$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \longrightarrow E(\underline{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} \longrightarrow \mathrm{cov}(\underline{X}) = \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \cdots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \cdots & \mathrm{var}(X_n) \end{pmatrix}$$

kovariančna matrika (simetrična, pozitivno semi-definitna)

$\underline{u} \otimes \underline{v} = \underline{u}\underline{v}^T$

$$\underline{X}\underline{X}^T = \begin{pmatrix} X_1^2 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2^2 & \cdots & X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ X_nX_1 & X_nX_2 & \cdots & X_n^2 \end{pmatrix}$$

$\mathrm{cov}(\underline{X}) = E[(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))^T] = E(\underline{X}\,\underline{X}^T) - E(\underline{X})E(\underline{X})^T$

korelacijski koeficient: $\mathrm{corr}(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2}{\sqrt{\mathrm{var}(X_1)\,\mathrm{var}(X_2)}}$

Če je A deterministična matrika (konstantna), velja: $E(A\underline{X}) = AE(\underline{X})$, $\mathrm{cov}(A\underline{X}) = A\,\mathrm{cov}(\underline{X})A^T$

$\mathrm{cov}(\langle \underline{X}, \underline{u}\rangle, \langle \underline{X}, \underline{v}\rangle) = \rangle\, \mathrm{cov}(\underline{X})\underline{u}, \underline{v}\rangle$, $\mathrm{cov}(\underline{u}^T\underline{X}, \underline{v}^T\underline{X}) = \underline{v}^T\,\mathrm{cov}(\underline{X})\underline{u}$

Standardna p-razsežna normalna porazdelitev je porazdelitev slučajnega vektorja $(Z_1, Z_2, \ldots, Z_n)$, kjer so $Z_1, \ldots, Z_p \sim N(0,1)$ in neodvisne.

Če je Q ortogonalna matrika in $\underline{Z}$ standarden normalen vektor, potem je $\underline{W} = Q\underline{Z}$ tudi standarden normalen.

Splošna n-razsežna normalna porazdelitev je vsaka porazdelitev slučajnega vektorja $\underline{W} = A\underline{Z} + \underline{u}$, kjer je $\underline{Z}$ standarden p-razsežni normalni vektor, A matrika $n \times p$ polnega ranga in $\underline{u} \in \mathbb{R}^n$.

$E(\underline{Z}) = 0$, $\mathrm{cov}(\underline{Z}) = I$, $E(\underline{W}) = \underline{u}$, $\mathrm{cov}(\underline{Z}) = AA^T$

Če $A \in \mathbb{R}^{n\times p}$ polnega ranga, je AA^T polnega ranga (in obrnljiva).

$\sigma > 0, X \sim N(\mu, \sigma^2) \Longrightarrow P(X \le a) = \Phi(\frac{a-\mu}{\sigma})$

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, potem $X_1 + \cdots + X_n \sim N(n\mu, \frac{\sigma^2}{n})$, $\overline{X} \sim N(\mu, \sigma^2)$, $\overline{X} - \mu \sim N(0, \frac{\sigma^2}{n})$

Pogojna gostota: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_y(y)}$

$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \Longrightarrow X_2|X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$

$||X||^2 = X^TX = sl(XX^T)$

Če poznamo porazdelitev slučajne spremenljivke Y in $f_{X|Y}$, potem velja $f_X(x) = E[f_{X|Y}(x)]$.

Pogojne pričakovane vrednosti: $E(X) = E[E(X|Y)]$, $\quad E[Xg(Y)|Y] = E(X|Y)g(X)$, v abstraktnem smislu definiramo $E[Y|X]$ kot funkcijo $\Psi(x)$, za katero za vsako omejeno zvezno funkcijo g velja $E[Yg(x)] = E[\Psi(x)g(x)]$.

$\mathrm{cov}(X,Y) = \mathrm{cov}(E(X|Z), E(Y|Z)) + E(\mathrm{cov}(X, Y|Z))$, med drugim $\mathrm{var}(X) = \mathrm{var}(E(X|Z)) + E(\mathrm{var}(X|Z))$

Če so $X_1, \ldots, X_n$ neodvisne med seboj in tudi od $Y_1, \ldots, Y_n)$, potem so $X_1, \ldots, X_n$ neodvisne tudi pogojno na Y.

Centralni limitni izrek

Izrek: Naj bodo $X_1, X_2, \ldots$ neodvisne, enako porazdeljene z $E(X_i^2) < \infty$ in $E(X_i) = \mu_1$ ter $\mathrm{var}(X_i) = \sigma_1^2$. $S_n = X_1 + X_2 + \cdots X_n$. Tedaj:

$$\frac{S_n - n\mu_1}{\sigma_1\sqrt{n}} \xrightarrow[n\to\infty]{\text{šibko}} N(0,1),$$

kjer $n\mu_1 = E(S_n)$ in $\sigma_1\sqrt{n} = \sigma(S_n)$.

Bolj ohlapno: n velik $\Longrightarrow S_n \dot\sim N(n\mu_1, n\sigma^2)$

$P(a \le S_n \le b) \approx \Phi(\frac{b-n\mu_1}{\sigma_1\sqrt{n}}) - \Phi(\frac{a-n\mu_1}{\sigma_1\sqrt{n}})$

Če slučajna spremenljivka živi v celih številih, lahko namesto $\le$ vzamemo $<$ in mejo povečamo za 1, ali pa vzamemo sredino.

Natančnost sredine je odvisna od asimetrije, ki jo meri $A(X) = \frac{E[(X-E(X))^3]}{(\mathrm{var}(X))^{\frac{3}{2}}}$.

Naj bodo $X_1, \ldots, X_n$ neodvisne in identično porazdeljene, $\sigma_1 = \sqrt{\mathrm{var}(X_i)}$, $\gamma_1 = E[|X_i - E(X_i)|^3]^{\frac{1}{3}}$, $S_n = X_1 + \cdots + X_n$. Ko $n \longrightarrow \infty$, $P(a_n \le S_n \le b_n)$ aproksimiramo z ustreznimi normalnimi. Zadosten pogoj, da gre:

- absolutna napaka $\to 0$: $n \gg \frac{\gamma_1^6}{\sigma_1^6}$
- relativna napaka $\to 0$: $n \gg \frac{\gamma_1^6}{\sigma_1^6}$ in $\min\{|a_n - E(S_n)|, |b_n - E(S_n)|\} \ll \frac{n^{\frac{2}{3}}\sigma_1^2}{\gamma_1}$

$\mu_1 = E(X_i)$, $\sup_{x\in\mathbb{R}} |P(S_n \le x) - \Phi(\frac{x-n\mu_1}{\sigma_1\sqrt{n}})| \le \frac{0.4774}{\sqrt{n}}\frac{\gamma_1^3}{\sigma_1^3}$

Porazdelitev χ^2:

Če so $Z_1, \ldots, Z_n$ neodvisne standardno normalne, potem je $Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n)$, $\chi^2(n) \xrightarrow{n\to\infty} N(0,1)$.

$\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$

Če $U \sim \Gamma(a, \lambda)$ in $V \sim \Gamma(b, \lambda)$, potem $U + V \sim \Gamma(a+b, \lambda)$

Če $U_1, \ldots, U_m \sim \Gamma(\frac{n}{2m}, \frac{1}{2})$ neodvisne, potem $U_1 + \cdots + U_m \sim \chi^2(n)$.

Razmerje Ljapunova:
$S = X_1 + \cdots + X_n$, $\mu = E(S)$, $\sigma^2 = \mathrm{var}(S)$, $X_1, \ldots, X_n$ neodvisne. $P(a \le S \le b) \approx \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$
$\sup_{x\in\mathbb{R}} |P(S \le x) - \Phi(\frac{x-\mu}{\sigma})| \le \frac{0.5591}{\sigma^3} \sum_{k=1}^n E[|X_k - E(X_k)|^3]$
Če desna stran konvergira k 0, imamo konvergenco k $N(0,1)$.

Dejanska statistika – vzročenje

$\widehat{a}$ je <u>nepristranska cenilka</u> za a, če je $E(\widehat{a}) = a$, <u>srednja kvadratična napaka</u>: $q(\widehat{a}) = E[(\widehat{a}-a)^2]$, <u>standardna napaka</u>: $\sqrt{q(\widehat{a})} = se(\widehat{a})$.

Slučajne spremenljivke $X_1, \ldots, X_n$ so <u>izmenljive</u>, če velja: $(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}) \stackrel{d}{=} (X_1, X_2, \ldots, X_n) \quad \forall \pi \in S_n$.
Za izmenljive sl. spr. $X_1, \ldots, X_n$ s pričakovano vrednosti $E(X_i) = \mu$, varianco $\mathrm{var}(X_i) = \sigma^2$, korelacijo $\mathrm{corr}(X_i, X_j) = \rho$, za $i \ne j$ je vzorčno povprečje $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ nepristranska cenilka za μ, $\mathrm{var}(\overline{X}) = \frac{\sigma^2}{n}(1+\rho(n+1))$, nepristranska cenilka za σ^2 pa je $\hat{\sigma}^2 = \frac{1}{(n-1)(1-\rho)} \sum_{i=1}^n (X_i - \overline{X})^2$

Enostavno slučajno vzorčenje
Populacija: $1, 2, \ldots, N$, vzorec: $K_1, K_2, \ldots, K_n$. Vrednosti spremenljivk na populaciji $x_1, x_2, \ldots, x_N$ (ne poznamo vseh). Poznamo vrednosti na vzorcu: $X_i = x_{K_i}$ (izmenljive, ker je vsaka n-terica enako verjetna)

Stratificirano vzorčenje:
Populacijo velikosti N razdelimo na k stratumov velikosti $N_1, \ldots, N_k$, kjer $w_1, w_2, \ldots, w_k$ ($w_i = \frac{N_i}{N}$ in $w_1 + \cdots + w_k = 1$) predstavljajo delež populacije v stratumih, $\mu_1, \ldots, \mu_k$ povprečja stratificiranih spremenljivk, $\sigma_1, \ldots, \sigma_k$ standardne odklone. Povprečje na celotni populaciji je $\mu = w_1\mu_1 + \cdots + w_k\mu_k$.
Varianca na celi populaciji: $\sigma^2 = \sigma_B^2 + \sigma_w^2$, kjer $\sigma_B^2 = \sum_{i=1}^k w_i(\mu_i - \mu)^2$ in $\sigma_w^2 = \sum_{i=1}^k w_i \sigma_i^2$.
Enostavni slučajni vzorci po stratumih:
$\begin{matrix} X_{11}, \ldots, X_{1n_1} & \overline{X}_1 \\ X_{21}, \ldots, X_{2n_2} & \overline{X}_2 \\ \vdots & \vdots \\ X_{k1}, \ldots, X_{kn_k} & \overline{X}_k \end{matrix}$, kjer so $\overline{X}_i$ vzorčna povprečja po stratumih. $\overline{X} = \sum_{i=1}^k w_i \overline{X}_i$ je nepristranska cenilka za μ,
$\sigma_i^2 = \frac{1}{n_i-1} \frac{N_i-1}{N_i} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$ nepristranska cenilka za σ_i^2, $\widehat{\sigma}_w^2 = \sum_{i=1}^k \frac{w_i}{n_i-1} \frac{N_i-1}{N_i} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$ nepristranska cenilka za σ_w^2 in $\widehat{\sigma}_B^2 = \sum_{i=1}^k w_i(\overline{X}_i - \overline{X})^2 - \sum_{i=1}^k (w_i - w_i^2) \frac{N_i - n_i}{N_i - 1} \frac{1}{n_i} \widehat{\sigma}_i^2$ nepristranska cenilka za σ_B^2.

intervali zaupanja

a bi radi ocenili: $a_{min} < a < a_{max}$, kjer je (a_{min}, a_{max}) interval zaupanja. $P(a_{min} < a < a_{max}) \ge 1 - \alpha$, kjer je $1 - \alpha$ stopnja zaupanja $(95\%, 99\%)$ in α stopnja tveganja $(5\%, 1\%)$.
Določanje IZ: pivotna funkcija $T(\underline{X}, a)$, kjer je $\underline{X}$ opažanje in a ocenjevani parameter.
IZ za μ, kjer je σ znan: $P(|\overline{X} - \mu| < M_\alpha) = 1 - \alpha$, $M_\alpha = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2})$
Če tudi σ ne poznamo in če so $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, potem $T(\underline{X}, \mu) = \frac{\overline{X}-\mu}{S^2}\sqrt{n}$, kjer $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$.
Za velike n je $\frac{\overline{X}-\mu}{S}\sqrt{n} \dot\sim N(0,1)$, v splošnem pa je $\frac{\overline{X}-\mu}{S}\sqrt{n} \sim \mathrm{Student}(n-1)$.

Ocenjevanje parametrov

Metoda momentov:
Vzorec $= (X_1, \ldots, X_n)^T$ je iz neke porazdelitve $f_X(x|\underline{\vartheta})$. Nastavimo parametre $\underline{\vartheta}$, da bo povprečje enako vzorčnemu povprečju, varianca vzorčni varianci.... Rešimo sistem enačb: $E[X] = \bar{X}$, $\mathrm{var}(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.
Metoda največjega verjetja: Postavimo se na stališče, da je vzorec $\underline{x} = (x_1, \ldots, x_n)^T$ neke porazdelitve $f_X(x|\underline{\vartheta})$, ki smo ga dobili verjetno eden izmed bolj verjetnih, in nastavimo parametre $\underline{\vartheta}$ tako, da bo najbolj verjeten. Kako verjeten je vzorec je podano s funkcijo verjetja L. Če so vzorci neodvisni, je $L(\underline{\vartheta}|\underline{x}) = \prod_{i=1}^n f_X(x_i|\underline{\vartheta})$. V diskretnem primeru je to $L(\underline{\vartheta}|\underline{x}) = \prod_{i=1}^n P_{\underline{\vartheta}}(X = x_i)$.
Definiramo tudi logaritemsko funkcijo verjetja $\ell(\underline{\vartheta}|\underline{x}) = \log(L(\underline{\vartheta}|\underline{x}))$.
Cenilka po metodi največjega verjetja (MLE) je vrednost $\hat{\underline{\vartheta}}$, kjer je dosežen maksimum L, ali ekvivalentno, ℓ. Cenilka je nepristranska. Ponavadi jo dobimo tako da rešimo sistem $\frac{\partial \ell}{\partial \underline{\vartheta}}(\underline{\vartheta}|\underline{x}) = 0$ za $\underline{\vartheta}$.
Za MLE velja: $\sqrt{n}(\vartheta - \hat{\underline{\vartheta}}) \dot\sim N(0, I^{-1}(\underline{\vartheta}))$.
Fisherjeva matrika informacije:
Varianca MLE je dana s Fisherjevo matriko informacije.
Matrike je dana z $I_{kl}(\underline{\vartheta}) = -E\left[\frac{\partial^2}{\partial\vartheta_k \partial\vartheta_l}(\log(f_X(x|\underline{\vartheta})))\right]$, $I(\underline{\vartheta}) = -E[\mathrm{hesse}(\log(f_X(x|\underline{\vartheta})))]$.
Velja: $\mathrm{var}(\hat{\underline{\vartheta}}_k) = \frac{(I^{-1})_{kk}}{n}$, $\mathrm{se}(\hat{\underline{\vartheta}}_k) = \sqrt{\frac{(I^{-1})_{kk}}{n}}$.
Za primer, ko je ϑ skalar, se formule poenostavijo: $I(\vartheta) = E[\frac{\partial^2}{\partial\vartheta^2} \log(f_X(x|\vartheta))]$,
$\mathrm{se}(\hat{\vartheta}) = \frac{1}{\sqrt{nI(\vartheta)}}$. Aproksimativni interval zaupanja: $\vartheta \in \left(\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{nI(\hat{\vartheta})}}\right)$.

Testiranje hipotez

Linearna regresija

Predpostavljamo, da so opaženi slučajni podatki $\underline{Y} = (Y_1, \ldots, Y_n)^T$ nastali kot $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, kjer je X znana deterministična $n \times m$ matrika, $\underline{\beta}$ neznan determinističen m vektor, $\underline{\varepsilon}$ neznan slučajen vektor napak, za katerega predpostavimo standardni regresijski model, ki pravi: $E[\underline{\varepsilon}] = 0$ in $\mathrm{var}(\underline{\varepsilon}) = \sigma^2 I$. Drugače povedano, napake ε_i so nekorelirane, varianca vsake posamezne pa je σ^2.
Za $\underline{\beta} = (\alpha, \beta)^T$ in $X = (1, x_1; 1, x_2; \ldots; 1, x_n)$ preidemo na standardno ocenjevanje s premico.
Nepristranska cenilka za $\underline{\beta}$ je cenilka po metodi najmanjših kvadratov $\underline{\hat{\beta}} = (X^T X)^{-1} X^T Y$.
Varianca cenilke: $\mathrm{var}(\underline{\beta}) = \sigma^2 (X^T X)^{-1} = \sigma^2 C$.
Nepristranska cenilka za σ^2 je $\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=1}^n \underline{\hat{\varepsilon}}_i^2$, kjer je $\underline{\hat{\varepsilon}} = \underline{Y} - X\underline{\beta}$.
Če gledamo samo posamezne komponente je cenilka za β_i enaka $\hat{\beta}_i = (\underline{\hat{\beta}})_i$ in $\mathrm{var}(\hat{\beta}_i) = \sigma^2 C_{ii}$, $\mathrm{se}(\hat{\beta}_i) = \sigma\sqrt{C_{ii}}$.
Za lin. kombinacijo komponent je cenilka enaka $a^T \underline{\hat{\beta}}$ in njena varianca je $\mathrm{var}(a^T \underline{\hat{\beta}}) = a^T C a$, za poljuben vektor a.
Izrek (Gauss-Markov): Cenilka po metodi najmanjših kvadratov je najboljša med vsemi linearnimi nepristranskimi cenilkami (ima najmanjšo varianco). Za vsako linearno cenilko $\underline{\tilde{\beta}} = L\underline{Y}$ mora veljati $LX = I$ in posledično $\mathrm{var}(\underline{\tilde{\beta}}) = \mathrm{var}(\underline{\tilde{\beta}} - \underline{\hat{\beta}}) + \mathrm{var}(\underline{\hat{\beta}})$, saj je $\mathrm{cov}(\underline{\tilde{\beta}} - \underline{\hat{\beta}}, \underline{\hat{\beta}}) = 0$. Enako velja tudi za cenilko $a^T \underline{\hat{\beta}}$ za vsak vektor a, v posebnem za $a = e_i$ tudi za cenilke po komponentah.
Če za $\underline{\varepsilon}$ predpostavljamo $\mathrm{var}(\underline{\varepsilon}) = \sigma^2 \Sigma$ za neko pd. matriko Σ, potem to prevedemo na standardni model z množenjem z leve s $(\Sigma)^{-\frac{1}{2}}$. Cenilka za $\underline{\beta}$ postane $\underline{\hat{\beta}} = (X^T \Sigma X)^{-1} X^T \Sigma^{-1} \underline{Y}$.

Za testiranje hipoteze $\beta_i = 0$ proti $\beta_i \neq 0$ uporabimo testno statistiko $t = \frac{\underline{\hat{\beta}}_i}{\hat{\sigma}\sqrt{C_{ii}}} \sim t_{n-m}$.

Inverzi matrik

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}$$