaava normaalijakaumaan perustuva luottamusväli.
- (ii) Oletetaan, että konstruoinme testit tarkastelun kohteena olevaa parametrin θ arvoa koskevalle nollahypoteesille määräämällä testien hyväksymisalueet *samalla* merkitsevyydestasolla α sekä *t*-jakaumasta että normaalijakaumasta. Tällöin *t*-jakaumaan perustuva hyväksymisalue on aina *suurempi* kuin vastaava normaalijakaumaan perustuva hyväksymisalue.

Oletetaan, että olemme määränneet testisuuren arvon parametria θ koskevalle nollahypoteesille. Tällöin testisuuren arvoa vastaava *t*-jakaumasta määrätty *p*-arvo on aina *suurempi* kuin normaalijakaumasta määrätty *p*-arvo.

7.3. Asymptoottiset testit

Testiprobleema

Olkoot

S^n = kaikkien mahdollisten havaintopisteiden $\mathbf{x} = (x_1, x_2, \dots, x_n)$ joukko

H = yleinen hypoteesi, joka kiinnittää otoksen jakauman

H_0 = nollahypoteesi, joka kiinnittää jotkin yleisessä hypoteesissa kiinnitetyn jakauman parametreista

Tilastollisella testillä tarkoitetaan päätössääntöä, joka kertoo milloin nollahypoteesi jätetään voimaan ja milloin nollahypoteesi hylätään. Testin päätössääntö jakaa kaikkien mahdollisten havaintopisteiden joukon S^n hyväksymisalueeseen ja hylkäysalueeseen.

Jos havaintopiste $\mathbf{x} = (x_1, x_2, \dots, x_n)$ joutuu hylkäys- eli kriittiselle alueelle silloin, kun nollahypoteesi pätee, tehdään 1. lajin virhe eli hylkäysvirhe. Jos havaintopiste $\mathbf{x} = (x_1, x_2, \dots, x_n)$ joutuu hyväksymisalueelle silloin, kun nollahypoteesi ei päde, tehdään 2. lajin virhe eli hyväksymisvirhe.

Testin merkitsevyytason valinta tarkoittaa 1. lajin virheen todennäköisyyden kiinnittämistä etukäteen, ennen testin tekemistä. Jos merkitsevyytaso on valittu etukäteen, voidaan niiden testien joukosta, joilla on sama 1. lajin virheen todennäköisyys, pyrkiä löytämään se, joka minimoi 2. lajin virheen todennäköisyyden. Jos tällainen testi on olemassa, sitä kutsutaan *tasaisesti voimakkaimmaksi* testiksi.

Osamäärätesti

Olkoot

$\hat{\theta}$ = parametrin θ SU-estimaattori, joka on määrätty olettaen, että testin yleinen hypoteesi H pätee

$L(\hat{\theta})$ = otoksen uskottavuusfunktion arvo pisteessä $\hat{\theta}$

$\hat{\theta}_0$ = parametrin θ rajoitettu SU-estimaattori, joka on määrätty olettaen, että testin nollahypoteesi H_0 pätee

$L(\hat{\theta}_0)$ = otoksen uskottavuusfunktion arvo pisteessä $\hat{\theta}_0$

Tällöin satunnaismuuttuja

$$\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

on (uskottavuus-) osamäärätestisuure nollahypoteesille H_0 .

Huomaa, että

$$0 \leq \lambda \leq 1$$

Osamäärätestisuureta λ muodostettaessa mallin parametrit joudutaan estimoimaan suurimman uskottavuuden menetelmällä kahdella eri tavalla:

- (i) Osamäärätestisuureen *nimittäjä* saadaan etsimällä uskottavuusfunktion maksimi *yleisen hypoteesin* H pätiessä. Tämä merkitsee *vapaan ääriarvottehtävän* ratkaisemista.
- (ii) Osamäärätestisuureen *osoittaja* saadaan etsimällä uskottavuusfunktion maksimi *nollahypoteesin* H_0 pätiessä. Tämä merkitsee *sidotun ääriarvottehtävän* ratkaisemista.

Koska osamäärätestisuureella λ on taipumus saada *pieniä* arvoja, jos nollahypoteesi H_0 *ei päde*, osamäärätestin *hylkäysalueeksi* valitaan kaikkien mahdollisten havaintopisteiden joukon S^n alue

$$\{\mathbf{x} \in S^n \mid \lambda < \lambda(\alpha)\}$$

jossa *kriittinen arvo* $\lambda(\alpha)$ valitaan siten, että

$$\Pr(\mathbf{x} \in S^n \mid \lambda < \lambda(\alpha)) = \alpha$$

jos nollahypoteesi H_0 *pätee*. Jos osamäärätestisuureen λ arvo joutuu hylkäysalueelle, *nollahypoteesi* H_0 *hylätään*.

Osamäärätestin käytössä on kuitenkin usein se käytännön ongelma, että osamäärätestisuureen λ jakauma *ei välttämättä ole mitään tunnettua tyyppiä*.

Osamäärätestisuureelle λ pätee kuitenkin tiettyjen (melko lievien) ehtojen vallitessa ja *nollahypoteesin* H_0 *pätiessä* seuraava *asymptoottinen jakaumatulos*:

$$-2 \log \lambda = -2 \log \frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} = 2l(\hat{\boldsymbol{\theta}}) - 2l(\hat{\boldsymbol{\theta}}_0) \sim_a \chi^2(m)$$

jossa

$$m = \text{nollahypoteesin kiinnittämien parametrien lukumäärä}$$

Jos nollahypoteesi H_0 *ei päde*, testisuureella $-2 \log \lambda$ on taipumus saada *suuria* arvoja.

Testin *hylkäysalueeksi* valitaan kaikkien mahdollisten havaintopisteiden joukon S^n alue

$$\{\mathbf{x} \in S^n \mid -2 \log \lambda > \chi^2_\alpha(m)\}$$

jossa *kriittinen arvo* $\chi^2_\alpha(m)$ valitaan siten, että

$$\Pr(\mathbf{x} \in S^n \mid -2 \log \lambda > \chi^2_\alpha(m)) = \alpha$$

jos nollahypoteesi H_0 *pätee*. Jos testisuureen $-2 \log \lambda$ arvo joutuu hylkäysalueelle, *nollahypoteesi* H_0 *hylätään*.

Osamäärätestisuuretta muodostettaessa joudutaan mallin parametrit estimoimaan, so. uskottavuusfunktio joudutaan maksimoimaan sekä yleisen hypoteesin H että nollahypoteesin H_0 pätiessä. Monissa testausasetelmissä toinen näistä ääriarvottehtävistä on *vaikea*, mutta toinen on *helppo* ratkaista. Tämä havainto on johtanut seuraavien (asymptoottisten) testien konstruointiin:

- (i) *Waldin testissä* parametrit estimoidaan olettaen, että yleinen hypoteesi H pätee.
- (ii) *Lagrangen kertojatestissä* parametrit estimoidaan olettaen, että nollahypoteesi H_0 pätee.

Waldin testi lineaarisille rajoituksille

Olkoon

$\hat{\boldsymbol{\theta}}$ = parametrin $\boldsymbol{\theta}$ SU-estimaattori, joka on määrätty olettaen, että testin yleinen hypoteesi H pätee

Olkoon *nollahypoteesina* lineaarinen hypoteesi

$$H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$$

jossa \mathbf{R} on $m \times p$ -matriisi, jolle

$$r(\mathbf{R}) = m$$

Määritellään satunnaismuuttuja

$$W = (\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r})'[\mathbf{R}\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r})$$

jossa matriisi

$$\mathbf{I}(\boldsymbol{\theta}) = -E[\mathbf{D}^2l(\boldsymbol{\theta})] = E[(\mathbf{D}l(\boldsymbol{\theta}))(\mathbf{D}l(\boldsymbol{\theta}))']$$

on Fisherin informaatiomatriisi. Satunnaismuuttuja W on **Waldin testisuure** nollahypoteesille H_0 .

Jos nollahypoteesi H_0 ei päde, Waldin testisuureella W on taipumus saada *suuria* arvoja.

Waldin testisuureelle W pätee tiettyjen (hyvin lievien) ehtojen vallitessa ja *nollahypoteesin* H_0 pätiessä seuraava *asymptoottinen jakaumatulos*:

$$W \underset{a}{\sim} \chi^2(m)$$

jossa

$$m = \text{nollahypoteesin kiinnittämien parametrien lukumäärä}$$

Testin *hylkäysalueeksi* valitaan kaikkien mahdollisten havaintopisteiden joukon S^n alue

$$\{\mathbf{x} \in S^n \mid W > \chi_\alpha^2(m)\}$$

jossa *kriittinen arvo* $\chi_\alpha^2(m)$ valitaan siten, että

$$\Pr(\mathbf{x} \in S^n \mid W > \chi_\alpha^2(m)) = \alpha$$

jos nollahypoteesi H_0 *pätee*. Jos Waldin testisuuren W arvo joutuu hylkäysalueelle, *nollahypoteesi* H_0 *hylätään*.

Waldin testi epälineaarisille rajoituksille

Olkoon

$\hat{\boldsymbol{\theta}}$ = parametrin $\boldsymbol{\theta}$ SU-estimaattori, joka on määrätty olettaen, että testin yleinen hypoteesi H pätee

Olkoon *nollahypoteesina*

$$H_0 : \mathbf{h}(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}), \dots, h_m(\boldsymbol{\theta})) = \mathbf{0}$$

jossa funktiot h_i ovat epälineaarisia reaaliarvoisia funktioita.

Määritellään $p \times m$ -matriisi

$$\mathbf{H} = [h_{ij}] = \left[\frac{\partial}{\partial \theta_i} h_j(\boldsymbol{\theta}) \right]$$

Määritellään satunnaismuuttuja

$$W = [\mathbf{h}(\hat{\boldsymbol{\theta}})]' [\hat{\mathbf{H}} \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}) \hat{\mathbf{H}}]^{-1} [\mathbf{h}(\hat{\boldsymbol{\theta}})]'$$

jossa matriisi

$$\mathbf{I}(\boldsymbol{\theta}) = -E[\mathbf{D}^2 l(\boldsymbol{\theta})]$$

on Fisherin informaatiomatriisi ja

$$\hat{\mathbf{H}} = \left[\frac{\partial h_j(\boldsymbol{\theta})}{\partial \theta_i} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

Satunnaismuuttuja W on **Waldin testisuure** nollahypoteesille H_0 .

Jos nollahypoteesi H_0 ei päde, Waldin testisuureella W on taipumus saada *suuria* arvoja.

Waldin testisuurelle W pätee tiettyjen (melko lievien) ehtojen vallitessa ja *nollahypoteesin* H_0 pätiessä seuraava *asymptoottinen jakaumatulos*:

$$W \sim_a \chi^2(m)$$

jossa

$$m = \text{nollahypoteesin kiinnittämien parametrien lukumäärä}$$

Testin *hylkäysalueeksi* valitaan kaikkien mahdollisten havaintopisteiden joukon S^n alue

$$\{\mathbf{x} \in S^n \mid W > \chi_\alpha^2(m)\}$$

jossa *kriittinen arvo* $\chi_\alpha^2(m)$ valitaan siten, että

$$\Pr(\mathbf{x} \in S^n \mid W > \chi_\alpha^2(m)) = \alpha$$

jos nollahypoteesi H_0 *pätee*. Jos Waldin testisuureen W arvo joutuu hylkäysalueelle, *nollahypoteesi* H_0 *hylätään*.

Lagrangen kertojatesti

Olkoon

$$\hat{\boldsymbol{\theta}}_0 = \text{parametrin } \boldsymbol{\theta} \text{ rajoitettu SU-estimaattori, joka on määrätty olettaen, että testin nollahypoteesi } H_0 \text{ pätee}$$

Määritellään satunnaismuuttuja

$$LM = [\mathbf{D}(\hat{\boldsymbol{\theta}}_0)]' \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_0) [\mathbf{D}(\hat{\boldsymbol{\theta}}_0)]$$

jossa matriisi

$$\mathbf{I}(\boldsymbol{\theta}) = -E[\mathbf{D}^2 l(\boldsymbol{\theta})] = E[(\mathbf{D}l(\boldsymbol{\theta}))(\mathbf{D}l(\boldsymbol{\theta}))']$$

on Fisherin informaatiomatriisi ja

D(θ)

on *tehokkaan pistemäärän* vektori. Satunnaismuuttuja LM on **Lagrangen kertojatestisuure** nollahypoteesille H_0 .

Lagrangen kertojatestisuurelle LM pätee tiettyjen (melko lievien) ehtojen vallitessa ja *nollahypoteesin* H_0 pätiessä seuraava *asymptoottinen jakaumatulos*:

$$LM \sim_a \chi^2(m)$$

jossa

$$m = \text{nollahypoteesin kiinnittämien parametrien lukumäärä}$$

Testin *hylkäysalueeksi* valitaan kaikkien mahdollisten havaintopisteiden joukon S^n alue

$$\{\mathbf{x} \in S^n \mid LM > \chi^2_\alpha(m)\}$$

jossa *kriittinen arvo* $\chi^2_\alpha(m)$ valitaan siten, että

$$\Pr(\mathbf{x} \in S^n \mid LM > \chi^2_\alpha(m)) = \alpha$$

jos nollahypoteesi H_0 *pätee*. Jos Lagrangen kertojatestisuureen LM arvo joutuu hylkäysalueelle, *nollahypoteesi* H_0 *hylätään*.

Lagrangen kertojatesti ja pienimmän neliösumman menetelmä

Oletetaan, että logaritminen uskottavuusfunktio

$$l(\theta) = \log L(\theta)$$

on *verrannollinen neliösummaan*:

$$l(\theta) \propto \frac{1}{2\sigma^2} \sum_{i=1}^n \varepsilon_i^2(\theta)$$

Tällöin parametrin θ *suurimman uskottavuuden estimaattori* voidaan määrätä *pienimmän neliösumman menetelmällä*. Tämä on tavanomainen tilanne sekä *lineaaristen* että *epälineaaristen regressiomallien* soveltamisen yhteydessä.

Muodostetaan *apuregressio*

$$\hat{\varepsilon}_{0i} = \hat{\mathbf{z}}'_{0i} \boldsymbol{\gamma} + \delta_i, i = 1, 2, \dots, n$$

jossa

$$\varepsilon_i = \varepsilon_i(\theta)$$

$$\mathbf{z}_i = \mathbf{z}_i(\theta) = -\mathbf{D}\varepsilon_i(\theta)$$

$$\hat{\varepsilon}_{0i} = \varepsilon_i(\hat{\theta}_0)$$

$$\hat{\mathbf{z}}_{0i} = \mathbf{z}_i(\hat{\theta}_0)$$

ja $\hat{\theta}_0$ on parametrin θ *rajoitettu suurimman uskottavuuden estimaattori*, joka on siis määrätty maksimoimalla (logaritminen) uskottavuusfunktio, kun nollahypoteesin H_0 asettamat rajoitukset parametriavaruudelle on otettu huomioon.

Lagrangen kertojatestisuure LM voidaan määrätä tämän apuregression *selitysasteen* avulla:

$$LM = nR_0^2$$

Kuvattu tapa muodostaa LM-testisuure on erityisen kätevä silloin, kun nollahypoteesin mukaiset rajoitukset ovat *nollarajoituksia*, jotka yksinkertaistavat yleisen hypoteesin kiinnittämää mallia.

Toinen tapa käyttää yo. apuregressiota hyväksi on testata tavanomaisella *lineaaristen mallien* (ks. lukua 8) yhteydessä sovellettavalla *F-testillä* nollahypoteesia

$$H_0 : \boldsymbol{\gamma} = \mathbf{0}$$

Näin saatava testi on *asymptoottisesti ekvivalentti* ym. apuregression selitysasteeseen perustavan testin kanssa. Kirjallisuudessa esitettyjen tietojen mukaan LM-testin *F*-testimuoto toimii kuitenkin *pienissä otoksissa* paremmin kuin selitysasteeseen perustuva muoto.

Monet tilastollisten mallien *diagnostisista testeistä* voidaan tulkita LM-testeiksi. Tällaisia ovat esimerkiksi tavanomaiset testit regressiomallien jäännöstermien *homoskedastisuudelle*, *autokorreloimattomuudelle* ja *normaalisuudelle* tai ARMA-mallien jäännöstermien *autokorreloimattomuudelle*.

Osamäärätestin, Waldin testin ja Lagrangen kertojatestin vertailua

Kuten edellä on todettu *osamäärätestisuuretta* muodostettaessa joudutaan mallin parametrit estimoimaan (ts. uskottavuusfunktio tai sen logaritmi maksimoimaan) sekä yleisen hypoteesin H että nollahypoteesin H_0 pätiessä. Monissa testausasetelmissä toinen näistä ääriarvotehtävistä on *vaikea*, mutta toinen on *helppo* ratkaista.

Juuri tämä havainto on johtanut *Waldin testin* ja *Lagrangen kertojatestin* konstruointiin:

- (i) Waldin testissä mallin parametrit estimoidaan olettaen, että yleinen hypoteesi H pätee.
- (ii) Lagrangen kertojatestissä mallin parametrit estimoidaan olettaen, että nollahypoteesi H_0 pätee.

Oletetaan, että sovellamme osamäärätestiä, Waldin testiä ja Lagrangen kertojatestiä *samassa testausasetelmassa* (so. saman nollahypoteesin testaamiseen).

Olkoon

$$LR = \text{osamäärätestisuure}$$

$$W = \text{Waldin testisuure}$$

$$LM = \text{Lagrangen testisuure}$$

Aina pätee:

$$0 \leq LR \leq 1$$

Jos nollahypoteesi H_0 pätee, niin

$$\left. \begin{array}{l} -2 \log LR \\ W \\ LM \end{array} \right\} \sim_a \chi^2(m)$$

jossa

$m =$ nollahypoteesin H_0 kiinnittämien parametrien lukumäärä

Siten osamäärätesti, Waldin testi ja Lagrangen kertojatesti johtavat *asymptoottisesti* (suurissa otoksissa) samaan johtopäätökseen nollahypoteesin hylkäämisestä. *Pienissä otoksissa* testit saattavat kuitenkin johtaa erilaisiin johtopäätöksiin.

Testi valitaan yleensä käytännöllisten aspektien, kuten laskutyön helppouden, perusteella. Juuri tästä syystä monet tilastollisten mallien käytön yhteydessä sovellettavista *diagnostisista testeistä* ovat *Lagrangen kertojatestejä*.

7.4. Numeerinen optimointi

Funktion minimointi

Oletetaan, että haluamme *minimoida reaaliarvoisen kriteerifunktion*

$$g(\boldsymbol{\theta})$$

muuttujan $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ *suhteen.*

Funktio $g(\boldsymbol{\theta})$ saavuttaa (tietyin ehdoin) *lokaalin minimin* pisteessä $\hat{\boldsymbol{\theta}}$, jos

$$\mathbf{D}g(\hat{\boldsymbol{\theta}}) = 0$$

ja

$$\mathbf{D}^2g(\hat{\boldsymbol{\theta}}) > 0$$

jossa

$$\mathbf{D}g(\hat{\boldsymbol{\theta}}) = \text{funktion } g(\boldsymbol{\theta}) \text{ gradientti pisteessä } \hat{\boldsymbol{\theta}}$$

$$\mathbf{D}^2g(\hat{\boldsymbol{\theta}}) = \text{funktion } g(\boldsymbol{\theta}) \text{ Hessen matriisi pisteessä } \hat{\boldsymbol{\theta}}$$

Funktion $g(\boldsymbol{\theta})$ minimi voidaan pyrkiä määräämään ratkaisemalla *normaaliyhtälö*

$$\mathbf{D}g(\boldsymbol{\theta}) = 0$$

Jos normaaliyhtälön ratkaiseminen ei onnistu *suljetussa muodossa*, funktion $g(\boldsymbol{\theta})$ minimoinnissa on turvaututtava *numeerisiin optimointimenetelmiin*.

Huomautus:

Monissa tilastollisissa ongelmissa normaaliyhtälön ratkaiseminen suljetussa muodossa on mahdollista. Tästä on tärkeänä esimerkkinä yleisen lineaarisen mallin pienimmän neliösumman estimoinnissa syntyvän normaaliyhtälön ratkaiseminen.

Funktion minimointi iteratiivisten menetelmien avulla

Tilanteessa, jossa normaaliyhtälöä

$$\mathbf{D}g(\boldsymbol{\theta}) = 0$$

ei pystytä ratkaisemaan suljetussa muodossa, funktion $g(\boldsymbol{\theta})$ minimin antava piste voidaan pyrkiä löytämään *iteratiivisesti algoritmilla*

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \lambda \mathbf{d}(\hat{\boldsymbol{\theta}}_t)$$

jossa

$$\hat{\boldsymbol{\theta}}_t = \text{minimipisteen approksimaatio iteraatioaskeleessa } t$$

$$\lambda = \text{askelpituus}$$

$$\mathbf{d}(\hat{\boldsymbol{\theta}}_t) = \text{suuntavektorin arvo pisteessä } \hat{\boldsymbol{\theta}}_t$$

$$\hat{\boldsymbol{\theta}}_{t+1} = \text{minimipisteen approksimaatio iteraatioaskeleessa } t+1$$

Iteraatioaskeleessa t saatu arvio $\hat{\boldsymbol{\theta}}_t$ minimin paikasta päivitetään askeleessa $t+1$ ottamalla λ :n mittainen askel suuntavektorin $\mathbf{d}(\hat{\boldsymbol{\theta}}_t)$ suuntaan. Jos suuntavektorit $\mathbf{d}(\hat{\boldsymbol{\theta}}_t)$ valitaan sopivasti (ja funktiolla $g(\boldsymbol{\theta})$ on sopivat ominaisuudet) kuvattu iteraatioprosessi konvergoi kohti funktion $g(\boldsymbol{\theta})$ minimiä. Erilaiset valinnat suuntavektoriksi $\mathbf{d}(\hat{\boldsymbol{\theta}}_t)$ johtavat erilaisiin minimointialgoritmeihin.

Optimaalisen askelpituuden λ määrittämisessä käytetään tavallisesti jotakin erillistä hakumenetelmää. Nämä hakumenetelmät perustuvat tavallisesti funktion

$$\phi(\lambda) = g(\boldsymbol{\theta}_t + \lambda \mathbf{d}(\boldsymbol{\theta}_t))$$

minimointiin askelpituuden λ suhteen.

Jyrkimmän laskun menetelmä

Jyrkimmän laskun menetelmä perustuu *iteraatioon*

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t - \lambda \mathbf{D}(\hat{\boldsymbol{\theta}}_t)$$

jossa

$$\mathbf{D}(\hat{\boldsymbol{\theta}}_t) = \text{funktion } g(\boldsymbol{\theta}) \text{ gradientti pisteessä } \hat{\boldsymbol{\theta}}_t$$

Jyrkimmän laskun menetelmä perustuu siihen, että *funktio vähenee nopeimmin gradienttivektorin vastavektorin suuntaan.*

Koska jyrkimmän laskun menetelmä konvergoi usein hyvin hitaasti, sitä ei tavallisesti käytetä ääriarvojen etsintään. Se muodostaa kuitenkin perustan useille kohdefunktion *derivaattoihin perustuville optimointimenetelmille.*

Newtonin ja Raphsonin algoritmi

Newtonin ja Raphsonin algoritmi perustuu *iteraatioon*

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t - \lambda [\mathbf{D}^2(\hat{\boldsymbol{\theta}}_t)]^{-1} \mathbf{D}(\hat{\boldsymbol{\theta}}_t)$$

jossa

$$\mathbf{D}(\hat{\boldsymbol{\theta}}_t) = \text{funktion } g(\boldsymbol{\theta}) \text{ gradientti pisteessä } \hat{\boldsymbol{\theta}}_t$$

$$\mathbf{D}^2(\hat{\boldsymbol{\theta}}_t) = \text{funktion } g(\boldsymbol{\theta}) \text{ Hessen matriisi pisteessä } \hat{\boldsymbol{\theta}}_t$$

Newtonin ja Raphsonin algoritmi *konvergoi*, jos minimoitava funktio $g(\boldsymbol{\theta})$ on *kupera alaspäin*, minkä takaa se, että matriisi

$$\mathbf{D}^2(\boldsymbol{\theta}_t) > 0$$

kaikille θ . Newtonin ja Raphsonin algoritmi konvergoi yleensä paljon nopeammin kuin jyrkimmän laskun algoritmi.

Newtonin ja Raphsonin algoritmista on kehitetty muunnoksia, joissa ei haittaa, vaikka funktio $g(\theta)$ ei olisikaan kaikkialla kupera alaspäin, kunhan se on kupera alaspäin minimipisteen lähellä. Ehkä tunnetuin tällainen algoritmi (jossa Hessen matriisin $\mathbf{D}^2(\theta_t)$ diagonaalille lisätään sopiva, iteraatioaskeleesta toiseen muuttuva vakio) on *Marquardt'in algoritmi*.

Newtonin ja Raphsonin algoritmi ja suurimman uskottavuuden estimointi

Olkoot

$$X_1, X_2, \dots, X_n$$

joukko satunnaismuuttujia, joiden yhteisjakauman pistetodennäköisyys- tai tiheysfunktio

$$f(x_1, x_2, \dots, x_n; \theta)$$

riippuu parametrasta θ .

Otoksen X_1, X_2, \dots, X_n uskottavuusfunktion

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta)$$

maksimi voidaan löytää etsimällä *logaritmisen uskottavuusfunktion vastaluvun*

$$-l(\theta) = -\log L(\theta) = g(\theta)$$

minimi. Tällöin *Newtonin ja Raphsonin algoritmi* saa muodon

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \lambda[\mathbf{D}^2 l(\hat{\theta}_t)]^{-1} \mathbf{D}l(\hat{\theta}_t)$$

Olkoon

$$\theta_{Max} = \theta_{Max}(x_1, x_2, \dots, x_n)$$

uskottavuusfunktion maksimin antava piste. Tällöin parametrin θ *suurimman uskottavuuden estimaattori* on

$$\hat{\theta}_{Max} = \theta_{Max}(X_1, X_2, \dots, X_n)$$

Huomaa, että estimaattorin $\hat{\theta}$ *kovarianssimatriisina* voidaan käyttää matriisiä

$$-[\mathbf{D}^2 l(\hat{\theta})]^{-1}$$

mikä saadaan sivutuotteena Newtonin ja Raphsonin algoritmista.

Scoring-algoritmi ja suurimman uskottavuuden estimointi

Korvataan *Newtonin ja Raphsonin algoritmista* matriisi

$$-\mathbf{D}^2 l(\theta)$$

Fisherin informaatiomatriisilla $\mathbf{I}(\theta)$:

$$\mathbf{I}(\theta) = -E[\mathbf{D}^2 l(\theta)] = E[(\mathbf{D}l(\theta))(\mathbf{D}l(\theta))']$$

Scoring-algoritmi perustuu *iteraatioon*

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \lambda [\mathbf{I}(\hat{\boldsymbol{\theta}}_t)]^{-1} \mathbf{D}l(\hat{\boldsymbol{\theta}}_t)$$

jossa

$$\mathbf{D}l(\hat{\boldsymbol{\theta}}_t) = \text{funktion } g(\boldsymbol{\theta}) \text{ gradientti pisteessä } \hat{\boldsymbol{\theta}}_t$$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_t) = \text{Fisherin informaatiomatriisi pisteessä } \hat{\boldsymbol{\theta}}_t$$

Scoring-algoritmi konvergoi yleensä nopeammin kuin Newtonin ja Raphsonin algoritmi, mikä johtuu siitä, että matriisi

$$\mathbf{I}(\boldsymbol{\theta}) = -E[\mathbf{D}^2l(\boldsymbol{\theta})]$$

on parametrin $\boldsymbol{\theta}$ funktiona tavallisesti paljon yksinkertaisempaa muotoa kuin matriisi

$$\mathbf{D}^2l(\boldsymbol{\theta})$$

ja vaatii siten vähemmän prosessointia.

Huomaa, että estimaattorin $\hat{\boldsymbol{\theta}}$ kovarianssimatriisina voidaan käyttää matriisia

$$[\mathbf{I}(\hat{\boldsymbol{\theta}})]^{-1}$$

mikä saadaan sivutuotteena Scoring-algoritmista.

Gaussin ja Newtonin algoritmi ja pienimmän neliösumman estimointi

Parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattori voidaan monissa tilanteissa määrätä pienimmän neliösumman menetelmällä minimoimalla neliösummaa

$$f(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) = \sum_{j=1}^n \varepsilon_j^2(\boldsymbol{\theta})$$

Tällöin Newtonin ja Raphsonin algoritmi voidaan muokata ns. Gaussin ja Newtonin algoritmiksi.

Gaussin ja Newtonin algoritmi perustuu *iteraatioon*

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t - \lambda \left[\sum_{j=1}^n \frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}} \cdot \frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}'} \right]^{-1} \sum_{j=1}^n \frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}} \varepsilon_j(\hat{\boldsymbol{\theta}}_t)$$

Otetaan käyttöön merkinnät

$$\mathbf{z}_j = -\frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}}$$

$$\varepsilon_j(\hat{\boldsymbol{\theta}}_t) = \varepsilon_j$$

Tällöin Gaussin ja Newtonin algoritmi saa muodon

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \lambda \left[\sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j' \right]^{-1} \sum_{j=1}^n \mathbf{z}_j \varepsilon_j$$

mikä voidaan tulkita sarjaksi peräkkäisiä *apuregressioita*

$$\varepsilon_t = \mathbf{z}'_t \boldsymbol{\gamma} + \delta_t, t = 1, 2, \dots, n$$

joiden avulla minimipisteen approksimaatiota $\hat{\boldsymbol{\theta}}_t$ päivitetään. Tämä nähdään siitä, että apu-regression regressiokertoimien $\boldsymbol{\gamma}$ pienimmän neliösumman estimaattori \mathbf{c} on muotoa

$$\mathbf{c} = \left[\sum_{j=1}^n \mathbf{z}_j \mathbf{z}'_j \right]^{-1} \sum_{j=1}^n \mathbf{z}_j \varepsilon_j$$

Huomaa, että estimaattorin $\hat{\boldsymbol{\theta}}$ kovarianssimatriisina voidaan käyttää matriisiä

$$\left[\sum_{j=1}^n \mathbf{z}_j \mathbf{z}'_j \right]^{-1} = \left[\sum_{j=1}^n \frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}} \cdot \frac{\partial \varepsilon_j(\hat{\boldsymbol{\theta}}_t)}{\partial \boldsymbol{\theta}'} \right]^{-1}$$

mikä saadaan sivutuotteena Gaussin ja Newtonin algoritmista.