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

kovariančna matrika (simetrična, pozitivno semi-definitna)

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

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

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

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

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

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

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 $\underline{u} \in \mathbb{R}^n$.

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

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

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

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 kot funkcijo $\Psi(x)$, za katero za vsako omejeno zvezno funkcijo g velja $E[Yg(x)] = E[\Psi(x)g(x)]$.

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

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 $\text{var}(X_i) = \sigma_1^2$. $S_n = X_1 + X_2 + \cdots X_n$. 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 $\Longrightarrow S_n \dot{\sim} N(n\mu_1, n\sigma^2)$

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

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

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

Naj bodo $X_1, \ldots, X_n$ neodvisne in identično porazdeljene, $\sigma_1 = \sqrt{\text{var}(X_i)}$, $\gamma_1 = E[|X_i - E(X_i)|^3]^{\frac{1}{3}}$, $S_n = X_1 + \cdots + X_n$. Ko $n \longrightarrow \infty$, $P(a_n \leq S_n \leq b_n)$ 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 $\min\{|a_n - E(S_n)|, |b_n - E(S_n)|\} \ll \frac{n^{\frac{2}{3}}\sigma_1^2}{\gamma_1}$

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

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. $P(a \leq S \leq b) \approx \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$
$\sup_{x \in \mathbb{R}} |P(S \leq x) - \Phi(\frac{x-\mu}{\sigma})| \leq \frac{0.5591}{\sigma^3} \sum_{k=1}^{n} E[|X_k - E(X_k)|^3]$
Če desna stran konvergira k 0, imamo konvergenco k $N(0,1)$.

Dejanska statistika

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

Slučajne spremenljivke $X_1, \ldots, X_n$ so izmenljive, 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 $i \neq j$ je vzorčno povprečje $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ nepristranska cenilka za μ, $\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}(1 + \rho(n+1))$, nepristranska cenilka za σ^2 pa je $\hat{\sigma}^2 = \frac{1}{(n-1)(1-\rho)} \sum_{i=1}^{n} (X_i - \bar{X})^2$

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)