Introdução à Estatística Multivariada

Disciplina: Modelagem Estatística
Instrutor: Luiz Max Carvalho
Monitores: Eduardo Adame & Ezequiel Braga

Março 2024

Introdução

A ideia desse documento é auxiliar na transição para a Estatística Multivariada. É importante estar bem afiado em Álgebra Linear e (obviamente) nos conteúdos vistos em Inferência Estatística. As referências principais serão o Petersen e Pedersen (2008) e o Soch et al. (2024).

Vetores Aleatórios

Ao invés de lidarmos com uma variável aleatória $X \in \mathcal{X} \subseteq \mathbb{R}$, com $\mu = \mathbb{E}[X]$ e $\sigma^2 = \mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$, introduzimos agora vetores aleatórios. Por exemplo, $\boldsymbol{X} \in \mathcal{X}^n$, com **vetor de média $\boldsymbol{\mu}$** tal que $\mu_i = \mathbb{E}[X_i]$ (denotamos $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$) e **matriz de covariância** $\Sigma = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right]$, que vem de um produto externo. Lembre-se que perdemos comutatividade ao trabalhar em $\mathcal{X}^{n \times n}$.

Formas Quadráticas

Uma forma quadrática é uma expressão polinomial em que cada termo possui grau 2. Por exemplo, $y_1^2 + y_2^2$ e $2y_1^2 + y_2^2 + 3y_1y_2$ são formas quadráticas em y_1 e y_2, mas $y_1^2 + y_2^2 + 2y_1$ e $y_1^2 + 3y_2^2 + 2$ não são.

Seja $\mathbf{A}$ uma matriz simétrica. Então, a expressão $\mathbf{y}^\top \mathbf{A} \mathbf{y} = \sum_i \sum_j a_{ij} y_i y_j$ é uma forma quadrática nos y_i's. Analogamente, a expressão $(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ é uma forma quadrática nos termos $(y_i - \mu_i)$. Quando $\mathbf{y}^\top \mathbf{A} \mathbf{y} > 0$, $\forall \mathbf{y} \neq \mathbf{0}$, diz-se que a forma quadrática (e a matriz $\mathbf{A}$) é positiva definida. Além disso, o posto da matrix $\mathbf{A}$ é chamado de número de graus de liberdade da forma quadrática.

Teorema (Cochran)**.** *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da distribuição normal com média 0 e varância σ^2. Sejam $Q = \sum_{i=1}^n X_i^2$ e $Q_1, \ldots, Q_k$ formas quadráticas tais que $Q = \sum_{i=1}^k Q_i$, onde Q_i possui m_i graus de liberdade. Então, $Q_1, \ldots, Q_k$ são variáveis aleatórias independentes com $\frac{Q_i}{\sigma^2} \sim \chi^2(m_i)$ $(i = 1, \ldots, k)$ se, e somente se, $\sum_{i=1}^k m_i = n$.*

Uma consequênsia do toerema acima é que se $X_1^2 \sim \chi^2(m)$ e $X_2^2 \sim \chi^2(k)$ são independentes, então, $X^2 = X_1^2 - X_2^2 \sim \chi^2(m-k)$, desde que $X^2 \geq 0$ e $m > k$.

Identidades Relevantes

Seja $\boldsymbol{X}$ variável aleatória com média $\boldsymbol{\mu}$ e matriz de covariância Σ

- $\mathbb{E}[A\boldsymbol{X} + B] = A\mu + B$
- $\mathbb{V}[A\boldsymbol{X}] = A\Sigma A^\top$
- $\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top] = \Sigma + \mu\mu^\top$

Matrizes de Covariância

A noção de variãncia passa a ser representada, em geral, em termos de covariância. Sabemos desde a parte univariadada que $\text{Cov}[X, X] = \mathbb{V}[X]$ e isso continua valendo aqui.

Sabemos também que essa matriz será, ao menos, semi-definida positiva, pois $\forall\, w \in \mathbb{R}^n$ temos que:

$$\begin{aligned}
\mathbb{V}[w^\top \boldsymbol{X}] = w^\top \mathbb{V}[\boldsymbol{X}] w = w^\top \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right] w, \\
= \mathbb{E}\left[w^\top (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top w\right], \\
= \mathbb{E}\left[\left[w^\top (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])\right]^2\right] \geq 0.
\end{aligned}$$

Principais Distribuições Multivariadas

Multivariada Normal

Seja $\boldsymbol{X} \in \mathbb{R}^n$. Então, $\boldsymbol{X}$ é normalmente distribuido com média $\boldsymbol{\mu} \in \mathbb{R}^n$ e matriz de covariância $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ ($\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) se, e somente se, sua p.d.f é dada por

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n \mid \boldsymbol{\Sigma} \mid}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mu)\right].$$

Dirichlet (Generalização da Beta)

Seja $\boldsymbol{X} \in \mathbb{R}^n$. Então, $\boldsymbol{X}$ segue a distribuição Dirichlet com parâmetros de concentração $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n]$ ($\boldsymbol{X} \sim \text{Dir}(\boldsymbol{\alpha})$) se, e somente se, sua p.d.f é dada por

$$f(\mathbf{x}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^n \alpha_i}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n x_i^{\alpha_i - 1},$$

onde $\alpha_i > 0$, $\forall i = 1, \ldots, n$ e a densidade é zero se $x_i \notin [0, 1]$ para algum $i = 1, \ldots, n$ ou $\sum_{i=1}^n x_i \neq 1$.

Multinomial (Generalização da Binomial)

Seja $\boldsymbol{X} \in \mathbb{R}^k$. Então, $\boldsymbol{X}$ segue a distribuição multinomial com número de tentativas n e probabilidades $p_1, \ldots, p_k$ ($\boldsymbol{X} \sim \text{Mult}(n, [p_1, \ldots, p_k])$) se $\boldsymbol{X}$ é o número de observações pertencentes a k categorias distintas em n ensaios independentes, cuja densidade é dada por

$$f(\mathbf{x}; n, p_1, \ldots, p_k) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i},$$

onde $\sum_{i=1}^k x_i = n$; $x_i \geq 0, \forall i = 1, \ldots, k$; e $\sum_{i=1}^k p_i = 1$.

Pareto

Seja $\boldsymbol{X} \in \mathbb{R}^k$. Então, $\boldsymbol{X}$ segue a distribuição de Pareto com parâmetros $a = (a_1, \ldots, a_k)$ e p se sua p.d.f é dada por

$$f(\mathbf{x}; a, p) = \frac{p(p+1)\ldots(p+k-1)}{\left(\prod_{i=1}^k a_i\right)\left[(\sum_{i=1}^k a_i^{-1} x_i) - k + 1\right]^{p+k}},$$

para $x_i > a_i > 0,\ i = 1, \ldots, k,\ p > 0$.

Normal Matricial

Seja $\boldsymbol{X} \in \mathbb{R}^{n \times p}$. Então, $\boldsymbol{X}$ segue a distribuição normal com média M, covariância entre linhas U e covariância entre colunas V ($\boldsymbol{X} \sim \mathcal{MN}(M, U, V)$) se, e somente se, sua p.d.f é dada por

$$f(X; M, U, V) = \frac{1}{\sqrt{(2\pi)^{np} \mid V \mid^n \mid U \mid^p}} \exp\left[-\frac{1}{2} tr(V^{-1}(X-M)^\top U^{-1}(X-M)\right].$$

Wishart (Generalização da Gamma)

Seja $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, com $\boldsymbol{X} \sim \mathcal{MN}(0, I_n, V)$. Tome $S = \boldsymbol{X}^\top X$. Então, S segue a distribuição de Wishart com matriz de escala V e n graus de liberdade ($S \sim \mathcal{W}(V, n)$), com $n > p - 1$ e V simétrica positiva definida. Sua p.d.f é dada por

$$f(X; V, n) = \frac{1}{2^{np/2} \mid V \mid^{n/2} \Gamma_p(\frac{n}{2})} \mid X \mid^{(n-p-1)/2} \exp\left(-\frac{1}{2} tr(V^{-1}X)\right),$$

onde $\Gamma_p(\frac{n}{2}) = \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma(\frac{n}{2} - \frac{j-1}{2})$.

Recursos para manipulação algébrica

Completar Quadrado Multidimensional

Tome $\boldsymbol{X}$ matriz simétrica positiva definida $d \times d$ e $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Vale que

$$\mathbf{u}^\top X\mathbf{u} - 2\mathbf{v}^\top\mathbf{u} = \left(\mathbf{u} - X^{-1}\mathbf{v}\right)^\top X\left(\mathbf{u} - X^{-1}\mathbf{v}\right) - \mathbf{v}^\top X^{-1}\mathbf{v}.$$

Para verificar isso, basta expandir a forma quadrática e usar a simetria da matriz X:

$$\begin{aligned}
\left(\mathbf{u} - X^{-1}\mathbf{v}\right)^\top X\left(\mathbf{u} - X^{-1}\mathbf{v}\right) &= \left(\mathbf{u}^\top X - \mathbf{v}^\top\left(X^\top\right)^{-1} X\right)\left(\mathbf{u} - X^{-1}\mathbf{v}\right), \\
&= \left(\mathbf{u}^\top X - \mathbf{v}^\top\right)\left(\mathbf{u} - X^{-1}\mathbf{v}\right), \\
&= \mathbf{u}^\top X\mathbf{u} - \mathbf{u}^\top\mathbf{v} - \mathbf{v}^\top\mathbf{u} + \mathbf{v}^\top X^{-1}\mathbf{v}, \\
&= \mathbf{u}^\top X\mathbf{u} - 2\mathbf{u}^\top\mathbf{v} + \mathbf{v}^\top X^{-1}\mathbf{v}.
\end{aligned}$$

Distribuição Normal Multivariada

Distribuição Marginal

Seja $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com $\boldsymbol{\mu} \in \mathbb{R}^d$ e $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. Suponha que $\boldsymbol{X}$ pode ser particionado de forma que tenha-se

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix},\ \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix},\ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_2^\top & \boldsymbol{\Sigma}_3 \end{bmatrix}$$

com $\boldsymbol{\mu}_1 \in \mathbb{R}^k$, $\boldsymbol{\mu}_2 \in \mathbb{R}^{d-k}$, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{k\times k}$, $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{k\times(d-k)}$ e $\boldsymbol{\Sigma}_3 \in \mathbb{R}^{(d-k)\times(d-k)}$, para $k < d$. Então, $\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ e $\boldsymbol{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$.

Distribuição Condicional

Considere $\boldsymbol{X}$ como anteriormente. Então,

$$\begin{aligned}
\boldsymbol{X}_1 \mid \boldsymbol{X}_2 \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1) \text{ com } &\begin{cases} \hat{\boldsymbol{\mu}}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_3^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2) \\ \hat{\boldsymbol{\Sigma}}_1 = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_3^{-1}\boldsymbol{\Sigma}_2^\top \end{cases}, \\
\boldsymbol{X}_2 \mid \boldsymbol{X}_1 \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2) \text{ com } &\begin{cases} \hat{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1) \\ \hat{\boldsymbol{\Sigma}}_2 = \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2^\top \end{cases}.
\end{aligned}$$

Combinação Linear

Tome $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ e $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$. Então,

$$A\boldsymbol{X} + B\boldsymbol{Y} + \mathbf{c} \sim \mathcal{N}\left(A\boldsymbol{\mu}_X + B\boldsymbol{\mu}_Y + \mathbf{c},\ A\boldsymbol{\Sigma}_X A^\top + B\boldsymbol{\Sigma}_Y B^\top\right).$$

Referências

Petersen, K. B. e M. S. Pedersen (out. de 2008). *The Matrix Cookbook*. Version 20081110. URL: http://www2.imm.dtu.dk/pubdb/p.php?3274.

Soch, Joram et al. (jan. de 2024). *StatProofBook/StatProofBook.github.io: StatProofBook 2023*. Versão 2023. DOI: 10.5281/zenodo.10495684. URL: https://doi.org/10.5281/zenodo.10495684.