A Distribuição Normal Multivariada

Prof. Caio Azevedo

Introdução

- 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) $\boldsymbol{Y} = (Y_1, ..., Y_p)'$ 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: $f_{\boldsymbol{Y}}(\boldsymbol{y})$
- Função de distribuição acumulada $F_{\boldsymbol{Y}}(\boldsymbol{y}) = P(Y_1 \leq y_1, ..., Y_p \leq y_p)$.
- Vetor de médias: $\boldsymbol{\mu} = \mathcal{E}(\boldsymbol{Y}) = \begin{bmatrix} \mathcal{E}(Y_1) \\ \mathcal{E}(Y_2) \\ \vdots \\ \mathcal{E}(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$.
- Matriz de covariâncias: $\boldsymbol{\Sigma} = Cov(\boldsymbol{Y}) = \mathcal{E}[(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})'] =$

$$\mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}') - \boldsymbol{\mu}\boldsymbol{\mu}' = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_p^2 \end{bmatrix}$$

- Função geradora de momentos:

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

- Função característica:

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

- Sejam $\boldsymbol{A}$ e $\boldsymbol{B}$ matrizes não aleatórias, então
    - $\mathcal{E}(\boldsymbol{AY}) = \boldsymbol{A}\mathcal{E}(\boldsymbol{Y})$.
    - $Cov(\boldsymbol{AY}) = \boldsymbol{A}Cov(\boldsymbol{Y})\boldsymbol{A}'$.
    - $Cov(\boldsymbol{AY}, \boldsymbol{BX}) = \mathcal{E}\left((\boldsymbol{AY} - \boldsymbol{A\mu})(\boldsymbol{BX} - \boldsymbol{B\mu})'\right) = \boldsymbol{A}Cov(\boldsymbol{Y}, \boldsymbol{X})\boldsymbol{B}'$.
    - Se $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{B}$, então $M_{\boldsymbol{Y}}(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{B}} M_{\boldsymbol{X}}(\boldsymbol{t}'\boldsymbol{A})$.

Esperança de um vetor aleatório

$$\begin{aligned}
\mathcal{E}(\boldsymbol{AY}) &= \mathcal{E}\left(\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q1} & A_{q2} & \dots & A_{qp} \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}\right) = \mathcal{E}\left(\begin{bmatrix} \sum_{i=1}^{p} A_{1i}Y_i \\ \sum_{i=1}^{p} A_{2i}Y_i \\ \vdots \\ \sum_{i=1}^{p} A_{qi}Y_i \end{bmatrix}\right) \\
&= \begin{bmatrix} \sum_{i=1}^{p} A_{1i}\mathcal{E}(Y_i) \\ \sum_{i=1}^{p} A_{2i}\mathcal{E}(Y_i) \\ \vdots \\ \sum_{i=1}^{p} A_{qi}\mathcal{E}(Y_i) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{p} A_{1i}\mu_i \\ \sum_{i=1}^{p} A_{2i}\mu_i \\ \vdots \\ \sum_{i=1}^{p} A_{qi}\mu_i \end{bmatrix} = \boldsymbol{A}\mathcal{E}(\boldsymbol{Y}) = \boldsymbol{A\mu}
\end{aligned}$$

Matriz de covariâncias e fgm de um vetor aleatório

- Matriz de covariâncias.

$$
\begin{aligned}
Cov(\boldsymbol{AY}) &= \mathcal{E}[(\boldsymbol{AY} - \boldsymbol{A}\mu)(\boldsymbol{AY} - \boldsymbol{A}\mu)'] = \mathcal{E}[\boldsymbol{A}(\boldsymbol{Y} - \mu)(\boldsymbol{Y} - \mu)'\boldsymbol{A}'] \\
&= \boldsymbol{A}\mathcal{E}[(\boldsymbol{Y} - \mu)(\boldsymbol{Y} - \mu)']\boldsymbol{A}' = \boldsymbol{A}Cov(\boldsymbol{Y})\boldsymbol{A}'
\end{aligned}
$$

- Se $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{B}$, então

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathcal{E}(e^{\boldsymbol{t}'(\boldsymbol{AX}+\boldsymbol{B})}) = \mathcal{E}(e^{\boldsymbol{t}'\boldsymbol{AX}+\boldsymbol{t}'\boldsymbol{B}}) = \mathcal{E}(e^{\boldsymbol{t}'\boldsymbol{AX}})\mathcal{E}(e^{\boldsymbol{t}'\boldsymbol{B}}) \\
&= M_{\boldsymbol{X}}(\boldsymbol{t}'\boldsymbol{A})e^{\boldsymbol{t}'\boldsymbol{B}} = M_{\boldsymbol{X}}(\boldsymbol{t}^*)e^{\boldsymbol{t}'\boldsymbol{B}},
\end{aligned}
$$

em que $\boldsymbol{t}^* = \boldsymbol{t}'\boldsymbol{A}$

Matriz de correlações

- Seja $\mathbf{\Sigma} = Cov(\boldsymbol{Y}) =$ e defina $\boldsymbol{D} = \text{diag}\,(\sigma_1, \sigma_2, ..., \sigma_p)$, em que $\sigma_i = \sqrt{\sigma_i^2}, i = 1, 2, ..., p$.
- Assim, temos que $\boldsymbol{\rho} = \boldsymbol{D}^{-1}\mathbf{\Sigma}\boldsymbol{D}^{-1}$ é a matriz de correlações associada a $\boldsymbol{Y}$, pois

Matriz de correlações (cont.)

$$\boldsymbol{\rho} = \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^{-1} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p^2 \end{bmatrix}$$

$$\times \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^{-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{bmatrix}$$

em que $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$

Variância generalizada

- A variância generalizada (VG) é definida como o determinante da matriz de covariâncias, e procura resumir a variabilidade associada à dados multivariados (matriz de dados).
- Considere uma matriz de covariâncias 2×2, ou seja:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

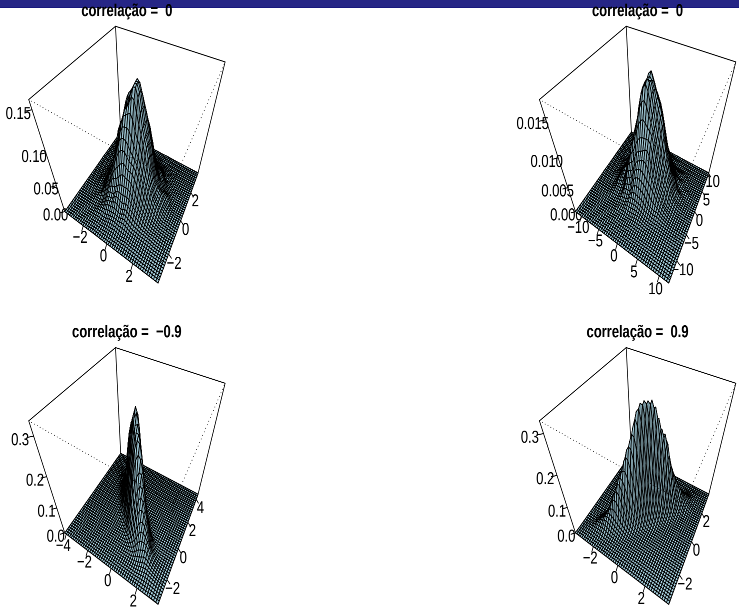
- Assim $VG = \sigma_1^2\sigma_2^2 - \sigma_1^2\sigma_2^2\rho^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$. Em geral, quanto maior/menor forem as variâncias de cada componente e/ou menores/maiores forem as correlações, maior/menor será a variância generalizada.

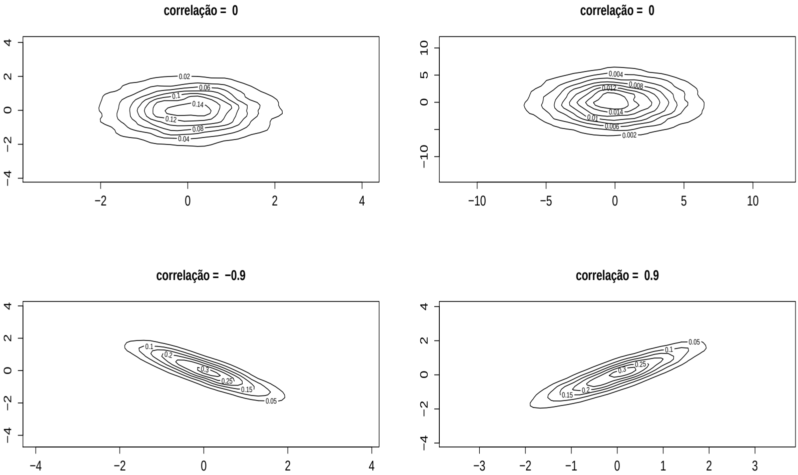
Distribuição Normal multivariada

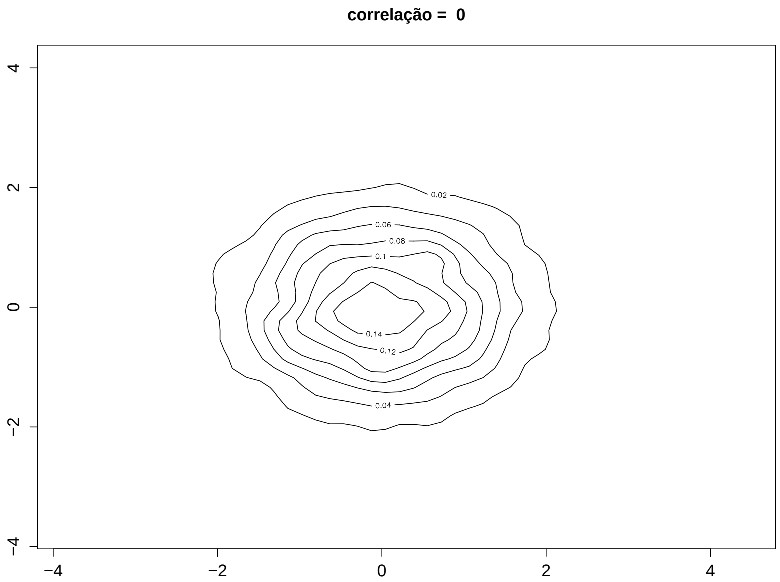
- Dizemos que $\boldsymbol{Y} = (Y_1, ..., Y_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se sua fdp é dada por

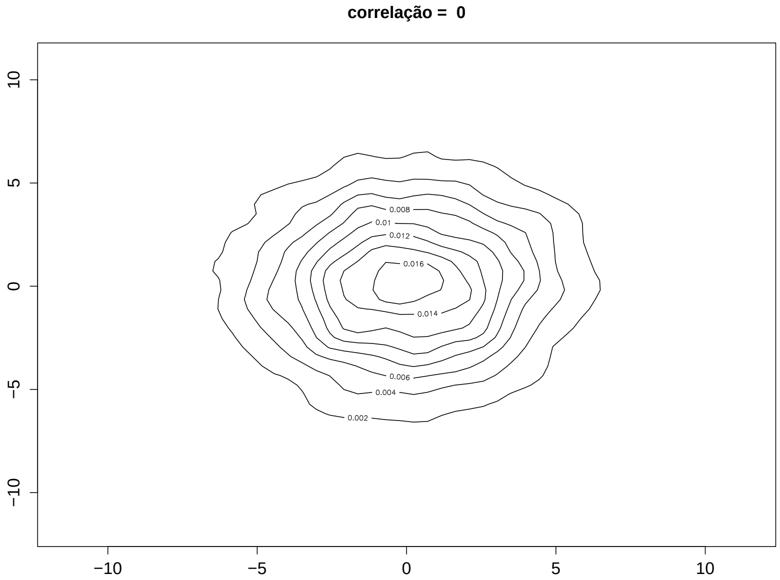
$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}\right)\right\} \mathbb{1}_{\mathcal{R}^p}(\boldsymbol{y})$$

- $\boldsymbol{\mu}$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz de covariâncias.
- $M_{\boldsymbol{Y}}(\boldsymbol{t}) = exp\left\{\boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right\}$.
- Se $p = 1$ então $f_Y(y) = \frac{1}{\sqrt{2}}(\sigma^2\pi)^{-1/2} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \mathbb{1}_{\mathcal{R}}(y)$.
- Nos dois gráficos seguintes: $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ ou (=9, no segundo gráfico), correlação $= \sigma_{11}$. VG, respectivamente (1;81;0,3429;0,3429).

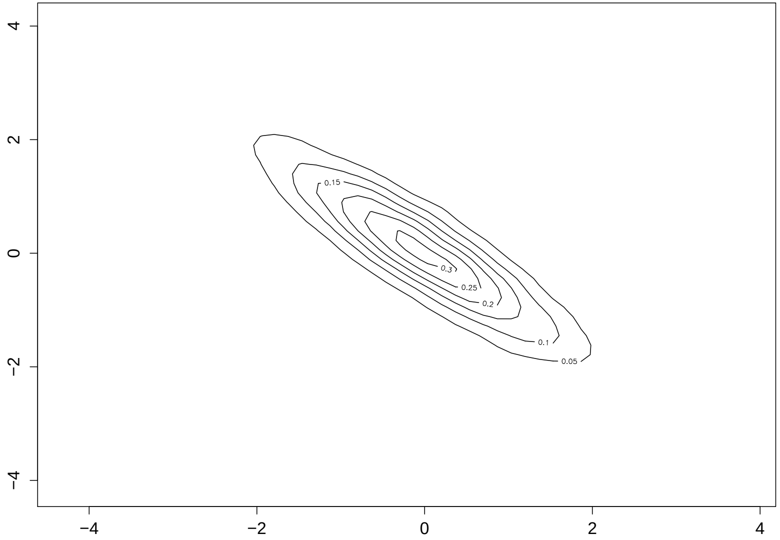

correlação = 0
correlação = 0
correlação = −0.9
correlação = 0.9

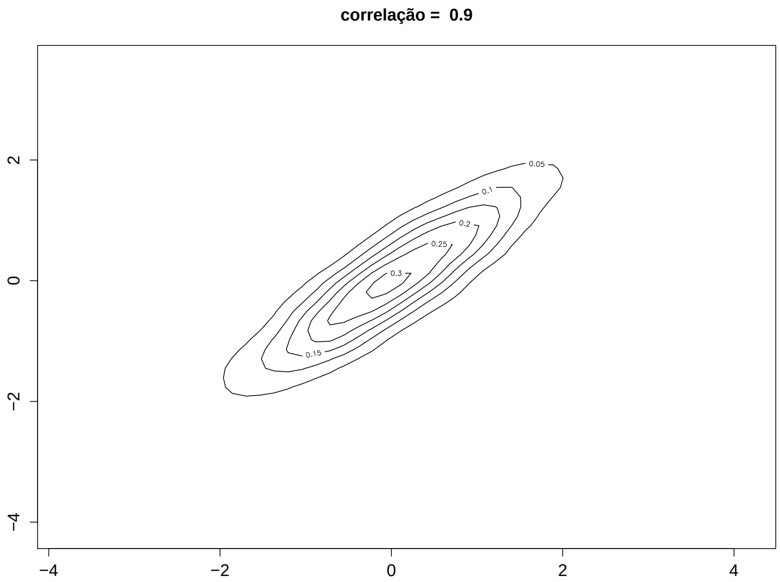

correlação = 0
0.02
0.06
0.1
0.14
0.12
0.08
0.04
correlação = 0
0.004
0.008
0.012
0.014
0.01
0.006
0.002
correlação = −0.9
0.1
0.2
0.3
0.25
0.15
0.05
correlação = 0.9
0.05
0.1
0.25
0.3
0.2
0.15


correlação = 0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
4
2
0
−2
−4
−4
−2
0
2
4


correlação = 0
0.002
0.004
0.006
0.008
0.01
0.012
0.014
0.016
10
5
0
−5
−10
−10
−5
0
5
10

correlação = −0.9




correlação = 0.9
0.05
0.1
0.2
0.25
0.3
0.15
2
0
−2
−4
−4
−2
0
2
4

Propiedades

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

Derivadas matriciais úteis

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

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

Derivadas matriciais úteis

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

  Então:

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

- Logo

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

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

Derivadas matriciais úteis

- Alguns resultados:

$$
\begin{aligned}
\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= \boldsymbol{A}, (\boldsymbol{A}'), \text{ se } \boldsymbol{A} \text{ for um vetor linha} \\
\frac{\partial \boldsymbol{x}'\boldsymbol{A}}{\partial \boldsymbol{x}} &= \boldsymbol{A}', (\boldsymbol{A}), \text{ se } \boldsymbol{A} \text{ for um vetor coluna} \\
\frac{\partial \boldsymbol{x}'\boldsymbol{x}}{\partial \boldsymbol{x}} &= 2\boldsymbol{x} \\
\frac{\partial \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}, 2\boldsymbol{A}\boldsymbol{x}, \text{ se } \boldsymbol{A} \text{ for simétrica}
\end{aligned}
$$

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

- Queremos demonstrar que

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

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

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

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

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

- Matriz Jacobiana

$$\boldsymbol{J} = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{z}} = \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} & \cdots & \frac{\partial y_1}{\partial z_p} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} & \cdots & \frac{\partial y_2}{\partial z_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_p}{\partial z_1} & \frac{\partial y_p}{\partial z_2} & \cdots & \frac{\partial y_p}{\partial z_p} \end{bmatrix}$$

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

- Além disso,

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

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

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

Obtenção da fgm

- Por simplicidade, suponha que $\boldsymbol{\mu} = \mathbf{0}$. Assim, temos que:

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = \int_{\mathcal{R}^p} |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - 2\boldsymbol{t}'\boldsymbol{y}\right)\right\} d\boldsymbol{y}$$

- Contudo, note que

$\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - 2\boldsymbol{t}'\boldsymbol{y} = (\boldsymbol{y} - \boldsymbol{\Sigma}\boldsymbol{t})'\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y} - \boldsymbol{\Sigma}\boldsymbol{t}) - \boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}$, Portanto,

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = e^{\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}} \underbrace{\int_{\mathcal{R}^p} |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\Sigma}\boldsymbol{t})'\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y} - \boldsymbol{\Sigma}\boldsymbol{t})\right\} d\boldsymbol{y}}_{1}$$

$$= \quad e^{\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}$$

Obtenção da fgm

- Logo, se $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ e $Y = \boldsymbol{\mu} + \boldsymbol{X}$, então

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathcal{E}(e^{\boldsymbol{t}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{X}}) = e^{\boldsymbol{t}'\boldsymbol{\mu}} M_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}$$

Obtenção do vetor de médias

- Temos que $\mathcal{E}(\boldsymbol{Y}) = \left.\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\right|_{\boldsymbol{t}=\boldsymbol{0}}$. Por outro lado, temos que:

$$\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}} = \frac{\partial e^{\boldsymbol{\mu}'\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}}{\partial \boldsymbol{t}} = M_{\boldsymbol{Y}}(\boldsymbol{t})\left[\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}\right]$$

- Assim

$$\mathcal{E}(\boldsymbol{Y}) = \left.\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\right|_{\boldsymbol{t}=\boldsymbol{0}} = M_{\boldsymbol{Y}}(\boldsymbol{0})\left[\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{0}\right] = \boldsymbol{\mu}$$

Obtenção do matriz de covariâncias

- Temos que $\mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}') = \left.\frac{\partial^2 M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}'}\right|_{\boldsymbol{t}=\boldsymbol{0}}$. Por outro lado, temos que:

$$
\begin{aligned}
\frac{\partial^2 M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}'} &= \frac{\partial}{\partial \boldsymbol{t}}\left[\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}'}\right] = \frac{\partial}{\partial \boldsymbol{t}}\left[\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\right]' = \frac{\partial}{\partial \boldsymbol{t}}\left[M_{\boldsymbol{Y}}(\boldsymbol{t})\left(\boldsymbol{\mu}' + \boldsymbol{t}'\boldsymbol{\Sigma}\right)\right] \\
&= \left(\frac{\partial M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\right)\left[\boldsymbol{\mu}' + \boldsymbol{t}'\boldsymbol{\Sigma}\right] + M_{\boldsymbol{Y}}(\boldsymbol{t})\boldsymbol{\Sigma} \\
&= M_{\boldsymbol{Y}}(\boldsymbol{t})\left[\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}\right]\left[\boldsymbol{\mu}' + \boldsymbol{t}'\boldsymbol{\Sigma}\right] + M_{\boldsymbol{Y}}(\boldsymbol{t})\boldsymbol{\Sigma}
\end{aligned}
$$

Obtenção do matriz de covariâncias

- Assim

$$
\begin{aligned}
\mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}') &= \left.\frac{\partial^2 M_{\boldsymbol{Y}}(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}'}\right|_{\boldsymbol{t}=\boldsymbol{0}} = M_{\boldsymbol{Y}}(\boldsymbol{0})\left[\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{0}\right]\left[\boldsymbol{\mu}' + \boldsymbol{0}'\boldsymbol{\Sigma}\right] + M_{\boldsymbol{Y}}(\boldsymbol{0})\boldsymbol{\Sigma} \\
&= \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}
\end{aligned}
$$

- Portanto, $Cov(\boldsymbol{Y}) = \mathcal{E}(\boldsymbol{Y}\boldsymbol{Y}') - \boldsymbol{\mu}\boldsymbol{\mu}' = \boldsymbol{\Sigma}$.

Obtenção das Marginais

- Note que, para um dado j, $M_{Y_j}(t_j) = M_{\boldsymbol{Y}}(\boldsymbol{t}^*)$, em que
  $\boldsymbol{t}^* = [0\ 0 \ldots \underbrace{t_j}_{\text{posição } j} \ldots 0\ 0]$.
- 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 $N(\mu_j, \sigma_j^2)$.

Observação sobre matrizes particionadas (em blocos)

- Seja $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ e $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{bmatrix}$
- Assim

$$\boldsymbol{X}'\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{A}_1 + \boldsymbol{X}_2'\boldsymbol{A}_3 & \boldsymbol{X}_1'\boldsymbol{A}_2 + \boldsymbol{X}_2'\boldsymbol{A}_4 \end{bmatrix}$$

e

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{A}_1\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{A}_3\boldsymbol{X}_1 + \boldsymbol{X}_1'\boldsymbol{A}_2\boldsymbol{X}_2 + \boldsymbol{X}_2'\boldsymbol{A}_4\boldsymbol{X}_2 \end{bmatrix}$$

Distribuições condicionais

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

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

- Para provar, basta usar a definição de distribuição condicional $\left(f_{\boldsymbol{Y}_1|\boldsymbol{Y}_2}(\boldsymbol{y}_1|\boldsymbol{y}_2) = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{\boldsymbol{Y}_2}(\boldsymbol{y}_2)}\right)$ e resultados relativos à matrizes particionadas.

Distribuições condicionais

- Temos que

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix} \\ &\times \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ \mathbf{0}' & \boldsymbol{I} \end{bmatrix}\end{aligned}$$

pois $\boldsymbol{\Sigma}$ é simétrica.

Distribuições condicionais

- Pode-se provar, portanto, que

$$
\begin{aligned}
(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) &= \left[\boldsymbol{y}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2-\boldsymbol{\mu}_2)\right]' \\
&\times \left(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} \\
&\times \left[\boldsymbol{y}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2-\boldsymbol{\mu}_2)\right] \\
&+ (\boldsymbol{y}_2-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2-\boldsymbol{\mu}_2) \qquad (1)
\end{aligned}
$$

Distribuições condicionais

- Pois

$$
\begin{aligned}
(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) &= \begin{bmatrix} \boldsymbol{y}_1-\boldsymbol{\mu}_1 \\ \boldsymbol{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}' \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I} \end{bmatrix} \\
&\times \begin{bmatrix} \left(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix} \\
&\times \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}^{-1}\boldsymbol{\Sigma}_{22} \\ \boldsymbol{0}' & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1-\boldsymbol{\mu}_1 \\ \boldsymbol{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}
\end{aligned}
$$

Distribuições condicionais

- Assim, basta usar (1) na definição $\left(f_{\boldsymbol{Y}_1|\boldsymbol{Y}_2}(\boldsymbol{y}_1|\boldsymbol{y}_2) = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{\boldsymbol{Y}_2}(\boldsymbol{y}_2)}\right)$, e o resultado da distribuição condicional segue (com um pouco de álgebra...).
- Além disso, considere que $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{22}||\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|$