Tilastollinen päättely

7. Suurimman uskottavuuden menetelmä ja asymptoottinen teoria

7.1. Suurimman uskottavuuden estimointimenetelmä

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

7.2. Suurimman uskottavuuden estimaattorin 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, Tyhjentävyys, Uskottavuusfunktio, Yhteisjakauma

7.3. Asymptoottiset testit

Asymptoottinen normaalisuus, Derivaatta, Diagnostinen testi, Estimaattori, Estimointi, Havainto, Havaintoarvo, Hessen matriisi, Hylkäysalue, Khi^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, Tyhjentävyys, 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.

7.5. Yleinen lineaarinen malli ja suurimman uskottavuuden menetelmä

Estimointi, *F*-jakauma, *F*-testi, Homoskedastisuus, Jakaumaoletus, Jäännösneliösumma, Jäännöstermi, Jäännösvarianssi, Kokonaisneliösumma, Korrelaatio, Korreloimattomuus, Kovarianssi, Kriittinen arvo, Lineaarinen regressiomalli, Luottamuskerroin, Luottamustaso, Luottamusväli, Mallineliösumma, Merkitsevyystaso, Nollahypoteesi, Normaalijakauma, Otos, Osamäärätesti, Parametri, Pienimmän neliösumman menetelmä, Regressiofunktio, Regressio-kerroin, Regressiomalli, Regressiotaso, Residuaali, Selitettävä muuttuja, Selittäjä, Selittävä muuttuja, Selitysaste, Sovite, Suurimman uskottavuuden menetelmä, *t*-jakauma, *t*-testi, Testi, Vakiotermi, Vapausasteet, Varianssianalyysihajotelma, Yleinen lineaarinen malli

7.1. Suurimman uskottavuuden estimointimenetelmä

Otos ja sen jakauma

Olkoot

$$X_1, X_2, \ldots, X_n$$

satunnaismuuttujia, joiden *yhteisjakauman pistetodennäköisyys-* tai *tiheysfunktio*

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

riippuu tuntemattomasta vektoriarvoisesta *parametrista*

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta$$

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

Oletetaan, että satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *havaitut arvot* ovat

$$x_1, x_2, \ldots, x_n$$

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

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

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

Havaintoarvot

$$x_1, x_2, \ldots, x_n$$

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

Uskottavuusfunktio

Otoksen

$$X_1, X_2, \ldots, X_n$$

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

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

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

Merkitsemme uskottavuusfunktiota seuraavalla tavalla:

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

Uskottavuusperiaatteen mukaan uskottavuusfunktio tiivistää yhteen kaiken olennaisen parametria $\boldsymbol{\theta}$ koskevan informaation otoksesta.

Suurimman uskottavuuden estimaattori

Olkoon

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

otoksen $X_1, X_2, \ldots, X_n$ *uskottavuusfunktio* ja olkoon

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

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

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

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

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

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

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

$$x_1, x_2, \ldots, x_n$$

mahdollisimman uskottaviksi.

Parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattori määrätään siis *maksimoimalla* uskottavuusfunktio

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

parametrin $\boldsymbol{\theta}$ suhteen. Säännöllisissä tapauksissa maksimi löydetään merkitsemällä uskottavuusfunktion $L(\boldsymbol{\theta})$ *derivaatta*

$$L'(\boldsymbol{\theta})$$

nollaksi ja ratkaisemalla $\boldsymbol{\theta}$ saadusta *normaaliyhtälöstä*

$$L'(\boldsymbol{\theta}) = 0$$

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

$$L''(\mathbf{t}) < 0$$

Uskottavuusfunktion maksimin antava parametrin $\boldsymbol{\theta}$ arvo voidaan etsiä myös maksimoimalla **logaritminen uskottavuusfunktio** eli **uskottavuusfunktion logaritmi**

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

parametrin $\boldsymbol{\theta}$ suhteen. Logaritmisen uskottavuusfunktion maksimointi on monissa tilanteissa helpompaa kuin uskottavuusfunktion itsensä maksimointi.

Huomautus:

Logaritminen uskottavuusfunktio ja uskottavuusfunktio saavuttavat maksiminsa *samassa pisteessä*.

Uskottavuusfunktio ja riippumattomat havainnot

Olkoon

$$X_1, X_2, \ldots, X_n$$

yksinkertaisen satunnaisotos jakaumasta $f_X(x; \boldsymbol{\theta})$, jossa $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ on jakauman muodon määräävä parametri.

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

$$X_1, X_2, \mathsf{K}, X_n \perp$$
$$X_i \quad f_X(x; \boldsymbol{\theta}), i = 1, 2, \mathsf{K}, n$$

Koska satunnaismuuttujat $X_1, X_2, \ldots, X_n$ oletettiin riippumattomiksi, otoksen $X_1, X_2, \ldots, X_n$ *uskottavuusfunktio* voidaan esittää muodossa

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

jossa

$$f(x_i; \boldsymbol{\theta}), \; i = 1, 2, \ldots, n$$

on havaintoarvoon x_i liittyvä pistetodennäköisyys- tai tiheysfunktio.

Siten vastaava *logaritminen uskottavuusfunktio* voidaan esittää muodossa

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\ &= \log\left(f(x_1; \boldsymbol{\theta}) \times f(x_2; \boldsymbol{\theta}) \times \mathsf{L} \times f(x_n; \boldsymbol{\theta}) \right) \\ &= \log f(x_1; \boldsymbol{\theta}) + \log f(x_2; \boldsymbol{\theta}) + \mathsf{L} + \log f(x_n; \boldsymbol{\theta}) \\ &= l(\boldsymbol{\theta}; x_1) + l(\boldsymbol{\theta}; x_2) + \mathsf{L} + l(\boldsymbol{\theta}; x_n) \end{aligned}$$

jossa

$$l(\boldsymbol{\theta}; x_i) = \log f(\boldsymbol{\theta}; x_i), \; i = 1, 2, \ldots, n$$

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

Suurimman uskottavuuden estimaattorin ominaisuudet

Suurimman uskottavuuden estimaattori ei välttämättä toteuta tavanomaisia hyvälle estimaattorille asetettavia kriteereitä, kun otoskoko on *äärellinen*:

(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 normaalinen*.

SU-estimaattorin tarkentuvuus tarkoittaa sitä, että SU-estimaattori toteuttaa *suurten lukujen lain*. Tämä merkitsee sitä, että estimaattorin arvo *konvergoi kohti "oikeata" parametrin arvoa*, kun havaintojen lukumäärän annetaan kasvaa rajatta.

SU-estimaattorin asymptoottinen normaalisuus tarkoittaa sitä, että SU-estimaattori toteuttaa *keskeisen raja-arvolauseen*. Tämä merkitsee se sitä, että 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 normaalisuus, 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*. Siten stokastisia prosesseja käytetään aikasarjojen tilastollisina malleina. Aikasarjamallien parametrien SU-estimaattoreiden *tarkentuvuus* ja *asymptoottinen normaalisuus voidaan todistaa* esimerkiksi silloin, kun aikasarjan generoinut stokastinen prosessi on jotakin seuraavista tyypeistä:

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

7.2. Suurimman uskottavuuden estimaattorin asymptoottiset ominaisuudet

1. kertaluvun osittaisderivaattojen vektori ja
2. kertaluvun osittaisderivaattojen matriisi

Olkoon

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

p-vektori.

Tällöin

$$\mathbf{D} = \left(\frac{\partial}{\partial \theta_1}, \mathrm{K}, \frac{\partial}{\partial \theta_p} \right) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \mathbb{M} \\ \frac{\partial}{\partial \theta_p} \end{bmatrix}$$

on 1. *kertaluvun osittaisderivaattaoperaattoreiden*

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

muodostama p-vektori ja

$$\mathbf{D}^2 = \mathbf{D}\mathbf{D}' = \begin{bmatrix} \frac{\partial^2}{\partial \theta_1^2} & \mathsf{L} & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} \\ \mathbb{M} & & \mathbb{M} \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} & \mathsf{L} & \frac{\partial^2}{\partial \theta_p^2} \end{bmatrix}$$

on 2. *kertaluvun osittaisderivaattaoperaattoreiden*

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

muodostama $p \times p$-matriisi.

Otos ja sen jakauma

Olkoot

$$X_1, X_2, \ldots, X_n$$

satunnaismuuttujia, joiden *yhteisjakauman pistetodennäköisyys-* tai *tiheysfunktio*

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

riippuu parametrista

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

Olkoot satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *havaitut arvot*

$$x_1, x_2, \ldots, x_n$$

Uskottavuusfunktio

Otoksen

$$X_1, X_2, \ldots, X_n$$

uskottavuusfunktio

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

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

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

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

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

vastaava **logaritminen uskottavuusfunktio**.

Suurimman uskottavuuden estimaattori

Olkoon

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

otoksen $X_1, X_2, \ldots, X_n$ *uskottavuusfunktio* ja olkoon

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

sellainen parametrin $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ arvo, joka maksimoi uskottavuusfunktion $L(\boldsymbol{\theta})$ arvon. Uskottavuusfunktion $L(\boldsymbol{\theta})$ maksimin antava parametrin $\boldsymbol{\theta}$ arvo $\mathbf{t}$ on havaintoarvojen $x_1, x_2, \ldots, x_n$ funktio.

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

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

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

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

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

Huomautus:

Merkintä

$$\mathbf{A} > 0$$

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

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

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

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

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

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

kaikille $\mathbf{z}$.

Tehokas pistemäärä ja Fisherin informaatiomatriisi

Vektoria

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

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

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

on 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

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

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

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

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

sanomme, että estimaattori $\hat{\boldsymbol{\theta}}$ on **tehokas**.

Suurimman uskottavuuden estimaattori ja tyhjentävyys

Olkoon

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

otoksen $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ yhteisjakauman pistetodennäköisyys- tai tiheysfunktio, joka riippuu parametrista $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. *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 $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ 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 parametrista $\boldsymbol{\theta}$.

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 $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. Tällöin estimaattori $\hat{\boldsymbol{\theta}}$ yhtyy parametrin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaattoriin.

Suurimman uskottavuuden estimaattorin tarkentuvuus

Estimaattoreiden stokastisia ominaisuuksia ei useinkaan tunneta *äärellisissä otoksissa* tai niiden ominaisuudet ovat ainakin vaikeasti hallittavia, 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 kohti "oikeata" parametrin arvoa*, kun havaintojen lukumäärän annetaan kasvaa rajatta.

Tunnuslukujen

$$W_n = W_n(X_1, X_2, \mathrm{K}, X_n),\ n = 1, 2, 3, \mathrm{K}$$

jono muodostaa **tarkentuvan jonon estimaattoreita** parametrille θ, jos

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

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

Ekvivalentti ehto estimaattorin W_n tarkentuvuudelle on, että

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

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

Tavallisesti tarkentuvuuden todistamisessa ei tarvitse vedota tarkentuvuuden määritelmään, sillä tarkentuvuuden todistamisessa voidaan usein vedota *Tshebyshevin epäyhtälöön*. **Tshebyshevin epäyhtälön** mukaan

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

Siten estimaattori W_n on *tarkentuva* parametrille θ, jos

$$\lim_{n\to\infty} \mathrm{E}_\theta[(W_n - \theta)^2] = 0$$

Koska

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

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

$$\lim_{n\to\infty} \mathrm{Var}_\theta(W_n) = 0$$

ja

$$\lim_{n\to\infty} \mathrm{Bias}_\theta(W_n) = 0$$

Tarkentuvien estimaattoreiden jonot eivät ole yksikäsitteisiä. Tämä nähdään seuraavasta:

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

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

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

Tällöin myös tunnuslukujen jono

$$U_n = a_n W_n + b_n,\ n = 1, 2, 3, \mathrm{K}$$

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 $\tau(\theta)$:

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

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

Olkoon tarkastelun kohteena olevana parametrina *p-vektori*

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

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

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

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

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

tarkentuvuuden kautta.

Suurimman uskottavuuden estimaattorin asymptoottinen tehokkuus ja normaalisuus

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

Voisi tuntua luontevalta määritellä estimaattorin suurten otosten varianssi estimaattorin varianssin raja-arvona, kun otoskoon annetaan kasvaa rajatta.

Olkoon tunnusluku T_n estimaattori parametrille θ ja olkoon

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

jono ei-satunnaisia vakioita. Jos

$$\lim_{n\to\infty} k_n \operatorname{Var}(T_n) = \tau^2 < \infty$$

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

Olkoon tunnusluku T_n estimaattori parametrille θ ja olkoon

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

jono ei-satunnaisia vakioita. Jos

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

jakaumakonvergenssin mielessä, niin sanomme, että τ^2 on tunnusluvun T_n **asymptoottinen varianssin**. Tämä merkitsee sitä, että tunnusluvun T_n rajajakaumana on otoskoon kasvaessa rajatta

normaalijakauma ja τ^2 on tämän rajajakauman varianssi. Monissa tilanteissa tunnusluvun asymptoottinen varianssi on samalla sen varianssin raja-arvo.

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

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

jakaumakonvergenssin mielessä ja

$$v(\theta) = \frac{[\tau'(\theta)]^2}{\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]}$$

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

Seuraavan lauseen mukaan *suurten uskottavuuden estimaattori* on *asymptoottisesti normaalinen* 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 normaalinen* ja *tehokas estimaattori* parametrille $\tau(\theta)$:
>
> $$\lim_{n\to\infty} \sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to \mathrm{N}(0, v(\theta))$$
>
> jossa $v(\theta)$ on Cramérin ja Raon alaraja.

Olkoon tarkastelun kohteena olevana parametrina nyt *p-vektori*

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

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

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

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

$$\hat{\boldsymbol{\theta}} \;\; {}_a \;\; \mathrm{N}_p\left(\boldsymbol{\theta}, \frac{1}{n}[\mathbf{IA}(\boldsymbol{\theta})]^{-1}\right)$$

jossa

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

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

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

$$\theta_i \text{ , } i = 1, 2, \ldots, p$$

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

(i) Parametreille

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

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

(ii) Parametrien

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

arvoja koskeville nollahypoteeseille voidaan konstruoida *testit* samaan tapaan kuin normaalijakauman odotusarvolle. Testien hylkäysalueet tai testisuureen arvoja vastaavat *p*-arvot voidaan tällöin määrätä normaalijakaumasta tai *t*-jakaumasta.

On syytä huomata, että SU-estimaattorin normaalisuutta koskeva tulos on luonteeltaan *asymptoottinen*, mutta otokset ovat todellisuudessa kooltaan aina *äärellisiä*. Siten SU-estimaattorin normaalisuuteen perustuvat *luottamusvälit* ja *testit* eivät ole *eksakteja*, vaan *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 normaalijakauman käyttö:

(i) Oletetaan, että konstruoimme luottamusvälit tarkastelun kohteena olevalle parametrille θ *samalla* luottamustasolla $(1-\alpha)$ sekä *t*-jakaumasta että normaalijakaumasta. Tällöin *t*-jakaumaan perustuva luottamusväli on *leveämpi* kuin vastaava normaalijakaumaan perustuva luottamusväli.

(ii) Oletetaan, että konstruoimme testit tarkastelun kohteena olevaa parametrin θ arvoa koskevalle nollahypoteesille määräämällä testien hyväksymisalueet *samalla* merkitsevyystasolla α sekä *t*-jakaumasta että normaalijakaumasta. Tällöin *t*-jakaumaan perustuva hyväksymisalue on *suurempi* kuin vastaava normaalijakaumaan perustuva hyväksymisalue.

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

7.3. Asymptoottiset testit

Testiprobleema

Olkoot

S^n = kaikkien mahdollisten havaintopisteiden $\mathbf{x} = (x_1, x_2, \ldots , x_n)$ 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 hylkäämme*. Testin päätössääntö jakaa kaikkien mahdollisten havaintopisteiden joukon S^n **hyväksymisalueeseen** ja **hylkäysalueeseen**.

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

Testin merkitsevyystason valinta tarkoittaa 1. *lajin virheen todennäköisyyden kiinnittämistä etukäteen*, ennen testin tekemistä. Jos merkitsevyystaso 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

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

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

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

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

Tällöin satunnaismuuttuja

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

on (uskottavuus-) **osamäärätestisuure** nollahypoteesille H_0.

Huomaa, että

$$0 \le \lambda \le 1$$

Osamäärätestisuuretta 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 ääriarvotehtävän* ratkaisemista.

(ii) Osamäärätestisuureen *osoittaja* saadaan etsimällä uskottavuusfunktion maksimi *nolla-hypoteesin* H_0 *pätiessä*. Tämä merkitsee *sidotun ääriarvotehtävän* ratkaisemista.

Jos nollahypoteesi H_0 *ei päde*, osamäärätestisuureella λ on taipumus saada *pieniä* arvoja.

Siksi testin *hylkäysalueeksi* valitaan otosavaruuden χ alue

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

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

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

jos nollahypoteesi H_0 *pätee*. Jos osamäärätestisuureen λ arvo joutuu hylkäysalueelle, *nolla-hypoteesi* 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 *nolla-hypoteesin* H_0 *pätiessä* seuraava *asymptoottinen jakaumatulos*:

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

jossa

m = 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 maksimoimaan sekä yleisen hypoteesin H että nollahypoteesin H_0 pätiessä. Monissa testausasetelmissa toinen näistä ääriarvotehtävistä 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

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

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 = nollahypoteesin kiinnittämien parametrien lukumäärä

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

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

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

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

jos nollahypoteesi H_0 *pätee*. Jos Waldin testisuureen 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*

$$\mathrm{H}_0 : \mathbf{h}(\boldsymbol{\theta}) = \left(h_1(\boldsymbol{\theta}), \mathrm{K}, h_m(\boldsymbol{\theta})\right) = \mathbf{0}$$

jossa funktiot h_i ovat epälineaarisia reaaliarvoisia funktioita.

Määritellään $p \times m$-matriisi

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

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

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

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

$$W \;\; {}_a \;\; \chi^2(m)$$

jossa

m = nollahypoteesin kiinnittämien parametrien lukumäärä

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

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

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

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

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

Lagrangen kertojatesti

Olkoon

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

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

on *Fisherin informaatiomatriisi* ja

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

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

Lagrangen kertojatestisuureelle LM pätee tiettyjen (hyvin lievien) ehtojen vallitessa ja *nolla-hypoteesin* H_0 *pätiessä* seuraava *asymptoottinen jakaumatulos*:

$$LM \;\; {}_a \;\; \chi^2(m)$$

jossa

m = 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

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

on *verrannollinen neliösummaan*:

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

Tällöin parametrin $\boldsymbol{\theta}$ *suurimman uskottavuuden estimaattori* voidaan määrätä *pienimmän neliösumman menetelmällä*. Tämä on tavanomainen tilanne sekä *lineaaristen* että *epälineaaristen regressiomallien* soveltamisen yhteydessä.

Muodostetaan *apuregressio*

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

jossa

$$\varepsilon_i = \varepsilon_i(\boldsymbol{\theta})$$
$$\mathbf{z}_i = \mathbf{z}_i(\boldsymbol{\theta}) = -\mathbf{D}\varepsilon_i(\boldsymbol{\theta})$$
$$\hat{\varepsilon}_{0i} = \varepsilon_i(\hat{\boldsymbol{\theta}}_0)$$
$$\hat{\mathbf{z}}_{0i} = \mathbf{z}_i(\hat{\boldsymbol{\theta}}_0)$$

ja $\hat{\boldsymbol{\theta}}_0$ on parametrin $\boldsymbol{\theta}$ *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.

Monet tilastollisten mallien *diagnostiset testit* voidaan tulkita LM-testeiksi. Tällaisia ovat esimerkiksi *homoskedastisuus*- ja *autokorreloimattomuustestit* regressiomalleille.

Osamäärätestin, Waldin testin ja Lagrangen kertojatestin vertailua

Osamäärätestisuuretta muodostettaessa joudutaan mallin parametrit estimoimaan (ts. uskottavuus-funktio tai sen logaritmi maksimoimaan) sekä yleisen hypoteesin H että nollahypoteesin H_0 pätiessä. Monissa testausasetelmissa toinen näistä ääriarvotehtävistä on *vaikea*, mutta toinen on *helppo* ratkaista.

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* (saman nollahypoteesin testaamiseen).

Olkoon

LR = osamäärätestisuure

W = Waldin testisuure

LM = Lagrangen testisuure

Aina pätee*:*

$$0 \le LR \le 1$$

Jos *nollahypoteesi* H_0 *pätee*, niin

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

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.

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

7.4. Numeerinen optimointi

Funktion minimointi

Oletetaan, että haluamme *minimoida reaaliarvoisen kriteerifunktion*

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

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

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

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

ja

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

jossa

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

$\mathbf{D}^2 g(\hat{\boldsymbol{\theta}})$ = funktion $g(\boldsymbol{\theta})$ *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 turvauduttava *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

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

λ = *askelpituus*

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

$\hat{\boldsymbol{\theta}}_{t+1}$ = 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 $\mathbf{d}(\boldsymbol{\theta}_t)$ suuntaan. Jos suuntavektorit $\mathbf{d}(\boldsymbol{\theta}_t)$ valitaan sopivasti (ja funktiolla $g(\boldsymbol{\theta})$ on sopivat ominaisuudet) iteraatioprosessi konvergoi kohti funktion $g(\boldsymbol{\theta})$ minimiä. Erilaiset valinnat suuntavektoriksi $\mathbf{d}(\boldsymbol{\theta}_t)$ johtavat *erilaisiin* minimointialgoritmeihin.

Optimaalisen askelpituuden λ määräämisessä käytetään tavallisesti jotakin erillistä haku-menetelmää. 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

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

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

Koska jyrkimmän laskun menetelmä konvergoi yleensä hitaasti, sitä ei tavallisesti käytetä funktion ääriarvojen etsintään, vaikka se muodostaakin perustan kaikille funktion derivaattoja käyttäville 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

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

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

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

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

kaikille $\boldsymbol{\theta}$. Newtonin ja Raphsonin algoritmi konvergoi yleensä nopeammin kuin jyrkimmän laskun algoritmi.

On hyvä tietää, että Newtonin ja Raphsonin algoritmista on kehitetty muunnoksia, joissa ei haittaa, jos funktio $g(\boldsymbol{\theta})$ ei ole kaikkialla konkaavi, kunhan se on konkaavi minimipisteen lähellä. Ehkä tunnetuin tällainen algoritmi (jossa matriisin $\mathbf{D}^2(\boldsymbol{\theta}_t)$ diagonaalille lisätään sopiva, iteraatio-askeleesta toiseen vaihteleva vakio) on *Marquardtin algoritmi*.

Newtonin ja Raphsonin algoritmi ja suurimman uskottavuuden estimointi

Olkoot

$$X_1, X_2, \ldots, X_n$$

joukko satunnaismuuttujia, joiden yhteisjakauman pistetodennäköisyys- tai tiheysfunktio

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

riippuu parametrista $\boldsymbol{\theta}$.

Otoksen $X_1, X_2, \ldots, X_n$ uskottavuusfunktion

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

maksimi voidaan löytää etsimällä *logaritmisen uskottavuusfunktion vastaluvun*

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

minimi. Tällöin *Newtonin ja Raphsonin algoritmi* saa muodon

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

Olkoon

$$\boldsymbol{\theta}_{Max} = \boldsymbol{\theta}_{Max}(x_1, x_2, \mathsf{K}, x_n)$$

uskottavuusfunktion maksimin antava piste. Tällöin parametrin $\boldsymbol{\theta}$ *suurimman uskottavuuden estimaattori* on

$$\hat{\boldsymbol{\theta}}_{Max} = \boldsymbol{\theta}_{Max}(X_1, X_2, \mathsf{K}, X_n)$$

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

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

mikä saadaan sivutuotteena Newtonin ja Raphsonin algoritmista.

Scoring-algoritmi ja suurimman uskottavuuden estimointi

Korvataan *Newtonin ja Raphsonin algoritmissa* matriisi

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

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

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

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

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

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

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

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

on parametrin $\boldsymbol{\theta}$ funktiona tavallisesti yksinkertaisempi kuin matriisi

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

ja vaatii siten vähemmän laskutyötä.

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*

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

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*

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

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 matriisia

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

mikä saadaan sivutuotteena Gaussin ja Newtonin algoritmista.

7.5. Yleinen lineaarinen malli ja suurimman uskottavuuden menetelmä

Yleinen lineaarinen malli ja sen parametrointi

Olkoon

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk} + \varepsilon_j \,,\ j = 1, 2, \ldots, n$$

yleinen lineaarinen malli, jossa

y_j = **selitettävän muuttujan** *y havaittu* ja *satunnainen arvo* havaintoyksikössä $j = 1, 2, \ldots, n$

x_{ji} = **selittävän muuttujan** eli **selittäjän** x_i *havaittu* ja *kiinteä* eli *ei-satunnainen arvo* havaintoyksikössä $j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, k$

β_0 = vakioselittäjän *regressiokerroin, tuntematon* ja *kiinteä* eli *ei-satunnainen vakio*

β_i = selittäjän x_i *regressiokerroin, tuntematon* ja *kiinteä* eli *ei-satunnainen vakio*, $i = 1, 2, \ldots, k$

ε_j = *ei-havaittava* ja *satunnainen* **jäännöstermi**, $j = 1, 2, \ldots, n$

Yleistä lineaarista mallia koskevat standardioletukset

(i) Selittäjän x_i havaitut arvot x_{ji} ovat *kiinteitä* eli *ei-satunnaisia vakioita*, $j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, k$

(ii) Selittäjien välillä ei ole *lineaarisia riippuvuuksia*.

(iii) $\mathrm{E}(\varepsilon_j) = 0 \,,\ j = 1, 2, \ldots, n$

(iv) $\mathrm{Var}(\varepsilon_j) = \sigma^2 \,,\ j = 1, 2, \ldots, n$

(v) $\mathrm{Cor}(\varepsilon_j, \varepsilon_l) = 0 \,,\ j \neq l$

(vi) $\varepsilon_j \sim \mathrm{N}(0, \sigma^2) \,,\ j = 1, 2, \ldots, n$

Oletusta (iv) on tapana kutsua *homoskedastisuusoletukseksi* ja sen mukaan kaikilla jäännöstermeillä $\varepsilon_j \,,\ j = 1, 2, \ldots, n$ on sama varianssi, jota kutsutaan *jäännösvarianssiksi*. Oletusta (v) on tapana kutsua *korreloimattomuusoletukseksi* ja sen mukaan jäännöstermit eivät korreloi keskenään.

Huomaa, että normaalisuusoletus (vi) sisältää oletukset (iv) ja (v).

Regressiotaso

Yhtälö

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

määrittelee *tason* avaruudessa $^{k+1}$. Tätä tasoa kutsutaan **regressiotasoksi**.

Odotusarvovektori ja kovarianssimatriisi

Olkoot

$$x_1, x_2, \ldots, x_p$$

satunnaismuuttujia, joiden *odotusarvot* ovat

$$\mathrm{E}(x_i) = \mu_i \,,\ i = 1, 2, \ldots, p$$

ja *kovarianssit* ovat

$$\begin{aligned}\mathrm{Cov}(x_i,x_j) &= \mathrm{E}[(x_i - \mathrm{E}(x_i))(x_j - \mathrm{E}(x_j))] \\ &= \mathrm{E}[(x_i - \mu_i)(x_j - \mu_j)] \\ &= \sigma_{ij} \,,\ i = 1, 2, \ldots, p, j = 1, 2, \ldots, p\end{aligned}$$

Huomaa, että

$$\mathrm{Cov}(x_i,x_j) = \mathrm{E}(x_i x_j) - \mu_i \mu_j$$

ja

$$\mathrm{Cov}(x_i,x_i) = \sigma_{ii} = \mathrm{Var}(x_i) = \sigma_i^2$$

Olkoon

$$\mathbf{x} = (x_1, x_2, \mathsf{K}, x_p) = \begin{bmatrix} x_1 \\ x_2 \\ \mathsf{M} \\ x_p \end{bmatrix}$$

satunnaismuuttujien $x_1, x_2, \ldots, x_p$ muodostama p-vektori.

Tällöin satunnaisvektorin $\mathbf{x}$ *odotusarvovektori* on p-vektori

$$\mathrm{E}(\mathbf{x}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mathsf{M} \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

ja satunnaisvektorin $\mathbf{x}$ *kovarianssimatriisi* on $p \times p$-matriisi

$$\mathrm{Cov}(\mathbf{x}) = [\sigma_{ij}] = \boldsymbol{\Sigma}$$

Huomaa, että

$$\begin{aligned}\mathrm{Cov}(\mathbf{x}) &= \mathrm{E}[(\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{x} - \mathrm{E}(\mathbf{x}))'] \\ &= \mathrm{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] \\ &= \mathrm{E}(\mathbf{x}\mathbf{x}') - \boldsymbol{\mu}\boldsymbol{\mu}'\end{aligned}$$

Yleisen lineaarisen mallin matriisiesitys

Muodostetaan *selitettävän muuttujan y* havaituista arvoista y_j (j = 1, 2, … , n) n-vektori

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Muodostetaan *selittäjien* x_i havaituista arvoista x_{ji} (j = 1, 2, … , n; i = 1, 2, … , k) ja ykkösistä $n\times(k+1)$-matriisi

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

Muodostetaan *regressiokertoimista* β_i (i = 0, 1, 2, … , k) (k+1)-vektori

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Muodostetaan *jäännöstermeistä* ε_j (j = 0, 1, 2, … , n) n-vektori

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Tällöin *yleinen lineaarinen malli* voidaan kirjoittaa *matriisein* muotoon

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Mallissa on seuraavat osat:

- $\mathbf{y}$ = *selitettävän muuttujan y havaittujen arvojen* y_j , j = 1, 2, … , n muodostama *satunnainen* n-vektori
- $\mathbf{X}$ = *selittävien muuttujien* eli *selittäjien* x_1, x_2, … , x_k *havaittujen arvojen* x_{ji} , j = 1, 2, … , n , i = 1, 2, … , k ja *ykkösten* muodostama *ei-satunnainen* $n\times(k + 1)$-matriisi
- $\boldsymbol{\beta}$ = *regressiokertoimien* β_i , i = 0, 1, 2, … , k muodostama *tuntematon* ja *kiinteä* eli *ei-satunnainen* (k + 1)-vektori
- $\boldsymbol{\varepsilon}$ = *jäännöstermien* ε_j , j = 1, 2, … , n muodostama *ei-havaittava* ja *satunnainen* n-vektori

Yleistä lineaarista mallia koskevat standardioletukset matriisimuodossa

(i) Matriisin $\mathbf{X}$ alkiot ovat *ei-satunnaisia vakioita*.

(ii) Matriisi $\mathbf{X}$ on *täysiasteinen*:

$$r(\mathbf{X}) = k + 1$$

(iii) $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$

(iv)&(v) *Homoskedastisuus-* ja *korreloimattomuusoletus*:

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

(vi) *Normaalisuusoletus*:

$$\boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$$

Kohdissa (iv), (v) ja (vi) oleva matriisi $\mathbf{I}$ on *n×n-yksikkömatriisi*.

Standardioletuksista (i) ja (iii) seuraa, että

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

Standardioletuksista (i), (iii), (iv) ja (v) seuraa, että

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}) &= \mathrm{E}[(\mathbf{y} - \mathrm{E}(\mathbf{y}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))'] \\ &= \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \\ &= \mathrm{Cov}(\boldsymbol{\varepsilon}) \\ &= \sigma^2\mathbf{I}\end{aligned}$$

Normaalisuusoletuksesta (vi) seuraa, että jäännöstermin $\boldsymbol{\varepsilon}$ tiheysfunktio on muotoa

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n; \sigma^2) = (2\pi)^{-\frac{1}{2}n}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right\}$$

ja edelleen, että selitettävän muuttujan $\mathbf{y}$ tiheysfunktio on

$$f(y_1, y_2, \ldots, y_n; \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-\frac{1}{2}n}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Yleisen lineaarisen mallin uskottavuusfunktio ja logaritminen uskottavuusfunktio

Standardioletuksesta (vi) seuraa, että havaintojen $y_1, y_2, \ldots, y_n$ *uskottavuusfunktio* on

$$L(\boldsymbol{\beta}, \sigma^2; y_1, y_2, \ldots, y_n) = (2\pi\sigma^2)^{-\frac{1}{2}n}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Vastaava *logaritminen uskottavuusfunktio* on

$$\begin{aligned}l(\boldsymbol{\beta}, \sigma^2; y_1, y_2, \ldots, y_n) &= \log L(\boldsymbol{\beta}, \sigma^2; y_1, y_2, \ldots, y_n) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\end{aligned}$$

Yleisen lineaarisen mallin parametrien suurimman uskottavuuden estimointi

Yleisen lineaarisen mallin

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

regressiokertoimien vektorin $\boldsymbol{\beta}$ *suurimman uskottavuuden estimaattori* yhtyy standardioletuksien (i)-(vi) pätiessä vektorin $\boldsymbol{\beta}$ *pienimmän neliösumman estimaattoriin*

$$\hat{\boldsymbol{\beta}} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Jäännösvarianssin σ^2 SU-estimaattori on

$$\hat{\sigma}^2 = \frac{1}{n} SSE$$

$$SSE = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Perustelu:

Maksimoidaan logaritminen uskottavuusfunktio

$$l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

parametrien $\boldsymbol{\beta}$ ja σ^2 suhteen.

Derivoidaan funktio $l(\boldsymbol{\beta}, \sigma^2; \mathbf{y})$ regressiokertoimien vektorin $\boldsymbol{\beta}$ suhteen ja merkitään derivaatta nollaksi:

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2}(-2\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Vektori $\boldsymbol{\beta}$ voidaan ratkaista tästä yhtälöstä, koska matriisi $\mathbf{X}$ oletettiin täysiasteiseksi, jolloin matriisi $\mathbf{X}'\mathbf{X}$ on *epäsingulaarinen*. Ratkaisuksi saadaan

$$\hat{\boldsymbol{\beta}} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Ratkaisu vastaa uskottavuusfunktion maksimia, koska matriisi

$$-\frac{\partial l(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \frac{1}{2\sigma^2}\mathbf{X}'\mathbf{X}$$

on *positiivisesti definiitti*.

Estimaattori $\mathbf{b}$ yhtyy regressiokertoimien vektorin $\boldsymbol{\beta}$ *pienimmän neliösumman estimaattoriin*, koska logaritmisen uskottavuusfunktion maksimointi parametrin $\boldsymbol{\beta}$ suhteen on selvästi ekvivalenttia sen kanssa, että minimoidaan neliömuoto

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

parametrin $\boldsymbol{\beta}$ suhteen.

Sijoitetaan ratkaisu $\hat{\boldsymbol{\beta}} = \mathbf{b}$ logaritmiseen uskottavuusfunktion lausekkeeseen, derivoidaan saatava lauseke parametrin σ^2 suhteen ja merkitään derivaatta nollaksi:

$$\frac{\partial l(\mathbf{b},\sigma^2;\mathbf{y})}{\partial \sigma^2} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) = 0$$

Parametri σ^2 voidaan ratkaista tästä yhtälöstä.

Ratkaisuksi saadaan

$$\hat{\sigma}^2 = \frac{1}{n}SSE$$

jossa

$$SSE = (\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb})$$

■

Estimoitu regressiotaso

Olkoon

$$\mathbf{b} = (b_0, b_1, b_2, \ldots, b_k)$$

regressiokertoimien vektorin $\boldsymbol{\beta}$ *suurimman uskottavuuden* (*pienimmän neliösumman*) *estimaattori*.

Yhtälö

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

määrittelee *tason* avaruudessa $^{k+1}$. Tätä tasoa kutsutaan **estimoiduksi regressiotasoksi**.

Yleisen lineaarisen mallin suurimman uskottavuuden estimaattoreiden ominaisuudet

Jos yleinen lineaarinen malli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

toteuttaa standardioletukset (i)-(vi), niin regressiokertoimien vektorin $\boldsymbol{\beta}$ SU-estimaattorilla

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

on seuraavat ominaisuudet:

(i) $\mathbf{b}$ on *harhaton* parametrille $\boldsymbol{\beta}$:

$$\mathrm{E}(\mathbf{b}) = \boldsymbol{\beta}$$

(ii) $\mathrm{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

(iii) $\mathbf{b}$ ja $\hat{\sigma}^2$ ovat yhdessä *tyhjentäviä* parametreille $\boldsymbol{\beta}$ ja σ^2.

(iv) $\mathbf{b}$ on *tehokas* eli *minimivarianssinen* estimaattori.

(v) $\mathbf{b}$ on *tarkentuva*.

(vi) $\mathbf{b}$ noudattaa *normaalijakaumaa*:

$$\mathbf{b} \sim \mathrm{N}_{k+1}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Jos yleinen lineaarinen malli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

toteuttaa standardioletukset (i)-(vi), niin jäännösvarianssin σ^2 SU-estimaattorilla

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

on seuraavat ominaisuudet:

(i) $\hat{\sigma}^2$ on *harhainen*, mutta estimaattori

$$s^2 = \frac{n-1}{n-k-1}\hat{\sigma}^2 = \frac{SSE}{n-k-1}$$

on *harhaton* parametrille σ^2.

(ii) $\mathbf{b}$ ja $\hat{\sigma}^2$ ovat yhdessä *tyhjentäviä* parametreille $\boldsymbol{\beta}$ ja σ^2.

(iii) $\hat{\sigma}^2$ ei ole *tehokas* eli *minimivarianssinen* estimaattori.

(iv) $\hat{\sigma}^2$ on *tarkentuva*.

(v) $(n-k-1)s^2/\sigma^2$ noudattaa χ^2*-jakaumaa* vapaustein $(n-k-1)$:

$$\frac{(n-k-1)s^2}{\sigma^2} \quad \chi^2(n-k-1)$$

Gaussin ja Markovin lause

Olkoot $\hat{\boldsymbol{\theta}}_1$ ja $\hat{\boldsymbol{\theta}}_2$ kaksi *harhatonta* estimaattoria parametrille $\boldsymbol{\theta}_1 = (\theta_1, \theta_2, \mathrm{K}, \theta_p)$. Sanomme, että estimaattori $\hat{\boldsymbol{\theta}}_1$ on **parempi** eli **tehokkaampi** kuin estimaattori $\hat{\boldsymbol{\theta}}_2$, jos

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}_1) \le \mathrm{Cov}(\hat{\boldsymbol{\theta}}_2)$$

jolla tarkoitetaan sitä, että

$$\mathbf{t}'\mathrm{Cov}(\hat{\boldsymbol{\theta}}_1)\mathbf{t} \le \mathbf{t}'\mathrm{Cov}(\hat{\boldsymbol{\theta}}_2)\mathbf{t}$$

kaikille $\mathbf{t} \in$ ^{p}.

Tärkeä perustelu *suurimman uskottavuuden* (*pienimmän neliösumman*) *menetelmän* käytölle yleisen lineaarisen mallin regressiokertoimien estimoimiseen on seuraava lause:

Gaussin ja Markovin lause:

Oletetaan, että yleinen lineaarinen malli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

toteuttaa standardioletukset (i)-(v). Tällöin regressiokertoimien vektorin $\boldsymbol{\beta}$ *suurimman uskottavuuden* (*pienimmän neliösumman*) *estimaattori*

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

on *paras* regressiokertoimien vektorin $\boldsymbol{\beta}$ *lineaaristen ja harhattomien estimaattoreiden joukossa*.

Perustelu:

Olkoon

$$\mathbf{c} = \mathbf{D}^*\mathbf{y}$$

jokin standardioletukset toteuttavan yleisen lineaarisen mallin

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

regressiokertoimien vektorin $\boldsymbol{\beta}$ *lineaarinen* ja *harhaton* estimaattori. Matriisi $\mathbf{D}^*$ on $(k+1)\times n$-matriisi, joka ei riipu vektorista **y**.

Olkoon

$$\mathbf{D} = \mathbf{D}^* - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Siten

$$\begin{aligned}\mathbf{c} &= \mathbf{D}^*\mathbf{y}\\ &= (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}\\ &= (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\\ &= (\mathbf{D}\mathbf{X} + \mathbf{I})\boldsymbol{\beta} + (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\varepsilon}\end{aligned}$$

Koska

$$\begin{aligned}\mathrm{E}(\mathbf{c}) &= (\mathbf{D}\mathbf{X} + \mathbf{I})\boldsymbol{\beta} + (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\,\mathrm{E}(\boldsymbol{\varepsilon})\\ &= (\mathbf{D}\mathbf{X} + \mathbf{I})\boldsymbol{\beta} \qquad\qquad | \ \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}\end{aligned}$$

niin estimaattori **c** on *harhaton* vain, jos

$$\mathbf{D}\mathbf{X} = \mathbf{0}$$

Siten

$$\begin{aligned}\mathbf{c} &= \mathbf{D}^*\mathbf{y}\\ &= \boldsymbol{\beta} + (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\varepsilon}\end{aligned}$$

Tarkastellaan nyt estimaattorin **c** *kovarianssimatriisia*:

$$\begin{aligned}\mathrm{Cov}(\mathbf{c}) &= \mathrm{E}[(\mathbf{c} - \mathrm{E}(\mathbf{c}))(\mathbf{c} - \mathrm{E}(\mathbf{c}))']\\ &= \mathrm{E}[(\mathbf{c} - \boldsymbol{\beta})(\mathbf{c} - \boldsymbol{\beta})']\\ &= \mathrm{E}[(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')']\\ &= (\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\,\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')(\mathbf{D}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})\\ &= \sigma^2[\mathbf{D}\mathbf{D}' + \mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}' + (\mathbf{X}'\mathbf{X})^{-1}] \qquad | \ \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}\\ &= \sigma^2[\mathbf{D}\mathbf{D}' + (\mathbf{X}'\mathbf{X})^{-1}] \qquad\qquad | \ \mathbf{D}\mathbf{X} = \mathbf{0}\end{aligned}$$

Koska

$$\begin{aligned}\mathbf{t}'\mathrm{Cov}(\mathbf{c})\mathbf{t} &= \sigma^2[t'\mathbf{D}\mathbf{D}'t + t'(\mathbf{X}'\mathbf{X})^{-1}t]\\ &\geq \sigma^2[t'(\mathbf{X}'\mathbf{X})^{-1}t] \qquad | \ t'\mathbf{D}\mathbf{D}'t \geq 0\\ &= \mathbf{t}'\mathrm{Cov}(\mathbf{b})\mathbf{t}\end{aligned}$$

niin

$$\mathrm{Cov}(\mathbf{c}) \geq \mathrm{Cov}(\mathbf{b})$$

ja siten suurimman uskottavuuden (pienimmän neliösumman) estimaattori **b** on *paras* regressiokertoimien vektorin $\boldsymbol{\beta}$ *lineaaristen ja harhattomien estimaattoreiden joukossa*.

■

Sovite ja residuaali

Olkoon

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

yleisen lineaarisen mallin

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

regressiokertoimien vektorin $\boldsymbol{\beta}$ SU-estimaattori.

Määritellään estimoidun mallin **sovite** $\hat{\mathbf{y}}$ kaavalla

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\mathbf{b} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{M}\mathbf{y}\end{aligned}$$

jossa

$$\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Matriisi **M** on *projektio* (*symmetrinen* ja *idempotentti*):

$$\mathbf{M}' = \mathbf{M}$$

$$\mathbf{M}^2 = \mathbf{M}$$

Määritellään estimoidun mallin **residuaali e** kaavalla

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{X}\mathbf{b} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{I} - \mathbf{M})\mathbf{y} \\ &= \mathbf{P}\mathbf{y}\end{aligned}$$

jossa

$$\mathbf{P} = \mathbf{I} - \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Matriisi **P** on *projektio* (*symmetrinen* ja *idempotentti*):

$$\mathbf{P}' = \mathbf{P}$$

$$\mathbf{P}^2 = \mathbf{P}$$

Lisäksi projektiot **M** ja **P** ovat *ortogonaalisia* sekä sarakkeittensa että riviensä suhteen:

$$\mathbf{M}\mathbf{P} = \mathbf{P}\mathbf{M} = \mathbf{0}$$

Sovitteen ja residuaalin ominaisuudet

Sovitteella $\hat{\mathbf{y}}$ on seuraavat *stokastiset ominaisuudet*:

$$\mathrm{E}(\hat{\mathbf{y}}) = \mathbf{X}\boldsymbol{\beta}$$
$$\mathrm{Cov}(\hat{\mathbf{y}}) = \sigma^2\mathbf{M} = \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Lisäksi, jos jäännöstermi $\boldsymbol{\varepsilon}$ on normaalijakautunut, niin

$$\hat{\mathbf{y}} \quad \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{M})$$

Sovitteella **e** on seuraavat *stokastiset ominaisuudet*:

$$\mathrm{E}(\mathbf{e}) = \mathbf{0}$$
$$\mathrm{Cov}(\mathbf{e}) = \sigma^2\mathbf{P} = \sigma^2(\mathbf{I} - \mathbf{M}) = \sigma^2(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

Lisäksi, jos jäännöstermi $\boldsymbol{\varepsilon}$ on normaalijakautunut, niin

$$\mathbf{e} \quad \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{P})$$

Varianssianalyysihajotelma

Olkoon

$$SST = (\mathbf{y} - \bar{y}\mathbf{1})'(\mathbf{y} - \bar{y}\mathbf{1}) = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

selitettävän muuttujan *y* havaittujen arvojen $y_1, y_2, \ldots, y_n$ **kokonaisneliösumma**, jossa

$$\mathbf{y} = (y_1, y_2, \mathrm{K}, y_n)$$

on havaintoarvojen $y_1, y_2, \ldots, y_n$ muodostama *n*-vektori,

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

on havaintoarvojen $y_1, y_2, \ldots, y_n$ aritmeettinen keskiarvo ja

$$\mathbf{1} = (1,1, \mathrm{K}, 1)$$

on ykkösten muodostama *n*-vektori. Kokonaisneliösumma *SST* kuvaa selitettävän muuttujan *y* havaittujen arvojen $y_1, y_2, \ldots, y_n$ vaihtelua.

Huomaa, että selitettävän muuttujan *y* havaittujen arvojen $y_1, y_2, \ldots, y_n$ otosvarianssi saadaan kaavalla

$$s_y^2 = \frac{1}{n-1}SST$$

Määritellään **residuaalien neliösumma** eli **jäännösneliösumma** *SSE* kaavalla

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \sum_{i=1}^{n} e_i^2$$

jossa

$$\mathbf{e} = (e_1, e_2, \mathrm{K}, e_n)$$

on residuaalien muodostama *n*-vektori. Jäännösneliösumma *SSE* kuvaa residuaalien $e_1, e_2, \ldots, e_n$ vaihtelua.

Huomaa, että jäännösvarianssin σ^2 harhaton estimaattori saadaan kaavalla

$$s^2 = \frac{1}{n-k-1} SSE$$

Voidaan osoittaa, että aina pätee

$$SSE \leq SST$$

Tarkastellaan erotusta

$$SSM = SST - SSE \geq 0$$

Voidaan osoittaa, että

$$SSM = (\hat{\mathbf{y}} - \bar{y}\mathbf{1})'(\hat{\mathbf{y}} - \bar{y}\mathbf{1}) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

jossa

$$\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \mathsf{K}, \hat{y}_n)$$

on sovitteiden $\hat{y}_1, \hat{y}_2, \mathsf{K}, \hat{y}_n$ muodostama *n*-vektori,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

on havaintoarvojen $y_1, y_2, \ldots, y_n$ aritmeettinen keskiarvo ja

$$\mathbf{1} = (1,1, \mathsf{K}, 1)$$

on ykkösten muodostama *n*-vektori. Koska erotus *SSM* = *SST* – *SSE* voidaan esittää neliösummana, erotusta *SSM* kutsutaan **mallineliösummaksi**.

Huomaa, että

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i = \bar{\hat{y}}$$

Edellä esitetyn nojalla kokonaisneliösumma *SST* voidaan hajottaa neliösummien *SSM* ja *SSE* summaksi:

$$SST = SSM + SSE$$

Kokonaisneliösumman *SST* hajotelmaa mallineliösumman *SSM* ja jäännösneliösumman *SSE* summaksi kutsutaan **varianssianalyysihajotelmaksi**.

On ilmeistä, että mitä pienempi on jäännösneliösumman *SSE* = **e´e** osuus kokonaisneliösummasta *SST*, sitä suurempi on mallineliösumman *SSM* osuus ja sitä paremmin malli selittää selitettävän muuttujan havaittujen arvojen vaihtelun.

Selityaste ja sen ominaisuudet

Määritellään estimoidun lineaarisen regressiomallin **selitysaste** R^2 kaavalla

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST}$$

Koska *varianssianalyysihajotelman* mukaan

$$SST = SSM + SSE$$

niin

$$0 \le R^2 \le 1$$

Voidaan osoittaa, että seuraavat ehdot ovat *yhtäpitäviä*:

(i) $R^2 = 1$

(ii) Kaikki residuaalit häviävät:

$e_j = 0$ kaikille $j = 1, 2, \ldots, n$

(iii) Kaikki havaintopisteet

$$(x_{j1}, x_{j2}, \mathrm{K}, x_{jk}, y_j),\ j = 1, 2, \mathrm{K}, n$$

asettuvat *samalle tasolle*.

(iv) Malli *selittää täydellisesti* selitettävän muuttujan y arvojen vaihtelun.

Edelleen voidaan osoittaa, että myös seuraavat ehdot ovat *yhtäpitäviä*:

(i) $R^2 = 0$

(ii) $b_1 = b_2 = \cdots = b_k = 0$

(iii) Malli *ei ollenkaan selitä* selitettävän muuttujan arvojen vaihtelua.

Siten selitysaste R^2 kuvaa *estimoidun mallin selittämää osuutta selitettävän muuttujan y arvojen vaihtelusta*.

Selitysaste ilmoitetaan tavallisesti prosentteina, jolloin sanotaan, että *estimoitu malli on selittänyt*

$$100 \times R^2\ \%$$

selitettävän muuttujan y arvojen vaihtelusta.

Voidaan osoittaa, että

$$R^2 = [\mathrm{Cor}(y, \hat{y})]^2$$

jossa

$$\mathrm{Cor}(y, \hat{y})$$

on selitettävän muuttujan y havaittujen arvojen $y_1, y_2, \ldots, y_n$ ja estimoidun mallin sovitteiden $\hat{y}_1, \hat{y}_2, \mathrm{K}, \hat{y}_n$ tavanomainen *otoskorrelaatiokerroin*.

PNS-estimaattoreiden varianssien estimointi

Jos standardioletukset (i)-(v) pätevät, regressiokertoimen

$$\beta_i\ ,\ i = 0,\ 1,\ 2,\ \ldots\ ,\ k$$

suurimman uskottavuuden estimaattorin b_i varianssin

$$\mathrm{D}^2(b_i) = \sigma_{b_i}^2 = \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}]_{i+1,i+1}$$

harhaton estimaattori on

$$\hat{\mathrm{D}}^2(b_i) = s^2[(\mathbf{X}'\mathbf{X})^{-1}]_{i+1,i+1}$$

jossa

$$s^2 = \frac{1}{n-1}SSE = \frac{1}{n-1}(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) = \frac{1}{n-1}\mathbf{e}'\mathbf{e} = \frac{1}{n-1}\sum_{i=1}^{n} e_i^2$$

on *jäännösvarianssin σ^2 harhaton estimaattori*.

Regressiokertoimien luottamusvälit

Valitaan luottamustasoksi $(1 - \alpha)$. Jos standardioletukset (i)-(vi) pätevät, niin regressiokertoimen

$$\beta_i\ ,\ i = 0,\ 1,\ 2,\ \ldots\ ,\ k$$

luottamusvälit ovat muotoa

$$b_i \pm t_{\alpha/2}\,\hat{\mathrm{D}}(b_i)\ ,\ i = 0,1,2,\mathrm{K}\ ,k$$

jossa

b_i = regressiokertoimen β_i *SU-estimaattori*

$\pm t_{\alpha/2}$ = luottamustasoa $(1 - \alpha)$ vastaavat *luottamuskertoimet t*-jakaumasta, jonka vapausasteiden lukumäärä on $(n - k - 1)$

$\hat{\mathrm{D}}^2(b_i)$ = regressiokertoimen β_i SU-estimaattorin *varianssin harhaton estimaattori*

Regressiokertoimen β_i luottamusväli voidaan helposti johtaa käyttämällä hyväksi sitä, että satunnaismuuttuja

$$t_i = \frac{b_i - \beta_i}{\hat{\mathrm{D}}(b_i)}\ ,\ i = 0,1,2,\mathrm{K}\ ,k$$

noudattaa *t*-jakaumaa vapausastein $(n - k - 1)$:

$$t_i \quad t(n-1)\ ,\ i = 1,2,\mathrm{K}\ ,n$$

mikä merkitsee sitä, että satunnaismuuttujan t_i jakauma ei riipu estimoinnin kohteena olevista parametreista ja siten satunnaismuuttuja t_i on *saranasuure*.

Regressiokertoimen β_i *luottamusväli*

$$b_i \pm t_{\alpha/2}\,\hat{\mathrm{D}}(b_i)\ ,\ i = 0,1,2,\mathrm{K}\ ,k$$

luottamustasolla $(1 - \alpha)$ *peittää regressiokertoimen β_i todellisen arvon todennäköisyydellä* $(1 - \alpha)$:

$$\Pr(b_i - t_{\alpha/2}\,\hat{\mathrm{D}}(b_i) \le \beta_i \le b_i + t_{\alpha/2}\,\hat{\mathrm{D}}(b_i)) = 1-\alpha$$

Jos otantaa toistetaan, niin luottamustason *frekvenssitulkinnan* mukaan otoksista konstruoiduista luottamusväleistä

$$100\times(1-\alpha)\ \%$$

peittää regressiokertoimen β_i todellisen arvon ja

$$100\times\alpha\,\%$$

väleistä ei peitä regressiokertoimen β_i todellista arvoa.

Lineaaristen rajoitusten testaus

Olkoon

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

standardioletukset (i)-(vi) toteuttava yleinen lineaarinen malli, jossa $\mathbf{X}$ on ei-satunnainen $n\times(k+1)$-matriisi, $n \ge k+1$, $r(\mathbf{X}) = k+1$.

Oletetaan, että regressiokertoimien vektorille $\boldsymbol{\beta}$ on asetettu *nollahypoteesi*

$$\mathrm{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

jossa $\mathbf{R}$ on ei-satunnainen $m\times(k+1)$-matriisi, $m \le k+1$, $r(\mathbf{R}) = m$. Nollahypoteesin H_0 mukaan regressiokertoimia *sitoo m lineaarisia rajoitusta* eli *side-ehtoa*. Side-ehtoa

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

kutsutaan usein *yleiseksi lineaariseksi hypoteesiksi*.

Regressiokertoimien vektorin $\boldsymbol{\beta}$ *rajoittamaton suurimman uskottavuuden* (*pienimmän neliösumman*) *estimaattori* on

$$\mathbf{b} = \mathbf{U}\mathbf{X}'\mathbf{y}$$

jossa

$$\mathbf{U} = (\mathbf{X}'\mathbf{X})^{-1}$$

Regressiokertoimien vektorin $\boldsymbol{\beta}$ *rajoitettu* eli *sidottu suurimman uskottavuuden* (*pienimmän neliösumman*) *estimaattori* on

$$\mathbf{b}_R = \mathbf{b} + \mathbf{U}\mathbf{R}'\mathbf{S}(\mathbf{r} - \mathbf{R}\mathbf{b})$$

jossa

$$\mathbf{b} = \mathbf{U}\mathbf{X}'\mathbf{y}$$

on vektorin $\boldsymbol{\beta}$ *rajoitettu suurimman uskottavuuden estimaattori*, matriisi $\mathbf{U}$ on kuten edellä ja

$$\mathbf{S} = (\mathbf{R}\mathbf{U}\mathbf{R}')^{-1}$$

Määritellään *jäännösneliösummat*

$$SSE = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$
$$SSE_R = (\mathbf{y} - \mathbf{X}\mathbf{b}_R)'(\mathbf{y} - \mathbf{X}\mathbf{b}_R)$$

Tällöin **osamäärätestisuure** *nollahypoteesille*

$$\mathrm{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

on muotoa

$$\lambda = \left(\frac{SSE}{SSE_R} \right)^{-\frac{n}{2}}$$

Osamäärätestisuureen λ jakaumaa ei tunneta, mutta osamäärätestisuureen *asymptoottista jakaumaa* koskevasta yleisestä tuloksesta seuraa, että *nollahypoteesin pätiessä*

$$-2\log\lambda = -2\log\left(\frac{SSE}{SSE_R} \right)^{-\frac{n}{2}} = n\log(SSE_R) - n\log(SSE) \quad {}_a\chi^2(m)$$

jossa

m = rajoitusten eli side-ehtojen lukumäärä

= rivien lukumäärä matriisissa **R**

Suuret testisuureen $-2\log\lambda$ arvot merkitsevät sitä, että *nollahypoteesi on asetettava kyseenalaiseksi*.

Osamäärätestisuureen λ jakaumaa ei siis tunneta, kun otoskoko on *äärellinen*. Sen sijaan testisuure

$$F = \frac{n-k-1}{m}\left(\frac{1}{\lambda^{\frac{2}{n}}} - 1 \right)$$

noudattaa eksaktisti *F-jakaumaa* vapausastein m ja $(n-k-1)$ nollahypoteesin

$$\mathrm{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

pätiessä:

$$F \quad {}_{\mathrm{H}_0} F(m, n-k-1)$$

Testisuure F voidaan kirjoittaa seuraaviin muotoihin:

$$F = \frac{n-k-1}{m} \cdot \frac{SSE_R - SSE}{SSE}$$
$$= \frac{(\mathbf{r}-\mathbf{Rb})'\mathbf{S}(\mathbf{r}-\mathbf{Rb})}{ms^2}$$

jossa

$$(n-k-1)s^2 = (\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}) = SSE$$

Testisuure F voidaan johtaa myös *Waldin testinä* tai *Lagrangen kertojatestinä*.

Testi regression olemassaololle

Olkoon testattavana nollahypoteesina

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \mathsf{L} = \beta_k = 0$$

Jos nollahypoteesi H_0 *pätee*, selitettävä muuttuja *y ei riipu* lineaarisesti yhdestäkään selittäjästä x_1, x_2, … , x_k. Sen sijaan, jos nollahypoteesi H_0 *ei päde*, selitettävä muuttuja *y riippuu* lineaarisesti ainakin yhdestä selittäjästä x_1, x_2, … , x_k.

Määritellään ***F*-testisuure**

$$\begin{aligned} F &= \frac{n-k-1}{k} \cdot \frac{SSM}{SSE} \\ &= \frac{n-k-1}{k} \cdot \frac{SST - SSE}{SSE} \\ &= \frac{n-k-1}{k} \cdot \frac{R^2}{1-R^2} \end{aligned}$$

jossa

SST = selitettävän muuttujan *y* havaittujen arvojen kokonaisvaihtelua kuvaava neliösumma

SSM = estimoidun mallin *mallineliösumma*

SSE = estimoidun mallin *jäännösneliösumma*

R^2 = estimoidun mallin *selitysaste*

Testisuure *F* vertaa toisiinsa *residuaalivarianssia*

$$s^2 = \frac{1}{n-k-1} SSE$$

ja *mallivarianssia*

$$s_M^2 = \frac{1}{k} SSM$$

Oletetaan, että standardioletukset (i)-(vi) pätevät. Tällöin testisuure *F* noudattaa *nollahypoteesin*

$$H_0 : \beta_1 = \beta_2 = \mathsf{L} = \beta_k = 0$$

pätiessä F-jakaumaa vapausastein *k* ja $(n - k - 1)$:

$$F \underset{H_0}{\sim} F(k, n-k-1)$$

Testisuureen *F normaaliarvo* eli odotusarvo nollahypoteesin H_0 pätiessä on approksimatiivisesti (suurille *n*) = 1:

$$E(F) \underset{H_0}{\approx} 1$$

Suuret testisuureen *F* arvot viittaavat siihen, että nollahypoteesi H_0 *ei päde*.

Testi on erikoistapaus edellä esitetystä testistä *yleiselle lineaarisille hypoteesille*. Testisuure *F* voidaan johtaa *osamäärätestinä*, *Waldin testinä* tai *Lagrangen kertojatestinä*.

Testit regressiokertoimille

Olkoon nollahypoteesina

$$H_{0i} : \beta_i = 0 \, , \; i = 0,1,2, \mathsf{K} \, , k$$

Jos nollahypoteesi H_{00} *pätee*, mallissa *ei ole vakiota*. Jos nollahypoteesi H_{0i} *pätee*, selitettävä muuttuja *y ei riipu* lineaarisesti selittäjästä x_i , $i = 1, 2, \ldots, k$. Jos nollahypoteesi H_{0i} *ei päde*, selitettävä muuttuja *y riippuu* lineaarisesti selittäjästä x_i , $i = 1, 2, \ldots, k$.

Määritellään ***t*-testisuureet**

$$t_i = \frac{b_i}{\hat{\mathrm{D}}(b_i)}$$

jossa

b_i = regressiokertoimen β_i *SU-estimaattori*

$\hat{\mathrm{D}}^2(b_i)$ = regressiokertoimen β_i SU-estimaattorin *varianssin harhaton estimaattori*

Oletetaan, että standardioletukset (i)-(vi) pätevät. Tällöin testisuure t_i noudattaa *nollahypoteesin*

$$H_{0i} : \beta_i = 0 , i = 0,1,2,\mathrm{K},k$$

pätiessä t-jakaumaa vapausastein $(n - k - 1)$:

$$t_i \underset{H_0}{\sim} t(n-k-1) , i = 1,2,\mathrm{K},n$$

Testisuureen t_i normaaliarvo eli odotusarvo nollahypoteesin H_{0i} pätiessä on

$$\mathrm{E}(t_i) =_{H_{0i}} 0 , i = 0,1,2,\mathrm{K},k$$

Itseisarvoltaan suuret testisuureen t_i arvot viittaavat siihen, että nollahypoteesi H_{0i} ei päde.

Testit ovat erikoistapauksia edellä esitetystä testistä *yleiselle lineaarisille hypoteesille*. Testisuureet t_i , $i = 1,2, \ldots, k$ voidaan johtaa *osamäärätesteinä, Waldin testeinä* tai *Lagrangen kertojatesteinä*.