**CC0282 - Probabilidade I**

**Material Didático - 21/07/2021 Prof. Maurício Mota**

# 1 Vetor Aleatório Discreto Bidimensional (X,Y).

Seja $(X, Y)$ um vetor aleatório discreto bidimensional com função densidade de probabilidade conjunta dada por $f(x, y)$ com suporte $A$, isto é,

$$A = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) > 0\}.$$

. Sejam $f_X(x)$ e $f_Y(y)$ as marginais.

A função $f(x, y)$ satisfaz as seguintes propriedades:

i. $f(x, y) = P(X = x, Y = y) \geq 0$

ii. $\sum_{(x,y)\in A} f(x, y) = 1.$

iii Seja $E$ um evento então
$P(E) = \sum_{(x,y)\in E} f(x, y).$

## 1.1 Função de Probabilidade marginal de X.

$$f_X(x) = \sum_x f(x, y) \;\; I_{A_X}(x),$$

onde $A_X = \{x \in R \mid f_X(x) > 0\}$.

## 1.2 Função de Probabilidade marginal de Y.

$$f_Y(y) = \sum_y f(x, y) \;\; I_{A_Y}(y),$$

onde $A_Y = \{y \in R \mid f_Y(y) > 0\}$.

## 1.3 Independência entre X e Y.

Dizemos que $X$ e $Y$ são independentes se e só se a conjunta puder ser fatorada no produto das marginais, isto é,

$$f(x, y) = f_X(x) \times f_Y(y), \;\; \forall(x, y) \in \mathbf{R}^2.$$

## 1.4 Distribuição Condicional de X dado Y=y.

Seja $y \in A_Y$ então a distribuição condicional de $X \mid Y = y$ é dada por:

$$f(x|Y = y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x,y)}{f_Y(y)}\, I_{A_{XY}}\,(x),$$

onde $A_{XY} = \{x \in \mathbf{R} \mid f(x|y) > 0\}$.

## 1.5 Distribuição Condicional de Y dado X=x.

Seja $x \in A_X$ então a distribuição condicional de $Y \mid X = x$ é dada por:

$$f(y|X = x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x,y)}{f_X(x)}\, I_{A_{YX}}\,(y),$$

onde $A_{YX} = \{y \in \mathbf{R} \mid f(y|x) > 0\}$.
Com as definições das condicionais temos uma outra definição de independência entre $X$ e $Y$.
Dizemos que $X$ e $Y$ são independentes se e só se

$$f(y|x) = f_Y(y) \quad e \quad f(x|y) = f_X(x).$$

## 1.6 Valor Esperado de g(X,Y).

Seja $g(x,y)$ uma função do $\mathbf{R}^2$ em $\mathbf{R}$ com esperança finita, então

$$E(g(X,Y)) = \sum_{(x,y)\in A} g(x,y) \times f(x,y).$$

## 1.7 Covariância entre Duas Variáveis Aleatórias X e Y.

Sejam $X$ e $Y$ variáveis aleatórias com momentos de segunda ordem em relação à origem finitos, isto é, $E(X^2) < \infty$, $E(Y^2) < \infty$ e $E(XY) < \infty$. A covariância ente elas é definida por:

$$Cov(X,Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = E(XY) - E(X)E(Y).$$

Definição: Quando $Cov(X,Y) = 0$ dizemos que $X$ e $Y$ são não correlacionadas.
Fato: Se $X$ e $Y$ forem independentes $Cov(X,Y) = 0$. Mas a recíproca não é verdadeira

Propriedades da Covariância:

1. $Cov(X,c) = 0$, onde $c$ é uma constante.
2. $Cov(X,X) = V(X)$.
3. $Cov(X,Y) = Cov(Y,X)$.
4. $Cov(aX + b, cY + d) = ac\, Cov(X,Y)$, onde $a, b, c, d$ são constantes.
5. $Cov(X + Y, Z + W) = Cov(X,Z) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W)$.

Fato: Para duas variáveis aleatórias $X$ e $Y$ quaisquer temos:

$$V(X+Y) = V(X) + V(Y) + 2\,Cov(X,Y),$$

e se elas forem independentes:

$$V(X+Y) = V(X) + V(Y).$$

Prova:

$$\begin{aligned} V(X+Y) &= Cov(X+Y, X+Y) \\ &= Cov(X,X) + Cov(X,Y) + Cov(Y,X) + Cov(Y,Y) \\ &= V(X) + Cov(X,Y) + Cov(X,Y) + V(Y) \\ &= V(X) + V(Y) + 2\,Cov(X,Y). \end{aligned}$$

Fato: Para duas variáveis aleatórias $X$ e $Y$ quaisquer temos:

$$V(X-Y) = V(X) + V(Y) - 2\,Cov(X,Y),$$

## 1.8 Correlação entre Duas Variáveis Aleatórias X e Y.

Para medir o grau da associação linear entre $X$ e $Y$ usamos o coeficiente de correlação que é definido por:

$$\rho(X,Y) = \rho = \frac{Cov(X,Y)}{\sigma(X) \times \sigma(Y)}.$$

Propriedades do coeficiente de correlação.

1. $-1 \leq \rho(X,Y) \leq 1$, ou $|\rho(X,Y)| \leq 1$.
2. $\rho(X,X) = 1$ e $\rho(X,-X) = -1$.
3. Ele é adimensional.
4. $\rho(X+a, Y+b) = \rho(X,Y)$, onde $a, b$ são constantes.
5. $\rho(aX, bY) = sinal(ab)\rho(X,Y)$, onde $sinal(c) = 1$ se $c$ for positivo , $sinal(c) = -1$ se $c$ for negativo e $sinal(c) = 0$ quando $c = 0$.

Quando $\rho = 1$ há uma relação linear crescente entre $Y = a + bX$ com $b > 0$. Quando $\rho = -1$ há uma relação linear decrescente entre $Y = c + dX$ com $d < 0$.

## 1.9 Esperança Condicional de $Y|X = x$.

Ela é definida por

$$E(Y|X=x) = \sum_y y f(y|x).$$

Perceba que $E(Y|x)$ é uma função de $x$, isto é, $E(Y|x) = g(x)$ e é denominada curva de regressão de $Y$ sobre $x$. Na realidade $E(Y|x)$ é um valor da variável aleatória $V = E(Y|X)$.

Assim,

a.

$$E(V) = E[E(Y|X)] = E(Y).$$

b.
$$V(Y) = E[Var(Y|X)] + Var[E(Y|X)].$$

c.
$$Cov(X,Y) = Cov(X,V) = cov(X, E(Y|X)).$$

Na realidade podemos calcular a esperança de qualquer função de $Y$, $h(Y)$, através da condicional ,isto é,

$$E[h(Y)] = E[E(h(Y)|X)] = \sum_y h(y)\ f(y|x).$$

Note que $h(Y) = t^Y$ temos que:

$$G_Y(t) = E(t^Y) = E[E(t^Y|X)] = E[G_{Y|X}(t)] = \sum_y t^y\ f(y|x).$$

## 1.10 Esperança Condicional de $X|Y = y$.

Ela é definida por

$$E(X|Y = y) = \sum_x x\ f(x|y).$$

Perceba que $E(X|y)$ é uma função de $y$, isto é, $E(X|y) = h(y)$ e é denominada curva de regressão de $X$ sobre $y$. Na realidade $E(X|y)$ é um valor da variável aleatória $W = E(X|Y)$.

Assim,

a.
$$E(W) = E[E(X|Y)] = E(X).$$

b.
$$V(X) = E[Var(X|Y)] + Var[E(X|Y)].$$

c.
$$Cov(X,Y) = Cov(X,V) = cov(X, E(Y|X)).$$

Na realidade podemos calcular a esperança de qualquer função de $X$, $h(X)$, através da condicional, isto é,

$$E[h(X)] = E[E(h(X)|Y)] = \sum_x h(x)\ f(x|y).$$

Note que $h(X) = e^{tX}$ temos que:

$$M_X(t) = E(e^{tX}) = E[E(e^{tX}|Y)] = E[M_{X|Y}(t)] = \sum_y e^{tx}\ f(x|y).$$

Vamos enunciar um teorema do Meyer-página 178:

Seja $(X,Y)$ um vetor aleatório bidimensional e suponha que:

$$E(X) = \mu_x,\ \ E(Y) = \mu_y,\ \ V(X) = \sigma_x^2,\ \ e\ \ V(Y) = \sigma_y^2.$$

Seja $\rho$ o coeficiente de correlação entre $X$ e $Y$. Se a regressão de $Y$ em $X$ for linear teremos:

$$E(Y|x) \quad = \quad \mu_y + \rho\, \frac{\sigma_y}{\sigma_x}\, (x - \mu_x). \tag{1}$$

Se a regressão de $X$ em $Y$ for linear teremos:

$$E(X|y) \quad = \quad \mu_x + \rho\, \frac{\sigma_x}{\sigma_y}\, (y - \mu_y). \tag{2}$$

Fato: Se $b_1$ for o coeficiente angular da regressão linear de $Y$ em $X$ e se $b_2$ for o coeficiente angular da regressão linear de $X$ em $Y$ então

$$b_1 \times b_2 = \rho^2,$$

e o sinal de $\rho$ e´o mesmo sinal de $b_1$ e $b_2$.

### 1.11 Função Geradora de probabilidades Bivariada de (X,Y).

Ela é definida por:

$$G_{(X,Y)}(t_1, t_2) = E(t_1^X \times t_2^Y) = \sum_{(x,y)\in A} t_1^x \times t_2^y \, f(x, y).$$

Propriedades:

a. $G_X(t) = G_{(X,Y)}(t, 1)$.

b. $G_Y(t) = G_{(X,Y)}(1, t)$

c. Se $X$ e $Y$ forem independentes então $G_{(X,Y)}(t_1, t_2) = G_X(t_1) \times G_y(t_2)$.

d. A função geradora de probabilidades de $S = X + Y$ é dada por:

$$G_S(t) = E(t^S) = E(t^{X+Y}) = E(t^X \times t^Y) = G_{(X,Y)}(t, t),$$

note que não precisa de independência.

e. A derivada parcial de primeira ordem de $G$ com relação a $t_1$ é dada por

$$\frac{\partial G}{\partial t_1}(t_1, t_2) = E\left[X \times t_1^{X-1} \times t_2^Y\right].$$

Assim,

$$E(X) = \frac{\partial G}{\partial t_1}(t_1, t_2) \Big|_{(1,1)}.$$

f. A derivada parcial de primeira ordem de $G$ com relação a $t_2$ é dada por

$$\frac{\partial G}{\partial t_2}(t_1, t_2) = E\left[t_1^X \times Y \times t_2^{Y-1}\right].$$

Assim,

$$E(Y) = \frac{\partial G}{\partial t_2}(t_1, t_2) \Big|_{(1,1)}.$$

g. A derivada parcial de segunda ordem de $G$ com relação a $t_1$ é dada por

$$\frac{\partial^2 G}{\partial t_1^2}(t_1, t_2) = E\left[X(X-1) \times t_1^{X-2} \times t_2^Y\right].$$

Assim,

$$E(X(X-1)) = \frac{\partial^2 G}{\partial t_1^2}(t_1, t_2) \Big|_{(1,1)}.$$

h. A derivada parcial de segunda ordem de $G$ com relação a $t_2$ é dada por

$$\frac{\partial^2 G}{\partial t_2^2}(t_1, t_2) = E\left[t_1^X \times Y(Y-1)t_2^{Y-2}\right].$$

Assim,

$$E(Y(Y-1)) = \frac{\partial^2 G}{\partial t_2^2}(t_1, t_2)\ \Big|_{(1,1)}.$$

i. A derivada parcial de segunda ordem de $G$ com relação a $t_1$ e $t_2$ é dada por

$$\frac{\partial^2 G}{\partial t_1 t_2}(t_1, t_2) = E\left[XY \times t_1^{X-1} \times t_2^{Y-1}\right].$$

Assim,

$$E(XY)) = \frac{\partial^2 G}{\partial t_1 t_2}(t_1, t_2)\ \Big|_{(1,1)}.$$

## 1.12 Função Geradora de Momentos Bivariada de (X,Y).

Ela é definida por:

$$M_{(X,Y)}(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right) = \sum_{(x,y)\in A} e^{t_1 x + t_2 y} \times\ f(x, y).$$

Propriedades:

a. $M_X(t) = M_{(X,Y)}(t, 0)$.

b. $M_Y(t) = M_{(X,Y)}(0, t)$

c. Se $X$ e $Y$ forem independentes então $M_{(X,Y)}(t_1, t_2) = M_X(t_1) \times M_y(t_2)$.

d. A função geradora de momentos de $S = X + Y$ é dada por:

$$M_S(t) = E(e^{tS}) = E(e^{t(X+Y)} = E(e^{tX+tY}) = M_{(X,Y)}(t, t),$$

note que não precisa de independência.

e. A derivada parcial de primeira ordem de $M$ com relação a $t_1$ é dada por

$$\frac{\partial M}{\partial t_1}(t_1, t_2) = E\left[X \times e^{t_1 X + t_2 Y}\right].$$

Assim,

$$E(X) = \frac{\partial M}{\partial t_1}(t_1, t_2)\ \Big|_{(0,0)}.$$

f. A derivada parcial de primeira ordem de $M$ com relação a $t_2$ é dada por

$$\frac{\partial M}{\partial t_2}(t_1, t_2) = E\left[Y \times e^{t_1 X + t_2 Y}\right].$$

Assim,

$$E(Y) = \frac{\partial M}{\partial t_2}(t_1, t_2)\ \Big|_{(0,0)}.$$

g. A derivada parcial de segunda ordem de $M$ com relação a $t_1$ é dada por

$$\frac{\partial^2 M}{\partial t_1^2}(t_1, t_2) = E\left[X^2 \times e^{t_1 X + t_2 Y}\right].$$

Assim,

$$E(X^2)) = \frac{\partial^2 M}{\partial t_1^2}(t_1, t_2) \Big|_{(0,0)}.$$

h. A derivada parcial de segunda ordem de $M$ com relação a $t_2$ é dada por

$$\frac{\partial^2 M}{\partial t_2^2}(t_1, t_2) = E\left[Y^2 \times e^{t_1 X + t_2 Y}\right].$$

Assim,

$$E(Y^2) = \frac{\partial^2 m}{\partial t_2^2}(t_1, t_2) \Big|_{(0,0)}.$$

i. A derivada parcial de segunda ordem de $M$ com relação a $t_1$ e $t_2$ é dada por

$$\frac{\partial^2 M}{\partial t_1 t_2}(t_1, t_2) = E\left[XY \times e^{t_1 X + t_2 Y}\right].$$

Assim,

$$E(XY)) = \frac{\partial^2 M}{\partial t_1 t_2}(t_1, t_2) \Big|_{(0,0)}.$$

Calculados estes momentos podemos calcular as variâncias e a covariância e apresentá-los em forma matricial através do vetor de médias $\mu^{'} = (E(X), E(Y))$ e da matriz de variâncias e covariâncias $\Sigma$.

## 1.13 Função de Distribuição Acumulada Bivariada

Seja $(X, Y)$ um vetor aleatório discreto bidimensional com função densidade conjunta dada por $f_{(X,Y)}(x, y)$. A Função de Distribuição Conjunta Bidimensional Discreta do vetor aleatório $(X, Y)$ é definida por:

$$F : R^2 \to [0, 1]$$

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{(u,v)\in E} f_{(X,Y)}(u, v),$$

onde

$$E = \{(u, v) \in R^2 | u \leq x, v \leq y\}.$$

Propriedades da Função de Distribuição Acumulada Bivariada de Qualquer Tipo.

a. $F(x, y)$ é uma função não decrescente de $x$. $F(x, y)$ é uma função não decrescente de $y$.

b. $F(x, y)$ é uma função contínua à direita de $x$. $F(x, y)$ é uma função contínua à direita de $y$. Isto é:

$$\lim_{0<h\to 0} F(x + h, y) = \lim_{0<h\to 0} F(x, y + h) = F(x, y).$$

c. $\lim\limits_{x\to -\infty} F(x, y) = 0$, $\lim\limits_{y\to -\infty} F(x, y) = 0$, $\lim\limits_{\substack{x\to -\infty \\ y\to -\infty}} F(x, y) = F(-\infty\ , -\infty) = 0$

e $\lim\limits_{\substack{x\to \infty \\ y\to \infty}} F(x, y) = F(\infty\ , \infty) = 1.$

d. Se $x_1 < x_2$ e $y_1 < y_2$ , então:

$$P(x_1 < X \leq x_2,\ y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0.$$

Além disso:

a. $F_X(x) = \lim_{y \to \infty} F_{(X,Y)}(x, y)$.

b. $F_Y(y) = \lim_{x \to \infty} F_{(X,Y)}(x, y)$.

c. Se $X$ e $Y$ são independentes então: $F_{(X,Y)}(x, y) = F_X(x)F_Y(y), \forall\ (x, y) \in R^2$.

Vamos exercitar teoricamente os conceitos apresentados:

Prove a seguinte desigualdade:

$$F_X(x) + F_Y(y) - 1 \leq F_{(X,Y)}(x, y) \leq \sqrt{F_X(x)F_Y(y)}, \quad \forall\ (x, y) \in R^2.$$

Seja $F_0 : R^2 \to R$ uma função definida por:

$$F_0(x, y) = \begin{cases} 1 & \text{se } x \geq 0 \text{ e } y \geq 0 \text{ e } x + y \geq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Será que $F_0$ é uma função de distribuição de um vetor aleatório $(X, Y)$?
Mostre que $P(0 < X \leq 1, 0 < Y \leq 1) = -1$. E aí? A propriedade $d$ é fundamental? As outras são verdadeiras!!!!!!!!!!!!!!!!!!!!!!?

Exemplo: Seja $(X, Y)$ um vetor aleatório discreto com

$$f(x, y) = \frac{1}{4}\ I_A\ (x, y), \quad A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}.$$

calcule a função de de distribuição acumulada de $(X, Y)$.
Solução: Vamos calcular inicialmente

$$\begin{aligned}
F(1, 1) &= P(X \leq 1, Y \leq 1) = P(X = 1, Y = 1) = \frac{1}{4} \\
F(1, 2) &= P(X \leq 1, Y \leq 2) = P(X = 1, Y = 1) + P(X = 1, Y = 2) = \frac{1}{2} \\
F(2, 1) &= P(X \leq 2, Y \leq 1) = P(X = 1, Y = 1) + P(X = 2, Y = 1) = \frac{1}{2} \\
F(2, 2) &= P(X \leq 2, Y \leq 2) = P(X = 1, Y = 1) + P(X = 1, Y = 2) + \\
&\quad P(X = 2, Y = 1) + P(X = 2, Y = 2) = 1.
\end{aligned}$$

Outros valores de $(x, y)$ o valor da acumulada é obtido assim:

$$\begin{aligned}
F(3/2, 3/2) &= P(X \leq 3/2, Y \leq 3/2) = P(X \leq 1, Y \leq 1) = F(1, 1) = \frac{1}{4} \\
F(1, 4; , 2, 7) &= P(X \leq 1, 4; Y \leq 2, 7) = P(X \leq 1, Y \leq 2) = F(1, 2) = \frac{1}{2} \\
F(3; 1, 9) &= P(X \leq 3, Y \leq 1, 9) = P(X \leq 2, Y \leq 1) = F(2, 1) = \frac{1}{2} \\
F(4, 5) &= P(X \leq 4, Y \leq 5) = P(X \leq 2, Y \leq 2) = F(2, 2) = 1.
\end{aligned}$$

A acumulada bivariada é dada para qualquer ponto $(x, y) \in \mathbf{R}^2$:

$$F(x,y)=\begin{cases} 0 & \text{se } \ x<1 \text{ ou } y<1 \\ \frac{1}{4} & \text{se } \ 1\le x<2 \quad ,1\le y<2 \\ \frac{1}{2} & \text{se } \ 1\le x<2 \quad ,y\ge 2 \\ \frac{1}{2} & \text{se } \ x\ge 2 \quad ,1\le y<2 \\ 1 & \text{se } x\ge 2 \ \ e\ y\ge 2 \end{cases}$$

## 1.14 Distribuição Trinomial

Seja $(X,Y)$ com distribuição bivariada (trinomial) de parâmetros $n$, $p_1$ e $p_2$.

A função de probabilidade conjunta de $(X,Y)$ é dada por:

$$f(x,y)=\frac{n!}{x!y!(n-x-y)!}\, p_1^x\, p_2^y\, (1-p_1-p_2)^{n-x-y}\, I_{\{0,1,\ldots,n\}}\,(x)\;\; I_{\{0,1,\ldots,n-x\}}\,(y).$$

Sua função geradora de probabilidades bivariada é dada por:

$$G_{(X,Y)}(t_1,t_2)=(p_1t_1+p_2\,t_2+1-p_1-p_2)^n$$

onde $t_1$ e $t_2$ são reais.

As marginais de $X$ e de $Y$ são dadas por:

$$\begin{aligned} G_X(t) &= G_{(X,Y)}(t,1)=(p_1t+p_2+1-p_1-p_2)^n \\ &= (p_1t+1-p_1)^n\,, \end{aligned}$$

assim $X\sim Bin(n,p_1)$ e $Y\sim Bin(n,p_2)$.

A função geradora de momentos bivariada de $(X,Y)\sim Trinomial\ (n,p_1,p_2)$, é dada por:

$$M_(t_1,t_2)=\left(p_1e^{t_1}+p_2\,e^{t_2}+1-p_1-p_2\right)^n$$

onde $t_1$ e $t_2$ são reais.

As condicionais de $X|Y=y$ e de $Y|X=x$ são dadas por:

$$X|Y=y\sim Bin\left(n-y,p=\frac{p_1}{1-p_2}\right)$$

$$Y|X=x\sim Bin\left(n-x,p=\frac{p_2}{1-p_1}\right)$$

A esperança de $XY$ é dada por:

$$E(XY)=n(n-1)p_1\,p_2.$$

A covariância entre $X$ e $Y$ é dada por:

$$Cov(X,Y)=-np_1\,p_2.$$

A correlação entre $X$ e $Y$ é dada por:

$$\rho = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}.$$

A distribuição de $S = X + Y$ é dada por:

$$\begin{aligned} G_S(t) &= E[t^S] \\ &= E[t^X t^Y] \\ &= G_{(X,Y)}(t,t) \\ &= (p_1 t + p_2 t + 1 - p_1 - p_2)^n \\ &= ((p_1 + p_2)t + 1 - p_1 - p_2)^n, \end{aligned}$$

que é a função geradora de probabilidades da Binomial $(n, p_1 + p_2)$.

## 1.15 Distribuição HiperGeométrica Bivariada

A função de probabilidade conjunta de $(X, Y)$ é dada por:

$$f(x,y) = P(X = x, Y = y) = \frac{\binom{A_1}{x}\binom{A_2}{y}\binom{N-A_1-A_2}{n-x-y}}{\binom{N}{n}},$$

com $0 \leq x + y \leq n$.

$$X \sim \text{Hipergeométrica}(N, A, n).$$

$$Y \sim \text{Hipergeométrica}(N, A_2, n).$$

$$Y|X = x \sim HG(N - A_1, A_2, n - x).$$

$$X|Y = y \sim HG(N - A_2, A_1, n - y).$$

$$E(XY) = \frac{n(n-1)\ A_1\ A_2}{N(N-1)}.$$

$$Cov(X,Y) = -n\frac{A_1}{N}\frac{A_2}{N}\frac{N-n}{N-1}.$$

$$\rho = -\sqrt{\frac{A_1 A_2}{(N-A_1)(N-A_2)}}.$$

$$S = X + Y \sim HG(N, A_1 + A_2, n).$$

## 1.16 Desigualdade de Cauchy -Schwarz

Desigualdade de Cauchy-Schwarz: Para quaisquer duas variáveis aleatórias com variâncias finitas,

$$E^2(XY) \leq E(X^2)E(Y^2),$$

ou

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}.$$