5 Vetores aleatórios

5.1 Distribuição conjunta

O que vimos atrás sobre a função de distribuição generaliza-se facilmente para o estudo simultâneo de duas variáveis aleatórias.

Definição

Seja (X_1, X_2) uma variável aleatória bidimensional. Então

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2),\ \forall (x_1, x_2) \in \mathbb{R}^2$$

diz-se a **função de distribuição conjunta** de (X_1, X_2).

Algumas propriedades

1. $0 \leq F_{X_1,X_2}(x_1, x_2) \leq 1,\ \forall (x_1, x_2) \in \mathbb{R}^2$
2. $F_{X_1,X_2}(x_1 + \Delta_{x_1}, x_2 + \Delta_{x_2}) \geq F_{X_1,X_2}(x_1, x_2), \forall \Delta_{x_1}, \Delta_{x_2} \geq 0$
3. $\lim\limits_{x_1,x_2 \to +\infty} F_{X_1,X_2}(x_1, x_2) = 1$
4. $\lim\limits_{x_i \to -\infty} F_{X_1,X_2}(x_1, x_2) = 0$
5. $\lim\limits_{x_i \to +\infty} F_{X_1,X_2}(x_1, x_2) = F_{X_j}(x_j)$, com $i \neq j$

Definição

Seja (X_1, X_2) uma variável aleatória bidimensional discreta. Então

$$f_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2),\ \forall (x_1, x_2) \in \mathbb{R}^2$$

diz-se a **função de probabilidade conjunta** de (X_1, X_2).

Notas

1. $f_{X_1,X_2}(x_1, x_2) \geq 0, \forall (x_1, x_2) \in \mathbb{R}^2$

2. $\sum_{x_1} \sum_{x_2} f_{X_1,X_2}(x_1, x_2) = 1$

Exemplo – Ases e figuras

Tiram-se duas cartas ao acaso de um baralho de 52 cartas. Sejam X o número de figuras (reis, damas ou valetes) e Y o número de ases obtidos.

a. Determine a função de probabilidade conjunta do par aleatório (X, Y).

b. Calcule os valores da função de distribuição conjunta nos pontos (0,1.5) e (3,1).

Exemplo – Ases e figuras (continuação)

$Y \backslash X$	**0**	**1**	**2**
0	$\frac{105}{221}$	$\frac{72}{221}$	$\frac{11}{221}$
1	$\frac{24}{221}$	$\frac{8}{221}$	0
2	$\frac{1}{221}$	0	0

Nota

Não confundir a função de probabilidade conjunta (definida em $\mathbb{R}^2$) com a sua representação abreviada na tabela acima!

Definição

Seja (X_1, X_2) uma variável aleatória bidimensional contínua. A função

$$f_{X_1,X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2},$$

$\forall (x_1, x_2) \in \mathbb{R}^2$ onde F_{X_1,X_2} é diferenciável, diz-se a **função densidade de probabilidade conjunta** de (X_1, X_2).

Notas

1. $f_{X_1,X_2}(x_1, x_2) \geq 0, \forall (x_1, x_2) \in \mathbb{R}^2$

2. $\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X_1,X_2}(x_1, x_2)\, dx_2\, dx_1 = 1$

5.2 Distribuições marginais e condicionais

Definição

Seja (X_1, X_2) uma variável aleatória bidimensional. Então

$$f_{X_i}(x_i) = \sum_{x_j} f_{X_1,X_2}(x_1, x_2) \left(= \int_{\mathbb{R}} f_{X_1,X_2}(x_1, x_2)\, dx_j \right),$$

com $i \neq j$ e $\forall x_i \in \mathbb{R}$, diz-se a **função (densidade) de probabilidade marginal** de X_i.

Definição

Seja (X_1, X_2) uma variável aleatória bidimensional. Então

$$f_{X_i|X_j=x_j}(x_i) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_j}(x_j)},$$

$\forall x_i \in \mathbb{R}$ e $\forall x_j \in \mathbb{R} : f_{X_j}(x_j) > 0$, diz-se a **função (densidade) de probabilidade condicional** de X_i dado $X_j = x_j$ para $i \neq j$.

5.3 Independência e correlação

Definição

Duas variáveis aleatórias, X e Y, dizem-se **independentes** se para todo $A,\ B \subset \mathbb{R}$ os acontecimentos $X \in A$ e $Y \in B$ são independentes, isto é, se

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

Teorema

As variáveis aleatórias X e Y são independentes se e só se

$$F_{X,Y}(x, y) = F_X(x)F_Y(y),\ \forall (x, y) \in \mathbb{R}^2.$$

Teorema

As variáveis aleatórias X e Y são independentes se e só se

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y),\ \forall (x,y) \in \mathbb{R}^2.$$

Qual o efeito da independência sobre as distribuições condicionais das variáveis aleatórias?

Valor esperado de uma função de um par aleatório

Teorema

Sejam X e Y duas variáveis aleatórias. O **valor esperado** de uma função $g(X,Y)$, caso exista, é dado por

$$E[g(X,Y)] = \begin{cases} \sum\limits_x \sum\limits_y g(x,y)\ f_{X,Y}(x,y), & \text{(discreto)} \\ \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} g(x,y)\ f_{X,Y}(x,y)\, dy\, dx, & \text{(contínuo)} \end{cases}$$

Teorema

Sendo X e Y duas variáveis aleatórias então

$$E[aX + bY] = aE[X] + bE[Y],\ \forall (a,b) \in \mathbb{R}^2,$$

ou seja, o valor esperado é um **operador linear**.

Teorema

Se X e Y são variáveis aleatórias independentes então

$$E[g(X)\ h(Y)] = E[g(X)]\ E[h(Y)],$$

para quaisquer funções g e h.

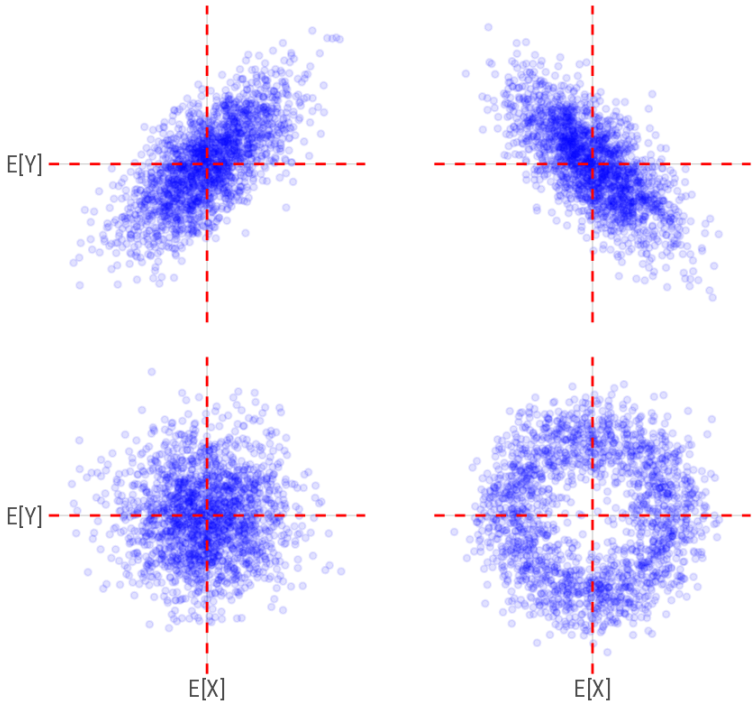
Covariância e coeficiente de correlação linear

Definição

A **covariância** de X e Y, caso exista, é definida por

$$Cov[X,Y] = E\left[(X - E[X])(Y - E[Y])\right]$$

Interpretação da covariância



Algumas propriedades da covariância

1. $Cov[X,Y] = Cov[Y,X]$
2. $Cov[X,X] = Var[X]$
3. $Cov[X,Y] = E[XY] - E[X]\,E[Y]$

Teorema

Se X e Y são variáveis aleatórias independentes então

$$Cov[g(X), h(Y)] = 0,$$

para quaisquer funções g e h.

Definição

O **coeficiente de correlação linear** de Pearson entre X e Y é definido por

$$Corr[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

Algumas propriedades do coeficiente de correlação linear

1. $-1 \leq Corr[X, Y] \leq 1$
2. $Corr[X, Y] = \pm 1 \iff Y = aX + b$, com $a \neq 0$ e $b \in \mathbb{R}$
3. $Corr[aX, Y + b] = \frac{a}{|a|} Corr[X, Y]$, $a \neq 0$ e $b \in \mathbb{R}$

5.4 Combinações lineares de variáveis aleatórias

Definição

Seja $X_1, \ldots, X_n$ uma sucessão de variáveis aleatórias. Uma **combinação linear** dessas variáveis é uma variável aleatória Y definida por

$$Y = \sum_{i=1}^{n} c_i X_i, \text{ com } (c_1, \ldots, c_n) \in \mathbb{R}^n.$$

Algumas propriedades

1. $E[Y] = \sum_{i=1}^{n} c_i \, E[X_i]$

2. $Cov\left[\sum_{i=1}^{m} c_i \, X_i, \sum_{j=1}^{n} d_j \, Y_j\right] = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i \, d_j \, Cov[X_i, Y_j]$

3. $Var[Y] = \sum_{i=1}^{n} c_i^2 \, Var[X_i] + 2 \sum_{i=1}^{n} \sum_{j>i} c_i \, c_j \, Cov[X_i, X_j]$

Teorema

Sejam $X_i \sim Bi(n_i, p)$, $i = 1, \ldots, n$, variáveis aleatórias independentes. Então

$$Y = \sum_{i=1}^{n} X_i \sim Bi\left(\sum_{i=1}^{n} n_i, p\right).$$

Teorema

Sejam $X_i \sim Poi(\lambda_i)$, $i = 1, \ldots, n$, variáveis aleatórias independentes. Então

$$Y = \sum_{i=1}^{n} X_i \sim Poi\left(\sum_{i=1}^{n} \lambda_i\right).$$

Teorema

Sejam $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, $i = 1, \ldots, n$, variáveis aleatórias independentes. Então

$$Y = \sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} (c_i \sigma_i)^2\right).$$

Em geral, não é fácil determinar a distribuição da soma de uma sucessão de variáveis aleatórias independentes!

Definição

Uma sucessão de variáveis aleatórias $X_1, \ldots, X_n$, com funções de distribuição $F_1, \ldots, F_n$, **converge em distribuição** para uma variável aleatória X $\left(X_n \xrightarrow{\mathcal{D}} X\right)$, quando $n \to +\infty$, se $F_n \to F_X$ em todo o ponto de continuidade de F_X.

Teorema do limite central

Seja $X_1, \ldots, X_n, \ldots$ uma sucessão de variáveis aleatórias não correlacionadas e identicamente distribuídas com valor esperado e variância finitos. Sendo $S_n = \sum_{i=1}^{n} X_i$, então, quando $n \to +\infty$,

$$Z_n = \frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} \xrightarrow{\mathcal{D}} Z \sim N(0,1)$$

Notas

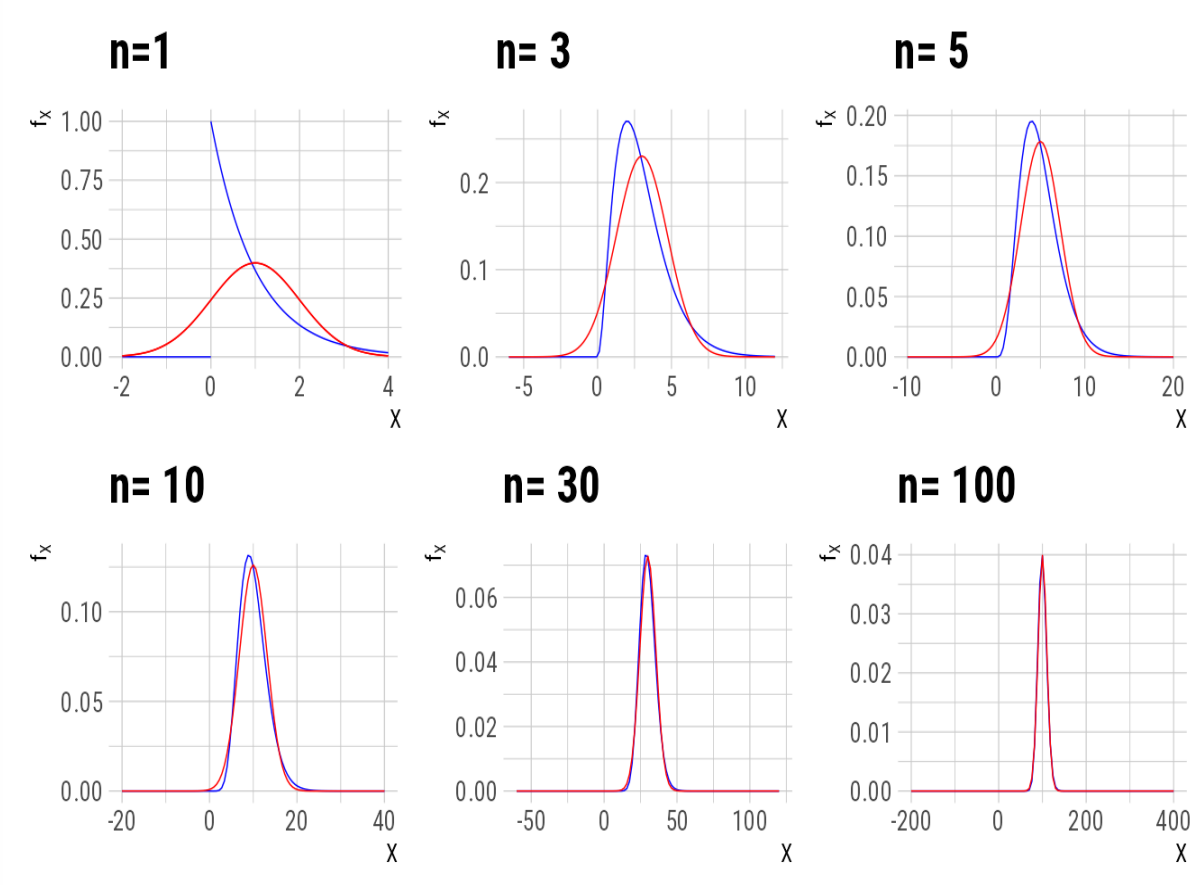
1. $E[S_n] = \sum\limits_{i=1}^{n} E[X_i] = nE[X]$;

2. $Var[S_n] = \sum\limits_{i=1}^{n} Var[X_i] = nVar[X]$;

3. Para n suficientemente grande,

$$P\left(\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} \leq x\right) \simeq \Phi(x)$$

Exemplo – Soma de exponenciais independentes



Exemplo – Uma aplicação do TLC

Sejam $X_i \sim Ber(p)$, $i = 1, \ldots, n$, variáveis aleatórias independentes e $S_n = \sum_{i=1}^{n} X_i$.

Tem-se que $E[S_n] = np$, $Var[S_n] = np(1-p)$ e, pelo TLC,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} N(0,1).$$

Note-se que $S_n \sim Bi(n,p)$. Então, para n suficientemente grande

$$S_n \overset{a}{\sim} N(np, np(1-p)).$$