

# ME720 - Modelos Lineares Generalizados

## Parte 2 - A Distribuição Normal Multivariada

Profa. **Larissa Avila Matos**

# Brevíssima revisão de cálculo de probabilidades

Como usual, denotaremos por uma letra maiúscula, e.g.  $Y$ , uma variável aleatória (v.a.) e por uma letra minúscula,  $y$ , um valor observado (realização de um experimento aleatório) desta v.a..

Um vetor aleatório (vea)  $\mathbf{Y} = (Y_1, \dots, Y_p)$  é uma coleção (arranjo) de variáveis aleatórias.

As v.a.'s que compõem um vea podem apresentar alguma estrutura de dependência e/ou serem de diferentes tipos (discretas, contínuas ou mistas).

A função de densidade de probabilidade ou função de probabilidade é denotada por  $f_{\mathbf{Y}}(\mathbf{y})$ .

A função de distribuição acumulada é denotada por

$$F_{\mathbf{Y}}(\mathbf{y}) = P(Y_1 \leq y_1, \dots, Y_p \leq y_p).$$

Vetor de médias: O valor esperado de um vetor aleatório  $\mathbf{Y}_{(p \times 1)}$  é definido como o vetor de valores esperados das  $p$  variáveis aleatórias  $Y_1, Y_2, \dots, Y_p$  de  $\mathbf{Y}$ :

$$\boldsymbol{\mu} = \mathcal{E}(\mathbf{Y}) = \begin{bmatrix} \mathcal{E}(Y_1) \\ \mathcal{E}(Y_2) \\ \vdots \\ \mathcal{E}(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix},$$

onde  $E(Y_i) = \mu_i$  é obtido a partir da distribuição marginal de  $\mathbf{Y}_i$ .

**Matriz de covariâncias:** As variâncias  $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$  de  $Y_1, Y_2, \dots, Y_p$  e as covariâncias  $\sigma_{ij}, \forall i \neq j$ , podem ser convenientemente arranjadas em uma matriz de covariâncias, denotada por  $\Sigma$ , da seguinte forma:

$$\Sigma = Cov(\mathbf{Y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}.$$

A  $i$ -ésima linha de  $\Sigma$  contém a variância de  $Y_i$  e as covariâncias de  $Y_i$  com cada uma das outras v.a.'s.

Em muitas aplicações assumimos que  $\Sigma$  seja positiva definida. Isso realmente acontece quando as  $Y$ 's são v.a.'s contínuas e não existe qualquer relação linear entre elas. Se existe alguma relação linear entre as  $Y$ 's, assumimos que  $\Sigma$  seja positiva semidefinida.

A matriz de covariâncias  $\Sigma$  pode ser expressa como o valor esperado de uma matriz aleatória,  $\Sigma = \mathcal{E}((\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})') = \mathbf{Y}\mathbf{Y}' - \boldsymbol{\mu}\boldsymbol{\mu}'$ .

**Matriz de correlações:** A matriz de correlações é definida como:

$$\boldsymbol{\rho} = \text{Corr}(\mathbf{Y}) = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix},$$

onde  $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$  é a correlação de  $Y_i$  e  $Y_j$ .

Definindo,  $\mathbf{V} = \text{diag}(\Sigma) = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ , podemos reescrever  $\boldsymbol{\rho}$  como

$$\begin{aligned}\boldsymbol{\rho} &= \mathbf{V}^{-1/2}\Sigma\mathbf{V}^{-1/2}, \\ \Sigma &= \mathbf{V}^{-1/2}\boldsymbol{\rho}\mathbf{V}^{-1/2}.\end{aligned}$$

# Funções lineares de vetores aleatórios

Sejam  $\mathbf{X}$  e  $\mathbf{Y}$  vetores aleatórios, e  $\mathbf{A}$  e  $\mathbf{B}$  matrizes de constantes (não aleatórias). Então, assumindo que as matrizes e vetores em cada produto são conformes, temos

- $\mathcal{E}(\mathbf{AY}) = \mathbf{A}\mathcal{E}(\mathbf{Y})$ ,
- $Cov(\mathbf{AY}, \mathbf{BX}) = \mathbf{A}Cov(\mathbf{Y}, \mathbf{X})\mathbf{B}'$ ,
- $Cov(\mathbf{AY}) = \mathbf{A}Cov(\mathbf{Y})\mathbf{A}'$ ,
- $Cov(\mathbf{AY} + \mathbf{B}) = \mathbf{A}Cov(\mathbf{Y})\mathbf{A}'$ .

# Função geradora de momentos e característica

Seja  $\mathbf{Y}$  um vetor aleatório  $p \times 1$ .

A **função geradora de momentos** é dada por

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

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

A **função característica** é dada por

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

# Distribuição Normal multivariada

Dizemos que  $\mathbf{Y}$  segue uma distribuição Normal com média  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ ,  $\mathbf{Y} = (Y_1, \dots, Y_p) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , se sua fdp é dada por

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

onde  $\boldsymbol{\mu}$  é o vetor de médias e  $\boldsymbol{\Sigma}$  é a matriz de covariâncias.



# Demonstração de 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:

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

Considere a transformação  $\mathbf{z} = \Psi^{-1}(\mathbf{y} - \mu) \Leftrightarrow \mathbf{y} = \Psi\mathbf{z} + \mu$ . Assim, temos que  $d\mathbf{y} = \Psi' d\mathbf{z}$  e  $|J| = |\Psi'|$ .

Além disso,

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

Assim,

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

# Função geradora de momentos

A função geradora de momentos da distribuição normal multivariada é dada por

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp \left\{ \boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} \right\}.$$

Prova: ??

# 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  $M_{Y_j}(t_j) = \exp \left\{ \mu_j t_j + \frac{\sigma_j^2 t_j}{2} \right\}$ .

A fgm acima corresponde à fgm de uma v.a. com distribuição  $N(\mu_j, \sigma_j^2)$ .

# Propiedades

- 1 Fechada sob marginalização:  $Y_i \sim N(\mu_i, \sigma_i^2)$ .
- 2 Quaisquer duas variáveis aleatórias  $Y_i$  e  $Y_j$  são independentes se  $\sigma_{ij} = 0$ , ou seja,  $Y_i \perp Y_j, \forall i \neq j \Leftrightarrow \sigma_{ij} = 0$ .
- 3 Se  $\mathbf{A}_{(q \times p)}$  for uma matriz não aleatória, então

$$\mathbf{V} = \mathbf{A}\mathbf{Y} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}').$$

- 4 Seja  $\mathbf{A}_{(q \times p)}$  uma matriz não aleatória e  $\mathbf{b}$  um vetor de constantes  $p \times 1$ , então

$$\mathbf{Z} = \mathbf{A}\mathbf{Y} + \mathbf{b} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}').$$

# Propiedades

- 5 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

$$V = \frac{1}{\sigma^2} \mathbf{Y}' \mathbf{A} \mathbf{Y} \sim \chi_r^2, \quad r = \text{tr}(\mathbf{A}).$$

Em particular, se  $\mathbf{A} = \mathbf{I}$ , então  $\frac{1}{\sigma^2} \mathbf{Y}' \mathbf{Y} \sim \chi_p^2$ .

- 6 Se  $\mathbf{A}_{(p \times p)}$  for uma matriz não aleatória, então

$$\mathcal{E}(\mathbf{Y}' \mathbf{A} \mathbf{Y}) = \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}.$$

Demonstração 1-4: Imediatas.

Demonstração 5: ??

Demonstração 6: ??

# Distribuições condicionais

Seja  $\mathbf{Y}_{(p \times 1)}$  um vetor aleatório particionado da seguinte forma:

$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)'$ , onde  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$  são vetores de dimensões  $p_1 \times 1$  e  $p_2 \times 1$ , respectivamente, com  $p = p_1 + p_2$ . Com essa partição, temos que

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

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

Então,  $\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ , onde

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_2 - \boldsymbol{\mu}_2) \quad \text{and} \quad \boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.$$

# Derivadas matriciais úteis

Sejam  $\mathbf{A}_{(m \times n)}$  e  $\mathbf{x}_{(n \times 1)}$ , então

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



# Estimadores de máxima verossimilhança

Estimadores de máxima verossimilhança (dada uma amostra aleatória):

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i$$

e

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \overline{\mathbf{Y}})' (\mathbf{Y}_i - \overline{\mathbf{Y}}) .$$

- Notas de aula do Prof. Caio Azevedo.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.