A Distribuição Normal Multivariada

Prof. Caio Azevedo

Brevíssima revisão de cálculo de probabilidades

- Como usual, denotaremos por uma letra maiúscula, e.g. Y, uma variável aleatória (va) e por uma letra minúscula, y, um valor observado (realização de um experimento aleatório) desta va.
- Um vetor aleatório (vea) $\mathbf{Y} = (Y_1, ..., Y_p)'$ é uma coleção (arranjo) de variáveis aleatórias.
- As va's que compõem um vea podem apresentar alguma estrutura de dependência e/ou serem de diferentes tipos (discretas, contínuas ou mistas).

- Função densidade de probabilidade ou função de probabilidade: $f_{\mathbf{Y}}(\mathbf{y})$
- Função de distribuição acumulada $F_{\mathbf{Y}}(\mathbf{y}) = P(Y_1 \leq y_1, ..., Y_p \leq y_p)$.
- Vetor de médias: $\boldsymbol{\mu} = \mathcal{E}(\mathbf{Y}) = \begin{bmatrix} \mathcal{E}(Y_1) \\ \mathcal{E}(Y_2) \\ \vdots \\ \mathcal{E}(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$
- Matriz de covariâncias: $\boldsymbol{\Sigma} = Cov(\mathbf{Y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_p^2 \end{bmatrix}$

- Função geradora de momentos:

  $$M_{\mathbf{Y}}(\mathbf{t}) = \int \left(\ldots \left(\int \left(\int e^{\mathbf{y}'\mathbf{t}} f_{\mathbf{Y}}(\mathbf{y}) dy_1 \right) dy_2 \right) \ldots \right) dy_p$$

- Função característica:

  $$\Phi_{\mathbf{Y}}(\mathbf{t}) = \int \left(\ldots \left(\int \left(\int e^{i\mathbf{y}'\mathbf{t}} f_{\mathbf{Y}}(\mathbf{y}) dy_1 \right) dy_2 \right) \ldots \right) dy_p$$

- Sejam **A** e **B** matrizes não aleatórias, então

  - $\mathcal{E}(\mathbf{AY}) = \mathbf{A}\mathcal{E}(\mathbf{Y})$.
  - $Cov(\mathbf{AY}) = \mathbf{A}Cov(\mathbf{Y})\mathbf{A}'$.
  - $Cov(\mathbf{AY}, \mathbf{BX}) = \mathbf{A}Cov(\mathbf{Y}, \mathbf{X})\mathbf{B}'$.
  - Se $\mathbf{Y} = \mathbf{AX} + \mathbf{B}$, então $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{B}} M_{\mathbf{X}}(\mathbf{t}'\mathbf{A})$.

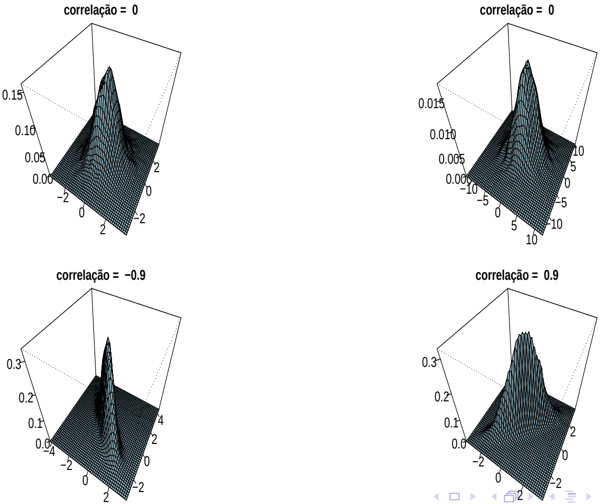
Distribuição Normal multivariada

- Dizemos que $\mathbf{Y} = (Y_1, ..., Y_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se sua fdp é dada por

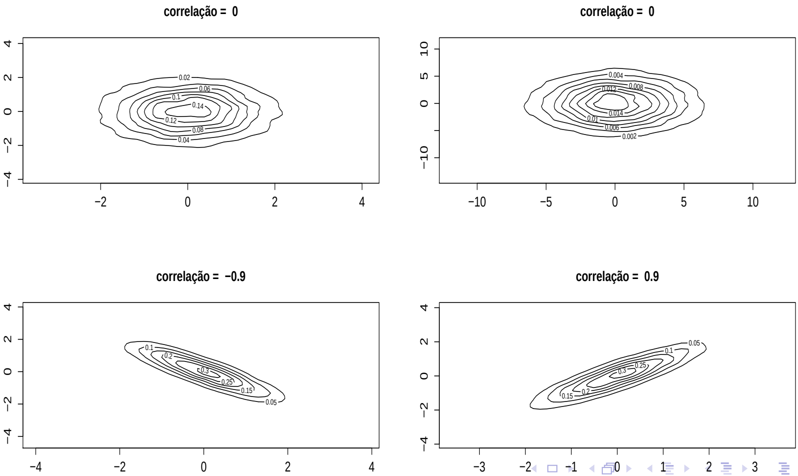
$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \mathbb{1}_{\mathcal{R}^p}(\mathbf{y})$$

- $\boldsymbol{\mu}$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz de covariâncias.
- $M_{\mathbf{Y}}(\mathbf{t}) = exp\left\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\}$

$\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, correlação $= \sigma_{11}$



$\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, correlação $= \sigma_{11}$



Propiedades

- Fechada sob marginalização: $Y_i \sim N(\mu_i, \sigma_i^2)$.
- $Y_i \perp Y_j, \forall i \neq j \Leftrightarrow \text{Cov}(Y_i, Y_j) = \sigma_{ij} = 0$.
- Se $\mathbf{A}_{(q \times p)}$ e $\mathbf{B}_{(q \times 1)}$ forem matrizes não aleatórias, então $\mathbf{V} = \mathbf{AY} + \mathbf{B} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{B}, \mathbf{A\Sigma A}')$.
- Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, simétrica e idempotente de rank = p, $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{(p \times p)}$, então $V = \frac{1}{\sigma^2}\mathbf{Y}'\mathbf{AY} \sim \chi_r^2, r = rank(\mathbf{A})$. Em particular, se $\mathbf{A} = \mathbf{I}$, então $\frac{1}{\sigma^2}\mathbf{Y}'\mathbf{Y} \sim \chi_p^2$.
- Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, então $\mathcal{E}(\mathbf{Y}'\mathbf{AY}) = tr(\mathbf{A\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

Derivadas matriciais úteis

- Sejam $\mathbf{A}_{(m \times n)}$ e $\mathbf{x}_{(n \times 1)}$ tais que

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}; \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Derivadas matriciais úteis

- Sejam $\mathbf{A}_{(m\times n)}$ e $\mathbf{x}_{(n\times 1)}$, e defina $\mathbf{y} = \mathbf{A}\mathbf{x}$ ($\mathbf{A}$ não depende de $\mathbf{x}$).

  Então:

$$\mathbf{y} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \sum_{k=1}^{n} a_{2k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{bmatrix}$$

- Logo

$$
\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \sum_{k=1}^{n} a_{1k}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^{n} a_{1k}x_k}{\partial x_2} & \dots & \frac{\partial \sum_{k=1}^{n} a_{1k}x_k}{\partial x_n} \\ \frac{\partial \sum_{k=1}^{n} a_{2k}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^{n} a_{2k}x_k}{\partial x_2} & \dots & \frac{\partial \sum_{k=1}^{n} a_{2k}x_k}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{k=1}^{n} a_{mk}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^{n} a_{mk}x_k}{\partial x_2} & \dots & \frac{\partial \sum_{k=1}^{n} a_{mk}x_k}{\partial x_n} \end{bmatrix}
$$

$$
\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}
$$

Derivadas matriciais úteis

- Alguns resultados:

$$
\begin{aligned}
\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \mathbf{A} \\
\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} &= \mathbf{A}' \\
\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{x} \\
\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{A} + \mathbf{A}')\mathbf{x}
\end{aligned}
$$

Dem. de que $f_{\mathbf{Y}}(.)$ é uma fdp

- Queremos demonstrar que

$$I = \int_{\mathcal{R}^p} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = \int_{\mathcal{R}}\int_{\mathcal{R}} ... \int_{\mathcal{R}} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = 1$$

Note que, se $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$ (decomposição de Cholesky), temos que:

$$\begin{aligned} I &= \int_{\mathcal{R}^p} |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\left(\boldsymbol{\Psi}\boldsymbol{\Psi}'\right)^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} d\mathbf{y} \\ &= \int_{\mathcal{R}^p} |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\Psi}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]'(\boldsymbol{\Psi})^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} d\mathbf{y} \end{aligned}$$

considere a transformação $\mathbf{z} = \boldsymbol{\Psi}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \Leftrightarrow \mathbf{y} = \boldsymbol{\Psi}\mathbf{z} + \boldsymbol{\mu}$.

Dem. de que $f_{\mathbf{Y}}(.)$ é uma fdp

- Matriz Jacobiana

$$\mathbf{J} = \frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} & \cdots & \frac{\partial y_2}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \frac{\partial y_n}{\partial z_2} & \cdots & \frac{\partial y_n}{\partial z_n} \end{bmatrix}$$

- Temos que $\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \boldsymbol{\Psi}\mathbf{z}+\boldsymbol{\mu}}{\partial \mathbf{z}} = \boldsymbol{\Psi}$.

- Além disso,

$$|\mathbf{J}| = |\mathbf{\Psi}| = |\mathbf{\Psi}|^{1/2}|\mathbf{\Psi}|^{1/2} = |\mathbf{\Psi}|^{1/2}|\mathbf{\Psi}'|^{1/2} = |\mathbf{\Psi}\mathbf{\Psi}'|^{1/2} = |\mathbf{\Sigma}|^{1/2}$$

- Lembrando que $\int_{\mathcal{R}^p} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = \int_{\mathcal{R}^p} f_{\mathbf{Y}}(\mathbf{\Psi}\mathbf{z} + \boldsymbol{\mu})|\mathbf{J}|d\mathbf{z}$
- Assim,

$$\begin{aligned} I &= \int_{\mathcal{R}^p} |\mathbf{\Sigma}|^{-1/2}|\mathbf{\Sigma}|^{1/2}(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{z}\right\}d\mathbf{z} \\ &= \prod_{i=1}^{p}\underbrace{\int_{\mathcal{R}}\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}z_i^2\right\}dz_i}_{1} = 1 \end{aligned}$$

Obtenção das Marginais

- Note que, para um dado j, $M_{Y_j}(t_j) = M_{\mathbf{Y}}(\mathbf{t}^*)$, em que $\mathbf{t}^* = [0\ 0 ... \underbrace{1}_{\text{posição } j} ... 0\ 0]$.
- Logo, temos que $M_{Y_j}(t_j) = \exp\left\{\mu_j t_j + \frac{\sigma_j^2 t_j}{2}\right\}$.
- A fgm acima corresponde à fgm de uma va com distribuição $N(\mu_j, \sigma_j^2)$.

Distribuições condicionais

- Seja $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)'$ e

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

  em que $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$.

- Então $\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim N(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}})$, em que

$$\overline{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{y}_2 - \boldsymbol{\mu}_2\right); \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

- Estimadores de máxima verossimilhança (dada uma amostra aleatória) $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_i$ e $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{Y}_i - \overline{\mathbf{Y}}\right)'\left(\mathbf{Y}_i - \overline{\mathbf{Y}}\right)$.