

Probabilidade (PPGECD000000001)

Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD)

Sessão 6

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

Variáveis Aleatórias Multidimensionais

Muitas vezes estamos interessados na descrição probabilística de mais de um característico numérico de um experimento aleatório. Por exemplo, podemos estar interessados na distribuição de alturas e pesos de indivíduos de uma certa classe. Para tanto precisamos estender a definição de variável aleatória para o caso multidimensional.

Definição 1

Seja (Ω, \mathcal{A}, P) um espaço de probabilidade. Uma função $\vec{X} : \Omega \rightarrow \mathbb{R}^n$ é chamada de um vetor aleatório se para todo evento B Boreliano de \mathbb{R}^n , $\vec{X}^{-1}(B) \in \mathcal{A}$.

Onde um evento é Boreliano em \mathbb{R}^n pertence a menor σ -álgebra que contém todas as regiões da seguinte forma: $C_{\vec{a}} = \{(X_1, X_2, \dots, X_n) : X_i \leq a_i, 1 \leq i \leq n\}$.

Dado um vetor aleatório \vec{X} , pode-se definir uma probabilidade induzida $P_{\vec{X}}$ no espaço mensurável $(\mathbb{R}^n, \mathcal{B}^n)$ da seguinte maneira: para todo $A \in \mathcal{B}^n$, definimos $P_{\vec{X}}(A) = P(\vec{X}^{-1}(A))$. Por definição de vetor aleatório, tem-se que $\vec{X}^{-1}(A) \in \mathcal{A}$, então $P_{\vec{X}}$ está bem definida. Para um vetor aleatório \vec{X} , uma maneira simples e básica de

descrever a probabilidade induzida $P_{\vec{X}}$ é utilizando sua *função de distribuição acumulada conjunta*.

Função de Distribuição Acumulada Conjunta

Definição 2

A função de distribuição acumulada conjunta de um vetor aleatório \vec{X} , representada por $F_{\vec{X}}$ ou simplesmente por F , é definida por

$$\begin{aligned} F_{\vec{X}}(\vec{x}) &= P(C_{\vec{x}}) \\ &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \forall \vec{x} \in \mathbb{R}^n. \end{aligned}$$

Propriedades da Função de Distribuição Acumulada Conjunta e Cálculo da Probabilidade em um Hipercubo n -dimensional

Consideremos um vetor aleatório $\vec{X} = (X_1, X_2, \dots, X_n)$ definido em um espaço de probabilidade, com função de distribuição acumulada conjunta dada por:

$$F_{\vec{X}}(\vec{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad \forall \vec{x} \in \mathbb{R}^n.$$

A função $F_{\vec{X}}$ satisfaz as seguintes propriedades:

F1. (Monotonicidade) Se $x_i \leq y_i, \forall i \leq n$, então $F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y})$.

Prova: Para cada i , $x_i \leq y_i$ implica que o evento $\{X_i \leq x_i\}$ está contido em $\{X_i \leq y_i\}$, ou seja:

$$\{X_i \leq x_i\} \subseteq \{X_i \leq y_i\}.$$

Considerando a interseção desses eventos para todos os i :

$$C_{\vec{x}} = \bigcap_{i=1}^n \{X_i \leq x_i\} \subseteq \bigcap_{i=1}^n \{X_i \leq y_i\} = C_{\vec{y}}.$$

Como $C_{\vec{x}} \subseteq C_{\vec{y}}$ e a probabilidade é uma medida não negativa e aditiva, temos:

$$P(C_{\vec{x}}) \leq P(C_{\vec{y}}),$$

o que implica:

$$F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y}).$$

F2. (Continuidade à Direita em cada variável (coordenada)) A função $F_{\vec{X}}(\vec{x})$ é contínua à direita em cada uma das variáveis, ou seja, para cada $i \leq n$ e para toda sequência $y_m \downarrow x_i$ (com $y_m > x_i$ e $y_m \rightarrow x_i$):

$$\lim_{y_m \downarrow x_i} F_{\vec{X}}(x_1, \dots, x_{i-1}, y_m, x_{i+1}, \dots, x_n) = F_{\vec{X}}(\vec{x}).$$

Prova: Considere uma sequência y_m tal que $y_m \downarrow x_i$. Para cada m , temos:

$$\{X_i \leq x_i\} \subseteq \{X_i \leq y_m\}.$$

Assim, os eventos $C_m = \bigcap_{j \neq i} \{X_j \leq x_j\} \cap \{X_i \leq y_m\}$ formam uma sequência decrescente de conjuntos $C_{m+1} \subseteq C_m$.

Pela continuidade da medida de probabilidade em seqüências decrescentes de conjuntos (continuidade monótona), temos:

$$\lim_{m \rightarrow \infty} P(C_m) = P \left(\bigcap_{j \neq i} \{X_j \leq x_j\} \cap \{X_i \leq x_i\} \right) = F_{\vec{x}}(\vec{x}).$$

Portanto:

$$\lim_{y_m \downarrow x_i} F_{\vec{x}}(x_1, \dots, x_{i-1}, y_m, x_{i+1}, \dots, x_n) = F_{\vec{x}}(\vec{x}).$$

F3a. (Comportamento em $x_i \rightarrow -\infty$) Se para algum $i \leq n$, $x_i \rightarrow -\infty$, então:

$$\lim_{x_i \rightarrow -\infty} F_{\vec{x}}(\vec{x}) = 0.$$

Prova: À medida que $x_i \rightarrow -\infty$, o evento $\{X_i \leq x_i\}$ se torna o conjunto vazio, pois X_i não pode assumir valores menores do que $-\infty$. Assim, a interseção dos eventos é

$$C_{\vec{x}} = \bigcap_{i=1}^n \{X_i \leq x_i\} = \emptyset.$$

Como a probabilidade do conjunto vazio é zero: $F_{\vec{x}}(\vec{x}) = P(\emptyset) = 0$. Portanto:

$$\lim_{x_i \rightarrow -\infty} F_{\vec{x}}(\vec{x}) = 0.$$

F3b. (Comportamento em $x_i \rightarrow \infty$) Se $x_i \rightarrow \infty$, então:

$$\lim_{x_i \rightarrow \infty} F_{\vec{X}}(\vec{x}) = F_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Prova: À medida que $x_i \rightarrow \infty$, o evento $\{X_i \leq x_i\}$ se torna o espaço inteiro, pois $X_i \leq \infty$ é sempre verdadeiro. Assim, a restrição sobre X_i é removida, e temos:

$$\bigcap_{j=1}^n \{X_j \leq x_j\} = \left(\bigcap_{j \neq i} \{X_j \leq x_j\} \right) \cap \{X_i \leq x_i\} \xrightarrow{x_i \rightarrow \infty} \bigcap_{j \neq i} \{X_j \leq x_j\}.$$

Portanto, a função de distribuição tende a:

$$\lim_{x_i \rightarrow \infty} F_{\vec{X}}(\vec{x}) = P \left(\bigcap_{j \neq i} \{X_j \leq x_j\} \right) = F_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Isso mostra que a função de distribuição conjunta de n variáveis tende para a função de distribuição conjunta das $n - 1$ variáveis restantes quando $x_i \rightarrow \infty$.

Observação: Em particular, quando todos os $x_i \rightarrow \infty$, temos:

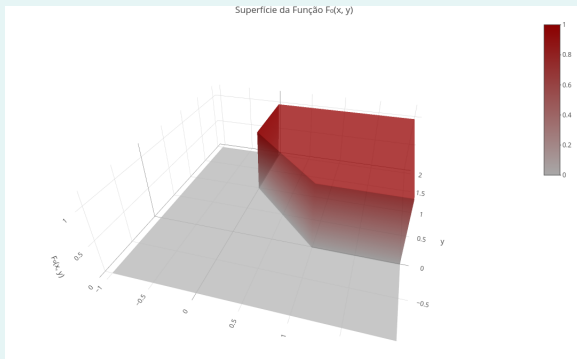
$$\lim_{\vec{x} \rightarrow \infty} F_{\vec{X}}(\vec{x}) = 1,$$

pois estamos considerando o evento certo (todo o espaço de probabilidade).

O próximo exemplo mostra que para $n \geq 2$ as propriedades F1, F2, e F3 não são suficientes para que F seja uma função de distribuição.

Exemplo 1

Seja $F_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ uma função definida no plano tal que $F_0(x, y) = 1$ se $x \geq 0, y \geq 0$, e $x + y \geq 1$, e $F_0(x, y) = 0$, caso contrário.



F1. Se $x_1 \leq x_2$ e $y_1 \leq y_2$, então:

$$F_0(x_1, y_1) \leq F_0(x_2, y_2)$$

- Caso 1: Se $F_0(x_1, y_1) = 0$, então $F_0(x_1, y_1) \leq F_0(x_2, y_2)$, pois $F_0(x_2, y_2) \geq 0$.
- Caso 2: Se $F_0(x_1, y_1) = 1$, então temos que $x_1 \geq 0$, $y_1 \geq 0$ e $x_1 + y_1 \geq 1$.

Como $x_1 \leq x_2$ e $y_1 \leq y_2$, temos que: $x_2 \geq x_1 \geq 0$, $y_2 \geq y_1 \geq 0$ e $x_2 + y_2 \geq x_1 + y_1 \geq 1$. Portanto, $F_0(x_2, y_2) = 1$. Assim, em ambos os casos, $F_0(x_1, y_1) \leq F_0(x_2, y_2)$, o que prova a propriedade F1.

F2. Analisemos a continuidade pela direita em x :

- Se $x_0 < 0$: Como $F_0(x_0, y) = 0$ para qualquer y , e $F_0(x, y) = 0$ para x próximo de x_0 , a função é constante e, portanto, contínua.
- Se $x_0 + y \geq 1$: Então $F_0(x, y) = 1$ para x próximo de x_0 à direita, mantendo y fixo. Portanto, o limite pela direita é $1 = F_0(x_0, y)$.
- Se $x_0 + y < 1$: Como $x \downarrow x_0$, mas $x_0 + y < 1$, $F_0(x, y) = 0$, e $F_0(x_0, y) = 0$. Logo, a função é contínua pela direita.

O mesmo raciocínio vale para y . Portanto, $F_0(x, y)$ é contínua pela direita em cada variável, comprovando a propriedade F2.

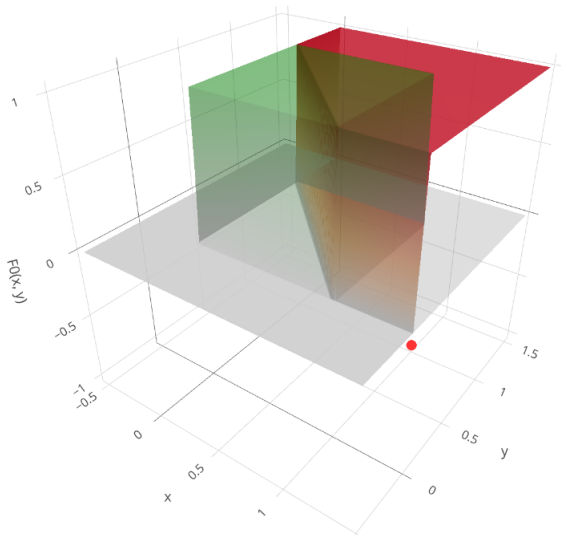
F3. Quando $x \rightarrow -\infty$ ou $y \rightarrow -\infty$, temos:

$$\lim_{x \rightarrow -\infty} F_0(x, y) = \lim_{y \rightarrow -\infty} F_0(x, y) = 0$$

Como $F_0(x, y) = 0$ para $x < 0$ ou $y < 0$, quando $x \rightarrow -\infty$ ou $y \rightarrow -\infty$, a função tende a zero. Agora, para $x \rightarrow \infty$, mantendo y fixo e $y \geq 0$: Se $y \geq 0$ e $x + y \geq 1$, então $F_0(x, y) = 1$. Logo, $\lim_{x \rightarrow \infty} F_0(x, y) = 1$ para $y \geq 0$. Similarmente para $y \rightarrow \infty$. Portanto, a função F_0 possui limites adequados nos infinitos, satisfazendo a propriedade F3.

É claro que F1, F2, e F3 são satisfeitas, mas F_0 não é função de distribuição de nenhum vetor aleatório (X, Y) . De fato, calculemos a probabilidade do retângulo bidimensional

$$\begin{aligned} 0 &\leq P(0 < X \leq 1, 0 < Y \leq 1) \\ &= F_0(1, 1) - F_0(1, 0) - F_0(0, 1) + F_0(0, 0) \\ &= 1 - 1 - 1 + 0 = -1 \end{aligned}$$



F4. (Cálculo da Probabilidade em um Hipercubo n -dimensional) Calcular a probabilidade de um evento retangular definido em termos de um vetor aleatório bidimensional $\vec{X} = (X_1, X_2)$ e generalizar o resultado para um vetor aleatório n -dimensional $\vec{X} = (X_1, X_2, \dots, X_n)$.

Prova: Caso Bidimensional Considere um espaço de probabilidade (Ω, \mathcal{F}, P) e um vetor aleatório $\vec{X} : \Omega \rightarrow \mathbb{R}^2$. Definimos os seguintes eventos: Para $i = 1, 2$: $A_i(x_i) = \{X_i \leq x_i\}$; $B_i = A_i(y_i) \setminus A_i(x_i) = \{x_i < X_i \leq y_i\}$, com $x_i < y_i$
As probabilidades são:

$$P(B_i) = F_{X_i}(y_i) - F_{X_i}(x_i)$$

O evento retangular de interesse é:

$$B = B_1 \cap B_2 = \{x_1 < X_1 \leq y_1, x_2 < X_2 \leq y_2\}$$

Logo

$$P(B) = [F_{\vec{X}}(y_1, y_2) - F_{\vec{X}}(x_1, y_2)] - [F_{\vec{X}}(y_1, x_2) - F_{\vec{X}}(x_1, x_2)]$$

Para mostrar isto, podemos reescrever o evento B em termos de eventos cumulativos (Borelianos convenientes) da seguinte forma pelo princípio da inclusão-exclusão para eventos :

$$B = (X_1 \leq y_1, X_2 \leq y_2) \setminus [(X_1 \leq x_1, X_2 \leq y_2) \cup (X_1 \leq y_1, X_2 \leq x_2)] \cup (X_1 \leq x_1, X_2 \leq x_2)$$

Utilizando o princípio da inclusão-exclusão, podemos escrever:

$$P(B) = P(X_1 \leq y_1, X_2 \leq y_2) - P(X_1 \leq x_1, X_2 \leq y_2) - P(X_1 \leq y_1, X_2 \leq x_2) + P(X_1 \leq x_1, X_2 \leq x_2)$$

e substituindo as probabilidades pelas funções de distribuição acumulada conjunta temos:

$$P(B) = F_{\vec{X}}(y_1, y_2) - F_{\vec{X}}(x_1, y_2) - F_{\vec{X}}(y_1, x_2) + F_{\vec{X}}(x_1, x_2)$$

Podemos reorganizar a expressão de forma que

$$P(B) = [F_{\vec{X}}(y_1, y_2) - F_{\vec{X}}(x_1, y_2)] - [F_{\vec{X}}(y_1, x_2) - F_{\vec{X}}(x_1, x_2)]$$

Generalizemos para o caso n -dimensional:

Consideremos um vetor aleatório $\vec{X} = (X_1, X_2, \dots, X_n)$. Para cada $i = 1, 2, \dots, n$, definimos: $A_i(x_i) = \{X_i \leq x_i\}$, $B_i = A_i(y_i) \setminus A_i(x_i) = \{x_i < X_i \leq y_i\}$, com $x_i < y_i$. Desta forma, o evento retangular é:

$$B = \bigcap_{i=1}^n B_i = \{x_i < X_i \leq y_i, \forall i = 1, 2, \dots, n\}$$

Precisamos mostrar que

$$P(B) = \sum_{\vec{\delta} \in \{0,1\}^n} (-1)^{\delta_1 + \delta_2 + \dots + \delta_n} F_{\vec{X}}(t_1, t_2, \dots, t_n)$$

Para cada i : $t_i = x_i$ se $\delta_i = 1$ e $t_i = y_i$ se $\delta_i = 0$ em que a soma é sobre todos os 2^n vetores binários $\vec{\delta} = (\delta_1, \delta_2, \dots, \delta_n)$.

Note que $B_i = \{X_i \leq y_i\} \setminus \{X_i \leq x_i\}$ Portanto, B pode ser escrito como:

$$B = \left(\bigcap_{i=1}^n \{X_i \leq y_i\} \right) \setminus \left(\bigcup_{S \subseteq \{1, \dots, n\}, S \neq \emptyset} \left(\bigcap_{i \in S} \{X_i \leq x_i\} \cap \bigcap_{j \notin S} \{X_j \leq y_j\} \right) \right)$$

Se aplicamos o princípio da inclusão-exclusão para a probabilidade da união de eventos. A probabilidade do evento B é:

$$P(B) = P\left(\bigcap_{i=1}^n \{X_i \leq y_i\}\right) - \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(\bigcap_{l=1}^k \{X_{i_l} \leq x_{i_l}\} \cap \bigcap_{j \notin \{i_1, \dots, i_k\}} \{X_j \leq y_j\}\right)$$

Cada termo da soma pode ser escrito como uma função de distribuição acumulada conjunta avaliada em pontos específicos: Para cada subconjunto $S \subseteq \{1, 2, \dots, n\}$ definimos $t_i = x_i$ se $i \in S$ e $t_i = y_i$ se $i \notin S$. O tamanho do subconjunto S é $|S|$

Assim, a probabilidade pode ser escrita como:

$$P(B) = \sum_{S \subseteq \{1, 2, \dots, n\}} (-1)^{|S|} F_{\vec{X}}(t_1, t_2, \dots, t_n)$$

em que a soma é sobre todos os 2^n subconjuntos de $\{1, 2, \dots, n\}$.

Nota sobre os sinais: O expoente $|S|$ em $(-1)^{|S|}$ provém da aplicação do princípio da inclusão-exclusão, onde cada interseção de k eventos é somada ou subtraída de acordo com $(-1)^k$.

Para $n = 2$, os subconjuntos S são:

1. $S = \emptyset$, $|S| = 0$, $t_i = y_i$ para todos i :

$$(-1)^0 F_{\bar{X}}(y_1, y_2) = +F_{\bar{X}}(y_1, y_2)$$

2. $S = \{1\}$, $|S| = 1$, $t_1 = x_1$, $t_2 = y_2$:

$$(-1)^1 F_{\bar{X}}(x_1, y_2) = -F_{\bar{X}}(x_1, y_2)$$

3. $S = \{2\}$, $|S| = 1$, $t_1 = y_1$, $t_2 = x_2$:

$$(-1)^1 F_{\bar{X}}(y_1, x_2) = -F_{\bar{X}}(y_1, x_2)$$

4. $S = \{1, 2\}$, $|S| = 2$, $t_i = x_i$ para todos i :

$$(-1)^2 F_{\bar{X}}(x_1, x_2) = +F_{\bar{X}}(x_1, x_2)$$

Somando todos os termos:

$$P(B) = F_{\bar{X}}(y_1, y_2) - F_{\bar{X}}(x_1, y_2) - F_{\bar{X}}(y_1, x_2) + F_{\bar{X}}(x_1, x_2)$$

Que coincide com o resultado obtido no caso bidimensional.

Seguindo a mesma lógica, para n dimensões, a probabilidade é:

$$P(B) = \sum_{\vec{\delta} \in \{0,1\}^n} (-1)^{\delta_1 + \delta_2 + \dots + \delta_n} F_{\vec{X}}(t_1, t_2, \dots, t_n)$$

em que: $\vec{\delta} = (\delta_1, \delta_2, \dots, \delta_n)$ é um vetor binário representando se usamos x_i ou y_i em cada posição. Para cada i , temos $t_i = x_i$ se $\delta_i = 1$, $t_i = y_i$ se $\delta_i = 0$ e o sinal $(-1)^{\delta_1 + \delta_2 + \dots + \delta_n}$ resulta da aplicação do princípio da inclusão-exclusão.

Assim, a probabilidade do evento retangular B em n dimensões é expressa em termos das funções de distribuição acumulada conjunta avaliadas em todos os pontos possíveis combinando x_i e y_i , com os sinais determinados pelo número de x_i na combinação.

Note que se a propriedade F4 é válida temos que

$$P(x_1 < X_1 \leq x_1, \dots, x_n < X_n \leq y_n) = \sum_{\vec{\delta} \in \{0,1\}^n} (-1)^{\delta_1 + \delta_2 + \dots + \delta_n} F_{\vec{X}}(t_1, t_2, \dots, t_n) \geq 0$$

Distribuição marginal

A função de distribuição acumulada de X_i que se obtém a partir da função acumulada conjunta de

X_1, \dots, X_n fazendo $x_j \rightarrow \infty$ para $j \neq i$ é conhecida como *função de distribuição marginal* de X_i .

Exemplo

Seja $F_{X_1, X_2}(x_1, x_2)$ a função de distribuição acumulada conjunta de uma variável aleatória normal bivariada com vetor de médias $\mu = (0, 0)$ e matriz de covariância:

$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, onde $\rho = 0.5$ é o coeficiente de correlação. Para encontrar a distribuição marginal acumulada de X_1 , usamos o limite:

$$F_{X_1}(x_1) = \lim_{x_2 \rightarrow \infty} F_{X_1, X_2}(x_1, x_2).$$

A distribuição acumulada conjunta é:

$$F_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right) du dv.$$

Tomando o limite $x_2 \rightarrow \infty$, temos:

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du,$$

que é a função de distribuição acumulada da normal padrão univariada:

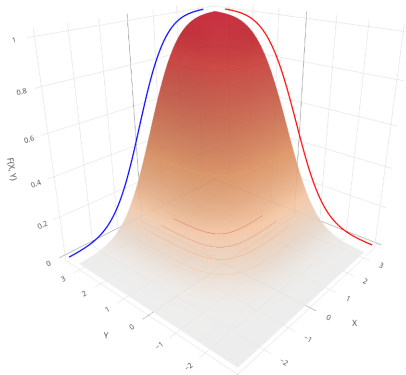
$$F_{X_1}(x_1) = \Phi(x_1),$$

onde $\Phi(x)$ é a CDF da normal padrão. Analogamente, para a marginal acumulada de X_2 :

$$F_{X_2}(x_2) = \lim_{x_1 \rightarrow \infty} F_{X_1, X_2}(x_1, x_2).$$

Tomando o limite, obtemos:

$$F_{X_2}(x_2) = \Phi(x_2).$$



Tipos de Vetores Aleatórios

Os tipos discretos e contínuos de variáveis aleatórias têm os seguintes análogos no caso multivariado.

- (a) Se \vec{X} for um vetor aleatório discreto, ou seja assumir um número enumerável de valores $\{\vec{x}_1, \vec{x}_2, \dots\}$, podemos definir uma função de probabilidade de massa conjunta, p tal que

- ▶ $p(\vec{x}_i) \geq 0$.
- ▶ $\sum_{i=1}^{\infty} p(\vec{x}_i) = 1$.

Neste caso, pode-se definir a *função probabilidade de massa marginal* de X_i como sendo

$$p_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

- (b) Seja $\vec{X} = (X_1, \dots, X_n)$ um vetor aleatório e F sua função de distribuição. Se existe uma função $f(x_1, \dots, x_n) \geq 0$ tal que

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n,$$
$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

então f é chamada de densidade conjunta das variáveis aleatórias X_1, \dots, X_n , e neste caso, dizemos que \vec{X} é (absolutamente) contínuo. Neste caso, define-se a *densidade marginal* de X_i como sendo

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} \cdots \int f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

Comentário: Distribuição condicional de X dada Y discreta

Seja X uma variável aleatória no espaço de probabilidade (Ω, \mathcal{A}, P) , e seja A um evento aleatório tal que $P(A) > 0$. Usando o conceito de probabilidade condicional, podemos definir a distribuição condicional de X dado o evento A por

$$P(X \in B|A) = \frac{P([X \in B] \cap A)}{P(A)},$$

para B boreliano. Pode-se verificar facilmente que isto define uma probabilidade nos borelianos verificando-se os axiomas de Kolmogorov. Podemos interpretar a distribuição condicional de X dado A como a nova distribuição que se atribui a X quando sabe-se da ocorrência do evento A . A função de distribuição associada à distribuição condicional é chamada função distribuição condicional de X dado A :

$$F_X(x|A) = P(X \leq x|A).$$

Agora suponhamos que os eventos aleatórios A_1, A_2, \dots formem uma partição (finita ou enumerável) de Ω . Pelo Teorema da Probabilidade Total, temos

$$P(X \in B) = \sum_n P(A_n)P(X \in B|A_n), \forall B \in \mathcal{B},$$

e

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_n P(A_n)P(X \leq x|A_n) \\ &= \sum_n P(A_n)F_X(x|A_n), \forall x. \end{aligned}$$

Em outras palavras, a distribuição de X (resp., função de distribuição) é uma média ponderada da distribuição condicional (resp., função de distribuição condicional) de X dado A_n , onde os pesos são as probabilidades dos membros A_n da partição.

Consideremos agora o caso em que a partição do espaço amostral é gerada por uma variável aleatória discreta. Para tanto, seja Y uma variável aleatória discreta em (Ω, \mathcal{A}, P) , tomando somente os valores y_1, y_2, \dots . Então, os eventos $A_n = [Y = y_n]$ formam uma partição de Ω . Neste caso, a distribuição

$$P(X \in B | Y = y_n) = P(X \in B | A_n),$$

para B boreliano, é chamada de distribuição condicional de X dado que $Y = y_n$, e valem as fórmulas

$$P(X \in B) = \sum_n P(Y = y_n)P(X \in B | Y = y_n), \quad B \text{ boreliano}$$

$$F_X(x) = \sum_n P(Y = y_n)F_X(x | Y = y_n).$$

Independência entre Variáveis Aleatórias

Sejam X_1, X_2, \dots, X_n variáveis aleatórias definidas no mesmo espaço de probabilidade (Ω, \mathcal{A}, P) . Informalmente, as variáveis aleatórias X_i 's são independentes se, e somente se, quaisquer eventos determinados por qualquer grupo de variáveis aleatórias distintas são independentes. Por exemplo, $[X_1 < 5]$, $[X_2 > 9]$, e $0 < X_5 \leq 3$ são independentes. Formalmente,

Definição 3

Um conjunto de variáveis aleatórias $\{X_1, \dots, X_n\}$ é mutuamente independente se, e somente se, para quaisquer eventos borelianos A_1, \dots, A_n ,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i).$$

O próximo teorema estabelece três critérios para provar que um conjunto de variáveis aleatórias é mutuamente independente.

Teorema 1

As seguintes condições são necessárias e suficientes para testar se um conjunto $\{X_1, \dots, X_n\}$ de variáveis aleatórias é mutuamente independente:

(a) $F_{\vec{X}}(\vec{x}) = \prod_{i=1}^n F_{X_i}(x_i).$

(b) Se \vec{X} for um vetor aleatório discreto,

$$p_{\vec{X}}(\vec{x}) = \prod_{i=1}^n p_{X_i}(x_i).$$

(c) Se \vec{X} for um vetor aleatório contínuo,

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n f_{X_i}(x_i), \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Demonstração

Para parte (a), note que se $\{X_1, \dots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i), \forall (x_1, \dots, x_n) \end{aligned}$$

A prova da suficiência da parte (a) será omitida pois envolve argumentos de teoria da medida. Para parte (b), se $\{X_1, \dots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$\begin{aligned} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 = x_1, \dots, X_n = x_n) \\ &= \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p_{X_i}(x_i), \forall (x_1, \dots, x_n) \end{aligned}$$

Reciprocamente, se a função de probabilidade de massa conjunta fatora e se $\{x_{i1}, x_{i2}, \dots, x_{in}, \dots\}$ são os possíveis valores assumidos pela variável aleatória X_i , temos que

$$\begin{aligned} & P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) \\ &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} P(X_1 = x_{1i}, \dots, X_n = x_{ni}) \\ &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} p_{X_1, \dots, X_n}(x_{1i}, \dots, x_{ni}) \\ &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} \prod_{j=1}^n p_{X_j}(x_{ji}) = \prod_{j=1}^n P(X_j \in B_j) \end{aligned}$$

A parte (c) é uma consequência direta da parte (a) e da definição de função de densidade. Omitimos os detalhes.

Nota 1

É fácil observar que utilizando, a definição de probabilidade condicional que se X e Y são independentes, então para todo A e B boreliano tal que $P(Y \in B) > 0$:

$$P(X \in A | Y \in B) = P(X \in A),$$

ou seja, se X e Y são independentes o conhecimento do valor de Y não altera a descrição probabilística de X .

Exemplos de Distribuições Multivariadas

A Distribuição Multinomial

Esta distribuição pode ser considerada como uma generalização da distribuição binomial. Considere um experimento aleatório qualquer e suponha que o espaço amostral deste experimento é particionado em k eventos $\{A_1, A_2, \dots, A_k\}$, onde o evento A_i tem probabilidade p_i . Suponha que se repita este experimento n vezes de maneira independente e seja X_i o número de vezes que o evento A_i ocorreu nestas n repetições. Então,

$$\begin{aligned} P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) \\ = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}, \end{aligned} \quad (1)$$

onde $\sum_{i=1}^k n_i = n$. (Relembre que o número de maneiras de arranjar n objetos, n_1 dos quais é de uma espécie, n_2 dos quais é de uma segunda espécie, \dots , n_k dos quais são de uma k -ésima espécie é dado pelo coeficiente multinomial $\frac{n!}{n_1! n_2! \dots n_k!}$.)

Exemplos de Distribuições Multivariadas

A Distribuição Normal Bivariada

O vetor aleatório (X, Y) possui distribuição normal bivariada quando tem densidade dada por

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

onde $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1, \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}$.

Se $\rho = 0$, esta densidade fatora e temos que X e Y são independentes. Se $\rho \neq 0$, esta densidade não fatora e X e Y não são independentes.

Funções de Variáveis Aleatórias

- Muitas vezes sabemos a distribuição de probabilidade que descreve o comportamento de uma variável aleatória X definida no espaço mensurável (Ω, \mathcal{A}) , mas estamos interessados na descrição de uma função $Y = H(X)$.
- Nosso problema é determinar $P(Y \in A)$, onde A é um evento Boreliano, dado P_X . Para determinarmos esta probabilidade, estaremos interessados na imagem inversa da função H , ou seja, a probabilidade do evento $\{Y \in A\}$ será por definição igual a probabilidade do evento $\{X \in H^{-1}(A)\}$, onde $H^{-1}(A) = \{x \in \mathbb{R} : H(x) \in A\}$.
- Precisamos restringir H tal que $H^{-1}(A)$ seja um evento boreliano para todo A boreliano, caso contrário não poderemos determinar $P(\{X \in H^{-1}(A)\})$; uma função que satisfaz esta condição é conhecida como *mensurável com respeito a \mathcal{A} e \mathcal{B}* . Note que Y também pode ser vista como uma função do espaço amostral Ω , $Y(\omega) = H(X(\omega))$ para todo $\omega \in \Omega$.
- Visto dessa maneira Y é uma variável aleatória definida em (Ω, \mathcal{A}) , pois para todo boreliano A , $Y^{-1}(A) = X^{-1}(H^{-1}(A))$ e como por suposição $H^{-1}(A)$ é boreliano e X é uma variável aleatória, temos que $X^{-1}(H^{-1}(A)) \in \mathcal{A}$ e portanto satisfaz a definição de uma variável aleatória.

Caso Discreto

Neste caso, para qualquer função H , temos que $Y = H(X)$ é uma variável aleatória discreta.

Suponha que X assumam os valores x_1, x_2, \dots e seja H uma função real tal que $Y = H(X)$ assumam os valores y_1, y_2, \dots . Vamos agrupar os valores que X assume de acordo com os valores de suas imagens quando se aplica a função H , ou seja, denotemos por $x_{i1}, x_{i2}, x_{i3}, \dots$ os valores de X tal que $H(x_{ij}) = y_i$ para todo j . Então, temos que

$$\begin{aligned} P(Y = y_i) &= P(X \in \{x_{i1}, x_{i2}, x_{i3}, \dots\}) \\ &= \sum_{j=1}^{\infty} P(X = x_{ij}) = \sum_{j=1}^{\infty} p_X(x_{ij}), \end{aligned}$$

ou seja, para calcular a probabilidade do evento $\{Y = y_i\}$, acha-se o evento equivalente em termos de X , isto é, todos os valores x_{ij} de X tal que $H(x_{ij}) = y_i$ e somam-se as probabilidades de X assumir cada um desses valores.

Funções de Variáveis Aleatórias

Exemplo 2

Caso Discreto Admita-se que X tenha os valores possíveis $1, 2, 3, \dots$ e suponha que $P(X = n) = (1/2)^n$. Seja $Y = 1$ se X for par e $Y = -1$ se X for ímpar. Então, temos que

$$P(Y = 1) = \sum_{n=1}^{\infty} (1/2)^{2n} = \sum_{n=1}^{\infty} (1/4)^n = \frac{1/4}{1 - 1/4} = 1/3.$$

Consequentemente,

$$P(Y = -1) = 1 - P(Y = 1) = 2/3.$$

Caso Discreto Vetorial

Podemos estender este resultado para uma função de um vetor aleatório \vec{X} de forma análoga. Neste caso se $\vec{Y} = H(\vec{X})$, denotemos por $\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \dots$ os valores de \vec{X} tal que $H(\vec{x}_{ij}) = \vec{y}_i$ para todo j . Então, temos que

$$\begin{aligned} P(\vec{Y} = \vec{y}_i) &= P(\vec{X} \in \{\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \dots\}) \\ &= \sum_{j=1}^{\infty} P(\vec{X} = \vec{x}_{ij}) = \sum_{j=1}^{\infty} p_{\vec{X}}(\vec{x}_{ij}), \end{aligned}$$

ou seja, para calcular a probabilidade do evento $\{\vec{Y} = \vec{y}_i\}$, acha-se o evento equivalente em termos de \vec{X} , isto é, todos os valores \vec{x}_{ij} de \vec{X} tal que $H(\vec{x}_{ij}) = \vec{y}_i$ e somam-se as probabilidades de \vec{X} assumir cada um desses valores.

Funções de Variáveis Aleatórias

Caso Contínuo

Vamos ver agora um exemplo no caso em que \tilde{X} é contínuo.

Exemplo 3

Se $X \sim U[0, 1]$, qual a distribuição de $Y = -\log(X)$? Como

$$0 < Y < \infty \Leftrightarrow 0 < X < 1$$

e $P(0 < X < 1) = 1$, temos $F_Y(y) = 0, y \leq 0$. Se $y > 0$, então

$$P(Y \leq y) = P(-\log(X) \leq y) = P(X \geq e^{-y}) = 1 - e^{-y},$$

ou seja, $Y \sim \text{Exp}(1)$.

Jacobiano de uma Função

Dado um conjunto de n equações em n variáveis x_1, \dots, x_n ,

$$y_1 = f_1(x_1, \dots, x_n), \dots, y_n = f_n(x_1, \dots, x_n),$$

a matriz Jacobiana é definida por

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \dots & \frac{\partial y_n}{\partial x_n} \end{pmatrix}$$

O determinante de J é chamado de *Jacobiano*.

Jacobiano de uma Função

Pode-se provar que o módulo Jacobiano nos dá a razão entre volumes n -dimensionais em \vec{y} e \vec{x} quando a maior dimensão Δx_i tende a zero. Deste modo, temos que o módulo do Jacobiano aparece quando queremos mudar as variáveis de integração em integrais múltiplas, ou seja, existe um teorema do cálculo que afirma que se $f : G_0 \rightarrow G$ for uma bijeção entre G_0 e G , f e as derivadas parciais que aparecem na matriz Jacobiana forem funções contínuas em G_0 , e o Jacobiano for diferente de zero para todo $x \in G_0$

$$\begin{aligned} & \int \cdots \int_A g(y_1, \dots, y_n) dy_1 \cdots dy_n \\ &= \int \cdots \int_{f^{-1}(A)} g(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)) |J| dx_1 \cdots dx_n, \end{aligned}$$

para qualquer função g integrável em $A \subseteq G$.

Vamos agora utilizar mudança de variáveis para resolver o seguinte exemplo da soma de duas variáveis aleatórias.

Jacobiano de uma Função

Exemplo

Suponha que (X, Y) tenha densidade conjunta $f(x, y)$ e seja $Z = X + Y$. Neste caso,

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = P((X, Y) \in B_z),$$

onde $B_z = \{(x, y) : x + y \leq z\}$. Portanto,

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) dx dy.$$

Jacobiano de uma Função

Exemplo (cont.)

Fazendo a mudança de variáveis $s = x + y$, $t = y$, que tem jacobiano igual a 1, temos

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^z f(s - t, t) ds dt = \int_{-\infty}^z \int_{-\infty}^{\infty} f(s - t, t) dt ds.$$

Logo, $\int_{-\infty}^{\infty} f(s - t, t) dt$ é a densidade da soma $Z = X + Y$, ou seja,

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - t, t) dt = \int_{-\infty}^{\infty} f(s, z - s) ds,$$

onde fizemos a troca de variáveis $s = z - t$ para obter a última expressão.

Jacobiano de uma Função

Exemplo (cont.)

Se X e Y forem variáveis aleatórias independentes com densidades f_X e f_Y , temos que $f(x, y) = f_X(x)f_Y(y)$, então,

$$\begin{aligned}f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-t)f_Y(t)dt \\&= \int_{-\infty}^{\infty} f_X(t)f_Y(z-t)dt = f_X * f_Y,\end{aligned}$$

onde $f_X * f_Y$ é conhecida como a *convolução das densidades* f_X e f_Y .

Jacobiano de uma Função

Vamos agora descrever o método do Jacobiano para funções mais gerais H . Suponha que $G_0 \subseteq \mathbb{R}^n$, $G \subseteq \mathbb{R}^n$ sejam regiões abertas, e que $H : G_0 \rightarrow G$ seja uma bijeção entre G_0 e G . Logo, existe a função inversa H^{-1} em G , de modo que $\vec{X} = H^{-1} \vec{Y}$. Suponha ainda que f é a densidade conjunta de \vec{X} e que $P(\vec{X} \in G_0) = 1$. Se as derivadas parciais de H^{-1} existirem e o Jacobiano J de H^{-1} for diferente de zero para todo $\vec{y} \in G$, podemos utilizar o teorema da mudança de variáveis e obter que para $B \subseteq G$, B boreliano, temos

$$\begin{aligned} P(\vec{Y} \in B) &= P(\vec{X} \in H^{-1}(B)) \\ &= \int \cdots \int_{H^{-1}(B)} f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int \cdots \int_B f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J| dy_1 \cdots dy_n. \end{aligned}$$

Jacobiano de uma Função

Como $P(\vec{Y} \in G) = P(\vec{X} \in H^{-1}(G)) = P(\vec{X} \in G_0) = 1$, temos que para todo boreliano B no \mathbb{R}^n ,

$$\begin{aligned} P(\vec{Y} \in B) &= P(\vec{Y} \in B \cap G) \\ &= \int \cdots \int_{B \cap G} f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J| dy_1 \cdots dy_n. \end{aligned}$$

Esta última integral é igual a integral sobre o conjunto B da função que toma o valor

$$f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J|,$$

para $\vec{y} \in G$, e zero no caso contrário. Portanto, pela definição de densidade temos que

$$f_{\vec{Y}}(y_1, \dots, y_n) = \begin{cases} f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J|, \\ \text{se } \vec{y} \in G, \\ 0, \text{ caso contrário.} \end{cases}$$

Jacobiano de uma Função

Observação

- (a) Note que J é o Jacobiano da função inversa H^{-1} , em alguns casos pode ser útil obter J a partir do Jacobiano J' da função H através da relação

$$J = \frac{1}{J'}|_{\vec{x}=H^{-1}(\vec{y})}.$$

- (b) Para obter a distribuição de $\vec{Y} = H(\vec{X})$ quando a dimensão de \vec{Y} é menor que a dimensão de \vec{X} muitas vezes é possível definir outras variáveis aleatórias Y'_1, \dots, Y'_m , utilizar o método do Jacobiano para determinar a densidade conjunta de $\vec{Y}, Y'_1, \dots, Y'_m$ e, finalmente, obter a densidade marginal conjunta de \vec{Y} . Considere o seguinte exemplo:

Jacobiano de uma Função

Observação

Exemplo 4

Suponha que X_1, X_2 tem densidade conjunta dada por $f(x, y)$ e que estamos interessados na distribuição de $Y_1 = X_1^2 + X_2$. Como esta não é uma transformação 1-1, ela não possui inversa. Vamos definir uma nova variável $Y_2 = X_1$ de modo que a função $(Y_1, Y_2) = H(X_1, X_2) = (X_1^2 + X_2, X_1)$ possua uma função inversa diferenciável, $(X_1, X_2) = H^{-1}(Y_1, Y_2) = (Y_2, Y_1 - Y_2^2)$. Deste modo temos que

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -2y_2 \end{pmatrix} = -1$$

Então temos que, $f_{Y_1, Y_2}(y_1, y_2) = f(y_2, y_1 - y_2^2)$. Finalmente, para encontrarmos f_{Y_1} integramos sobre todos os possíveis valores da variável Y_2 que introduzimos:

$$f_{Y_1} = \int_{-\infty}^{\infty} f(y_2, y_1 - y_2^2) dy_2.$$

Jacobiano de uma Função

Observação

- (c) Podemos utilizar o método do Jacobiano em outros casos em que a função H não é 1-1. Para tanto, suponha que G, G_1, \dots, G_k sejam subregiões abertas do \mathbb{R}^n tais que G_1, \dots, G_k sejam disjuntas e $P(\vec{X} \in \cup_{i=1}^k G_i) = 1$, tais que a função $H|_{G_i}$, a restrição de H a G_i , seja um correspondência 1-1 entre G_i e G , para $i = 1, \dots, k$. Suponha que para todo i , a função inversa de $H|_{G_i}$ satisfaça as hipóteses do caso anterior, e seja J_i o Jacobiano da inversa de $H|_{G_i}$. Pode-se provar que

$$f_{\vec{Y}}(y_1, \dots, y_n) = \begin{cases} \sum_{i=1}^k f(H|_{G_i}^{-1}(y_1, \dots, y_n)) |J_i|, \\ \text{se } \vec{y} \in G, \\ 0, \text{ caso contrário.} \end{cases}$$