Amostras da Distribuição Normal p-variada e estimadores de máxima verossimilhança

$\underset{\sim}{X}_1, \underset{\sim}{X}_2 \dots \underset{\sim}{X}_n$ vetores aleatórios p-dimensionais, constituindo uma amostra aleatória de uma distribuição normal com vetor de médias μ e matriz de covariância Σ.

Densidade de $\underset{\sim}{X}_1, \underset{\sim}{X}_2 \dots \underset{\sim}{X}_n$

$$\prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right\}$$

Problema:

μ e Σ desconhecidos, tomar essa função como função de μ e Σ e estimar μ e Σ por máxima verossimilhança.

$$L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right\}$$

$L(\mu,\Sigma)$ é a função de verossimilhança.

Sejam $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{X}_i$ e $S = \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{X})(\underset{\sim}{X}_i - \bar{X})'$.

Resultado

Seja A matriz simétrica $p \times p$ e x um vetor $p \times 1$.

Então

$$x'Ax = tr(\underset{1\times p}{x'}\underbrace{Ax}_{p\times 1}) = tr(Axx')$$

a) $x'Ax = tr(x'Ax)$ porque $x'Ax$ é um escalar.

b) $B_{m\times k}$ $C_{k\times m} \Longrightarrow tr(BC) = tr(CB)$

Tomando $B = x'$ e $C = Ax \Rightarrow tr(x'Ax) = tr(Axx')$

Cálculo dos Estimadores de máxima Verossimilhança de μ e Σ.

$$(x-\mu)'\Sigma^{-1}(x-\mu) = tr\left[(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right]$$

$$= tr\left[\Sigma^{-1}(x_i-\mu)(x_i-\mu)'\right] \quad *1$$

i,

$$\sum_{i=1}^{n} (x_i-\mu)'\Sigma^{-1}(x_i-\mu) = \sum_{i=1}^{n} tr\left[\Sigma^{-1}\underbrace{(x_i-\mu)(x_i-\mu)'}_{p\times p}\right]$$

$$= tr\left[\Sigma^{-1}\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'\right)\right] \quad ** ②$$

$$\hookrightarrow tr\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} tr(A_i)$$

Além disso

$$\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)' ③ = \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)(x_i-\bar{x}+\bar{x}-\mu)'$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' + \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\mu)'}_{\left[\sum_{i=1}^{n}(x_i-\bar{x})\right](\bar{x}-\mu)' \to 0} + \underbrace{\sum_{i=1}^{n}(\bar{x}-\mu)(x_i-\bar{x})'}_{(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) \to 0}$$

$$+ \sum_{i=1}^{n}(\bar{x}-\mu)(\bar{x}-\mu)' =$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)' ④$$

$$\therefore tr\left[\Sigma^{-1}\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'\right)\right]^{⑤} = \Sigma(x_i-\mu)'\Sigma^{-1}(x_i-\mu$$

$$= tr\left\{\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right]\right\}$$

$$= tr\left\{\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right]\right\} + n\, tr\left[\Sigma^{-1}(\bar{x}-\mu)(\bar{x}-\mu)'\right]$$

$$= tr\left\{\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right]\right\} + n\, tr\left[\underbrace{(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)}_{n^{o}}\right]$$

$$= tr\left\{\Sigma^{-1}\underbrace{\left[\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right]}_{nS}\right\} + n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

$$\| \quad \sum (x_i-\mu)'\Sigma^{-1}(x_i-\mu)$$

~~$l(\mu,\Sigma)$~~

$$l(\mu,\Sigma) = \ln L(\mu,\Sigma) = -\frac{1}{2}np\ln(2\pi) - \frac{1}{2}n\ln|\Sigma|$$

$$-\frac{1}{2}tr\{n\Sigma^{-1}S\} - \frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

Como Σ^{-1} é positiva definida, $(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$ é mínima para $\mu = \bar{x}$. Portanto

$\hat{\mu} = \bar{X}$ é o estimador de máxima verossimilhança de μ.

Cálculo de $\hat{\Sigma}$

Para $\hat{\mu} = \bar{x}$,

$$l(\hat{\mu}, \Sigma) = \underbrace{-\frac{1}{2} np \ln(2\pi)}_{} - \frac{1}{2} n \ln|\Sigma| - \frac{1}{2} tr\left[n \Sigma^{-1} S\right]$$

$$= c - \frac{1}{2} n \left\{ \ln|\Sigma| + tr\left[\Sigma^{-1} S\right] \right\}$$

Lema

Seja $f(\Sigma) = \log|\Sigma| + tr\left[\Sigma^{-1} A\right]$.

Se $A > 0$ (A é positiva definida) e $\Sigma > 0$ (Σ é positiva definida), então $f(\Sigma)$ é mínima para $\Sigma = A$.

Demonstração: Seber - Multivariate Observations pag 523 (resultado A.7.1)

Pelo lema, $l(\hat{\mu}, \Sigma)$ é mínima para $\hat{\Sigma} = S$.

Verifica-se que S é positiva definida pelo resultado A.5.13 Seber pag 522.

Portanto o estimador de máxima verossimilhança de Σ é $S = \frac{1}{n} \sum_{i=1}^{n} (\underset{\sim}{X}_i - \bar{X})(\underset{\sim}{X}_i - \bar{X})'$.

Propriedades dos Estimadores

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \frac{1}{n}\Sigma$$

$$E(S) = \frac{n-1}{n}\Sigma$$

Do tópico 2

se $\underset{\sim}{X}_1 \dots \underset{\sim}{X}_n$ i.i.d $\sim N_p(\mu, \Sigma)$

através do resultado5

Verifica-se que $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$

Qual a distribuição de S?

Sejam $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2 \dots \underset{\sim}{Y}_n$ vetores aleatórios p dimensionais independentes com $\underset{\sim}{Y}_i \sim N_p(0, \Sigma)$. A matriz aleatória

$Y = \sum_{i=1}^{n} \underbrace{\underset{\sim}{Y}_i \underset{\sim}{Y}_i'}_{\substack{p\times 1 \ \ 1\times p \\ p\times p}}$ Tem distribuição de Wishart, cuja densidade é

y matriz $p \times p$ de valores de Y

$\Gamma \to$ função gama, $W(y, \Sigma, n)$ é unidim. Esta função não será utilizada

$$W(y, \Sigma, n) = \begin{cases} \dfrac{|y|^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2} tr\, y \Sigma^{-1}\right)}{2^{np/2} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma\left(\frac{n+1-i}{2}\right)} & \text{para } y \text{ positiva definida} \\ 0 & \text{caso contrário} \end{cases}$$

n - denominado número de graus de liberdade

Para $p=1$, $\underset{\sim}{Y}_i = Y_i \sim N(0, 1)$, $Y = \sum_{i=1}^{n} Y_i Y_i' =$

$= \sum_{i=1}^{n} Y_i^2 \sim \chi^2_n$. A dist. de Wishart é a extensão

da dist. qui-quadrado para p qualquer. Essa dist. tem várias propriedades da qui-quadrado. Em particular,

Se $Z_1, Z_2 \dots Z_K$ são matrizes aleatórias independentes com distribuição de Wishart com parâmetros Σ e graus de liberdade $n_1, n_2, \dots n_K$, então

$$\sum_{j=1}^{K} Z_j \sim W(\Sigma, n_1 + n_2 + \dots n_K)$$

dist de Wishart com parâmetros Σ, e $gl = n_1 + n_2 + \dots + n_K$.

Se $X_i \sim N_p(\mu, \Sigma)$, e são independentes

Verifica-se que, se $n-1 \geq p$ então

$$(n-1)\underbrace{S^*}_{=nS} \sim W(\Sigma, n-1)$$

Resultado

Sejam $\underset{\sim}{X}_1, \underset{\sim}{X}_2 \cdots \underset{\sim}{X}_n$ amostra aleatória de tamanho n da distribuição normal p-variada com vetor de médias μ e matriz de covariância Σ. Então

a) $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$ $\quad n-1 > p$.

b) $(n-1) S^* = nS \sim W(\Sigma, n-1)$ $\quad S^* = \frac{1}{n-1} \sum_{i=1}^{n} (\underset{\sim}{X}_i - \bar{X})(\underset{\sim}{X}_i - \bar{X})'$

c) $\bar{X}$ e S são independentes.

Obs: Como Σ é desconhecido, a distribuição de $\bar{X}$ não pode ser usada diretamente na inferência sobre μ. A distribuição de S não depende de μ e dessa forma, sendo S independente de $\bar{X}$, é possível construir uma estatística, que será a base para a inferência sobre μ.

$\underset{\sim}{X}_i \sim N_p \Rightarrow \bar{X} \sim N_p$.

Se $\underset{\sim}{X}_i$ não tiver distribuição normal?

Teorema do Limite Central Multivariado

Sejam $\underset{\sim}{X}_1, \underset{\sim}{X}_2 \dots \underset{\sim}{X}_n$ vetores aleatórios p dimensionais independentes com média μ e matriz de covariância Σ. Para $n \to \infty$,

$$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{d} N_p(0, \Sigma).$$

Demonstração:

T.W. Anderson - An Introduction to Multivariate Statistical Analysis.

Correlação e Regressão na Distribuição Normal multivariada

Obtidos os estimadores de máxima verossimilhança de σ_{jk}, $j, k = 1, 2, \dots p$, queremos determinar os estimadores de máxima verossimilhança de ρ_{jk}, $j \neq k$.

Lema

Se a função distribuição de uma variável aleatória ou vetor aleatório depende dos parâmetros $\theta_1, \theta_2, \dots, \theta_m$ e λ_j, $j = 1, 2 \dots m$ são parâmetros univocamente determinados por

$$\lambda_1 = h_1(\theta_1, \theta_2 \dots \theta_m)$$

$$\vdots$$

$$\lambda_m = h_m(\theta_1, \theta_2 \dots \theta_m)$$

e se $\hat{\theta}_1 \dots \hat{\theta}_m$ são os estimadores de máxima verossimilhança de $\theta_1, \theta_2 \dots \theta_m$ então, os estimadores de máxima verossimilhança de $\lambda_1, \lambda_2 \dots \lambda_m$ são

$$\hat{\lambda}_1 = h_1(\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_m) \quad \dots \quad \hat{\lambda}_m = h_m(\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_m).$$

Como consequência, se $\underset{\sim}{X}_i$ $i = 1, 2 \dots n$ é uma amostra aleatória da dist. $N(\mu, \Sigma)$, os estimadores de máxima verossimilhança de $\rho_{jk} = Corr(X_j, X_k)$ são

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\,\sqrt{s_{kk}}} = \frac{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{\sqrt{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2}\,\sqrt{\sum_{i=1}^{n}(X_{ik} - \bar{X}_k)^2}}$$

$$j, k = 1, 2 \dots p \qquad j \neq k.$$

Em particular, para $p=2$, amostras de n pares da distribuição normal bivariada,

$\rho = Corr(X_1, X_2)$ $\qquad$ $\hat{\rho} = r \rightarrow$ coef. de correlação amostral

$$r = \frac{\sum_{i=1}^{n} (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\sqrt{\sum_{i=1}^{n} (X_{i1} - \bar{X}_1)^2} \sqrt{\sum_{i=1}^{n} (X_{i2} - \bar{X}_2)^2}}$$

Verifica-se que se $\rho = 0$, a variável aleatória r tem densidade

$$f(r) = \begin{cases} \dfrac{1}{\sqrt{\pi}} \dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left(1-r^2\right)^{\frac{n-4}{2}}, & -1 \leq r \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

e

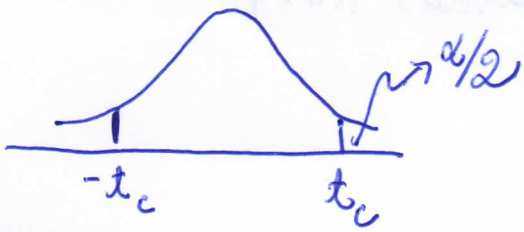
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$ (distribuição t-student com $n-2$ gl)

$H_0: \rho = 0$

$H_a: \rho \neq 0$

Rejeitamos Ho ao nível α se $|t_{obs}| \geq t_c$,

$P(t_{n-2} \geq t_c) = \frac{\alpha}{2}$



A partir daí, obtemos os correspondentes testes unicaudais.

Se $\rho \neq 0$, a distribuição de r não tem uma forma simples. Verifica-se no entanto que

$$\left(W - \frac{1}{2}\ln\frac{1+\rho}{1-\rho}\right)\sqrt{n-3} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1)$$

onde $W = \frac{1}{2}\ln\frac{1+r}{1-r}$.

$H_0: \rho = \rho_0$

$H_a: \rho \neq \rho_0$

Rejeitamos Ho se $\left|W - \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}\right|\sqrt{n-3} \geq z_c$,

$z_c \mid P(Z \geq z_c) = \frac{\alpha}{2}$, $Z \sim N(0,1)$

Teste de igualdade de dois coeficientes de correlação

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2 \qquad \underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N_2$$

$\underset{\sim}{X}_1, \underset{\sim}{X}_2 \dots \underset{\sim}{X}_{n_1}$ e $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2 \dots \underset{\sim}{Y}_{n_2}$ duas amostras aleatórias independentes entre si das distribuições de $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ respectivamente.

$H_0: \rho_{X_1X_2} = \rho_{Y_1Y_2}$

$H_a: \rho_{X_1X_2} \neq \rho_{Y_1Y_2}$

$$D = \frac{\frac{1}{2}\ln\frac{1+r_{X_1X_2}}{1-r_{X_1X_2}} - \frac{1}{2}\ln\frac{1+r_{Y_1Y_2}}{1-r_{Y_1Y_2}}}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}}$$

Rejeitamos H_0 ao nível α se

$$|D| \geq z_c \qquad z_c \mid P(Z \geq z_c) = \frac{\alpha}{2}, \quad Z \sim N(0,1)$$

Ex: X_1 - altura homens $\qquad$ Y_1 - altura mulheres

X_2 - peso homens $\qquad$ Y_2 - peso mulheres

$\rho_{X_1X_2} = \rho_{Y_1Y_2}$ a correlação entre peso e altura para homens é a mesma que para mulheres.

$$X^{(1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{bmatrix} \qquad X^{(2)} = \begin{bmatrix} X_{q+1} \\ \vdots \\ X_p \end{bmatrix}$$

$$X \sim N_p(\mu, \Sigma) \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$X^{(1)} | X^{(2)} = x_2 \sim N_q\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right)$$

Os elementos da matriz $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$ são denominados variâncias e covariâncias parciais pois medem a variabilidade e a dependência entre os elementos de $X^{(1)}$ fixados os valores de $X^{(2)}$.

Se

$\sigma_{jk \cdot q+1, q+2 \ldots p}$, $\quad j, k = 1, 2 \ldots q$

é o elemento de ordem jk da matriz $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$

o valor

$$\rho_{jk \cdot q+1, q+2 \ldots p} = \frac{\sigma_{jk \cdot q+1, q+2 \ldots p}}{\sqrt{\sigma_{jj \cdot q+1\, q+2 \ldots p}\ \sigma_{kk \cdot q+1\, q+2 \ldots p}}}$$

é denominado coeficiente de correlação parcial entre X_j e X_k fixados $X_{q+1}, X_{q+2} \dots X_p$.

Se $X^{(1)}$ é unidimensional $X^{(1)} = [X_1]$ e $X^{(2)} = \begin{bmatrix} X_{q+1} \\ X_{q+2} \\ \vdots \\ X_p \end{bmatrix}$,

seja $Y = \underbrace{\underset{1\times p-q}{\beta'} \underset{p-q\times 1}{X^{(2)}}}_{1\times 1} = \beta_{q+1} X_{q+1} + \beta \quad X_{q+2} + \dots + \beta_p X_p$

com $\beta' = [\beta_{q+1} \;\; \beta_{q+2} \dots \beta_p]$.

Verifica-se que o máximo valor de $Corr(X_1, Y)$ é obtido para

$$\beta = \Sigma_{22}^{-1} \delta_{12} \quad \text{onde}$$

$$\delta_{12}' = Cov(X_1, X^{(2)}) = [\sigma_{1q+1} \;\; \sigma_{1q+2} \dots \sigma_{1p}]$$

e

$$Y = \delta_{12}' \Sigma_{22}^{-1} X^{(2)}.$$

$$\mathrm{Cov}(X_1, Y) = \mathrm{Cov}\left(X_1, \sigma_{12}' \Sigma_{22}^{-1} X^{(2)}\right) =$$

$$= \mathrm{Cov}\left(X_1, X^{(2)}\right) \Sigma_{22}^{-1} \sigma_{12} = \underset{1\times(p-q)}{\sigma_{12}'} \underset{(p-q)\times(p-q)}{\Sigma_{22}^{-1}} \underset{(p-q)\times 1}{\sigma_{12}} =$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\beta' X^{(2)}\right) = \mathrm{Var}\left(\sigma_{12}' \Sigma_{22}^{-1} X^{(2)}\right) =$$

$$= \sigma_{12}' \Sigma_{22}^{-1} \Sigma_{22} \Sigma_{22}^{-1} \sigma_{12} = \underbrace{\sigma_{12}' \Sigma_{22}^{-1} \sigma_{12}}_{1\times 1} =$$

$$\therefore \mathrm{Corr}(X_1, Y) = \frac{\sigma_{12}' \Sigma_{22}^{-1} \sigma_{12}}{\sqrt{\sigma_{11}} \sqrt{\sigma_{12}' \Sigma_{22}^{-1} \sigma_{12}}} = \frac{\sqrt{\sigma_{12}' \Sigma_{22}^{-1} \sigma_{12}}}{\sqrt{\sigma_{11}}}$$

$\mathrm{Corr}(X_1, Y)$ é o coeficiente de correlação múltipla entre X_1 e os elementos de $X^{(2)}$.

$\beta = \Sigma_{22}^{-1} \sigma_{12}$ é o vetor de coeficientes de regressão da variável X_1 em função dos elementos de $X^{(2)}$.