DGI
2019

4.1. Definição de Distribuição Conjunta

Para um ***mesmo espaço amostral*** (S) ***podem definir-se várias variáveis aleatórias***. As diferentes variáveis aleatórias possuem ***características próprias que as individualizam*** mas ***podem também apresentar características de natureza conjunta que merecem ser identificadas.***

4.2. Função de Probabilidade Conjunta de Duas Variáveis Discretas

$$p(x,y) = PROB\,(X = x \;\wedge\; Y = y)$$

- $\displaystyle\sum_X \sum_Y p(x,y) = 1$

DGI
2019

4.3. Função de Probabilidade Conjunta de Duas Variáveis Contínuas

$$f(x,y) = PROB\left\{X \in [x,\ x+dx]\,,\ Y \in [y,\ y+dy]\right\}$$

A probabilidade de um par de valores $(X,\ Y)$ se situar no domínio D é dada por:

$$PROB\left\{(X,Y) \in D\right\} = \int\int_D f(x,y)\,dx\,dy$$

- $\int\int_S f(x,y)\,dx\,dy = 1$

DGI
2019

4.4. Distribuições Marginais

Quando as distribuições $p_X(x)$ e $p_Y(y)$ são obtidas a partir de $p(x,y)$, designam-se por *distribuições marginais*.

4.4.1. Função de Probabilidade Marginal de Duas Variáveis Discretas

$$p_X(x) = \sum_Y p(x,y)$$

$$p_Y(y) = \sum_X p(x,y)$$

4.4.2. Função de Probabilidade Marginal de Duas Variáveis Contínuas

$$f_X(x) = \int_Y f(x,y)\,dy$$

$$f_Y(y) = \int_X f(x,y)\,dx$$

Exemplo

Considere duas variáveis aleatórias X e Y cuja função de probabilidade conjunta $p(x,y)$ se apresenta no quadro seguinte:

$X \backslash Y$	1	2	3
2	$2k$	0.18	0.18
3	0.16	0.12	k

1. Determine o valor de k?

$$\sum_x \sum_y p(x,y) = 1 \quad \Leftrightarrow \quad k = 0.12$$

		Y			
		1	2	3	$p_X(x)$
X	2	0.24	0.18	0.18	0.60
	3	0.16	0.12	0.12	0.40
$p_Y(y)$		0.40	0.30	0.30	**1.00**

Exemplo (cont.)

2. Deduza a função de probabilidade de Y.

$$p_Y(y) = \sum_X p(x, y)$$

$\boldsymbol{y}$	$\boldsymbol{p(y)}$
1	0.4
2	0.3
3	0.3
$\sum$	1

4.5. Distribuições Condicionais

4.5.1. Função de Probabilidade Condicional de duas Variáveis Discretas

$$p_{X|Y}(x|y) = \frac{p(x,y)}{P_Y(y)}$$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{P_X(x)}$$

4.5.2. Função de Probabilidade Condicional de duas Variáveis Contínuas

$$f_{X|Y}(x|y) = \frac{f(x,y)}{\int_X f(x,y)\,dx}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{\int_Y f(x,y)\,dy}$$

DGI
2019

4.6. Independência

Sejam X e Y duas variáveis aleatórias discretas. Tais variáveis dizem-se ***independentes se e só se, para todos os valores de X e de Y, os acontecimentos***:

$$X = x \quad \text{e} \quad Y = y \quad \text{forem independentes}$$

Tal implica que:

$$PROB\,(X = x | Y = y) = PROB\,(X = x) \quad ,\forall_{x,y}$$

$$PROB\,(Y = y | X = x) = PROB\,(Y = y) \quad ,\forall_{x,y}$$

donde

$$p_{X|Y}(x|y) = \frac{p(x,y)}{P_Y(y)} = p_X(x), \quad \forall_{x,y}$$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{P_X(x)} = p_Y(y), \quad \forall_{x,y}$$

ou

$$p(x,y) = p_X(x) \times p_Y(y), \quad \forall_{x,y}$$

O conceito de independência pode estender-se ao caso de variáveis contínuas. Se X e Y forem *duas variáveis contínuas*, elas são *independentes* se e só se:

$$\forall_{x,y} : f(x,y) = f_X(x) \times f_Y(y)$$

DGI
2019

Exemplo (cont.)

3. Verifique se as variáveis são independentes.

As variáveis são independentes se:

$$\forall_{x,y}\ p(x,y) = p(x) \times p(y)$$

		Y			
		1	2	3	$p_X(x)$
X	2	0.24	0.18	0.18	0.6
	3	0.16	0.12	0.12	0.4
$p_Y(y)$		0.4	0.3	0.3	**1**

$0.24 = 0.6 \times 0.4$ ✓	$0.18 = 0.6 \times 0.3$ ✓	$0.18 = 0.6 \times 0.3$ ✓
$0.16 = 0.4 \times 0.4$ ✓	$0.12 = 0.4 \times 0.3$ ✓	$0.12 = 0.4 \times 0.3$ ✓

$\rightarrow$ são independentes

4.7. Covariância e Correlação

Se duas variáveis são independentes então, entre elas, não existe qualquer relação.

É de esperar que as seguintes variáveis sejam independentes:

- *Cor dos olhos dos professores do IPB*
- *Disciplinas que leccionam*
- No que diz respeito às variáveis altura e peso dos professores, é de esperar que elas não sejam independentes. Embora possa haver excepções, é de esperar que o peso tenda a aumentar com o aumento da altura.

Existem **várias possibilidades de relacionamento** entre o peso e a altura que satisfaçam a condição enunciada (relação linear, quadrática, exponencial, etc.). De todas elas a *mais simples é a linear, que pode ser utilizada como aproximação das restantes.* **A Covariância e a Correlação são medidas do grau de relacionamento linear entre duas variáveis.**

DGI
2019

4.7.1. Covariância

$$\sigma_{xy} = \sum_X \sum_Y (x-\mu_X)(y-\mu_Y) \times p(x,y)$$ **(variáveis discretas)**

$$\sigma_{xy} = \int \int (x-\mu_X)(y-\mu_Y) \times f(x,y)\, dx\, dy$$ **(variáveis contínuas)**

De uma forma geral, tanto para as variáveis discretas ou contínuas verifica-se que: $\sigma_{xy} = \mathrm{E}\left[(x-\mu_X)(y-\mu_Y)\right]$

4.7.2. Correlação

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\mathrm{E}\left[(x-\mu_X)(y-\mu_Y)\right]}{\sqrt{\mathrm{E}\left[(x-\mu_X)^2\right] \times \mathrm{E}\left[(y-\mu_Y)^2\right]}}$$

A correlação pode ser interpretada como a covariância das variáveis padronizadas:

$$X^* = \frac{x - \mu_X}{\sigma_X} \quad \text{e} \quad Y^* = \frac{y - \mu_Y}{\sigma_Y}$$

É de esperar que as seguintes variáveis sejam independentes:

- O coeficiente de correlação ρ_{XY} toma valores compreendidos entre -1 e 1

$$-1 \leq \rho_{xy} \leq 1$$

ρ_{xy}	**Interpretação**
1	Relacionamento linear perfeito *(directo)*
0	Ausência de relação linear
-1	Relacionamento linear perfeito *(inverso)*

Se $\varrho_{XY} = 0$ não se pode concluir que X e Y sejam independentes. $\rho_{XY} = 0$ significa apenas que não existe uma relação linear entre X e Y. Poderão existir relações não lineares entre elas.

Covariância positiva: quando uma das variáveis se desvia significativamente do seu valor esperado a outra também tenderá a desviar-se no mesmo sentido. Tal acarretará um *aumento da dispersão da soma das duas variáveis.*

Covariância negativa: Os desvios das duas variáveis serão mais frequentemente de sentido contrário, implicando uma *diminuição da dispersão da sua soma*.

4.8. Valor Esperado e Variância de uma Função de várias Variáveis Aleatórias

Considerem-se duas variáveis aleatórias quaisquer, discretas ou contínuas (X e Y)

e defina-se uma nova variável aleatória Z como função das anteriores

$$Z = g(X,\ Y)$$

Considere-se o caso: $Z = g(X,\ Y) = X + Y$

Neste caso, o *valor esperado e a variância de Z* vêm dados por:

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = \mu_X + \mu_Y$$

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + 2 \times \mathrm{COV}(X,Y) + \mathrm{VAR}(Y) = {\sigma^2}_X + 2 \times \sigma_{XY} + {\sigma^2}_Y$$

DGI
2019

Exemplo (cont.)

4. Considere uma nova variável $Z = X - Y$. Calcule a sua variância e o seu valor esperado

$$Z = X - Y$$

$$E(Z) = E(X) - E(Y) = \mu_X - \mu_Y = 0.5$$

$$VAR(Z) = VAR(X - Y) = VAR(X) + 2 \times COV(X, Y) + VAR(Y) =$$

X, Y independentes $\rightarrow$ $= VAR(X) + VAR(Y) =$

$$= 0.24 + 0.69 = 0.93$$

Quando X e Y forem não correlacionadas (*independentes*):

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + \mathrm{VAR}(Y) = \sigma^2{}_X + \sigma^2{}_Y$$

Quando X e Y forem *perfeitamente correlacionadas*:

$$\frac{\sigma_{XY}}{\sigma_X \times \sigma_Y} = 1 \quad \Rightarrow \quad \sigma_{XY} = \sigma_X \times \sigma_Y$$

Considere-se agora o caso em que a variável $Z = g(X, Y)$ é uma *combinação linear de* X *e* Y*:*

$$Z = g(X, Y) = a \times X + b \times Y$$

Neste caso, o *valor esperado* e a *variância* de *Z* vêm dados por:

$$\mathrm{E}(a \times X + b \times Y) = a \times \mathrm{E}(X) + b \times \mathrm{E}(Y) = a \times \mu_X + b \times \mu_Y$$

$$\begin{aligned}\mathrm{VAR}(a \times X + b \times Y) &= a^2 \times \mathrm{VAR}(X) + 2 \times a \times b \times \mathrm{COV}(X, Y) + b^2 \times \mathrm{VAR}(Y) = \\ &= a^2 \times {\sigma^2}_X + 2 \times a \times b \times \sigma_{XY} + b^2 \times {\sigma^2}_Y\end{aligned}$$

Considere-se finalmente o caso em que Z é uma função não linear em X e/ou em Y. Neste caso é possível obter *aproximações de $E(Z)$ e $Var(Z)$ a partir de uma aproximação linear de $g(X,\ Y)$ em torno de (μ_X, μ_Y):*

$$g(X,\ Y) \approx g(\mu_X, \mu_Y) + (X - \mu_X) \cdot \left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}} + (Y - \mu_Y) \cdot \left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}$$

Neste caso o *valor esperado* e a *variância* de Z vêm dados por:

$$\mathrm{E}\left[g(X,Y)\right] \approx g\left[\mathrm{E}(X), \mathrm{E}(Y)\right] = g(\mu_X, \mu_Y)$$

$$\mathrm{VAR}\left[g(X,Y)\right] \approx \left(\left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right)^2 \times \sigma^2{}_X +$$

$$+ 2 \times \left(\left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right) \left(\left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}} +\right) \times \sigma_{XY} + \left(\left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right)^2 \times \sigma^2{}_Y$$