Distribuições Multivariadas

L. F. Bossa

UFSC

24/03/25

Section 1

Variável Aleatória unidimensional

Variável Aleatória unidimensional

Uma variável aleatória X é uma função que associa um número real a cada resultado de um experimento aleatório. Podemos caracterizar uma variável aleatória por sua função densidade de probabilidade $f(x)$, que satisfaz a propriedade

$$\mathbb{P}(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Esperança

Também conhecido como média e denotado por μ.

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x) dx$$

Em geral, dada $g : \mathbb{R} \to \mathbb{R}$ temos

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx$$

Variância

$$\mathrm{Var}(X) = \mathbb{E}((X - \mu)^2)$$

Mede o quão dispersos estão os dados de X: quanto maior a variância, mais longe da média estão os dados.

A raiz quadrada da variância é o *desvio-padrão*, geralmente denotado por σ.

Variáveis Gaussianas

São caracterizadas pela sua função de densidade

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma}\right)$$

Uma variável aleatória X com média μ e variância σ^2 é denotada por $X \sim N(\mu, \sigma^2)$.

Z-score

Dada uma variável aleatória $X \sim N(\mu, \sigma^2)$, podemos fazer uma transformação

$$Z = \frac{X - \mu}{\sigma}$$

de modo que $Z \sim N(0, 1)$.

Section 2

Variável Aleatória em várias dimensões

Motivação

- Dificilmente uma coleta de dados vai coletar apenas um dado de cada amostra estudada
- Fazer uma análise conjunta dos dados permite usar ferramentas de álgebra linear (análise de componentes principais, clustering)

- Temos um vetor $\vec{X} = (X_1, \ldots, X_n)$ cujas componentes são variáveis aleatórias.
- Temos uma função de densidade $f(x_1, \ldots, x_n)$ de modo que

$$\mathbb{P}(X_1 \leq a_a, \ldots, X_n \leq a_n) = \int_{-\infty}^{a_1} \ldots \int_{-\infty}^{a_n} f(x_1, \ldots, x_n) dx_n \ldots dx_1$$

Probabilidade marginal

Surge quando queremos estudar a distribuição geral de uma variável só, "ignorando"as outras. Nesse caso, integramos a variável de interesse no intervalo de interesse, e as outras variáveis são integradas em todo $\mathbb{R}$.

$$\mathbb{P}(X_1 \leq a_1) = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \dots dx_1$$

Podemos definir então as densidades marginais

$$f_1(x) = \int_{-\infty}^{x} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \dots dx_1$$

Esperança

No caso vetorial, o operador esperança atua entrada-a-entrada

$$\mathbb{E}(\vec{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))$$

em que cada esperança é calculada com respeito à distribuição marginal de cada variável.

Propriedades da esperança

- Sendo c um vetor de constantes,

$$\mathbb{E}(c^\top \vec{X}) = c^\top \vec{\mu}$$

- Sendo A uma matriz de constantes,

$$\mathbb{E}(A\vec{X}) = A\vec{\mu}$$

Covariância

Dadas variáveis X, Y unidimensionais com médias μ_X, μ_Y, definimos a covariância de X e Y como

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y))$$

Por vezes também é usada a notação σ_{XY} para denotar a covariância.

A covariância indica o quanto X e Y estão *linearmente* relacionadas.

Note que

$$\mathrm{Cov}(X, X) = \mathbb{E}((X - \mu_X)(X - \mu_X)) = \mathbb{E}((X - \mu_X)^2) = \mathrm{Var}(X)$$

Independência estatística

Duas variáveis X_1 e X_2 são ditas independentes se for possível escrever a função de densidade como $f(x_1, x_2) = f_1(x_1)f_2(x_2)$

Teorema

Se X_1 e X_2 são independentes, então $\text{Cov}(X, Y) = 0$

Contra-exemplo

A recíproca não é verdadeira. Considere X uma variável aleatória com média zero e densidade $f(x)$ sendo uma função par. Defina $Y = X^2$. Nesse caso, claramente Y e X não são independentes, mas

$$\mathrm{Cov}(X, Y) = 0$$

Correlação

É a covariância "normalizada"

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y},$$

com σ_X e σ_Y sendo os desvios-padrão de X e Y.

A correlação é também denotada por ρ_{XY}.

- Por quê é normalizada?

Matriz de Variância-Covariância

Para calcular a variância de uma variável unidimensional, temos que calcular

$$\mathbb{E}((X - \mu_X)^2)$$

Por causa do termo quadrático, temos que adaptar esse cálculo para o caso vetorial.

Sendo $\vec{X} = (X_1, \ldots, X_n)$ uma variável aleatória e seja $\mathbb{E}(\vec{X}) = \vec{\mu}_X$ seu vetor-médio. Calculamos a variância de $\vec{X}$ como

$$\Sigma = \mathbb{E}((\vec{X} - \vec{\mu}_X)(\vec{X} - \vec{\mu}_X)^\top)$$

$$\Sigma = \mathbb{E}\begin{pmatrix} (X_1-\mu_{X_1})(X_1-\mu_{X_1}) & (X_1-\mu_{X_1})(X_2-\mu_{X_2}) & \dots & (X_1-\mu_{X_1})(X_n-\mu_{X_n}) \\ (X_2-\mu_{X_2})(X_1-\mu_{X_1}) & (X_2-\mu_{X_2})(X_2-\mu_{X_2}) & \dots & (X_2-\mu_{X_2})(X_n-\mu_{X_n}) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n-\mu_{X_n})(X_1-\mu_{X_1}) & (X_n-\mu_{X_n})(X_2-\mu_{X_2}) & \dots & (X_n-\mu_{X_n})(X_n-\mu_{X_n}) \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) & \dots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \dots & \mathrm{Var}(X_n) \end{pmatrix}$$

Teorema

A matriz de variância-covariância é simétrica definida positiva.

Teorema

A matriz de variância-covariância é simétrica definida positiva.

Para simplificar os cálculos, seja $\vec{W} = \vec{X} - \vec{\mu}_X$. A matriz de variância-covariância pode ser reescrita como

$$\Sigma = \mathbb{E}(\vec{W}\vec{W}^\top)$$

Dado um vetor x qualquer, note que, por linearidade da esperança,

$$x^\top \Sigma x = x^\top \mathbb{E}(\vec{W}\vec{W}^\top)x = \mathbb{E}((\vec{W}x)^\top \vec{W}x) = \mathbb{E}(\|\vec{W}x\|) \geq 0.$$

Teorema

Dada uma matriz Σ simétrica definida positiva, existe um vetor aleatório $\vec{X}$ tal que Σ é sua matriz de variância-covariância.

Teorema

Dada uma matriz Σ simétrica definida positiva, existe um vetor aleatório $\vec{X}$ tal que Σ é sua matriz de variância-covariância.

Da teoria de diagonalização, sabemos que toda matriz simétrica definida positiva pode ser diagonalizada por uma matriz ortogonal. Assim, podemos escrever

$$\Sigma = QDQ^\top$$

em que Q é uma matriz ortogonal ($Q^\top Q = I$) e D é uma matriz diagonal $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ com λ_i sendo os autovalores de Σ.

Escreva $\sqrt{D} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. Note então que $\sqrt{D}$ é simétrica e

$$\Sigma = Q\sqrt{D}\sqrt{D}Q^\top = (Q\sqrt{D})(Q\sqrt{D})^\top$$

Sejam $Y_i \sim N(0,1)$ variáveis aleatórias independentes, para $i \in \{1, \dots n\}$, e forme $\vec{Y} = (Y_1, \dots, Y_n)$. Finalmente, faça $\vec{X} = Q\sqrt{D}\vec{Y}$, e vamos mostrar que $\vec{X}$ tem matriz de variância-covariância Σ.

$$
\begin{aligned}
\mathbb{E}(\vec{X}\vec{X}^\top) &= \mathbb{E}(Q\sqrt{D}\vec{Y}(Q\sqrt{D}\vec{Y})^\top) = \mathbb{E}(Q\sqrt{D}\vec{Y}\vec{Y}^\top(Q\sqrt{D})^\top) \\
&= Q\sqrt{D}\mathbb{E}(\vec{Y}\vec{Y}^\top)\sqrt{D}Q^\top = Q\sqrt{D}\sqrt{D}Q^\top = \Sigma
\end{aligned}
$$

Como resumir a variância?

Por vezes, não queremos apresentar a variância como uma matriz, mas como um número que resuma o quão dispersos estão nossos dados.

Para isso, temos a

Variância Total

$$\mathrm{tr}(\Sigma) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Variância Generalizada

$$\det(\Sigma)$$

Distância Mahalanobis

Dados vetores X e Y que pertençam a uma distribuição multivariada com matriz de variância-covariância Σ, definimos a distância Mahalanobi entre eles como

$$d_{XY}^2 = (X - Y)^\top \Sigma^{-1} (X - Y)$$

Essa é a generalização do Z-score.

Variáveis gaussianas multivariadas

Finalmente estamos aptos a entender a fórmula para a distribuição gaussiana multivariada.

Para tal, vamos reinterpretar a fórmula da gaussiana em 1 dimensão

A gaussiana padrão é dada por

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

Podemos aplicar uma mudança de variáveis para alterar seu centro e dispersão, lembrando fazer uma escala para que a integral sobre $\mathbb{R}$ continue sendo 1.

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Mas note que termos quadráticos não generalizam bem para vetores. Para tanto, vamos reescrever o lado direito como

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right)$$

Nessa expressão fica mais claro que podemos substituir σ^2 pela matriz de covariância, $x - \mu$ por vetores, e o termo de escala por um determinante.

Nesse caso, temos a distribuição gaussiana multivariada com média $\vec{\mu}$ e matriz de variância-covariância Σ, cuja expressão é dada por

$$f(\vec{X}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\left(\vec{X} - \vec{\mu}\right)^{\top} \Sigma^{-1} \left(\vec{X} - \vec{\mu}\right)\right)$$

Section 3

Análise de componentes principais

Análise de Componentes Principais

- Podemos usar a matriz de variância-covariância para encontrar a direção de maior variância dos dados.
- Embora tenhamos n direções nas quais os dados variam, com uma mudança de coordenadas, podemos encontrar as direções que concentram boa parte da variação dos dados.

Diagonalização

Lembre-se que Σ pode ser escrita como

$$\Sigma = QDQ^\top$$

em que Q é uma matriz ortogonal e D é uma matriz diagonal.

Faça então a transformação $\vec{Y} = Q^\top \vec{X}$.

Note que a matriz de variância-covariância de $\vec{Y}$ é dada por

$$\begin{aligned}\mathbb{E}(\vec{Y}\vec{Y}^\top) &= \mathbb{E}(Q^\top \vec{X}\vec{X}^\top Q) = Q^\top \mathbb{E}(\vec{X}\vec{X}^\top)Q = Q^\top \Sigma Q \\ &= Q^\top QDQ^\top Q = D\end{aligned}$$

Veja que $\vec{Y}$ é uma variável aleatória com covariâncias zeradas.

Note que a matriz de variância-covariância de $\vec{Y}$ e de $\vec{X}$ são similares, então podemos usar a propriedade de invariância do traço e do determinante e obtemos

$$\mathrm{tr}(\Sigma) = \mathrm{tr}(D) = \sum_{i=1}^{n} \lambda_i \qquad \text{(soma dos aulovalores)}$$

$$\det(\Sigma) = \det(D) = \prod_{i=1}^{n} \lambda_i \qquad \text{(produto dos autovalores)}$$

Ou seja, a variância total e a variância generalizada se mantém inalteradas quando mudamos nossa análise de $\vec{X}$ para $\vec{Y}$.

Exemplos interativos

$\nearrow$ Aplicação no Geogebra