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) $\boldsymbol{Y} = (Y_1, ..., Y_p)$ 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 $f_{\boldsymbol{Y}}(\boldsymbol{y})$.

A função de distribuição acumulada é denotada por

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

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

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

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

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

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

A i-ésima linha de $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ seja positiva semidefinida.

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

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

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

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

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

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

Funções lineares de vetores aleatórios

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

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

Função geradora de momentos e característica

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

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

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

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

A **função característica** é dada por

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

Distribuição Normal multivariada

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

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

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

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

Queremos demonstrar que

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

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

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

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

Além disso,

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

Assim,

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

Função geradora de momentos

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

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

Prova: ??

Obtenção das Marginais

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

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 $\boldsymbol{A}_{(q \times p)}$ for uma matriz não aleatória, então

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

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

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

Propiedades

5. Se $\boldsymbol{A}_{(p\times p)}$ for uma matriz não aleatória, simétrica e idempotente de rank = p, $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_{(p\times p)}$, então

$$V = \frac{1}{\sigma^2}\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} \sim \chi^2_r, \ r = tr(\boldsymbol{A}).$$

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

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

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

Demonstração 1-4: Imediatas.

Demonstração 5: ??

Demonstração 6: ??

Distribuições condicionais

Seja $\boldsymbol{Y}_{(p\times 1)}$ um vetor aleatório particionado da seguinte forma: $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)'$, onde $\boldsymbol{Y}_1$ e $\boldsymbol{Y}_2$ 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, $\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, onde

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

Derviadas matriciais úteis

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

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

Estimadores de máxima verossimilhança

Estimadores de máxima verossimilhança (dada uma amostra aleatória):

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

e

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

Referência

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