2.3 Suurima tõepära hinnang

Sageli on võimalik teha eelduseid ka uuritava tunnuse jaotuse kohta. Teades uuritava tunnuse jaotust võime otsitavad parameetrid leida suurima tõepära meetodil. Üks praktikas kõige sagedamini esinev jaotus on normaaljaotus. Antud peatükis vaatamegi, millise hinnangu parameetritele saame, kui uuritava tunnuse Y jaotuseks oleks normaaljaotus.

Vaatame lineaarset mudelit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{2.9}$$

kus mudeli jääkide jaotuseks on normaaljaotus:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}; \sigma^2\mathbf{I}). \tag{2.10}$$

Antud mudelit võime kirja panna ka kujul:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{I}). \tag{2.11}$$

Kui juhusliku suuruse $\mathbf{y}$ jaotuseks on mitmemõõtmelise normaaljaotus, $\mathbf{y} \sim N(\boldsymbol{\mu}; \boldsymbol{\Sigma})$, siis $\mathbf{y}$ tihedusfunktsioon on kirja pandav kujul

$$f(\mathbf{y}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp(-(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2). \tag{2.12}$$

Seega on valimi tõepärafunktsiooniks

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma) &= |2\pi\sigma^2\mathbf{I}|^{-1/2} \exp\left(-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\sigma^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right) \end{aligned}$$

ja log-tõepära

$$l(\boldsymbol{\beta}, \sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}. \tag{2.13}$$

Maksimiseerime log-tõepära saamaks suurima tõepära hinnangut. Leiame esmalt tuletise $\boldsymbol{\beta}$ järgi

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2}\frac{\partial(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \\ &= -\frac{1}{2\sigma^2}2(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\frac{\partial(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \\ &= -\frac{1}{2\sigma^2}2(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(-\mathbf{X}) \\ &= \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{X} \end{aligned} \tag{2.14}$$

ja võrdsustame saadud tuletise seejärel nulliga

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\beta}} &= 0 \\ \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X} &= 0 \\ \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^T \mathbf{y}. \end{aligned}$$

Saadud võrrandisüsteemi lahendiks sobib aga ju hinnang $\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$:

$$\mathbf{X}^T\mathbf{X}\hat{\beta}_{ML} = \mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{P}_{\mathbf{X}}\mathbf{y} = \mathbf{X}^T\mathbf{y}$$

Seega avaldub suurima tõepära meetodil saadud hinnang (mis ei pruugi olla ühene!) parameetervektorile $\boldsymbol{\beta}$ täpselt samal kujul, kui vähimruutude meetodil saadud hinnang. Kui hinnangud on üheselt määratud, siis $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}$.

Kui vähimruutude meetod ei pakkunud meile võimalust hinnata jääkide dispersiooni σ^2, siis suurima tõepära meetodil võime saada hinnangu ka sellele parameetrile. Leiame log-tõepärafunktsiooni tuletise σ^2 järgi, asendame $\boldsymbol{\beta}$ leitud suurima tõepära hinnanguga ning saamegi kätte lahendi.

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \sigma^2} &= -\frac{n}{2}(2\pi\sigma^2)^{-1}(2\pi) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2}(-1\sigma^{-4}) \\ \frac{\partial l(\hat{\boldsymbol{\beta}}, \sigma)}{\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{((\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y})^T((\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y})}{2}\frac{1}{(\sigma^2)^2} \end{aligned}$$

Paneme tähele, et $(\mathbf{I} - \mathbf{P}_{\mathbf{X}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})$. Seega saame tuletise võrdsustamisel nulliga tulemuseks

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \sigma^2} &= 0 \\ -\frac{n}{2}\frac{1}{\sigma^2} + \frac{((\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y})^T((\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y})}{2}\frac{1}{(\sigma^2)^2} &= 0 \\ \frac{n}{2}\sigma^2 &= \mathbf{y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}/2 \\ \sigma^2 &= \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}}{n} \end{aligned}$$

Märkus: Nagu hiljem näeme, on saadud suurima tõepära hinnang dispersioonile nihkega hinnang. Saadud hinnangut kasutatakse praktikas harva, pigem eelistatakse nihketa hinnangut.

2.4 Vaatlusvektori hindamisest — geomeetriline vaatenurk

Defineerides vektori normi valemiga

$$||\mathbf{x}|| = (\mathbf{x}^T\mathbf{x})^{1/2},$$

langeb kahe vektori $\mathbf{y}_1$ ja $\mathbf{y}_2$ vaheline kaugus $||\mathbf{y}_1 - \mathbf{y}_2||$ kokku tavalise eukleidilise kaugusega.

Otsime sellist vektorit $\boldsymbol{\theta}$ mida saaksime esitada sõltumatute tunnuste kaudu (sõltumatute tunnuste lineaarkombinatsiooni abil ehk teisisõnu, $\boldsymbol{\theta} \in \mathcal{C}(\mathbf{X})$) ja mis paikneks võimalikult lähedal vaatlusvektorile $\mathbf{y}$ (ehk kaugus $||\mathbf{y} - \boldsymbol{\theta}||$ oleks minimaalne).

Teoreem 2.5

1. *Minimaalne kaugus vektorite* $\mathbf{y}$ *ja* $\boldsymbol{\theta} \in \mathcal{C}(\mathbf{X})$ *vahel saavutatakse sellise vektori* $\hat{\boldsymbol{\theta}} \in \mathcal{C}(\mathbf{X})$ *korral, mille puhul*

$$(\mathbf{y} - \hat{\boldsymbol{\theta}}) \perp \mathcal{C}(\mathbf{X}). \tag{2.15}$$

Selgitus: Väide 2.15 tähendab, et $\mathbf{y} - \hat{\boldsymbol{\theta}}$ *on risti (ortogonaalne) kõigi vektoritega* $\mathbf{x} \in \mathcal{C}(\mathbf{X})$ *ehk* $\hat{\boldsymbol{\theta}}$ *on vektori* $\mathbf{y}$ *ortogonaalne (rist-) projektsioon vektorruumi* $\mathcal{C}(\mathbf{X})$.

2. *Leidub selline vektor* $\hat{\boldsymbol{\theta}} \in \mathcal{C}(\mathbf{X})$, *mis rahuldab nõuet 2.15, ta on ühene ja avaldub kujul:*

$$\hat{\boldsymbol{\theta}} = \mathbf{P}_{\mathbf{X}}\mathbf{y} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y},$$

kus $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ *on ortogonaalne projektor alamruumi* $\mathcal{C}(\mathbf{X})$.

Tõestus.
1. Olgu $\hat{\boldsymbol{\theta}} \in \mathcal{C}(\mathbf{X})$ selline, et $(\mathbf{y} - \hat{\boldsymbol{\theta}}) \perp \mathcal{C}(\mathbf{X})$, st $\mathbf{x}^T(\mathbf{y} - \hat{\boldsymbol{\theta}}) = 0$ iga $\mathbf{x} \in \mathcal{C}(\mathbf{X})$ korral. Siis $\boldsymbol{\theta} \in \mathcal{C}(\mathbf{X})$ korral

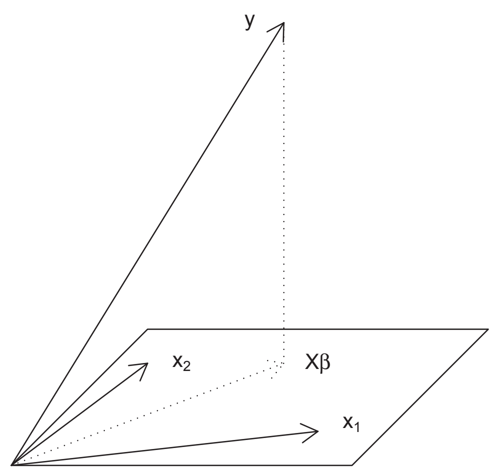
$$\begin{aligned} ||\mathbf{y} - \boldsymbol{\theta}||^2 &= (\mathbf{y} - \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\mathbf{y} - \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= (\mathbf{y} - \hat{\boldsymbol{\theta}})^T(\mathbf{y} - \hat{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \geq ||\mathbf{y} - \hat{\boldsymbol{\theta}}||^2, \end{aligned}$$

kuna $(\mathbf{y} - \hat{\boldsymbol{\theta}})^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = 0$ ortogonaalsuse nõudest tulenevalt. On selge, et $||\mathbf{y} - \boldsymbol{\theta}||^2$ saavutab miinimumi kui valime $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

2. Kuna $\hat{\boldsymbol{\theta}} \in \mathcal{C}(\mathbf{X})$, siis võime esitada ta maatriksi $\mathbf{X}$ veergude lineaarkombinatsioonina: $\hat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta}$. Lisanõudest $\mathbf{X}^T(\mathbf{y} - \hat{\boldsymbol{\theta}}) = 0$ saame, et $\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$ ehk $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$. Saadud võrrandisüsteem on lahenduv, lahendiks sobib $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$, ja seega

$$\hat{\boldsymbol{\theta}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} = \mathbf{P}_\mathbf{X}\mathbf{y}.$$

Seega vähimruutude meetodi puhul minimiseeritakse vaadeldud väärtuste vektori $\mathbf{y}$ ja tema prognoosi $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$ vaheline eukleidiline kaugus, $\hat{\boldsymbol{\theta}} := \mathbf{X}\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||$, vaata ka joonist 2.1.



Joonis 2.1: Vaatlusvektori $\mathbf{y}$ projektsioon maatriksi $\mathbf{X}$ veergude ruumile on lineaarse mudeli prognoos, $\hat{\mathbf{y}} = \mathbf{P}_\mathbf{X}\mathbf{y}\ (= \mathbf{X}\hat{\boldsymbol{\beta}})$

Märkus. Olles projektor alamruumi $\mathcal{C}(\mathbf{X})$ jätab $\mathbf{P}_\mathbf{X}$ muutmata mistahes vektori $\mathbf{x} \in \mathcal{C}(\mathbf{X})$, st $\mathbf{P_X x} = \mathbf{x}$. Veendume selles. Väide $\mathbf{x} \in \mathcal{C}(\mathbf{X})$ tähendab, et vektor $\mathbf{x}$ on esitatav kujul $\mathbf{x} = \mathbf{Xv}$. Seega $\mathbf{P_X x} = \mathbf{P_X X v} = \mathbf{Xv} = \mathbf{x}$.

Lemma 2.3 $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{P_X})$

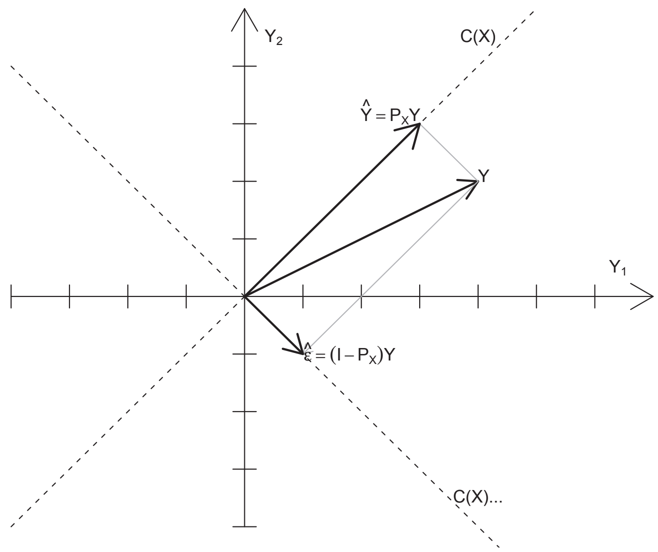
Tõestus: $\mathbf{x} \in \mathcal{C}(\mathbf{X}) \Leftrightarrow \exists \mathbf{v}, \mathbf{x} = \mathbf{Xv}$. Seega $\forall \mathbf{x} \in \mathcal{C}(\mathbf{X})$ kehtib: $\mathbf{x} = \mathbf{Xv} = \mathbf{P_X X v} \in \mathcal{C}(\mathbf{P_X})$.

$\forall \mathbf{x} \in \mathcal{C}(\mathbf{P_X})$ kehtib: $\mathbf{x} = \mathbf{P_X v} = \mathbf{X} \cdot (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T \mathbf{v} \in \mathcal{C}(\mathbf{X})$.
□

Järeldus:

$$\operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{P_X}). \tag{2.16}$$

Näide 2.1 *Vaatame erakordselt lihtsat mudelit ja andmestikku. Olgu tehtud kaks vaatlust,* $y_1 = 4$ *ja* $y_2 = 2$. *Mudeliks, mida soovime hinnata, on* $y_i = \mu + \varepsilon_i$. *Mudeli poolt määratud vektorruum* $\mathcal{C}(\mathbf{X})$ *on antud juhul kirja pandav järgmisel kujul:* $\{\mathbf{v} | \mathbf{v} = c \cdot (1,1)^T, c \in \mathbb{R}\}$. *Olukorda illustreerib joonis 2.2.*



Joonis 2.2: Vaatlusvektor $\mathbf{y}$, tema projektsioon $\hat{\mathbf{y}}$ maatriksi $\mathbf{X}$ veergudest moodustatud vektorruumi $\mathcal{C}(\mathbf{X})$ näites 2.1 toodud vaatluste ja mudeli korral