

2.5 Identifitseeritavus ja hinnatavus

Juhusliku suuruse jaotuse parameeter θ on identifitseeritav, kui kahe parameetri θ võimaliku väärtuse θ_1 ja θ_2 korral järeldeb jaotusfunktsioonide võrdsusest parameetrite võrdsus:

$$F_{\theta_1}(x) = F_{\theta_2}(x) \quad \forall x \Rightarrow \theta_1 = \theta_2.$$

Liikudes mudelite ja valimite juurde võime öelda: parameetervektor θ on identifitseeritav, kui valimi tõepärafunktsioonide võrdsusest järeldeb parameetrite võrdsus:

$$L(\theta_1|\mathbf{x}) = L(\theta_2|\mathbf{x}) \quad \forall \mathbf{x} \Rightarrow \theta_1 = \theta_2.$$

Kui valime sellise parameetervektori, mis ei ole miskitmoodi uuritava tunnusega seotud, siis parameetervektor pole identifitseeritav (parameetervektorit, mis sisaldab kojamehe vanaema sünniaastat, ei saa identifitseerida kirja-ja-kulli viskamise teel — kirjade-kullide jaotus on ikka ühesugune, olgu vanaema sünniaasta siis 1923 või 1932).

Parameetervektori θ funktsioon $g(\theta)$ on identifitseeritav, kui valimi tõepärafunktsioonide võrdsusest järeldeb funktsioonide võrdsus, $L(\theta_1|\mathbf{x}) = L(\theta_2|\mathbf{x}) \Rightarrow g(\theta_1) = g(\theta_2)$.

Lineaarsete mudelite ja normaaljaotuse eelduse kontekstis on lihtne näha, et funktsioon $h(\beta)$ on identifitseeritav vaid siis, kui tegemist on $\mathbf{X}\beta$ funktsiooniga, st kui leidub funktsioon g , nii et $h(\beta) = g(\mathbf{X}\beta)$.

Näitame esmalt, et kui $h(\beta)$ on kirja pandav kui $\mathbf{X}\beta$ funktsioon, $h(\beta) = g(\mathbf{X}\beta)$, siis on tegemist parameetrite β identifitseeritava funktsiooniga.

Võrdusest $h(\beta) = g(\mathbf{X}\beta)$ järeldeb, et kui $\mathbf{X}\beta_1 = \mathbf{X}\beta_2$ siis $h(\beta_1) = h(\beta_2)$, sest $h(\beta_1) = g(\mathbf{X}\beta_1) = g(\mathbf{X}\beta_2) = h(\beta_2)$.

Vaatame, kas sellisel juhul on $h(\beta)$ identifitseeritav?

Selgub, et mistahes $\mathbf{X}\beta$ funktsioon on identifitseeritav, sest:

$$\begin{aligned} L(\beta_1) &= L(\beta_2) \quad \forall \mathbf{y} \\ c \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta_1)^T(\mathbf{y} - \mathbf{X}\beta_1)\right) &= c \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta_2)^T(\mathbf{y} - \mathbf{X}\beta_2)\right) \\ (\mathbf{y} - \mathbf{X}\beta_1)^T(\mathbf{y} - \mathbf{X}\beta_1) &= (\mathbf{y} - \mathbf{X}\beta_2)^T(\mathbf{y} - \mathbf{X}\beta_2) \end{aligned}$$

Valides $\mathbf{y} = \mathbf{X}\beta_1$, saame

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\beta_1)^T(\mathbf{y} - \mathbf{X}\beta_1) &= (\mathbf{y} - \mathbf{X}\beta_2)^T(\mathbf{y} - \mathbf{X}\beta_2) \\ 0 &= (\mathbf{X}\beta_1 - \mathbf{X}\beta_2)^T(\mathbf{X}\beta_1 - \mathbf{X}\beta_2) \\ \mathbf{X}\beta_1 &= \mathbf{X}\beta_2 \end{aligned}$$

Seega tõepärafunktsioonide võrdsusest järeldus, et $\mathbf{X}\beta_1 = \mathbf{X}\beta_2$. Kui aga $h(\beta)$ on $\mathbf{X}\beta$ funktsioon, siis järeldub sellest ka soovitud tulemus — $h(\beta_1) = h(\beta_2)$ ehk $h(\beta)$ identifitseeritavus.

Kui vaatleme funktsiooni $h(\beta)$ mis pole vaadeldav $\mathbf{X}\beta$ funktsioonina, st. kui leiduvad β_1 ja β_2 , nii et $\mathbf{X}\beta_1 = \mathbf{X}\beta_2$ aga $h(\beta_1) \neq h(\beta_2)$, siis pole vastav funktsioon ka identifitseeritav:

$$\mathbf{X}\beta_1 = \mathbf{X}\beta_2 \Rightarrow L(\beta_1) = L(\beta_2),$$

aga eelduse järgi $h(\beta_1) \neq h(\beta_2)$, seega $h(\beta)$ pole identifitseeritav.

Parameetervektor β on identifitseeritav, kui $\mathbf{X}^T\mathbf{X}$ on pööratav (siis $\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta = g(\mathbf{X}\beta)$).

Identifitseeritavus on seotud uuritava tunnuse \mathbf{y} jaotusega. Vahel soovitakse kasutada identifitseeritavusega sarnast mõistet ilma uuritava tunnuse jaotust mängu toomata.

Definitsioon 2.1 *Parameetrite lineaarne funktsioon $\lambda^T\beta$ on hinnatav, kui leidub selline vektor \mathbf{v} , et $E(\mathbf{v}^T\mathbf{y}) = \lambda^T\beta$ (mistahes β väärtuse korral).*

Teisisõnu — parameetrite β funktsioon $\lambda^T\beta$ on hinnatav, kui on võimalik leida lineaarset nihketa hinnangut $\lambda^T\beta$ -le.

On lihtne näha, et parameetrite lineaarse funktsiooni hinnatavus on normaaljaotuse eelduse kehtides samaväärne parameetrite lineaarse funktsiooni identifitseeritavusega. Kui parameetervektor $\lambda^T\beta$ on hinnatav, siis

$$\begin{aligned} E(\mathbf{v}^T\mathbf{y}) &= \lambda^T\beta \quad \forall \beta \\ \mathbf{v}^T\mathbf{X}\beta &= \lambda^T\beta \quad \forall \beta \\ \mathbf{v}^T\mathbf{X} &= \lambda^T \end{aligned}$$

ja järelikult on $\lambda^T\beta$ esitatav kui $\mathbf{X}\beta$ (lineaarne) funktsioon (mis on identifitseeritav, kui vaatlusvektor on normaaljaotusega).

Kui aga \mathbf{y} on normaaljaotusega ja parameetrite lineaarne funktsioon $\lambda^T\beta$ on identifitseeritav, siis $\lambda^T\beta = \mathbf{v}^T\mathbf{X}\beta = E(\mathbf{v}^T\mathbf{y})$ (sest $\lambda^T\beta$ peab olema $\mathbf{X}\beta$ funktsioon) ja seega on ta ka hinnatav.

Hinnatavuse (estimability) eeliseks on see, et tema kontrollimiseks pole tarvis teha eelduseid uuritava tunnuse jaotuse kohta.

Meenutuseks:

- $\lambda^T\beta$ on hinnatav $\Leftrightarrow \exists \mathbf{v}$, nii et $\mathbf{v}^T\mathbf{X} = \lambda^T$.
- Kui maatriks $\mathbf{X}^T\mathbf{X}$ on pööratav, siis on parameetervektor β ise hinnatav ning hinnatavad on ka kõik võimalikud parameetrite lineaarkombinatsioonid.

- Mitte kõik parameetrite identifitseeritavad funktsioonid (näiteks β_1/β_2) ei pruugi olla hinnatavad (sest β_1/β_2 pole parameetrite lineaarne funktsioon). Kõik hinnatavad funktsioonid on aga (vähemalt normaaljaotuse kontekstis) identifitseeritavad.
- Kõik hinnatavad parameeterfunktsioonid on kirja pandavad kui suuruse $\mathbf{X}\beta$ ($=E\mathbf{y}$) funktsioonid.

Seega mistahes olemasolevate vaatluste lineaarkombinatsiooni keskvärtus on hinnatav. Näiteks kujuta endale ette lihtsat ühefaktorilist dispersioonanalüüsi, faktortunnuseks olgu sugu. Kui valimis on vähemalt 1 mees, on meeste keskvärtus hinnatav. Kui valimis on vähemalt 1 naine, on naiste keskvärtus hinnatav. Kui valimis on vähemalt 1 mees ja 1 naine, siis on meeste ja naiste erinevuse keskvärtus hinnatav.

2.6 Parim lineaarne nihketa hinnang - Best Linear Unbiased Estimator (BLUE)

Soovime hinnata hinnatavat parameeterfunktsiooni $\boldsymbol{\lambda}^T \boldsymbol{\beta}$. Soovime, et hinnang oleks nihketa hinnangute seas väiksema võimaliku varieeruvusega (veaga). Millist hinnangut peaksime kasutama?

Definitsioon 2.2 Hinnangut $\widehat{\boldsymbol{\lambda}^T \boldsymbol{\beta}}$ kutsutakse parimaks lineaarseks nihketa hinnanguks (BLUE), kui ta rahuldab järgmiseid nõudeid:

1. Lineaarsus

$$\widehat{\boldsymbol{\lambda}^T \boldsymbol{\beta}} = \mathbf{a}^T \mathbf{y}$$

2. Nihketus

$$\forall \boldsymbol{\beta} \quad E(\widehat{\boldsymbol{\lambda}^T \boldsymbol{\beta}}) = \boldsymbol{\lambda}^T \boldsymbol{\beta}$$

3. Minimaalne dispersioon (minimaalne keskmine ruutviga):

$$D(\widehat{\boldsymbol{\lambda}^T \boldsymbol{\beta}}) \leq D(\widetilde{\boldsymbol{\lambda}^T \boldsymbol{\beta}})$$

$\forall \widetilde{\boldsymbol{\lambda}^T \boldsymbol{\beta}}$ korral mis rahuldab nõudeid (1) ja (2). Teisisõnu:

$$D(\mathbf{a}^T \mathbf{y}) \leq D(\mathbf{b}^T \mathbf{y})$$

$\forall \mathbf{b}$ korral, mille puhul $E\mathbf{b}^T \mathbf{y} = \boldsymbol{\lambda}^T \boldsymbol{\beta}$ ($\forall \boldsymbol{\beta}$).

Märkus:

Nõude (3) võib kirja panna ka kujul

$$E(\mathbf{a}^T \mathbf{y} - \boldsymbol{\lambda}^T \boldsymbol{\beta})^2 \leq E(\mathbf{b}^T \mathbf{y} - \boldsymbol{\lambda}^T \boldsymbol{\beta})^2.$$

Seega otsime lineaarset nihketa hinnangut mille korral hinnangu oodatav ruutviga oleks minimaalne.

Teoreem 2.6 Gauss-Markovi teoreem

Vaatame mudelit

$$E\mathbf{y} = \mathbf{X}\boldsymbol{\beta}, \quad D(\mathbf{y}) = \sigma^2 \mathbf{I}.$$

Kui $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ on hinnatav, siis

$$\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} = \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{2.17}$$

on parim lineaarne nihketa hinnang $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ -le.

Tõestus

Esmalt paneme tähele, et nihketuse nõudest (2) järeldub:

$$\begin{aligned}\forall \boldsymbol{\beta} \quad \mathbb{E}(\widehat{\boldsymbol{\lambda}^T \boldsymbol{\beta}}) &= \boldsymbol{\lambda}^T \boldsymbol{\beta} \\ \forall \boldsymbol{\beta} \quad \mathbb{E}(\mathbf{a}^T \mathbf{y}) &= \boldsymbol{\lambda}^T \boldsymbol{\beta} \\ \forall \boldsymbol{\beta} \quad \mathbf{a}^T \mathbf{X} \boldsymbol{\beta} &= \boldsymbol{\lambda}^T \boldsymbol{\beta} \\ \mathbf{a}^T \mathbf{X} &= \boldsymbol{\lambda}^T\end{aligned}$$

Kuna sama nõue peab kehtima ka mingi teise lineaarse hinnangu $\mathbf{b}^T \mathbf{y}$ jaoks, saame:

$$(\mathbf{a}^T - \mathbf{b}^T) \mathbf{X} = 0.$$

Vaatame selle teise lineaarse nihketa hinnangu dispersiooni:

$$\begin{aligned}D(\mathbf{b}^T \mathbf{y}) &= D(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y} + \mathbf{a}^T \mathbf{y}) \\ &= D(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y}) + D(\mathbf{a}^T \mathbf{y}) + 2\text{cov}(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y}, \mathbf{a}^T \mathbf{y}) \\ &= D(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y}) + D(\mathbf{a}^T \mathbf{y}) + 2(\mathbf{b}^T - \mathbf{a}^T) \text{cov}(\mathbf{y}, \mathbf{y}) \mathbf{a}\end{aligned}$$

Juhul, kui $\mathbf{a}^T \mathbf{y} = \boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} = \mathbf{v}^T \mathbf{P}_{\mathbf{X}} \mathbf{y} = \mathbf{v}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, siis $\sigma^2 (\mathbf{b}^T - \mathbf{a}^T) \mathbf{a} = 0$ (sest $(\mathbf{b}^T - \mathbf{a}^T) \mathbf{X} = 0$) ja saame tulemuseks:

$$D(\mathbf{b}^T \mathbf{y}) = D(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y}) + D(\mathbf{a}^T \mathbf{y}) \quad (\geq D(\mathbf{a}^T \mathbf{y})).$$

□

Muuseas, saadud BLUE-hinnang on üheselt määratud:

$$\begin{aligned}D(\mathbf{b}^T \mathbf{y} - \mathbf{a}^T \mathbf{y}) &= 0 \\ \sigma^2 (\mathbf{b} - \mathbf{a})^T (\mathbf{b} - \mathbf{a}) &= 0 \\ \mathbf{b} &= \mathbf{a}.\end{aligned}$$

2.7 Parim nihketa hinnang (BUE)

H. Cramer ja C.R.Rao näitasid, et mistahes nihketa hinnangu dispersioon ei saa olla väiksem teatavast piirist (mida tuntakse kui Cramer-Rao alampiiri).

Ühe parameetri korral ütleb Cramer-Rao alampiir, et parameetri θ mistahes nihketa hinnangu $\hat{\theta}$ dispersioon ei saa olla suurem kui pöördväärtus Fisheri informatsioonist:

$$D(\hat{\theta}) \geq \mathcal{I}(\theta)^{-1}.$$

Sama tulemuse üldistus mitmemõõtmelisele juhule (ja sobib kasutamiseks ka nihkega hinnangute korral) on kirja pandav järgneval kujul.

Kui parameetervektori $\boldsymbol{\theta}$ hinnangu $\hat{\boldsymbol{\theta}}$ keskvärtus on kirja pandav kui $\boldsymbol{\theta}$ mingi funktsioon, $E\hat{\boldsymbol{\theta}} = \mathbf{g}(\boldsymbol{\theta})$, siis

$$D(\hat{\boldsymbol{\theta}}) \geq_L \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} [\mathcal{I}(\boldsymbol{\theta})]^{-1} \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \right]^T,$$

Kus \geq_L tähistab Löwneri osalist järjestust (kui $\mathbf{A} \geq_L \mathbf{B}$, siis $\mathbf{A} - \mathbf{B}$ on mittenegatiivselt määratud maatriks, st. $\mathbf{x}^T (\mathbf{A} - \mathbf{B}) \mathbf{x} \geq 0 \forall \mathbf{x}$).

Kui tegemist on nihketa hinnanguga, $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, siis $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} = \mathbf{I}$ ja

$$D(\hat{\boldsymbol{\theta}}) \geq_L [\mathcal{I}(\boldsymbol{\theta})]^{-1},$$

kus

$$\mathcal{I}(\boldsymbol{\theta}) = E \left\{ \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \right)^T \frac{\partial l}{\partial \boldsymbol{\theta}} \right\}.$$

Normaaljaotuse eeldusel, lineaarsete mudelite kontekstis (vaata valemit 2.14):

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X} / \sigma^2$$

ja seega on Fisheri informatsioon (juhul kui parameetervektor $\boldsymbol{\beta}$ on hinnatav) kirja pandav kui

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} \frac{1}{\sigma^2} E \{ [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X}]^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X} \} \\ &= \frac{1}{\sigma^4} \mathbf{X}^T E [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T] \mathbf{X} \\ &= \frac{1}{\sigma^4} \mathbf{X}^T \mathbf{I} \sigma^2 \mathbf{X} \\ &= \mathbf{X}^T \mathbf{X} / \sigma^2 \end{aligned}$$

Seega ei saa parameetervektori $\boldsymbol{\beta}$ ükski nihketa hinnang olla väiksema dispersiooniga kui $\mathcal{I}^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2$. Kuna aga lineaarse mudeli korral

$$\begin{aligned} D(\hat{\boldsymbol{\beta}}) &= D((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T D(\mathbf{y}) [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2, \end{aligned}$$

siis järelikult on meie poolt kasutatav hinnang parimaks nihketa hinnanguks parameetervektorile β .

Kokkuvõtteks: Kui andmed on normaaljaotusega, $\varepsilon \sim N$, siis on vähimruutude/suurima tõepära hinnang $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ parimaks nihketa hinnanguks.

Kui andmed pole normaaljaotusega, on sama hinnang parimaks lineaarseks nihketa hinnanguks (st. sellisel juhul võib olla võimalik leida hinnangut, mis on parem kui meie poolt pakutud hinnang, kuid seda hinnangut ei tasu otsida lineaarsete hinnangute klassist).