Tópico 3: Vetores Aleatórios

Ben Dêivide

6 de outubro de 2021

Descrição sobre o ponto - Distribuições marginais; Independência estocástica; Distribuições de transformações de vetores aleatórios; Esperança condicional; Principais distribuições de probabilidade.

1 Conceitos iniciais

Nos assuntos anteriores enfatizamos os modelos probabilísticos que envolveram somente uma variável, os chamados modelos univariados. Agora, iremos tratar de modelos que envolvem mais de uma variável aleatória, os chamados modelos multivariados. De fato muitos dos conceitos serão dados com base em duas variáveis aleatórias, podendo ser estendido para mais variáveis.

Denotemos um conjunto de p variáveis aleatórias $X_1, X_2, \ldots, X_p$ por meio de um vetor $\boldsymbol{X} \in \mathbb{R}^p$ definidos no espaço de probabilidade.

Definição 1 (Espaço de probabilidade). *O espaço de probabilidade é a tripla $(\Omega, \mathcal{F}, P)$, em que Ω é o espaço amostral, $\mathcal{F}$ é a σ-álgebra e P é a medida de probabilidade com domínio em $\mathcal{F}$.* □

Dessa forma, calculamos a probabilidade de um evento $A \in \mathcal{F}$, do qual $\mathcal{F}$ é uma coleção de subconjuntos de Ω. Entretanto, quando $\Omega = \mathbb{R}$, poderemos ter problemas quanto a medida de probabilidade de eventos não-contáveis, isto é, dado que A é um evento não-contável tal que $A \in \mathbb{R}$, como medir a probabilidade A, se o espaço de probabilidade não pode ser facilmente obtido? Para solucionar esse problema, vamos definir a σ-álgebra de Borel.

Definição 2 (σ-álgebra de Borel). *A σ-álgebra de Borel, denotada por $\mathcal{B}(\mathbb{R})$ ou $\mathcal{B}$, é a menor σ-álgebra que contém todos os intervalos abertos dos reais (a, b), isto é, $(-\infty \leq a < b \leq \infty)$. Os conjuntos $B \subseteq \mathbb{R}$ tais que $B \in \mathcal{B}$ são chamados conjuntos de Borel.* □

Definição 3 (Variável aleatória p-dimensional). *Uma variável aleatória p-dimensional $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$ é uma função vetorial mensurável, $\boldsymbol{X} : \Omega \to \mathbb{R}^p$, que mapeia o espaço mensurável $(\Omega, \mathcal{F})$ em $(\mathbb{R}^p, \mathcal{B}^p)$, tal que*

$$\boldsymbol{X}^{-1}(B) = \{w \in \Omega : \boldsymbol{X}(w) \in B\} \in \mathcal{F},$$

sendo $B = [B_1, B_2, \ldots, B_p]'$ o conjunto de Borel e $\mathcal{B}^p$ é a σ-álgebra p-dimensional de Borel, $\boldsymbol{X}^{-1}$ é a imagem inversa de conjuntos $B \in \mathcal{B}^p$ que pertencem a σ-álgebra $\mathcal{F}$. □

A Definição 3 implica que

$$\boldsymbol{X}(\omega) = [X_1(\omega), X_2(\omega), \ldots, X_p(\omega)]'$$

é tal que para todo $i = 1, 2, \ldots, p$ e todo $B_i \subset \mathbb{R}$, tem-se $X_i^{-1}(B_i) \in \mathcal{F}$.

2 Distribuição de probabilidade conjunta

Com o objetivo de caracterizar uma variável aleatória p-dimensional, definimos o seu modelo probabilístico.

Definição 4 (Distribuição de probabilidade conjunta). *A probabilidade conjunta* $P_{\boldsymbol{X}}(B) = P(\boldsymbol{X} \in B) = P(\boldsymbol{X}^{-1}(B) \in \mathcal{F})$ *é chamada de distribuição da variável aleatória p-dimensional* $\boldsymbol{X}$*, para todo evento* B *tal que* $B \in \mathcal{B}^p$*, sendo* $\mathcal{B}^p$ *uma* σ*-álgebra de Borel p-dimensional.*

Para o conjunto $B_i = (-\infty, x_i]$ para $i = 1, 2, \ldots, p$, definimos a função de distribuição de $\boldsymbol{X}$.

Definição 5 (Função de Distribuição Conjunta). *Seja* $\boldsymbol{X}$ *uma variável aleatória p-dimensional definida em* $(\Omega, \mathcal{F}, P)$*. A função de distribuição conjunta de* $\boldsymbol{X}$*, denotada por* $F_{\boldsymbol{X}}$*, é uma função* $F_{\boldsymbol{X}} : \mathbb{R}^p \to [0, 1]$ *definida por*

$$F_{\boldsymbol{X}} = P(\boldsymbol{X} \leq \boldsymbol{x}) = P(X_1 \leq x_1, \ldots, X_p \leq x_p), \tag{1}$$

para todo $x_i \in \mathbb{R}$*,* $i = 1, 2, \ldots, p$*.*

As propriedades da Função de distribuição Conjunta são apresentadas a seguir.

Teorema 1 (Propriedades da Função de distribuição Conjunta). *Seja* $\boldsymbol{X}$ *uma variável aleatória p-dimensional em* $(\Omega, \mathcal{F}, P)$ *então, para qualquer* $\boldsymbol{x} \in \mathbb{R}^p$*, temos que*

i) Se para algum i*,* $x_i \to -\infty$*, então* $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 0$*, e ainda se para todo* i*,* $x_i \to +\infty$ *então* $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 1$*;*

ii) $F_{\boldsymbol{X}}(\boldsymbol{x})$ *é não decrescente em cada uma de suas coordenadas* x_i*,* $i = 1, \ldots, p$*;*

iii) $F_{\boldsymbol{X}}(\boldsymbol{x})$ *é contínua em cada uma das suas coordenadas* x_i*,* $i = 1, \ldots, p$*;*

iv) $F_{\boldsymbol{X}}(\boldsymbol{x})$ *é tal que,* $\forall a_i, b_i \in \mathbb{R}$*,* $a_i < b_i$*,* $1 \leq i \leq p$*, temos* $P(a_1 \leq X_1 \leq b_1, \ldots, a_p \leq X_p \leq b_p) \geq 0$*.*

Demonstração. Para (i), se algum $x_i \to -\infty$, então $\{w : X_i(w) \leq x_i(w)\} \downarrow \emptyset$, e portanto, $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 0$. Se todos os x_is são tais que $x_i \to \infty$, então cada um dos eventos $\{w : X_i(w) \leq x_i(w)\}$ cresce para Ω. Logo a interseção desses eventos é o próprio Ω, e portanto, $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 1$.

Para (ii), seja $\boldsymbol{y} \leq \boldsymbol{x}$ para $\boldsymbol{x} \in \mathbb{R}^p$ e $\boldsymbol{y} \in \mathbb{R}^p$. Considerando $A = \{\boldsymbol{X} \leq \boldsymbol{y}\} \subseteq B = \{\boldsymbol{X} \leq \boldsymbol{x}\}$, então $P(\boldsymbol{X} \in A) = F_{\boldsymbol{X}}(\boldsymbol{y}) \leq P(\boldsymbol{X} \in B) = F_{\boldsymbol{X}}(\boldsymbol{x})$.

Para (iii), seja $\boldsymbol{x}^* \in \mathbb{R}^p$ tal que $\boldsymbol{x} \downarrow \boldsymbol{x}^*$, então $\{\boldsymbol{X} \leq \boldsymbol{x}\} \downarrow \{\boldsymbol{X} \leq \boldsymbol{x}^*\}$. Logo $F_{\boldsymbol{X}(\boldsymbol{x})} \downarrow F_{\boldsymbol{X}}(\boldsymbol{x})^*$. Como o resultado vale para qualquer vetor $\boldsymbol{x}^*$, a propriedade está demonstrada.

Para verificar o item (iv), basta observar que $F_{\boldsymbol{X}}(\boldsymbol{x})$ é uma medida de probabilidade e deve atender aos axiomas de Kolmogorov. Logo, $F_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$. $\square$

Para o item (iv) tem um exemplo em Magalhães p. 118-119 que mostra a importância desse item.

A partir da função distribuição conjunta, podemos obter o comportamento de cada variável isoladamente.

Definição 6 (Função de Distribuição Marginal)**.** *Seja $\boldsymbol{X}$ uma variável aleatória p-dimensional definida em $(\Omega, \mathcal{F}, P)$. Para cada k, $k = 1, 2, \ldots, p$, a função de distribuição marginal de X_k é definida por:*

$$F_{X_k}(x_k) = \lim_{\substack{x_i \to \infty \\ \forall i \neq k}} F_{\boldsymbol{X}}(\boldsymbol{x}), \tag{2}$$

em que o limite é aplicado em todas as coordenadas, exceto k. □

A atribuição de probabilidades para eventos envolvendo uma variável aleatória p-dimensional depende da característica dessa variável. Diremos que o suporte $S_{\boldsymbol{X}} = \{x_1, x_2, \ldots, x_p\}$ é o subconjunto dos valores que o vetor $\boldsymbol{X}$ pode assumir.

3 Tipos de variáveis aleatórias multidimensionais

Assim como no caso unidimensional, iremos apresentar a natureza das variáveis aleatórias p-dimensionais. Considere $S_{\boldsymbol{X}}$ o suporte dos valores que $\boldsymbol{X}$ pode assumir.

Definição 7 (Variável aleatória p-dimensional discreta)**.** *Uma variável aleatória p-dimensional $\boldsymbol{X}$ em $(\Omega, \mathcal{F}, P)$, é discreta se o suporte $S_{\boldsymbol{X}}$ é um subconjunto contável de $\mathbb{R}^p$ tal que a função de probabilidade, $P_{\boldsymbol{X}} : \mathbb{R}^p \to [0, 1]$, é definida por:*

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \tag{3}$$

sendo $\sum_{\boldsymbol{x} \in S_{\boldsymbol{X}}} P_{\boldsymbol{X}}(\boldsymbol{x}) = 1$, e

$$P_{X_k}(x_k) = \sum_{\substack{x_i \\ \forall i \neq k}} P_{\boldsymbol{X}}(\boldsymbol{x}) \tag{4}$$

é a função de probabilidade marginal. □

Definição 8 (Variável aleatória p-dimensional absolutamente contínua)**.** *Uma variável aleatória p-dimensional $\boldsymbol{X}$ em $(\Omega, \mathcal{F}, P)$, é absolutamente contínua se existir uma função $f_{\boldsymbol{X}}(\boldsymbol{x}) : \mathbb{R}^p \to [0, 1]$ tal que*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_p} f_{\boldsymbol{X}}(\boldsymbol{t}) dt_1 dt_2, \ldots dt_p, \tag{5}$$

sendo $f_{\boldsymbol{X}}(\boldsymbol{x})$ a função densidade de probabilidade, em que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x}) dx_1 dx_2, \ldots dx_p = 1, \tag{6}$$

e a função densidade marginal é dada por:

$$f_{X_k}(x_k) = \underbrace{\int_{x_1} \ldots \int_{x_p}}_{x_i \in S_{\boldsymbol{X}} \forall i \neq k} f_{\boldsymbol{X}}(\boldsymbol{x}) \, \underbrace{dx_1 dx_2, \ldots dx_p}_{x_i \in S_{\boldsymbol{X}} \forall i \neq k}. \tag{7}$$

□

4 Distribuição condicional bidimensional

Vamos restringir ao caso bidimensional, mas as ideias podem ser estendidas para o caso p-dimensional.

Definição 9 (Distribuição condicional bidimensional)**.** *Sejam X e Y duas variáveis aleatórias definidas em $(\Omega, \mathcal{F}, P)$, então a função de probabilidade condicional de X dado $Y = y$ é definido por*

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}, \tag{8}$$

para X e Y discretas e $P_Y(Y) > 0$, e a função densidade de probabilidade de X dado $Y = y$ é definido por

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \tag{9}$$

para X e Y contínuas e $f_Y(y) > 0$. □

Para o caso contínuo, dado $Y = y$ entendamos que a probabilidade condicional é o limite de probabilidades condicionais em $[y - \epsilon \leq Y \leq y + \epsilon]$ quando $\epsilon \to 0$. Podemos observar também que (8) e (9) obedecem as propriedades de uma função de probabilidade e função densidade de probabilidade, respectivamente. Para (8), observamos que claramente que $P_{X|Y}(x|y) > 0$ e que

$$\sum_i \frac{P_{X,Y}(x_i,y)}{P_Y(Y)} = \frac{1}{P_Y(Y)} \sum_i P_{X,Y}(x_i,y) = \frac{1}{P_Y(Y)} P_Y(Y) = 1.$$

Para (9), observamos também que $f_{X|Y}(x|y) > 0$ e que

$$\int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(Y)} dx = \frac{1}{f_Y(Y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = \frac{1}{f_Y(Y)} f_Y(Y) = 1.$$

Definição 10 (Função de distribuição condicional bidimensional)**.** *Sejam X e Y variáveis aleatórias definidas em $(\Omega, \mathcal{F}, P)$, então a função de distribuição de X dado $Y = y$ é definido por:*

$$F_{X|Y}(x|y) = \sum_{\{i: x_i \leq x\}} P_{X|Y}(x|y), \qquad \text{para } P_X(x) > 0, \tag{10}$$

considerando o caso discreto, e

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y) dt, \qquad \text{para } f_X(x) > 0, \tag{11}$$

considerando o caso contínuo. □

5 Independência estocástica

Definição 11 (Independência estocástica). *Seja $\boldsymbol{X}$ uma variável aleatória p-dimensional em $(\Omega, \mathcal{F}, P)$, então $X_1, X_2, \ldots, X_p$ são variáveis aleatórias independentes estocasticamente se e somente se para todos os valores $x_1, x_2, \ldots, x_p$, a função de distribuição é dada por:*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} F_{X_i}(x_i), \tag{12}$$

ou a função de probabilidade conjunta é dada por:

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} P_{X_i}(x_i), \tag{13}$$

considerando o caso discreto, ou a função densidade de probabilidade dada por:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} f_{X_i}(x_i), \tag{14}$$

considerando o caso contínuo. □

Muitas vezes vamos considerar um conjunto de variáveis aleatórias que, além de serem independentes, têm a mesma distribuição, o que chamamos de independentes e identicamente distribuídas, ou simplesmente *iid*.

Teorema 2 (Independência e distribuição condicional bidimensional). *Se X e Y são duas variáveis aleatórias independentes, então a função de distribuição condicional de X dado $Y = y$ é dada por*

$$F_{X|Y} = F_X(x), \tag{15}$$

e função de probabilidade condicional

$$P_{X|Y}(x|y) = P_X(x), \tag{16}$$

para o caso discreto, e a função densidade condicional

$$f_{X|Y}(x|y) = f_X(x), \tag{17}$$

para o caso contínuo. □

Demonstração. Vamos mostrar para o caso contínuo. Segue que

$$\begin{aligned}
F_{X|Y}(x|y) &= \int_{-\infty}^{x} \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \\
&= \int_{-\infty}^{x} \frac{f_X(x) f_Y(Y)}{f_Y(y)} dx, \qquad \text{(Independência)} \\
&= \int_{-\infty}^{x} f_X(x) dx \\
&= F_X(x).
\end{aligned}$$

Para a função de probabilidade condicional, temos

$$
\begin{aligned}
P_{X|Y}(x|y) &= \frac{P_{X,Y}(x,y)}{P_Y(y)} \\
&= \frac{P_X(x)P_Y(Y)}{P_Y(y)}, \qquad \text{(Independência)} \\
&= P_X(x).
\end{aligned}
$$

Para a função densidade condicional, temos

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\
&= \frac{f_X(x)f_Y(Y)}{f_Y(y)}, \qquad \text{(Independência)} \\
&= f_X(x).
\end{aligned}
$$

□

6 Esperança de $g(\boldsymbol{X})$

Um conceito extremamente importante envolvendo as variáveis aleatórias é a esperança.

Definição 12 (Esperança de $g(\boldsymbol{X})$)**.** *Seja $\boldsymbol{X}$ uma variável aleatória p-dimensional em $(\Omega, \mathcal{F}, P)$ e $g(.)$ uma função com domínio e contradomínio nos reais. A esperança da função $g(.)$ de $\boldsymbol{X}$, denotada por $E[g(\boldsymbol{X})]$, é definida por:*

i) se $\boldsymbol{X}$ for discreta,

$$E[g(\boldsymbol{X})] = \sum g(\boldsymbol{x})P_{\boldsymbol{X}}(\boldsymbol{x}), \tag{18}$$

ii) se X for contínua,

$$E[g(\boldsymbol{X})] = \int_{-\infty}^{\infty} g(\boldsymbol{X})f_{\boldsymbol{X}}(\boldsymbol{x})dx_1dx_2\ldots dx_p. \tag{19}$$

□

Existe uma prova em Magalhães p. 219 para $Y = g(X)$ e em Daniel Estatística Matemática (versão 21.01.2016, p.244); Mood p.153

Teorema 3. *Se $g(\boldsymbol{X}) = x_i$, então $E[g(\boldsymbol{X})] = E[X_i] = \mu_{X_i}$ é a média. Se $g(\boldsymbol{X}) = (X_i - \mu_{X_i})^2$, então $E[g(\boldsymbol{X})] = Var[X_i]$.* □

Demonstração. Considerando $(X_i - \mu_{X_i})^2$ um vetor contínuo (Similar pode ser provado para o caso discreto).

$$
\begin{aligned}
E[g(\boldsymbol{X})] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_i f_{\boldsymbol{X}}(\boldsymbol{x})dx_1dx_2\ldots dx_p \\
&= \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)dx_i = E[X_i].
\end{aligned}
$$

$$E[g(\boldsymbol{X})] = E[(X_i - \mu_{X_i})^2] = Var[X_i]$$

□

Uma forma de verificar a relação entre duas variáveis é dado pela covariância e a correlação.

Definição 13 (Covariância)**.** *Seja X e Y duas variáveis aleatórias em $(\Omega, \mathcal{F}, P)$. A covariância de X e Y, denotado por $Cov[X,Y]$, é definida por:*

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])], \tag{20}$$

desde que as esperanças $E[X]$ e $E[Y]$ existam.

Definição 14 (Coeficiente de correlação)**.** *O coeficiente de correlação, denotado por $\rho_{X,Y}$, das variáveis aleatórias X e Y é definida por:*

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}, \tag{21}$$

desde que $Cov[X,Y]$, σ_X e σ_Y existam e $\sigma_X > 0$ e $\sigma_Y > 0$.

A covariância é uma medida de dependência linear e pode assumir qualquer valor na reta real. Por exemplo, se $Cov[X,Y] = 3$ essa informação não fornece a intensidade da relação entre X e Y. Em contrapartida, a correlação é sempre entre -1 e 1, sendo que estes valores indicam a perfeita relação linear entre X e Y. Se $Cov[X,Y] = 0$ implica que X e Y não tem dependência linear. Entretanto, poderá algum outro tipo de relação, e portanto, não podemos afirmar que X e Y são independentes. Uma forma alternativa da covariância é dada a seguir.

Teorema 4. *A covariância de X e Y pode ser expressa por:*

$$Cov[X,Y] = E[XY] - \mu_X \mu_Y. \tag{22}$$

Demonstração.

$$\begin{aligned} Cov[X,Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &= E[XY] - \mu_X \mu_Y. \end{aligned}$$

□

Teorema 5. *Se X e Y são independentes e $g_1(.)$ e $g_2(.)$ duas funções, então*

$$E[g_1(X) g_2(Y)] = E[g_1(X)] E[g_2(Y)]. \tag{23}$$

□

Demonstração. Fazendo a prova para o caso contínuo, temos

$$\begin{aligned} E[g_1(X) g_2(X)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x) g_2(y) f_{X,Y}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x) g_2(y) f_X(x) f_Y(y) dx dy \text{ (Independência)} \\ &= \int_{-\infty}^{\infty} g_1(x) f_X(x) dx \int_{-\infty}^{\infty} g_2(y) f_Y(y) dy \\ &= E[g_1(X)] E[g_2(X)]. \end{aligned}$$

□

Teorema 6. *Se X e Y são variáveis aleatórias independentes, então $Cov[X,Y] = 0$ e $\rho_{X,Y} = 0$.* □

Demonstração. Como X e Y são independentes, então $E[XY] = E[X]E[Y]$, logo

$$Cov[X,Y] = E[XY] - E[X]E[Y] = 0 \Rightarrow \rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} = 0.$$

□

O contrário não é verdade, isto é, se $Cov[X,Y] = 0$ então X e Y são independentes.

Teorema 7. *Se X e Y são duas variáveis aleatórias em $(\Omega, \mathcal{F}, P)$ e a e b são constantes quaisquer, então*

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov[X,Y]. \tag{24}$$

Se X e Y são independentes, então

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y]. \tag{25}$$

□

Demonstração. A esperança de $aX + bY$ é dada por $E[aX + bY] = aE[X] + bE[Y] = a\mu_X + b\mu_Y$. Pelo Teorema 3 temos

$$\begin{aligned}
Var[aX + bY] &= E[(aX + bY - (a\mu_X + b\mu_Y))^2] \\
&= E[(a(X - \mu_X) + b(Y + \mu_Y))^2] \\
&= E[a^2(X - \mu_X)^2 + b^2(Y + \mu_Y)^2 + 2ab(X - \mu_X)(Y + \mu_Y)] \\
&= a^2 E[(X - \mu_X)^2] + b^2 E[(Y + \mu_Y)^2] + 2abE[(X - \mu_X)(Y + \mu_Y)] \\
&= a^2 Var[X] + b^2 Var[Y] + 2abCov[X,Y].
\end{aligned}$$

□

Teorema 8 (Desigualdade de Cauchy-Schwarz)**.** *Sejam X e Y duas variáveis aleatórias com segundo momento finito. Então*

$$E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}, \tag{26}$$

e

$$E[|XY|] = \sqrt{E[X^2]E[Y^2]}, \tag{27}$$

se e somente se $Y = aX$, para alguma constante a. □

Demonstração. Assumimos $E[X^2]$ e $E[Y^2]$ não nulos. Logo,

$$
\begin{aligned}
E\left[\left(\frac{|X|}{\sqrt{E[X^2]}} - \frac{|Y|}{\sqrt{E[Y^2]}}\right)^2\right] &\geq 0 \qquad (28)\\
E\left[\frac{X^2}{E[X^2]} - 2\frac{|XY|}{\sqrt{E[Y^2]E[X^2]}} + \frac{Y^2}{E[Y^2]}\right] &\geq 0\\
\frac{E[X^2]}{E[X^2]} - 2\frac{E[|XY|]}{\sqrt{E[Y^2]E[X^2]}} + \frac{E[Y^2]}{E[Y^2]} &\geq 0\\
2 - 2\frac{E[|XY|]}{\sqrt{E[Y^2]E[X^2]}} &\geq 0 \qquad (29)\\
E[|XY|] &\geq \sqrt{E[Y^2]E[X^2]}.
\end{aligned}
$$

A expressão (28) será igual a 0 apenas se

$$
\begin{aligned}
\frac{|X|}{\sqrt{E[X^2]}} - \frac{|Y|}{\sqrt{E[Y^2]}} &= 0\\
|Y| &= \frac{E[Y^2]}{\sqrt{E[X^2]}}|X|.
\end{aligned}
$$

Portanto, $Y = aX$ com $a = \pm\frac{E[Y^2]}{\sqrt{E[X^2]}}$. □

Teorema 9. *Sejam X e Y duas variáveis aleatórias com variâncias finitas e não nulas. Então:*

i) $|\rho_{X,Y}| \leq 1$;

ii) $|\rho_{X,Y}| = 1$ *se e somente se uma variável for função linear da outra, isto é,* $P(Y = aX + b) = 1$, *para constantes* $a \neq 0$ *e* b.

Demonstração. Para a prova do item (i), iremos utilizar duas desigualdades: desigualdade de Jensen e desigualdade de Cauchy-Schwarz (Teorema 8). Considere $U = (X - \mu_X)$ e $V = (Y - \mu_Y)$, então pela desigualdade de Jensen temos

$$
\begin{aligned}
|E[UV]| &\leq E[|UV|]\\
|E[(X-\mu_X)(Y-\mu_Y)]| &\leq E[|(X-\mu_X)(Y-\mu_Y)|].
\end{aligned}
$$

Pela desigualdade de Cauchy-Schwarz temos

$$E[|(X-\mu_X)(Y-\mu_Y)|] \leq \sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}.$$

Logo

$$
\begin{aligned}
|E[(X-\mu_X)(Y-\mu_Y)]| &\leq \sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}\\
|Cov[X,Y]| &\leq \sigma_X\sigma_Y\\
|\rho_{X,Y}| &\leq 1.
\end{aligned}
$$

Para o item (ii), pelo Teorema 8 (expressão (27)) temos que

$$|E[UV]| = \sqrt{E[U^2]E[V^2]}$$

se $V = aU$. Considerando que $U = (X - \mu_X)$ e $V = (Y - \mu_Y)$, logo

$$\begin{aligned}(Y - \mu_Y) &= a(X - \mu_X)\\ Y &= aX - a\mu_X + \mu_Y\\ Y &= aX + b,\end{aligned}$$

sendo $a = \pm\frac{E[Y^2]}{\sqrt{E[X^2]}}$ e $b = \mu_Y - a\mu_X$.

Agora, considerando $Y = aX + b$ vamos calcular a correlação de X e Y. Sabendo que $\mu_X = a\mu_X + b$ e $\sigma_Y^2 = a^2 Var[X]$, então

$$\begin{aligned}Cov[X, Y] &= E[XY] - \mu_X\mu_Y\\ &= E[X(aX + b)] - \mu_X(a\mu_X + b)\\ &= aE[X^2] + b\mu_X - a\mu_X^2 - b\mu_X\\ &= a(E[X^2] - \mu_X^2)\\ &= a\sigma_X^2.\end{aligned}$$

Logo,

$$|\rho_{X,Y}| = \frac{a\sigma_X^2}{\sigma_X(|a|\sigma_X)} = \begin{cases} 1, & a > 0;\\ -1, & a < 0.\end{cases}$$

□

6.1 Esperança condicional bidimensional

Definição 15 (Esperança condicional bidimensional)**.** *Sejam X e Y variáveis aleatórias no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e $g(.)$ uma função das duas variáveis aleatórias. Então a esperança condicional de $g(X, Y)$ dado $Y = y$, denotada $E[g(X, Y)|Y = y]$, é definida por*

$$E[g(X, Y)|Y = y] = \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x|y)dx, \tag{30}$$

se (X, Y) forem conjuntamente contínuas, e

$$E[g(X, Y)|Y = y] = \sum_x g(x, y) P_{X|Y}(x|y), \tag{31}$$

se (X, Y) forem conjuntamente discreta.

Essa Definição pode ser estendida para o caso p-dimensional. (Mood, p.157; Daniel p. 164, 21.01.2016).

Teorema 10. *Seja X e Y duas variáveis aleatórias em $(\Omega, \mathcal{F}, P)$, então*

$$E[g(Y)] = E[E[g(Y)|X]] \tag{32}$$

e em particular

$$E[Y] = E[E[Y|X]]. \tag{33}$$

□

Demonstração.

$$\begin{aligned}
E[E[g(Y)|X]] = E[h(X)] &= \int_{-\infty}^{\infty} h(x) f_X(x) dx \\
&= \int_{-\infty}^{\infty} E[g(Y)|x] f_X(x) dx \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x) dy \right] f_X(x) dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x) f_X(x) dy dx \text{ (Definição 9)} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y) f_{X,Y}(x,y) dy dx \\
&= E[g(Y)].
\end{aligned}$$

Para o outro resultado, basta substituir $g(Y)$ por Y. □

A $E[Y|X]$ é conhecida como a curva de regressão de Y sobre x.

Definição 16. *A variância de Y dado $X = x$ é definida por*

$$Var[Y|X = x] = E[Y^2|X = x] - (E[Y|X = x])^2. \tag{34}$$

□

Teorema 11. $Var[Y] = E[Var[Y|X]] + Var[E[Y|X]]$. □

Demonstração.

$$\begin{aligned}
E[Var[Y|X]] &= E[E[Y^2|X]] - E[(E[Y|X])^2] \\
&= E[Y^2] - E[(E[Y|X])^2] \\
&= Var[Y] + (E[Y])^2 - E[(E[Y|X])^2] \\
&= Var[Y] + (E[E[Y|X]])^2 - E[(E[Y|X])^2] \\
&= Var[Y] - Var[E[Y|X]].
\end{aligned}$$

□

7 Função geradora de momentos conjunta, Função característica conjunta e Momentos

Definição 17 (Momentos conjuntos não centrais). *O momentos conjuntos não centrais de $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$, em que os r_i's são 0 ou qualquer inteiro positivo, são definidos por:*

$$E[X_1^{r_1} X_1^{r_2} \ldots X_p^{r_p}] = \sum_{x_1} \sum_{x_2} \ldots \sum_{x_p} x_1^{r_1} x_2^{r_2} \ldots x_p^{r_p} P_{\boldsymbol{X}}(\boldsymbol{x}), \tag{35}$$

para o caso discreto, e

$$E[X_1^{r_1} X_1^{r_2} \ldots X_p^{r_p}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_1^{r_1} x_2^{r_2} \ldots x_p^{r_p} f_{\boldsymbol{X}}(\boldsymbol{x}), \tag{36}$$

para o caso contínuo.

Definição 18 (Momentos Centrais). *O momentos conjuntos centrais de* $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$, *em que os* r_i*'s são 0 ou qualquer inteiro positivo, são definidos por:*

$$E[(X_1 - \mu_{X_1})^{r_1} \ldots (X_1 - \mu_{X_p})^{r_p}] = \sum_{x_1} \ldots \sum_{x_p} (x_1 - \mu_{X_1})^{r_1} \ldots (x_1 - \mu_{X_p})^{r_p} P_{\boldsymbol{X}}(\boldsymbol{x}), \quad (37)$$

para o caso discreto, e

$$E[(X_1 - \mu_{X_1})^{r_1} \ldots (X_1 - \mu_{X_p})^{r_p}] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_1 - \mu_{X_1})^{r_1} \ldots (x_1 - \mu_{X_p})^{r_p} f_{\boldsymbol{X}}(\boldsymbol{x}), \quad (38)$$

para o caso contínuo.

Se $r_i = r_j = 1$ e os demais r_m's são 0, então um caso particular do momento conjunto central bidimensional é $E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = Cov[X_i, X_j]$.

Definição 19 (Função geradora de momentos conjunta). *A função geradora de momentos conjunta de* $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$ *é definido por:*

$$m_{\boldsymbol{X}}(t_1, \ldots, t_k) = E\left[e^{\sum_{i=1}^{p} t_i X_i}\right], \quad (39)$$

se a esperança existe em todos os valores $t_1, \ldots, t_k$ *tal que* $-h < t_i < h$ *para algum* $h > 0$, $i = 1, 2, \ldots, p$.

O r-ésimo momento de X_i pode ser obtido pela função geradora $m_{\boldsymbol{X}}(t_1, \ldots, t_k)$ diferenciando r vezes com respeito a t_i e então tome os limites para todos os $t's$ em torno de 0. Podemos também obter $E[X_j^r X_j^s]$ pela diferenciação da função geradora de momentos conjunta r vezes com respeito a t_i e s vezes com respeito a t_j e então tomar os limites quando todos os t_s aproximam de 0.

Teorema 12 ($m_{\boldsymbol{X}}(t_1, \ldots, t_k)$ para variáveis aleatórias independentes). *Se* $X_1, X_2, \ldots, X_p$ *são variáveis aleatórias independentes, então a função geradora de momentos conjunta é dada por:*

$$m_{\boldsymbol{X}}(t_1, \ldots, t_k) = \prod_{i=1}^{p} m_{X_i}(t_i), \quad (40)$$

se a esperança existe em todos os valores $t_1, \ldots, t_k$ *tal que* $-h < t_i < h$ *para algum* $h > 0$, $i = 1, 2, \ldots, p$. □

Demonstração. Se $X_1, X_2, \ldots, X_p$, então

$$\begin{aligned} m_{\boldsymbol{X}}(t_1, \ldots, t_k) &= E\left[e^{\sum_{i=1}^{p} t_i X_i}\right] \\ &= E\left[e^{t_1 X_1 + t_2 X_2 + \ldots + t_p X_p}\right] \\ &= E\left[e^{t_1 X_1} e^{t_2 X_2} \ldots e^{t_p X_p}\right] \\ &= E\left[e^{t_1 X_1}\right] E\left[e^{t_2 X_2}\right] \ldots E\left[e^{t_p X_p}\right] \\ &= \prod_{i=1}^{p} m_{X_i}(t_i). \end{aligned}$$

□

Corolário 1 (fgm para o caso iid)**.** *content...*

Definição 20 (Função característica conjunta)**.** *A função característica conjunta de* $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$ *é definido por:*

$$\varphi_{\boldsymbol{X}}(t_1, \ldots, t_k) = E\left[e^{i(\sum_{i=1}^{p} t_i X_i)}\right], \tag{41}$$

em que $t_1, t_2, \ldots, t_p$ *são números reais qualquer.* □

O r-ésimo momento de X_i pode ser obtido pela função característica conjunta $\varphi_{\boldsymbol{X}}(t_1, \ldots, t_k)$ diferenciando r vezes com respeito a t_i e então tome os limites para todos os $t's$ em torno de 0. Podemos também obter $E[X_j^r X_j^s]$ pode ser obtido pela diferenciação da função característica conjunta r vezes com respeito a t_i e s vezes com respeito a t_j e então tomar os limites quando todos os t_s aproximam de 0.

Teorema 13 ($\varphi_{\boldsymbol{X}}(t_1, \ldots, t_k)$ para variáveis aleatórias independentes)**.** *Se* $X_1, X_2, \ldots, X_p$ *são variáveis aleatórias independentes, então a função geradora de momentos conjunta é dada por:*

$$\varphi_{\boldsymbol{X}}(t_1, \ldots, t_k) = \prod_{i=1}^{p} \varphi_{X_i}(t_i), \tag{42}$$

em que $t_1, t_2, \ldots, t_p$ *são números reais qualquer.* □

Demonstração. Se $X_1, X_2, \ldots, X_p$, então

$$\begin{aligned}
\varphi_{\boldsymbol{X}}(t_1, \ldots, t_k) &= E\left[e^{i(\sum_{i=1}^{p} t_i X_i)}\right] \\
&= E\left[e^{i(t_1 X_1 + t_2 X_2 + \ldots + t_p X_p)}\right] \\
&= E\left[e^{it_1 X_1} e^{it_2 X_2} \ldots e^{it_p X_p}\right] \\
&= E\left[e^{it_1 X_1}\right] E\left[e^{it_2 X_2}\right] \ldots E\left[e^{it_p X_p}\right] \\
&= \prod_{i=1}^{p} \varphi_{X_i}(t_i).
\end{aligned}$$

□

Corolário 2 (função característica para o caso iid)**.** *content...*

8 Transformações de variáveis aleatórias

A técnica da função de distribuição é apresentado a seguir

Teorema 14 (Distribuição força-bruta de uma função (DeGroot, p. 178))**.** *Seja* $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ *uma variável aleatória p-dimensional em* $(\Omega, \mathcal{F}, P)$*, com função densidade conjunta* $f_{\boldsymbol{X}}(\boldsymbol{x})$*. Considere ainda* $Y = g(\boldsymbol{X})$*. Para cada número real* y*, defina* $S_y = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq y\}$*. Então a função de distribuição de* Y *é dada por:*

$$F_Y(y) = \int \underset{S_y}{\cdots} \int f_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x} \tag{43}$$

Para explicar essa técnica, consideraremos o caso de duas variáveis aleatórias X_1 e X_2 com distribuição conjunta conhecida e assumindo que elas são mapeadas por Y_1 e Y_2 pela transformação $y_1 = g_1(x, y)$ e $y_2 = g_2(x, y)$, sendo $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}^1$.

Teorema 15 (Funções de variáveis aleatórias bidimensionais discreta)**.** *Seja $\boldsymbol{X} = [X_1, X_2]'$ uma variável aleatória bidimensional discreta, com $P_{\boldsymbol{X}}(\boldsymbol{x})$ conhecida. Considere ainda que $Y_1 = g_1(\boldsymbol{X})$ e $Y_2 = g_2(\boldsymbol{X})$, em que $g_j : \mathbb{R}^2 \to \mathbb{R}^1$. Então $\boldsymbol{Y} = [Y_1, Y_2]'$ é uma variável aleatória bidimensional também discreta, e sua função de probabilidade conjunta, denotada $p_{\boldsymbol{Y}}(\boldsymbol{y})$, é dada por:*

$$P_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{x} \in B} P_{\boldsymbol{X}}(\boldsymbol{x}) \tag{44}$$

em que $B = \{\boldsymbol{x} \in S_{\boldsymbol{X}} : g_j(\boldsymbol{x}) = y_j,\ j = 1, 2\}$ e $S_{\boldsymbol{X}}$ é o suporte de $\boldsymbol{X}$.

Demonstração. Considere $\boldsymbol{g}(\boldsymbol{X}) = \boldsymbol{Y}$, então

$$\begin{aligned} P_{\boldsymbol{Y}}(\boldsymbol{y}) &= P_{\boldsymbol{Y}}(\boldsymbol{Y} = \boldsymbol{y}) \\ &= P_{\boldsymbol{Y}}(\boldsymbol{g}(\boldsymbol{X}) = \boldsymbol{y}) \\ &= \sum_{\boldsymbol{g}^{-1}(\boldsymbol{y}) = \boldsymbol{x} \in B} P_{\boldsymbol{X}}(\boldsymbol{X} = \boldsymbol{g}^{-1}(\boldsymbol{y})). \end{aligned}$$

□

Inserir exemplo Rosenthal p. 110.,

Para o caso contínuo, iremos considerar também $\boldsymbol{X} = [X_1, X_2]'$ uma variável aleatória bidimensional.

Teorema 16 (Funções de variáveis aleatórias bidimensionais contínuas)**.** *Seja $\boldsymbol{X} = [X_1, X_2]'$ uma variável aleatória bidimensional contínua, com $f_{\boldsymbol{X}}(\boldsymbol{x})$ conhecida e suporte $S_{\boldsymbol{X}}$. Considere ainda que $Y_1 = g_1(\boldsymbol{X})$, $Y_2 = g_2(\boldsymbol{X})$ e $\boldsymbol{g}(\boldsymbol{X}) = [Y_1, Y_2]'$, em que $\boldsymbol{g} : \mathbb{R}^2 \to \mathbb{R}^2$ é uma função injetiva (um-para-um). Então se $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ cuja função inversa $\boldsymbol{X} = \boldsymbol{g}^{-1}(\boldsymbol{Y})$ existe e é diferenciável para todo $\boldsymbol{y} \in S_{\boldsymbol{Y}}$, então a sua função densidade de probabilidade conjunta, denotada $f_{\boldsymbol{Y}}(\boldsymbol{y})$, é dada por:*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{J}| f_{\boldsymbol{X}}(\boldsymbol{g}^{-1}(\boldsymbol{y})), \tag{45}$$

sendo $\boldsymbol{J}$ o Jacobiano da transformação definido por:

$$\boldsymbol{J} = \begin{vmatrix} \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \\ \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} \end{vmatrix} = \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} - \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1}. \tag{46}$$

□

Demonstração. Sabemos que se a função densidade de probabilidade conjunta de $\boldsymbol{X} = [X_1, X_2]$ existe então $F_{\boldsymbol{X}}(\boldsymbol{x})$ existe. Como a função $\boldsymbol{g}$ é injetiva sua inversa tem valor único. Logo,

$$F_{\boldsymbol{Y}}(\boldsymbol{y}) = P(\boldsymbol{Y} \leq \boldsymbol{y}) = P(\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{y}) = P(\boldsymbol{x} \leq g^{-1}(\boldsymbol{y})) = F_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y})).$$

Como desejamos a função densidade de $\boldsymbol{Y}$, logo

$$
\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{\partial}{\partial \boldsymbol{y}} F'_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y})) \\
&= \left| \frac{\partial g^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}} \right| F'_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y})) \\
&= |\boldsymbol{J}| f_{\boldsymbol{X}}(\boldsymbol{g^{-1}}(\boldsymbol{y})),
\end{aligned}
$$

sendo $|\boldsymbol{J}| = \left|\frac{\partial g^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}}\right|$. □

Outros métodos de encontrar distribuições de funções de variáveis aleatórias:

- técnica da função de distribuição acumulada (Mood. 181-182);
- Teorema da probabilidade total (Mood p. 202; material de prob I Devanil);
- Técnica pela função geradora de momentos (Mood p. 189).

9 Distribuições de transformações de variáveis aleatórias multidimensionais

Para mostrar as técnicas de como obter as distribuições de funções de variáveis aleatórias algumas demonstrações serão com base no caso bidimensional para facilitar o entendimento.

Teorema 17 (Distribuição de soma e diferença de duas variáveis aleatórias)**.** *Sejam X e Y duas variáveis aleatórias contínuas com função densidade conjunta $f_{X,Y}(x,y)$, e considere $Z = X + Y$ e $V = X - Y$, então*

$$
f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)dy, \tag{47}
$$

e

$$
f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, x-v)dx = \int_{-\infty}^{\infty} f_{X,Y}(v+y, y)dy. \tag{48}
$$

Demonstração.

$$
\begin{aligned}
F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) \\
&= \iint\limits_{x+y\leq z} f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_{X,Y}(x,y)dy \right] dx \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z} f_{X,Y}(x, u-x)du \right] dx. \qquad (y = u - x)
\end{aligned}
$$

Agora,

$$
\begin{aligned}
f_Z(z) &= \frac{d}{dz}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,u-x)du\right]dx\\
&= \int_{-\infty}^{\infty}\left[\frac{d}{dz}\int_{-\infty}^{z} f_{X,Y}(x,u-x)du\right]dx\\
&= \int_{-\infty}^{\infty}\left[\frac{d}{dz}(F_{X,Y}(x,z-x)-\lim_{u\to-\infty}F_{X,Y}(x,u-x))\right]dx\\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx.
\end{aligned}
$$

O mesmo pode ser feito para (48). $\square$

Corolário 3 (Convolução)**.** *Se X e Y são variáveis aleatórias independentes e $Z = X+Y$, então*

$$
f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy, \tag{49}
$$

$\square$

Demonstração. Aplicando a Definição de independência em (47) o resultado é imediato. $\square$

Exemplos: Hogg (p.87)

Teorema 18 (Distribuição do produto e quociente para duas variáveis aleatórias)**.** *Sejam X e Y duas variáveis aleatórias contínuas com função densidade conjunta $f_{X,Y}(x,y)$, e considere $Z = XY$ e $U = X/Y$, então*

$$
f_Z(z) = \int_{-\infty}^{\infty}\frac{1}{|x|}f_{X,Y}\left(x,\frac{z}{x}\right)dx = \int_{-\infty}^{\infty}\frac{1}{|y|}f_{X,Y}\left(\frac{z}{y},y\right)dy \tag{50}
$$

e

$$
f_U(u) = \int_{-\infty}^{\infty}|y|f_{X,Y}(uy,y)dy. \tag{51}
$$

$\square$

Demonstração.

$$
\begin{aligned}
F_Z(Z\leq z) &= \iint\limits_{xy\leq z} f_{X,Y}(x,y)dxdy\\
&= \int_{-\infty}^{0}\left[\int_{z/x}^{\infty} f_{X,Y}(x,y)dy\right]dx + \int_{0}^{\infty}\left[\int_{-\infty}^{z/x} f_{X,Y}(x,y)dy\right]dx\\
&= \int_{-\infty}^{0}\left[\int_{z}^{-\infty} f_{X,Y}(x,\frac{u}{x})\frac{du}{x}\right]dx + \int_{0}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\frac{du}{x}\right]dx, \qquad (u=xy)\\
&= \int_{-\infty}^{0}\left[\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\frac{du}{-x}\right]dx + \int_{0}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,\frac{u}{x})\frac{du}{x}\right]dx\\
&= \int_{-\infty}^{z}\left[\int_{-\infty}^{0} f_{X,Y}(x,\frac{u}{x})\frac{dx}{-x}\right]du + \int_{-\infty}^{z}\left[\int_{0}^{\infty} f_{X,Y}(x,\frac{u}{x})\frac{dx}{x}\right]du\\
&= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_{X,Y}(x,\frac{u}{x})\frac{dx}{|x|}\right]du,
\end{aligned}
$$

então

$$\begin{aligned} f_Z(z) &= \frac{d}{dz}F_Z(z) \\ &= \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}(x, \frac{z}{x})dx. \end{aligned}$$

$$\begin{aligned} F_U(U \leq u) &= \iint\limits_{x/y \leq u} f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{0}\left[\int_{uy}^{\infty} f_{X,Y}(x,y)dx\right]dy + \int_{0}^{\infty}\left[\int_{-\infty}^{uy} f_{X,Y}(x,y)dx\right]dy \\ &= \int_{-\infty}^{0}\left[\int_{u}^{-\infty} yf_{X,Y}(sy,y)ds\right]dy + \int_{0}^{\infty}\left[\int_{-\infty}^{u} yf_{X,Y}(sy,s)ds\right]dy, \qquad (s = x/y) \\ &= \int_{-\infty}^{0}\left[\int_{-\infty}^{u} (-y)f_{X,Y}(sy,y)ds\right]dy + \int_{0}^{\infty}\left[\int_{-\infty}^{u} yf_{X,Y}(sy,s)ds\right]dy \\ &= \int_{-\infty}^{u}\left[\int_{-\infty}^{0} (-y)f_{X,Y}(sy,y)ds\right]dy + \int_{-\infty}^{u}\left[\int_{0}^{\infty} yf_{X,Y}(sy,s)dy\right]ds \\ &= \int_{-\infty}^{u}\left[\int_{-\infty}^{\infty} |y|f_{X,Y}(sy,y)dy\right]ds \end{aligned}$$

então

$$\begin{aligned} f_U(u) &= \frac{d}{du}F_U(u) \\ &= \int_{-\infty}^{\infty} |y|f_{X,Y}(uy,y)dy. \end{aligned}$$

□

Exemplos DeGroot p.183

10 Estatísticas de ordem

Seja X_1, X_2, ..., X_n, uma amostra aleatória de tamanho n de uma função de distribuição $F(.)$. Então $Y_1 \leq Y_2 \leq \ldots \leq Y_\alpha \leq \ldots \leq Y_n$, em que Y_α é a X_i considerada em ordem crescente de magnitude chamada estatística de ordem correspondente à amostra X_1, X_2, ..., X_n. Considere $y \in \mathbb{R}$ fixado, e seja $Z_i = I_{(-\infty,y]}(X_i)$, isto é, $Z_i = 1$ se $X_i \in (-\infty, y]$ e zero caso contrário. Então, a variável aleatória Z_i tem distribuição de Bernoulli de parâmetro $p = F_X(y)$. Com isto,

$$\sum_{i=1}^{n} Z_i = \text{número de } X_i{}'s \leq y,$$

tem distribuição binomial com parâmetros n e $F_X(y)$. Assim, a função de distribuição marginal de Y_α, $\alpha = 1, 2, \ldots, n$ é dado por:

$$\begin{aligned}
F_{Y_\alpha}(y) &= P[Y_\alpha \leq y] = P\left[\sum_{i=1}^{n} Z_i \geq \alpha\right], \\
F_{Y_\alpha}(y) &= \sum_{j=\alpha}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}, \qquad (52)
\end{aligned}$$

ou seja, deve-se garantir que $Y_\alpha \leq y$, que em outras palavras, que ao menos α valores dos $X_i's$ sejam menores ou iguais a y, como observado na Figura 1.

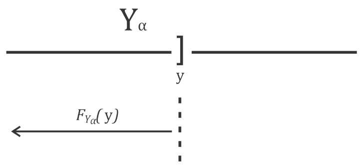

Figura 1: Reta da função distribuição

Para se determinar a função densidade de probabilidade de Y_α, será usado o auxílio da Figura 2.

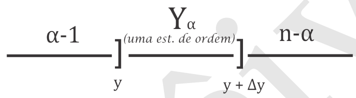

Figura 2: Reta da função densidade

Nos cálculos apresentados, será utilizada a função de probabilidade multinomial, com função conjunta dada por:

$$P(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1! k_2! \ldots k_m!} p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m},$$

com $\sum_{i=1}^{m} p_i = 1$ e $\sum_{i=1}^{m} k_i = n$, $k_i \in \mathbb{N}$, $0 \leq k_i \leq n$.

Assim, a probabilidade de $\alpha - 1$ $X_i's$ estarem contidos no intervalo $(-\infty, y]$ é $p_1 = F(y)$; a probabilidade de 1 X_i estar contido em $(y, y + \Delta y]$, é $p_2 = F(y + \Delta y) - F(y)$; e a probabilidade de $n - \alpha$ $X_i's$ estarem contidos no intervalo $(y + \Delta y, \infty)$ é $p_3 = 1 - F(y + \Delta y)$. Portanto,

$$\begin{aligned}
f_{Y_\alpha}(y) &= \lim_{\Delta y \to 0} \frac{F_{Y_\alpha}(y + \Delta y) - F_{Y_\alpha}(y)}{\Delta y} = \lim_{\Delta y \to 0} \frac{P[y < Y_\alpha \leq y + \Delta y]}{\Delta y}, \\
&= \lim_{\Delta y \to 0} P[(\alpha - 1) \text{ dos } X_i \leq y;\ 1\ X_i \text{ em } (y, y + \Delta y]; \\
&\quad (n - \alpha) \text{ dos } X_i > y + \Delta y] / \Delta y, \\
&= \frac{n!}{(\alpha - 1)!(n - \alpha)!} [F(y)]^{\alpha - 1} [1 - F(y)]^{n - \alpha} \times \\
&\quad \times \underbrace{\lim_{\Delta y \to 0} \frac{F(y + \Delta y) - F(y)}{\Delta y}}_{f(y)}, \\
f_{Y_\alpha}(y) &= \frac{n!}{(\alpha - 1)!(n - \alpha)!} [F(y)]^{\alpha - 1} [1 - F(y)]^{n - \alpha} f(y). \qquad (53)
\end{aligned}$$

Utilizando (52), pode-se obter a função de distribuição das estatísticas de ordem Y_1 e Y_n. Para Y_α com $\alpha = 1$, tem-se

$$F_{Y_1}(y) \quad = \quad \sum_{j=1}^{n} \binom{n}{1} [F(y)]^1 [1 - F(y)]^{n-1}.$$

Observando que

$$\binom{n}{0} [F(y)]^0 [1 - F(y)]^{n-0} + F_{Y_1}(y) = 1,$$

e que,

$$\binom{n}{0} [F(y)]^0 [1 - F(y)]^{n-0} = [1 - F(y)]^n,$$

então, a função de distribuição $F_{Y_1}(y)$ será:

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n. \tag{54}$$

Derivando (54) com relação a y, encontra-se a função densidade, ou seja,

$$f_{Y_1} = \frac{\partial F_{Y_1}(y)}{\partial y} = nf(y)[1 - F(y)]^{n-1} \tag{55}$$

Já a função de distribuição para $F_{Y_n}(y)$, utilizando (52), é

$$\begin{aligned} F_{Y_n}(y) \quad &= \quad \sum_{j=n}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}, \\ F_{Y_n}(y) \quad &= \quad \binom{n}{n} [F(y)]^n [1 - F(y)]^{n-n}, \\ F_{Y_n}(y) \quad &= \quad [F(y)]^n, \end{aligned} \tag{56}$$

em que derivando (56) com relação a y, encontra-se a função densidade,

$$f_{Y_n} = \frac{dF_{Y_n}(y)}{dy} = nf(y)[F(y)]^{n-1}. \tag{57}$$

Analogamente, pode-se deduzir a distribuição conjunta entre duas estatísticas de ordem Y_α e Y_β para $1 \leq \alpha < \beta \leq n$, observando a Figura 3.

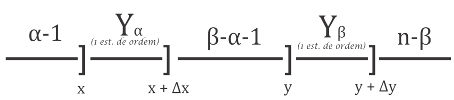

Figura 3: Reta da função densidade conjunta

Para isso, usando, a distribuição multinomial, observa-se que a probabilidade de $\alpha - 1$ $X_i's$ estarem contidos em $(-\infty, x]$ é $p_1 = F(x)$; a probabilidade de 1 X_i estar contido no intervalo $(x, x + \Delta x]$ é $p_2 = F(x + \Delta x) - F(x)$; a probabilidade de $\beta - \alpha - 1$

$X_i's$ estarem contidos em $(x+\Delta x, y]$ é $p_3 = F(y) - F(x+\Delta x)$; a probabilidade de 1 X_i estar contido em $(y, y+\Delta y]$ é $p_4 = F(y+\Delta y) - F(y)$; e a probabilidade de $n-\beta$ $X_i's$ estarem contidos no intervalo $(y+\Delta y, \infty)$ é $p_5 = 1 - F(y+\Delta y)$. Assim,

$$\begin{aligned}
f_{Y_\alpha,Y_\beta}(x,y)\Delta x\Delta y &\approx P[x < Y_\alpha \le x+\Delta x; y < Y_\beta \le y+\Delta y],\\
&\approx P[(\alpha-1) \text{ dos } X_i \le x;\ 1 \text{ dos } X_i \in (x, x+\Delta x]; \ldots\\
&\quad \ldots(\beta-\alpha-1) \text{ dos } X_i \in (x+\Delta x, y]; \ldots\\
&\quad \ldots\ 1 \text{ dos } X_i \in (y, y+\Delta y]; (n-\beta) \text{ dos } X_i > y+\Delta y],\\
f_{Y_\alpha,Y_\beta}(x,y) &= \lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}} \left\{\frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}\times\right.\\
&\quad \times[F(x)]^{\alpha-1}[F(y)-F(x+\Delta x)]^{\beta-\alpha-1}\times\\
&\quad \times[1-F(y+\Delta y)]^{n-\beta}\times\\
&\quad \left.\times\left[\frac{F(x+\Delta x)-F(x)}{\Delta x}\right]^1\left[\frac{F(y+\Delta y)-F(y)}{\Delta y}\right]^1\right\},\\
f_{Y_\alpha,Y_\beta}(x,y) &= \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}[F(x)]^{\alpha-1}\times\\
&\quad \times\lim_{\Delta x\to 0}[F(y)-F(x+\Delta x)]^{\beta-\alpha-1}\times\\
&\quad \times\lim_{\Delta y\to 0}[1-F(y+\Delta y)]^{n-\beta}\times\\
&\quad \times\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\left\{\left[\frac{F(x+\Delta x)-F(x)}{\Delta x}\right]^1\left[\frac{F(y+\Delta y)-F(y)}{\Delta y}\right]^1\right\},\\
f_{Y_\alpha,Y_\beta}(x,y) &= \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}[F(x)]^{\alpha-1}\times\\
&\quad \times[F(y)-F(x)]^{\beta-\alpha-1}[1-F(y)]^{n-\beta}f(x)f(y). \qquad (58)
\end{aligned}$$

A função densidade conjunta das estatísticas de ordem Y_1 e Y_n pode ser deduzida da função densidade (58) para $\alpha = 1$ e $\beta = n$, por:

$$\begin{aligned}
f_{Y_1,Y_n}(x,y) &= \frac{n!}{(1-1)!(n-1-1)!(n-n)!}\times\\
&\quad \times[F(x)]^{1-1}[F(y)-F(x+\Delta x)]^{n-1-1}\times\\
&\quad \times[1-F(y+\Delta y)]^{n-n}f(x)f(y),\\
f_{Y_1,Y_n}(x,y) &= n(n-1)f(x)f(y)[F(y)-F(x)]^{n-2}, \qquad (59)
\end{aligned}$$

para $x < y$.

Assim, o resumo das distribuições,

- Função de distribuição:
  - $F_{Y_1}(y) = 1 - [1-F(y)]^n$,
  - $F_{Y_n}(y) = [F(y)]^n$.
- Função densidade de probabilidade:
  - $f_{Y_1}(y) = nf(y)[1-F(y)]^{n-1}$,

– $f_{Y_n}(y) = nf(y)[F(y)]^{n-1}$.

- Função densidade de probabilidade conjunta:
  – $f_{Y_1,Y_n}(x,y) = n(n-1)f(x)f(y)[F(y)-F(x)]^{n-2}$ para $x < y$.

11 Principais distribuições

Definição 21 (Distribuição multinomial)**.** *Uma variável aleatória k-dimensional discreta $\boldsymbol{X} = [X_1, X_2, \ldots, X_k]'$ em $(\Omega, \mathcal{F}, P)$, tem distribuição multinomial se sua função de probabilidade é dada por:*

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{cases} \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} & \text{se } x_1 + x_2 + \ldots x_k = n, \\ 0 & \text{caso contrário,} \end{cases} \tag{60}$$

com $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$, $x_i \in \mathbb{N}$, $0 \leq x_i \leq n$, para $i = 1, 2, \ldots, k$. □

Teorema 19 (Esperança e variância da distribuição multinomial)**.** *Seja $\boldsymbol{X} = [X_1, X_2, \ldots, X_k]$ uma amostra aleatória k-dimensional com distribuição multinomial com parâmetros n e $\boldsymbol{p} = [p_1, p_2, \ldots, p_k]'$. A média e a variância de X_i são dadas por:*

$$E[X_i] = np_i \qquad e \qquad Var[X_i] = np_i(1-p_i), \qquad i = 1, 2, \ldots, k. \tag{61}$$

□

Demonstração. Pode ser encontrada DeGroot p. 336. □

Definição 22 (Distribuição normal multivariada)**.** *Sejam $X_1, X_2, \ldots, X_p$ variáveis aleatórias contínuas independentes tal que $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, 2, \ldots, p$. Então $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ distribuição normal multivariada se sua função densidade conjunta é dada por:*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \tag{62}$$

em que $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p]' \in \mathbb{R}^p$ e Σ é uma matriz positiva definida dada por:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma_3^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \sigma_p^2 \end{bmatrix}. \tag{63}$$

Teorema 20 (Esperança e variância da distribuição normal multivariada)**.** *Seja $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ uma amostra aleatória k-dimensional com distribuição normal multivariada com parâmetros $\boldsymbol{\mu} \in \mathbb{R}^p$ e Σ uma matriz positiva definida $p \times p$. A média e a variância de $\boldsymbol{X}$ são dadas por:*

$$E[\boldsymbol{X}] = \boldsymbol{\mu} \qquad e \qquad Var[\boldsymbol{\mu}] = \Sigma. \tag{64}$$

□