# 2 Paskaita. Suderintieji, nepaslinktieji ir mažiausios dispersijos įverčiai.

## 2.1 Suderintieji įverčiai

Vienas iš pagrindinių matematinės statistikos uždavinių yra nežinomų pasiskirstymo parametrų įvertinimas. Nežinomą parametrą $\theta$ reikia įvertinti, remiantis imtimi $X = (X_1, X_2, ..., X_n)$. Kiekvieną realiąją statistiką $\hat{\theta}(X) = \hat{\theta}(X_1, X_2, ..., X_n)$ vadinsime parametro $\theta$ *įverčiu*. Žinoma, ne kiekviena statistika tiks tam reikalui. Ji turi būti kokia nors prasme artima nežinomojo parametro reikšmei.

Išskirsime dvi R. Fišerio įvestas įverčių klases: suderintuosius ir nepaslinktuosius įverčius.

Tarkime, kad $\hat{\theta}_n(X_1, X_2, ..., X_n), n = 1, 2, \ldots$ yra įverčių seka. Sakysime, jog ji yra *suderinta*, jei kiekvienam $\varepsilon > 0$

$$\mathcal{P}_\theta(|\hat{\theta}_n(X_1, X_2, ..., X_n) - \theta| \geq \varepsilon) \to 0,$$

kai $n \to \infty$, t. y. $\hat{\theta}_n$ konverguoja į $\theta$ pagal tikimybę. Tai reiškia, kad su tikimybe, kiek norima artima 1, $\hat{\theta}_n(X)$ kiek norima mažai skirsis nuo $\theta$, jei tik $n$ bus pakankamai didelis. Paprastai ir kiekvienas $\hat{\theta}_n(X)$ vadinamas *suderintu*.

**Pavyzdys.** Sakykime, stebime atsitiktinį dydį su nežinomu vidurkiu $a \in \mathbf{R}$. Tada $\bar{X} = \dfrac{1}{n}(X_1 + X_2 + ... + X_n)$ yra suderintasis parametro $a$ įvertis.

**Pavyzdys.** Jei stebimasis atsitiktinis dydis turi žinomą vidurkį $a$, bet nežinomą dispersiją $\sigma^2$, tai statistika

$$S_0^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - a)^2$$

yra suderintasis $\sigma^2$ įvertis.

**Pavyzdys.** Jei stebimasis atsitiktinis dydis su nežinomu vidurkiu $a$, ir taip pat nežinoma dispersija $\sigma^2$, tai statistika

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

yra suderintasis $\sigma^2$ įvertis. Įrodysime šį teiginį.

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}((X_k - a) - (\bar{X} - a))^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - a)^2 - (\bar{X} - a)^2 = S_0^2 - (\bar{X} - a)^2.$$

Todėl

$$P(|S^2 - \sigma^2| \geq \varepsilon) \leq P(|S_0 - \sigma^2| \geq \frac{\varepsilon}{2}) + P(|\bar{X} - a| \geq \sqrt{\frac{\varepsilon}{2}}) \to 0,$$

kai $n \to \infty$.

## 2.2 Nepaslinktieji įverčiai

Sakykime, turime realiąją integruojamąją statistiką $\hat{\theta}(X) = \hat{\theta}(X_1, X_2, ..., X_n)$, kuria norime įvertinti nežinomą parametrą $\theta \in \Theta$. Dydį $E_\theta\hat{\theta}(X_1, X_2, ..., X_n) - \theta$ natūralu vadinti *įverčio poslinkiu*, arba *sistemine paklaida*. Jei visiems $\theta \in \Theta$

$$E_\theta\hat{\theta}(X_1, X_2, ..., X_n) = \theta,$$

tai sakome, kad įvertis $\theta$ yra *nepaslinktas*.

**Pavyzdys.** $\bar{X}$ yra nepaslinktasis nežinomo vidurkio $a$ įvertis, nes

$$E\bar{X} = \frac{1}{n}\sum_{k=1}^{n} EX_k = a.$$

**Pavyzdys.** Jei atsitiktinis dydis turi žinomą vidurkį $a$, bet nežinomą dispersiją $\sigma^2$, tai statistika

$$S_0^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - a)^2$$

yra nepaslinktas $\sigma^2$ įvertis. Iš tikrųjų,

$$ES_0^2 = \frac{1}{n}\sum_{k=1}^{n} E(X_k - a)^2 = \sigma^2.$$

**Pavyzdys.** Tarkime, kad atsitiktinis dydis turi nežinomą vidurkį $a \in \mathbf{R}$, ir taip pat nežinomą dispersiją $\sigma^2$, tai statistika

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{k=1}^{n}(X_k - a)^2 - \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - a)\right)^2 \\ &= \frac{1}{n}\left(1 - \frac{1}{n}\right)\sum_{k=1}^{n}(X_k - a)^2 - \frac{1}{n^2}\sum_{1\le j<k\le n}(X_j - a)(X_k - a). \end{aligned}$$

Iš čia

$$ES^2 = \left(1 - \frac{1}{n}\right)\sigma^2.$$

Taigi $S^2$ nėra nepaslinktas $\sigma^2$ įvertis. Kuo mažesnis $n$, tuo didesnis poslinkis (kai $n = 2, ES^2 = \sigma^2/2$. Dideliems $n$ tas poslinkis yra nedidelis: jis konverguoja į nulį, kai $n \to \infty$. Todėl sakoma, kad įvertis $S^2$ yra *asimptotiškai nepaslinktas*. Pastebėsime, kad

$$S_1^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

yra nepaslinktasis $\sigma^2$ įvertis.

## 2.3 Mažiausios dispersijos įverčiai.

Tam pačiam nežinomam parametrui įvertinti galime sudaryti daug nepaslinktųjų įverčių. Matėme, kad $\bar{X}$ yra nepaslinktasis vidurkio $a$ įvertis. Imkime kokius nors pastovius skaičius $q_1, q_2, \ldots, q_n$, tenkinančius sąlygą $q_1 + q_2 + \ldots + q_n = 1$, ir sudarykime statistiką

$$V = q_1 X_1 + q_2 X_2 + \ldots + q_n X_n.$$

Kadangi

$$EV = \sum_{k=1}^{n} q_k E X_k = a \sum_{k=1}^{n} q_k = a,$$

tai $V$ yra taip pat nepaslinktasis $a$ įvertis. $\bar{X}$ yra specialus jo atvejis, kai $q_1 = q_2 = \ldots = q_n = \dfrac{1}{n}$. Įvairiai parinkdami skaičius $q_1, q_2, \ldots, q_n$, galime gauti be galo daug nežinomo parametro $a$ įverčių. Kuris iš jų geriausias?

Jei turime du nežinomo parametro $\theta$ įverčius, tai natūralu laikyti geresniu tą, kurio reikšmės yra mažiau išsibarsčiusios apie parametrą $\theta$. Išsibarstymo matą galime parinkti įvairiai. Patogus ir dažniausiai vartojamas yra antrasis momentas $E(\hat{\theta} - \theta)^2$. Jei įvertis $\hat{\theta}$ yra nepaslinktasis, tai tas dydis yra įverčio dispersija. Kiekvienam $\theta_0 \in \Theta$ pažymėkime $H_{\theta_0}$ klasę visų nepaslinktųjų $\theta$ įverčių $T = T(X)$, turinčių dispersiją. Sakome, kad $\theta$ įvertis $T(\theta)$ turi *lokaliai mažiausią dispersiją* taške $\theta = \theta_0$, jei visiems $T \in H_{\theta_0}$ teisingos nelygybės

$$E(T_0 - \theta_0)^2 \leq E(T - \theta_0)^2.$$

Jei visiems $\theta \in \Theta$ ir visiems $T \in H_\theta$ teisingos nelygybės

$$E(T_0 - \theta)^2 \leq E(T - \theta)^2,$$

tai sakysime, kad įvertis $T(\theta)$ turi *tolygiai mažiausią dispersiją*.

Tegu $\hat{\theta}_1$ ir $\hat{\theta}_2$ yra du nepaslinktieji $\theta$ įverčiai. Tuomet $\hat{\theta}_1$ yra *efektyvesnis* už $\hat{\theta}_2$, jeigu

$$D\hat{\theta}_1 < D\hat{\theta}_2.$$

Tarkime, du kartus stebime atsitiktinį dydį $X$, kurio vidurkis $EX = \mu$, o dispersija $DX = \sigma^2$. Imtis $(X_1, X_2)$. Sudarome du $\mu$ įverčius: $\hat{\mu}_1 = X_1, \hat{\mu}_2 = (X_1 + X_2)/2$. Abu šie įverčiai yra nepaslinktieji. Tačiau $D\hat{\mu}_1 = DX_1 = \sigma^2$,

$$D\hat{\mu}_2 = \frac{1}{4}(DX_1 + DX_2) = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2},$$

t.y., $\hat{\mu}_2$ yra efektyvesnis už $\hat{\mu}_1$.

Įverčiai, tiesiškai priklausantys nuo $X_1, \ldots, X_n$, t.y. $T = a_1 X_1 + \ldots + a_n X_n, a_1, ..., a_n \in \mathbf{R}$, vadinami *tiesiniais*. Tiesinis nepaslinktasis įvertis, efektyvesnis už kitus tiesinius įverčius, vadinamas *geriausiu tiesiniu nepaslinktuoju įverčiu* (angl. BLUE - Best Linear Unbiased Estimator). Iš visų nepaslinktų tiesinių vidurkio įverčių $V = \sum q_i X_i, \sum q_i = 1$, geriausias tiesinis nepaslinktasis įvertis yra $\bar{X}$.

## 2.4 Įverčių sudarymo metodai

Įverčius galime sudaryti įvairiais būdais. Tačiau ne visi bet kaip sudaryti įverčiai bus tinkami. Istoriškai pirmasis iš dviejų metodų, kuriuos nagrinėsime, yra vadinamasis *momentų metodas*, pasiūlytas P. Čebyševo 1887 m. Tarkime, kad pasiskirstymų klasė $\{P_\theta, \theta \in \Theta\}, \Theta \in \mathbf{R}^s$ priklauso nuo vektorinio parametro $(\theta_1, \ldots, \theta_s)$ ir tie pasiskirstymai turi $s$ pirmųjų momentų $\alpha_r(\theta_1, \ldots, \theta_s), r = 1, \ldots, s$. Imame pirmuosius $s$ empirinių momentų $A_r, r = 1, \ldots, s$ ir prilyginame juos atitinkamiems teoriniams momentams. Gauname $s$ lygčių su $s$ nežinomųjų sistemą.

$$A_r = \alpha_r(\theta_1, \ldots, \theta_s), r = 1, \ldots, s.$$

Spręsime ją $\theta_1, \ldots, \theta_s$ atžvilgiu. Jei sprendiniai $\hat{\theta}_r = \hat{\theta}_r(X_1, \ldots, X_n), r = 1, \ldots, s$ egzistuoja, tai juos ir laikysime parametrų $\theta_1, \ldots, \theta_s$ įverčiais. Paaiškinsime šį metodą pavyzdžiu.

**Pavyzdys.** Tarkime, kad stebimasis atsitiktinis dydis yra pasiskirstęs pagal normalųjį dėsnį $\mathcal{N}(a, \sigma^2)$ su nežinomais parametrais $a \in R$ ir $\sigma > 0$. Imkime pirmąjį pradinį ir antrąjį centrinį momentus, lygindami teorinius ir empirinius momentus

$$\begin{cases} \bar{X} = a, \\ S^2 = \sigma^2. \end{cases}$$

Iš karto gauname $a$ ir $\sigma^2$ įverčius $\bar{X}$ ir $S^2$. Momentų metodas yra labai paprastas, bet jį taikant, ne visada galima gauti gerus rezultatus. Dažnai gaunami suderintieji įverčiai, bet jie nėra nepaslinktieji.

Geresni įverčiai gaunami didžiausio tikėtinumo metodu, pasiūlytu R. Fišerio 1912 m. Remiasi tikėtinumo principu: nežinomas parametras $\theta \in \Theta$ parenkamas taip, kad prie parinktos jo reikšmės būtų labiausiai tikėtina tai, kas įvyko (t.y. stebėtos imties $(X_1, \ldots, X_n)$ reikšmės). Taip parinkta parametro $\theta$ reikšmė vadinama didžiausio tikėtinumo įverčiu.

Tarkime, kad stebimas skirstinys $\{P_\theta, \theta \in \Theta\}, \Theta \in \mathbf{R}^s$ priklauso nuo vektorinio parametro $(\theta_1, \ldots, \theta_s)$ ir pasiskirstymo tankio funkcija yra $p_\theta$, t.y.

$$p_\theta(x) = p_\theta(x_1) \ldots p_\theta(x_n).$$

Visiems $\theta \in \Theta$ ir $x \in \mathbf{R}^n$ su sąlyga $p_\theta(x) > 0$ apibrėžkime funkcijas

$$h(\theta, x) = p_\theta(x_1) \ldots p_\theta(x_n), \qquad l(\theta, x) = \ln h(\theta, x) = \ln p_\theta(x_1) + \ldots + \ln p_\theta(x_n),$$

Funkcija $H(\theta) = H(\theta, X) = p_\theta(X_1) \ldots p_\theta(X_n)$, vadinama *tikėtinumo funkcija*. Nagrinėjamas ir jos logaritmas $L(\theta) = L(\theta, X) = \ln p_\theta(X_1) + \ldots + \ln p_\theta(X_n)$. $h(\theta, x)$ ir $l(\theta, x)$ yra jų realizacijos. Jei egzistuoja statistika $\hat{\theta} = \theta(X_1, \ldots, X_n)$, tenkinanti sąlygą

$$l(\hat{\theta}) = \max_{\theta \in \Theta} l(\theta),$$

tai ji vadinama parametro $\theta$ *didžiausio tikėtinumo įverčiu*. Didžiausio tikėtinumo įvertis yra vadinamųjų didžiausio tikėtinumo, lygčių

$$\frac{\partial H(\theta)}{\partial \theta_r}) = 0, r = 1, \ldots, s,$$

arba

$$\frac{\partial L(\theta)}{\partial \theta_r}) = 0, r = 1, \ldots, s,$$

sprendinys. Dažnai tos lygtys turi tik vieną sprendinį.

**Pavyzdys.** Tarkime, kad stebimasis atsitiktinis dydis, turintis Puasono skirstinį su nežinomu parametru $\lambda > 0$. Tada tikėtinumo funkcija

$$H(\lambda) = \frac{\lambda^{X_1+\ldots+X_n}}{X_1! \ldots X_n!} e^{-\lambda n}.$$

$$L(\lambda) = -\lambda n + \sum_{k=1}^{n} (X_k \ln \lambda - \ln X_k!).$$

Diferencijuodami pagal $\lambda$, gauname didžiausio tikėtinumo lygtį

$$\frac{\partial L}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{k=1}^{n} X_k = 0.$$

Iš jos randame didžiausio tikėtinumo įvertį, kuris yra efektyvusis $\lambda$ įvertis.

$$\hat{\lambda} = \bar{X}.$$

**Pavyzdys.** Panagrinėkime normalųjį dėsnį $\mathcal{N}(a, \sigma^2)$ su žinomu $\sigma > 0$, bet nežinomu parametru $a \in R$. Tikėtinumo funkcija

$$H(a) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (X_k - a)^2\}.$$

ir jos logaritmas

$$L(a) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (X_k - a)^2.$$

Didžiausio tikėtinumo lygtis yra

$$\frac{\partial L}{\partial a} = \frac{1}{\sigma^2} \sum_{k=1}^{n} (X_k - a) = 0.$$

Iš jos randame didžiausio tikėtinumo įvertį

$$\hat{a} = \bar{X}.$$

**Pavyzdys.** Jei turime normalųjį dėsnį $\mathcal{N}(a, \sigma^2)$ su žinomu $a$, bet nežinomu $\sigma^2$, tai diferencijuodami tikėtinumo funkcijos logaritmą

$$L(\sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (X_k - a)^2$$

pagal $\sigma^2$, gauname didžiausio tikėtinumo lygtį

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{n}(X_k - a)^2 = 0.$$

Iš jos gauname didžiausio tikėtinumo įvertį

$$\hat{\sigma^2} = S_0^2.$$

**Pavyzdys.** Jei turime normalųjį atsitiktinį dydį $\mathcal{N}(a, \sigma^2)$ su abiem nežinomais parametrais $a \in R$ ir $\sigma > 0$, tai išdiferencijavę tikėtinumo funkcijos logaritmą pagal $a$ ir $\sigma^2$, gauname lygčių sistemą

$$\frac{\partial L}{\partial a} = \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - a) = 0,$$
$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{n}(X_k - a)^2 = 0.$$

Iš pirmosios lygties

$$\hat{a} = \bar{X}.$$

Įrašę šį sprendinį į antrąją lygtį, gauname

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{n}(X_k - \bar{X})^2 = 0.$$

Iš jos

$$\hat{\sigma}^2 = S^2.$$

Efektyvieji $a$ ir $\sigma^2$ įverčiai yra $\bar{X}$ ir $S_1^2$. Pastebėsime, kad didžiausio tikėtinumo įverčiai, kai patenkintos gana bendros sąlygos, yra asimptotiškai efektyvūs.