

A Distribuição Normal Multivariada

Prof. Caio Azevedo

Dependência e Correlação

- **Dependência:** Em geral, está relacionada com o comportamento de elementos dentro de uma mesma população/grupo/fenômeno etc. Há vários tipos (aleatória, vetorial etc).
- **Correlação:** Está relacionada a tipos específicos de dependência (uma forma de quantificar). Em geral (mas não somente) está relacionada com algum tipo de relação (pe, linear) entre elementos.
- Assim, modelar dependência é uma abordagem (muito) mais ampla do que modelar correlação.
- Existência de dependência e/ou correlação \rightarrow **causalidade**.

Dependência

- Dentro do mundo probabilístico:
 - Tem relação com a ocorrência de dois ou mais eventos aleatórios.
 - Ausência de dependência \rightarrow correlação (de uma forma geral) nula.
 - Correlação nula $\not\rightarrow$ ausência de dependência (veja exemplos em [Livro do Barry James](#))
 - Correlação nula \rightarrow ausência de dependência, somente sob normalidade multivariada (veja Lista de Exercícios I).

Dependência e Correlação (cont.)

- Formas de modelar dependência probabilística:
 - Distribuições multivariadas.
 - Distribuições condicionais (univariadas e multivariadas).
 - Cópulas.
 - Modelos Mistos/Hierárquicos.
 - Equações (Funções) de Estimação Generalizadas.

Correlação

- De uma forma geral:
 - **Correlação de Pearson** (variáveis contínuas, discretas com suporte nos \mathbb{Z}).
 - **Correlações não paramétricas** (Kendall, Spearman etc) (qualquer tipo de variável).
 - **Correlação tetracórica** (variáveis binárias).
 - **Correlação policórica** (variáveis politômicas).
 - **Correlação intraclasse** (amostragem, dados hierárquicos).

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) $\mathbf{Y} = (Y_1, \dots, Y_p)'$ é uma coleção (arranjo) de variáveis aleatórias.
- As va's que compõem um vea podem apresentar alguma estrutura de dependência (correlação) 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 $F_{\mathbf{Y}}(\mathbf{y}) = P(Y_1 \leq y_1, \dots, Y_p \leq y_p)$.

- 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: $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y}) = \mathcal{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'] =$

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

- Função geradora de momentos:

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

- Função característica:

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

- Sejam \mathbf{A} e \mathbf{B} matrizes não aleatórias, então

- $\mathcal{E}(\mathbf{AY}) = \mathbf{A}\mathcal{E}(\mathbf{Y})$.

- $\text{Cov}(\mathbf{AY}) = \mathbf{A}\text{Cov}(\mathbf{Y})\mathbf{A}'$.

- $\text{Cov}(\mathbf{AY}, \mathbf{BX}) = \mathcal{E} [(\mathbf{AY} - \mathbf{A}\mu_{\mathbf{Y}})(\mathbf{BX} - \mathbf{B}\mu_{\mathbf{X}})'] = \mathbf{A}\text{Cov}(\mathbf{Y}, \mathbf{X})\mathbf{B}'$,

em que $\mu_{\mathbf{Y}} = \mathcal{E}(\mathbf{Y})$ e $\mu_{\mathbf{X}} = \mathcal{E}(\mathbf{X})$.

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

Esperança de um vetor aleatório

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

Matriz de covariâncias e fgm de um vetor aleatório

- Matriz de covariâncias.

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

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

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

Matriz de correlações

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

Matriz de correlações (cont.)

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

em que $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$

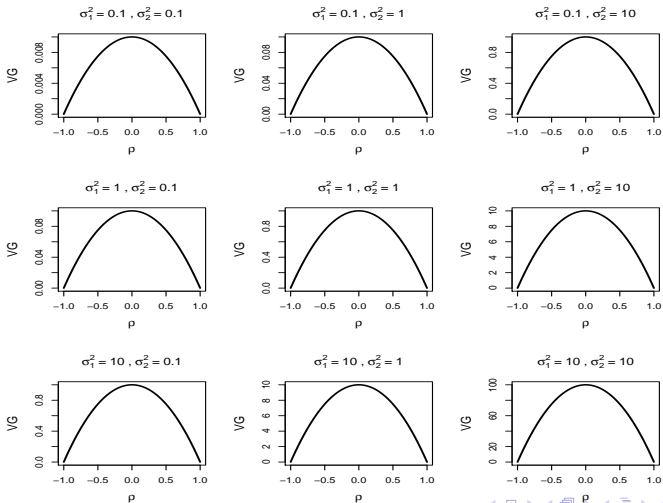
Variância generalizada

- 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:

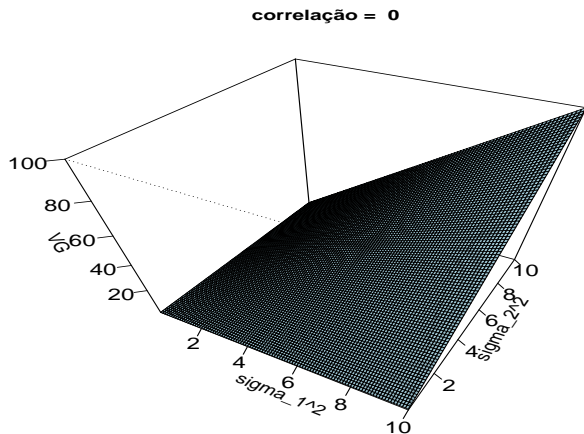
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

- 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.
- Com efeito, podemos notar que: $\lim_{|\rho| \rightarrow 1} VG = 0$, $\lim_{\rho \rightarrow 0} VG = \sigma_1^2\sigma_2^2$, $\lim_{\sigma_1^2 \rightarrow 0} VG = \lim_{\sigma_2^2 \rightarrow 0} VG = 0$, $\lim_{\sigma_1^2 \rightarrow \infty} VG = \lim_{\sigma_2^2 \rightarrow \infty} VG = \infty$

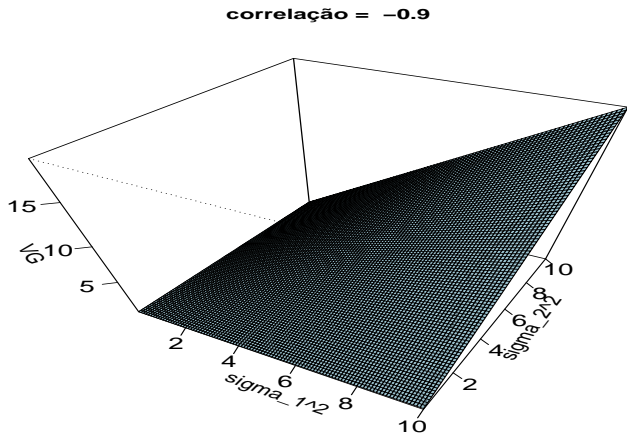
VG: fixando $(\sigma_1^2, \sigma_2^2)'$ e variando ρ



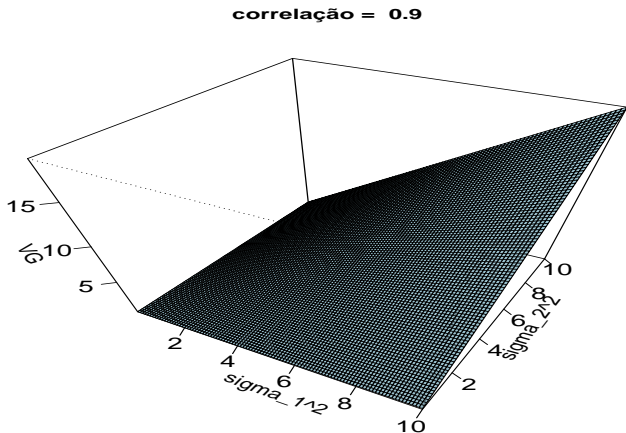
VG: fixando ρ e variando $(\sigma_1^2, \sigma_2^2)'$



VG: fixando ρ e variando $(\sigma_1^2, \sigma_2^2)'$



VG: fixando ρ e variando $(\sigma_1^2, \sigma_2^2)'$



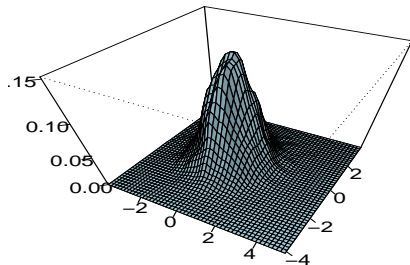
Distribuição Normal multivariada

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

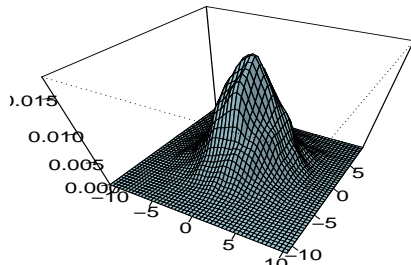
$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \mathbb{1}_{\mathcal{R}^p}(\mathbf{y})$$

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

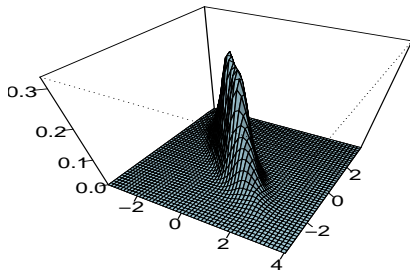
correlação = 0



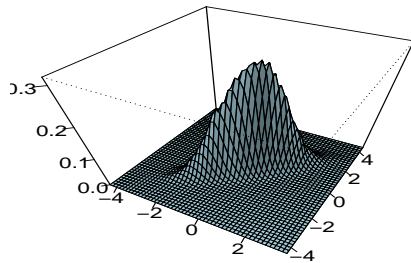
correlação = 0 , variâncias = 9



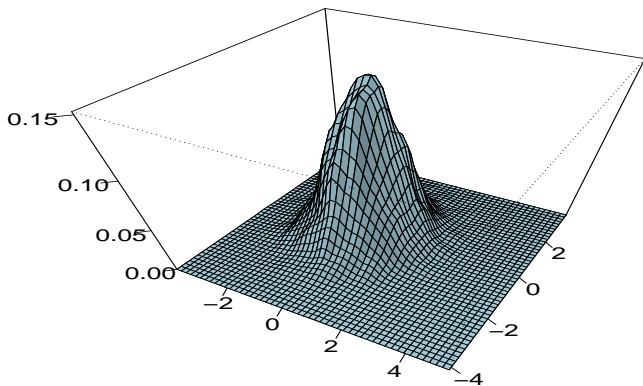
correlação = -0.9



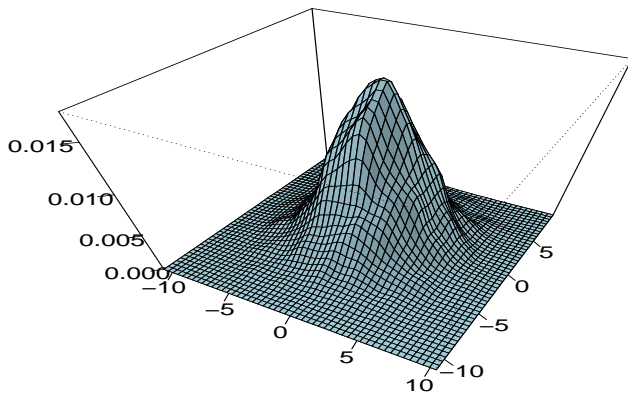
correlação = 0.9



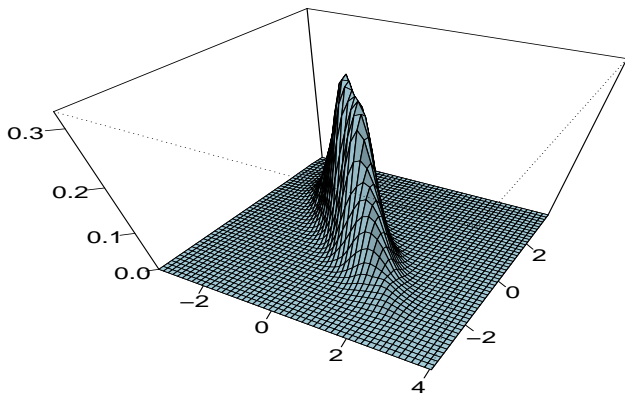
correlação = 0



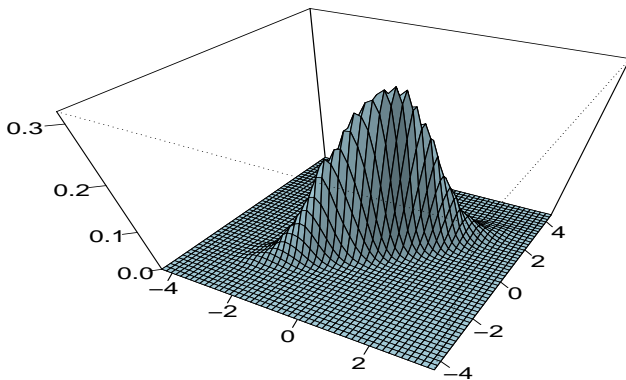
correlação = 0



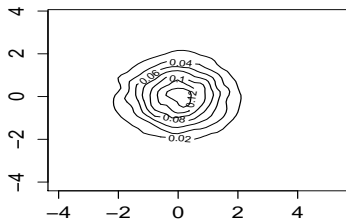
correlação = -0.9



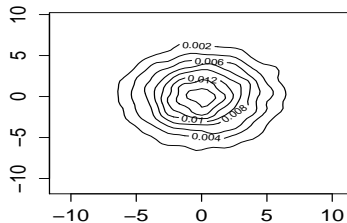
correlação = 0.9



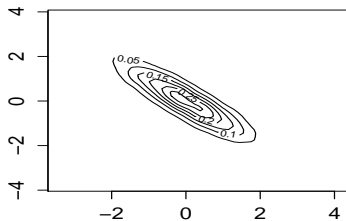
correlação = 0



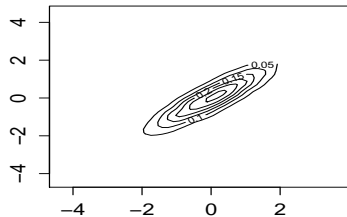
correlação = 0



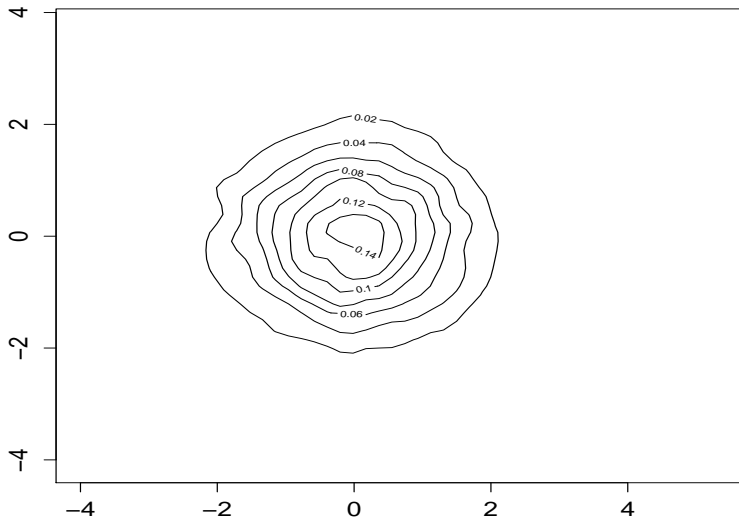
correlação = -0.9



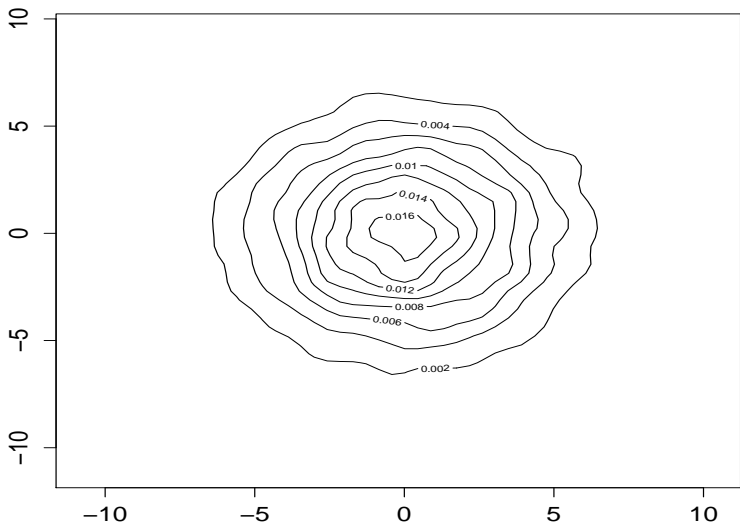
correlação = 0.9



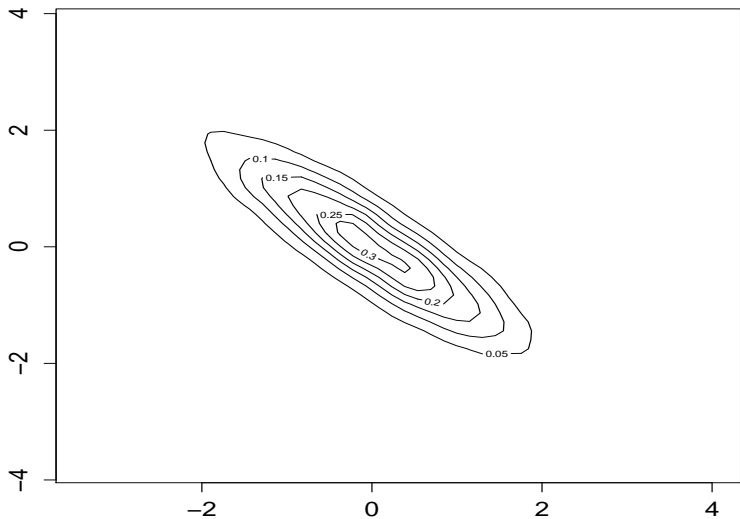
correlação = 0



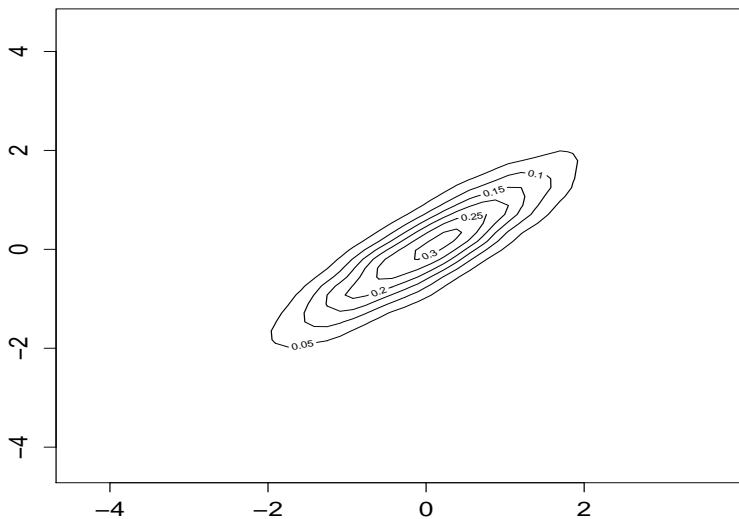
correlação = 0



correlação = -0.9



correlação = 0.9



Propiedades

- 1 Fechada sob marginalização: $Y_i \sim N(\mu_i, \sigma_i^2)$ (prova : usar o item 3)).
- 2 $Y_i \perp Y_j, \forall i \neq j \Leftrightarrow \text{Cov}(Y_i, Y_j) = \sigma_{ij} = 0$ (prova : fdp/fgm).
- 3 Se $\mathbf{A}_{(q \times p)}$ e $\mathbf{B}_{(q \times 1)}$ forem matrizes não aleatórias, então $\mathbf{V} = \mathbf{A}\mathbf{Y} + \mathbf{B} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{B}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ (prova : fgm).
- 4 Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, simétrica e idempotente e $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{(p \times p)}$, então $V = \frac{1}{\sigma^2} \mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_r^2, r = \text{rank}(\mathbf{A})$. Em particular, se $\mathbf{A} = \mathbf{I}$, então $\frac{1}{\sigma^2} \mathbf{Y}'\mathbf{Y} \sim \chi_p^2$ (prova : fgm).
- 5 Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, então $\mathcal{E}(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$ (prova: pesquisar).

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

- Agora 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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \sum_{k=1}^n a_{1k}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{1k}x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{1k}x_k}{\partial x_n} \\ \frac{\partial \sum_{k=1}^n a_{2k}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{2k}x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{2k}x_k}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{k=1}^n a_{mk}x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{mk}x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{mk}x_k}{\partial x_n} \end{bmatrix}$$
$$= \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Derivadas matriciais úteis

- Alguns resultados úteis [Searle e Khuri 2017, p. 176]:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}, (\mathbf{A}', \text{ se } \mathbf{A} \text{ for um vetor linha})$$

$$\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}', (\mathbf{A}, \text{ se } \mathbf{A} \text{ for um vetor coluna})$$

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}, 2\mathbf{Ax}, \text{ se } \mathbf{A} \text{ for simétrica}$$

Demonstração que $f_{\mathbf{Y}}(\cdot)$ 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 $\Sigma = \Psi\Psi'$ (**decomposição de Cholesky**), temos que:

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

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

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

- Matriz Jacobiana

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

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

- Além disso,

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

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

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

Obtenção da fgm

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

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

- Contudo, note que

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

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\frac{1}{2} \mathbf{t}' \boldsymbol{\Sigma} \mathbf{t}} \underbrace{\int_{\mathcal{R}^p} |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\Sigma} \mathbf{t})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\Sigma} \mathbf{t}) \right\} d\mathbf{y}}_1 \\ &= e^{\frac{1}{2} \mathbf{t}' \boldsymbol{\Sigma} \mathbf{t}} \end{aligned}$$

Obtenção da fgm

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

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

Obtenção do vetor de médias

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

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

- Assim

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

Obtenção do matriz de covariâncias

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

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

Obtenção do matriz de covariâncias

- Assim

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

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

Obtenção das Marginais

- Note que, para um dado j , $M_{Y_j}(t_j) = M_{\mathbf{Y}}(\mathbf{t}^*)$, em que

$$\mathbf{t}^* = [0 \ 0 \ \cdots \ \underbrace{t_j}_{\text{posição } j} \ \cdots \ 0 \ 0].$$

- Logo, temos que $M_{Y_j}(t_j) = \exp \left\{ \mu_j t_j + \frac{\sigma_j^2 t_j^2}{2} \right\}$.
- A fgm acima corresponde à fgm de uma va com distribuição $N(\mu_j, \sigma_j^2)$.

Observação sobre matrizes particionadas (em blocos)

■ Sejam $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ e $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix}$

■ Assim

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

e

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

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

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

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

Distribuições condicionais

- Temos que

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

pois Σ é simétrica.

Distribuições condicionais

- Pode-se provar, portanto, que

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

Distribuições condicionais

- Pois

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

Distribuições condicionais

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

Outras distribuições multivariadas

- Seja $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $Y_i = \exp(X_i)$, $i = 1, \dots, p$. Dizemos então que $\mathbf{Y} = (Y_1, \dots, Y_p)'$ tem distribuição log-normal multivariada com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ [Mardia, Kent e Bibby 1979, p. 24].
- Sejam $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $Y \sim \chi_\nu^2$ independentes. Tomando $T_i = X_i / \sqrt{Y/\nu}$, $i = 1, \dots, p$, então \mathbf{T} tem distribuição t de Student multivariada com parâmetros $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ e ν . Se $\nu = 1$, dizemos que \mathbf{T} tem distribuição Cauchy multivariada [Mardia, Kent e Bibby 1979, p. 24].

- Sejam $\mathbf{X}_1, \dots, \mathbf{X}_m \sim N_p(\mathbf{0}, \Sigma)$ independentes. Então dizemos que a matriz aleatória $p \times p$ dada por $\mathbf{M} = \sum_{i=1}^m \mathbf{X}_i \mathbf{X}_i'$ tem distribuição Wishart com parâmetro Σ e m graus de liberdade [Johnson e Wichern 2007, p. 174].

A função densidade de \mathbf{M} é

$$f(\mathbf{A}) = \frac{|\mathbf{A}|^{(n-p-2)/2} e^{-\text{tr}(\mathbf{A}\Sigma^{-1})/2}}{2^{p(n-1)/2} \pi^{p(p-1)/4} |\Sigma|^{(n-1)/2} \prod_{i=1}^p \Gamma((n-i)/2)} \mathbb{1}_{\{\mathbf{A} \text{ p.d.}\}}(\mathbf{A}),$$

em que $\Gamma(\cdot)$ denota a função Gama e p.d. significa positiva definida.

- Um vetor aleatório \mathbf{X} com distribuição Dirichlet com parâmetro $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p) \in (0, \infty)^{p+1}$ tem função densidade

$$f(x_1, \dots, x_p) = \frac{\Gamma(\sum_{i=0}^p \alpha_i)}{\prod_{i=0}^p \Gamma(\alpha_i)} \left(1 - \sum_{i=1}^p x_i\right)^{\alpha_0 - 1} \prod_{i=1}^p x_i^{\alpha_i - 1} \mathbb{1}_D(\mathbf{x})$$

em que $D = \{\mathbf{x} : \mathbf{x} \in [0, 1]^p, \sum_{i=0}^p x_i = 1\}$.

A distribuição Dirichlet é uma generalização multivariada da distribuição Beta [Mardia, Kent e Bibby 1979, p. 25].

- Um vetor aleatório \mathbf{X} com distribuição multinomial com parâmetros n e $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p) \in (0, \infty)^p$, $\sum_{i=1}^p \alpha_i = 1$, tem função de probabilidade

$$f(x_1, \dots, x_p) = \frac{n!}{x_1! \cdots x_p!} \prod_{i=1}^p \alpha_i^{x_i} \mathbb{1}_D(\mathbf{x})$$

em que $D = \{\mathbf{x} : \mathbf{x} \in \{0, 1, 2, \dots\}^p, \sum_{i=1}^p x_i = n\}$.

A distribuição multinomial é uma generalização multivariada da distribuição binomial [Mardia, Kent e Bibby 1979, p. 25].

- Um vetor aleatório $\mathbf{X}_{(n,1)}$ possui distribuição esférica se $\mathbf{R}\mathbf{X} \stackrel{d}{=} \mathbf{X}$ para qualquer matriz ortogonal $\mathbf{R}_{(n,n)}$, em que $\stackrel{d}{=}$ denota igualdade de distribuições.
- Um vetor aleatório $\mathbf{Y}_{(n,1)}$ possui distribuição elíptica multivariada com parâmetros de localização $\boldsymbol{\mu}_{(n,1)}$ e de escala $\boldsymbol{\Sigma}_{(n,n)}$ se

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}'\mathbf{X},$$

em que $\mathbf{A}_{(k,n)}$ é tal que $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$, $\text{rank}(\boldsymbol{\Sigma}) = k$ e $\mathbf{X}_{(k,1)}$ tem distribuição esférica.

- Essa classe de distribuições inclui as distribuições Normal, t de Student e Cauchy multivariadas [Fang, Kotz e Fang 1990, p. 27,31].

Referências I

-  FANG, K.; KOTZ, S.; FANG, K. W. *Symmetric multivariate and related distributions*. Boca Raton: CRC Press, 1990.
-  JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. Upper Saddle River: Prentice Hall, 2007.
-  MARDIA, K.; KENT, J.; BIBBY, J. *Multivariate analysis*. London: Academic Press, 1979.
-  SEARLE, S. R.; KHURI, A. I. *Matrix algebra useful for statistics*. Hoboken: John Wiley & Sons, 2017.