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 $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Uma função $\vec{X} : \Omega \to R^n$ 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 contem todas regiões da seguinte forma: $C_{\vec{a}} = \{(X_1, X_2, \ldots, X_n) : X_i \leq a_i, 1 \leq i \leq n\}$.

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

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

Propriedades

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

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

$$x_i \leq y_i \ \forall i \leq n \Rightarrow C_{\vec{x}} \subseteq C_{\vec{y}} \Rightarrow P(C_{\vec{x}}) \leq P(C_{\vec{y}}) \Rightarrow F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y}).$$

F2. $F(x_1, x_2, \ldots, x_n)$ é contínua a direita em cada uma das variáveis. Por exemplo, se $y_m \downarrow x_1$, então

$$F(y_m, x_2, \ldots, x_n) \downarrow F(x_1, x_2, \ldots, x_n), \text{ quando } m \to \infty.$$

F3a. Se para algum $i \leq n$ $x_i \to -\infty$, então $C_{\vec{x}}$ decresce monotonicamente para o conjunto vazio $\emptyset$. Logo, pela continuidade monotônica de probabilidade, temos que

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

F3b. Se $x_i \to \infty$, então $C_{\vec{x}}$ cresce monotonicamente para o conjunto

$$\{X_1 \leq x_1, \ldots X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n\},$$

ou seja a restrição em X_i é removida. Então, podemos escrever

$$\begin{aligned} &\lim_{x_i \to \infty} F_{\vec{X}}(\vec{x}) \\ &= F_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n). \end{aligned}$$

Portanto, a função de distribuição acumulada conjunta de $X_1, \ldots, X_{n-1}$ pode ser facilmente determinada da função de distribuição acumulada conjunta de $X_1, \ldots, X_n$ fazendo $x_n \to \infty$. Observe que *funções de distribuição acumuladas conjuntas de ordem maiores determinam as de ordem menores, mas o contrário não é verdadeiro*. Em particular, temos que

$$\lim_{\vec{x} \to \infty} F_{\vec{X}}(\vec{x}) = 1.$$

Distribuição marginal

A função de distribuição acumulada de X_i que se obtém a partir da função acumulada conjunta de
$X_1, \ldots, X_n$ fazendo $x_j \to \infty$ para $j \neq i$ é conhecida como *função de distribuição marginal* de X_i.

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 $F_0 : \mathbb{R}^2 \to \mathbb{R}$ 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. É claro que F1, F2, e F3 são satisfeitas, mas F_0 não é função de distribuição de nenhum vetor aleatório (X, Y). Se fosse, teríamos uma contradição

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

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 $\{\vec{x}_1, \vec{x}_2 \ldots, \}$, 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

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

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

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

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

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

Comentário: Distribuição condicional de X dada Y discreta

Seja X uma variável aleatória no espaço de probabilidade $(\Omega, \mathcal{A}, P)$, 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 $A_1, A_2, \ldots$ 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

$$F_X(x) = P(X \leq x) = \sum_n P(A_n)P(X \leq x|A_n)$$
$$= \sum_n P(A_n)F_X(x|A_n), \forall x.$$

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 $(\Omega, \mathcal{A}, P)$, tomando somente os valores $y_1, y_2, \ldots$. 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

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

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

Independência entre Variáveis Aleatórias

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. 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 $\{X_1, \ldots, X_n\}$ *é mutuamente independente se, e somente se, para quaisquer eventos borelianos* $A_1, \ldots, A_n$*,*

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

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 $\{X_1, \ldots, X_n\}$ 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,*

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

Demonstração

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

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

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

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

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

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

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 $\{A_1, A_2, \ldots, A_k\}$, 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,

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

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, ..., n_k dos quais são de uma k-ésima espécie é dado pelo coeficiente multinomial $\frac{n!}{n_1!n_2!\cdots n_k!}$.)

Exemplos de Distribuições Multivariadas

A Distribuição Normal Bivariada

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

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

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 $(\Omega, \mathcal{A})$, 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 $(\Omega, \mathcal{A})$, 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.

Funções de Variáveis Aleatórias

Caso Discreto

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

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

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

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 $1, 2, 3, \ldots$ 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,

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

Funções de Variáveis Aleatórias

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 $\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \ldots$ os valores de $\vec{X}$ tal que $H(\vec{x}_{ij}) = \vec{y}_i$ para todo j. Então, temos que

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

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 $\vec{X}$ 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, $Y \sim Exp(1)$.

Jacobiano de uma Função

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

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

a matriz Jacobiana é definida por

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

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 $f : G_0 \to G$ 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 $x \in G_0$

$$\int \cdots \int_A g(y_1, \ldots, y_n) dy_1 \cdots dy_n$$

$$= \int \cdots \int_{f^{-1}(A)} g(f_1(x_1, ..., x_n), \ldots, f_n(x_1, ..., x_n))|J| dx_1 \cdots dx_n,$$

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,

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

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 $H : G_0 \to G$ 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

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

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 $\mathbb{R}^n$,

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

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

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

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

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

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

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

(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 $Y'_1, \ldots, Y'_m$, utilizar o método do Jacobiano para determinar a densidade conjunta de $\vec{Y}, Y'_1, \ldots, Y'_m$ 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 $G, G_1, \ldots, G_k$ sejam subregiões abertas do $\mathbb{R}^n$ tais que $G_1, \ldots, G_k$ sejam disjuntas e $P(\vec{X} \in \cup_{i=1}^{k} G_i) = 1$, tais que a função $H|_{G_l}$, a restrição de H a G_l, seja um correspondência 1-1 entre G_l e G, para $l = 1, \ldots, k$. Suponha que para todo l, a função inversa de $H|_{G_l}$ satisfaça as hipóteses do caso anterior, e seja J_l o Jacobiano da inversa de $H|_{G_l}$. Pode-se provar que

$$f_{\vec{Y}}(y_1, \ldots, y_n) = \begin{cases} \sum_{l=1}^{k} f(H|_{G_l}^{-1}(y_1, \ldots, y_n))|J_l|, \\ \quad \text{se } \vec{y} \in G, \\ 0, \text{ caso contrário.} \end{cases}$$