# Probabilidade e Teoria da Informação

*Bartolomeu F. Uchôa Filho, Ph.D*

Grupo de Pesquisa em Comunicações – GPqCom
Departamento de Engenharia Elétrica
Universidade Federal de Santa Catarina
E-mail: `uchoa@eel.ufsc.br`

# Probabilidade I: Variáveis Aleatórias Discretas

Conceitos importantes:

1. modelo determinístico

2. fenômeno/experimento aleatório

3. modelo probabilístico

# Espaço Amostral

O *espaço amostral*, denotado por $\Omega$, é o conjunto de todos os possíveis resultados do experimento aleatório em questão.

# Exemplo

Se considerarmos o lançamento de um dado com seis faces, o espaço amostral dos possíveis resultados será $\Omega = \{1, 2, 3, 4, 5, 6\}$.

No caso do lançamento de uma moeda, teremos que $\Omega = \{cara, coroa\}$.

# Evento

Um *evento* é qualquer subconjunto do espaço amostral.

# Exemplo

No caso do dado, alguns dos possíveis eventos são:

- $A = \{4, 5, 6\} \subset \Omega$
- $B = \{1\} \subset \Omega$
- $\phi \subset \Omega$ (*evento impossível*)
- $\Omega$ (*evento certo*)

$\Omega$ e $\phi$ também são considerados eventos, os chamados *eventos triviais*.

Os eventos que contêm um único resultado de um experimento são chamados de *eventos elementares* ou *atômicos*.

# Elemento pertence, conjunto está contido

É importante distinguirmos entre um elemento de um conjunto com um único elemento e o próprio conjunto. Neste sentido, dizemos que

$$4 \in A$$

mas

$$\{4\} \subset A$$

# Medida de Probabilidade, $P$

Uma medida de probabilidade $P$ associa a cada evento um número real no intervalo [0,1].

# Exemplo

Supondo que um dado de seis faces seja equilibrado. Para os eventos

- $A = \{4, 5, 6\} \subset \Omega$
- $B = \{1\} \subset \Omega$

teremos:

$$P(A) = \frac{3}{6} = \frac{1}{2}\,,$$

$$P(B) = \frac{1}{6}, \quad P(\phi) = 0 \quad \text{e} \quad P(\Omega) = 1$$

# Os axiomas da probabilidade

A medida de probabilidade $P$ deve satisfazer as seguintes condições:

1. $P(A) \geq 0$

2. $P(\Omega) = 1$

3. Sejam dois eventos $A$ e $B$ tais que $A \cap B = \phi$. Então,

$$P(A \cup B) = P(A) + P(B)$$

# Exemplo

Para os eventos

- $A = \{4, 5, 6\} \subset \Omega$

- $B = \{1\} \subset \Omega$

teremos:

$$P(A \cup B) = P(\{1, 4, 5, 6\}) = P(A) + P(B) = \frac{1}{2} + \frac{1}{6} = \frac{4}{6}$$

# Probabilidade de Eventos Elementares

Para $\Omega = \{s_1, s_2, \ldots, s_n\}$, definimos a probabilidade de um evento elementar por:

$$p_i = P(\{s_i\}) \ \ , \ \ i = 1, 2, \ldots, n$$

# Variável Aleatória Discreta

Uma variável aleatória (v.a.) discreta é um mapeamento (função) do espaço amostral em um conjunto específico finito ou infinito, porém numerável (ou contável).

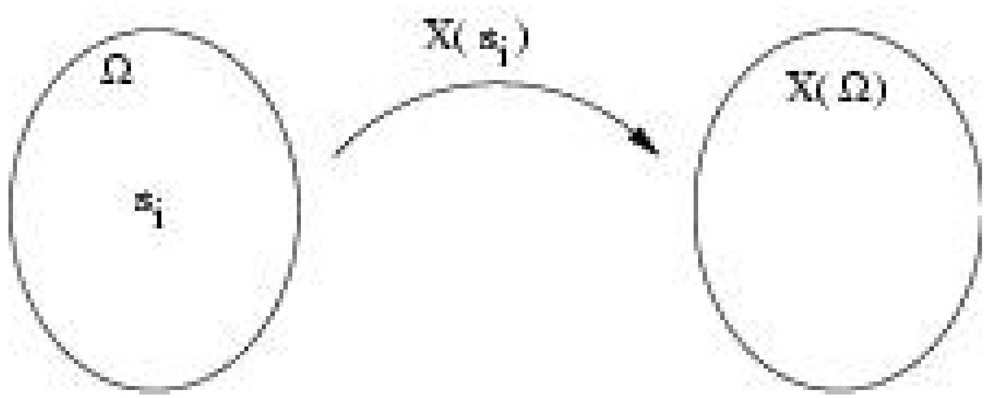


Figura 1: Geração de uma variável aleatória.

# Exemplo

Para $\Omega = \{\text{cara}, \text{coroa}\}$, podemos ter

$$\begin{aligned} X(cara) &= 0 \qquad X(\Omega) = \{0,1\} \ \text{(novo espaço amostral)} \\ X(coroa) &= 1 \end{aligned}$$

Para $\Omega = \{\text{peça defeituosa}, \text{peça perfeita}\}$, podemos ter

$$\begin{aligned} X(\text{peça defeituosa}) &= 0 \qquad X(\Omega) = \{0,1\} \ \text{(novo espaço amostral)} \\ X(\text{peça perfeita}) &= 1 \end{aligned}$$

Pelo conceito de v.a. binária, todos os experimentos com dois possíveis resultados são essencialmente o mesmo, com espaço amostral $X(\Omega) = \{0, 1\}$.

## Distribuição de Probabilidade de uma v.a. Discreta

Uma distribuição de probabilidades de uma v.a. discreta $X$ é um mapeamento, denotado por $P_X$, de $X(\Omega)$ para o intervalo [0:1].

$$\begin{aligned} P_X(x) &= P(X = x) \\ &= \text{probabilidade de } X = x \\ &= \text{probabilidade do evento } \{s : X(s) = x\} \end{aligned}$$

# Exemplo

Para $\Omega = \{sol, chuva, nublado\}$, podemos ter:

$$\begin{aligned} X(sol) &= 0 \\ X(chuva) &= 1 \\ X(nublado) &= 2 \end{aligned}$$

$P_X(1) = P(X = 1) =$ probabilidade de chuva

$P_X(x) =$ probabilidade do evento $\{s : X(s) = x\}$

# A partir dos axiomas:

Como $P_X$ é uma probabilidade relacionada aos eventos atômicos *sol*, *chuva* e *nublado*, temos:

$$P_X(x) \geq 0 \quad \text{e} \sum_{x \in X(\Omega)} P_X(x) = 1$$

# Outras conceitos sobre probabilidades:

- Probabilidade Conjunta
- Probabilidade Condicionada
- Independência Estatística

# Exemplo

Vamos considerar um experimento para determinar o clima na cidade de Florianópolis. Considere os três eventos $A$, $B$ e $C$:

$$\begin{aligned} A: \text{ num dado dia, a temperatura é } &\geq 10^o\text{C} \\ B: \text{ num dado dia, chove } &\geq 5 \text{ mm} \\ C: \text{ num dado dia, a temperatura é } &\geq 10^o\text{C e chove} \geq 5 \text{ mm} \end{aligned}$$

Note que $C = A \cap B$. A *interseção* é também representada por $C \triangleq AB$.

# Probabilidade Conjunta

A *probabilidade conjunta* é definida como:

$$P(C) = P(AB)$$

que é a probabilidade de os eventos $A$ e $B$ ocorrerem simultaneamente.

## Probabilidade conjunta p/ $\geq 3$ eventos:

Podemos facilmente estender este conceito para 3 ou mais eventos. Por exemplo, se $E$, $F$ e $G$ são eventos quaisquer, a probabilidade conjunta destes três eventos será escrita como

$$P(EFG)$$

ou

$$P(E, F, G).$$

# De volta ao exemplo sobre o clima e a chuva em Floripa

Seja $n_i$ o número de dias em que ocorreu o evento $i$, onde $i = A, B$ ou $C$.

Vamos supor que depois de observarmos o clima na cidade de Florianópolis por um período de $n = 1000$ dias, encontramos:

$$n_A = 811 \qquad n_B = 306 \qquad n_{AB} = 283$$

como pode ser observado no diagrama de Venn.

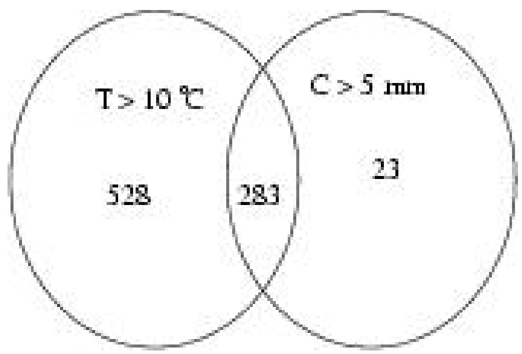


Figura 2: Diagrama de Venn.

# Obtendo as “probabilidades”

Assim,

$$P(A) \cong \frac{n_A}{n} = \frac{811}{1000} = 0,811$$
$$P(B) \cong \frac{n_B}{n} = 0,306$$
$$P(AB) \cong \frac{n_{AB}}{n} = 0,283$$

## Qual o significado de relação $\frac{n_{AB}}{n_A}$?

# Interpretando a relação $\frac{n_{AB}}{n_A}$?

Note que estamos computando a fração do número de dias em que a temperatura é maior que $10^o$C na qual tenha chovido pelo menos 5 mm.

Esta fração portanto representa aproximadamente a probabilidade de chover pelo menos 5 mm **dado que** a temperatura em um certo dia é maior ou igual a $10^o$C.

# Regra de BayesProbabilidade Condicionada

De modo mais geral, estamos considerando a probabilidade de ocorrer um evento $B$ dado que um evento $A$ tenha ocorrido. Esta é chamada de *probabilidade condicionada*.

Sucintamente, dizemos que esta é a probabilidade de $B$ dado $A$, e a denotamos por $P(B|A)$. Assim,

$$P(B|A) \cong \frac{n_{AB}}{n_A} = \frac{\frac{n_{AB}}{n}}{\frac{n_A}{n}} \triangleq \frac{P(AB)}{P(A)}$$

# Regra de Bayes

A *Regra de Bayes* é formalmente apresentada como:

$$P(B|A) \triangleq \frac{P(AB)}{P(A)}, \text{ para } P(A) > 0$$

ou

$$P(A|B) \triangleq \frac{P(AB)}{P(B)}, \text{ para } P(B) > 0$$

# Exemplo

Vamos considerar o *Canal Binário Simétrico* (ou BSC, do inglês "binary symmetric channel").

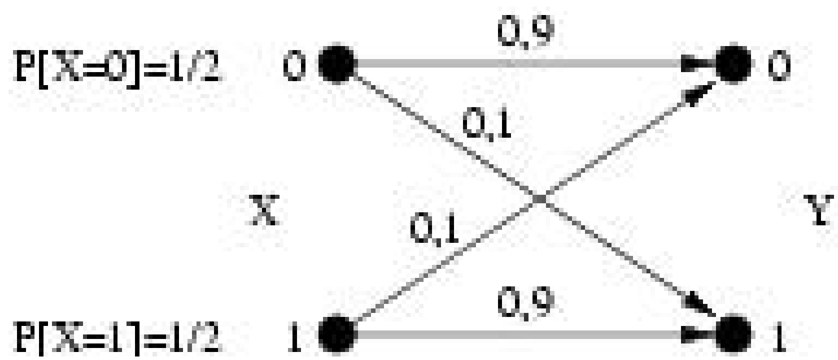


Figura 3: Canal binário simétrico com probabilidade de transição 0,1.

## Exemplo do BSC (cont.)

Esse canal é definido pelas seguintes probabilidades condicionadas:

$$P(Y = 1|X = 1) = P(Y = 0|X = 0) = 0,9 \quad \text{(probabilidade de acerto)}$$

e

$$P(Y = 1|X = 0) = P(Y = 0|X = 1) = 0,1 \quad \text{(probabilidade de erro)}$$

e pelas probabilidades *a priori*:

$$P(X = 0) = P(X = 1) = \frac{1}{2}.$$

Calcule a probabilidade conjunta: $P(X = 0, Y = 0)$.

## Solução:

Pela regra de Bayes, podemos escrever a probabilidade conjunta de duas maneiras:

$$P(X=0,\ Y=0) = P(X=0|Y=0)\ \ P(Y=0)$$

ou

$$P(X=0,\ Y=0) = P(Y=0|X=0)\ \ P(X=0)$$

Pelos dados do problema, devemos escolher a segunda maneira, o que fornece:

$$P(X=0,\ Y=0) = 0,9 \times \frac{1}{2} = 0,45.$$

# Teorema da Probabilidade Total

Sejam $A_1, A_2, \ldots, A_n$, $n$ eventos mutuamente excludentes e exaustivos, ou seja:

$$\bigcap_{i=1}^{n} A_i = \phi \quad \text{e} \quad \bigcup_{i=1}^{n} A_i = \Omega,$$

respectivamente.

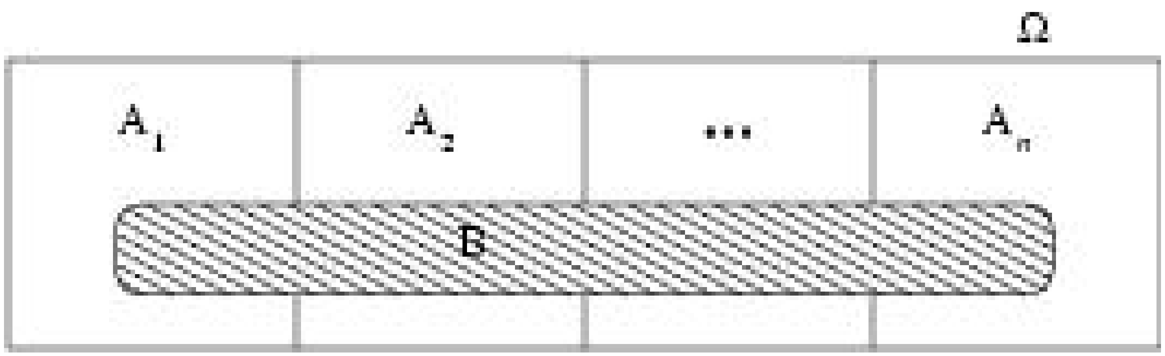


Figura 4: Teorema da probabilidade total.

# Teorema da probabilidade total (cont.)

Supondo que $P(A_i) \neq 0, \forall i$, temos que:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n).$$

Este resultado é conhecido como o *Teorema da Probabilidade Total*.

# Exemplo

Consideremos mais uma vêz o *Canal Binário Simétrico*

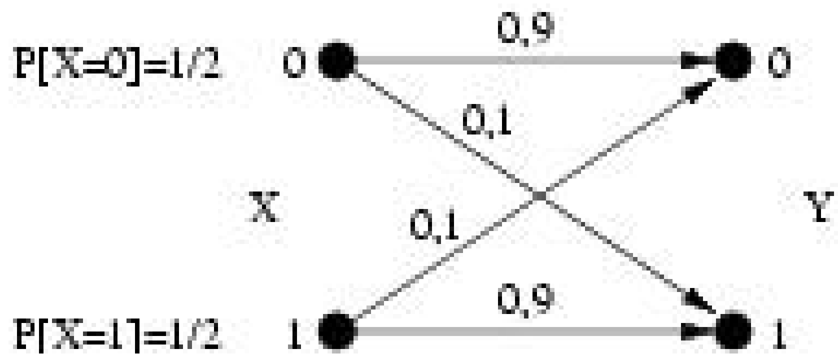


Figura 5: Canal binário simétrico com probabilidade de transição 0,1.

# Calcule:

1. Calcule $P(Y = 0)$.

   *Solução*:

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1)$$
$$= 0,1 \times \frac{1}{2} + 0,9 \times \frac{1}{2} = \frac{1}{2}$$

2. Calcule $P(X = 0|Y = 0)$.

   *Solução*:

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{0,45}{\frac{1}{2}} = 0,9$$

# Independência Estatística

Dois eventos $A \subset \Omega$ e $B \subset \Omega$ com $P(A) > 0$ e $P(B) > 0$ são *estatisticamente independentes* se e somente se

$$P(AB) = P(A)P(B)$$

# Independência Estatística (cont.)

Note que, como

$$P(AB) = P(B|A)P(A)$$

temos que se os eventos $A$ e $B$ são estatisticamente independentes, então

$$P(B|A) = P(B)$$

Ou seja, o fato de o evento $A$ ter ocorrido não muda a probabilidade de o evento $B$ ocorrer.

Similarmente, como

$$P(AB) = P(A|B)P(B)$$

também temos que

$$P(A|B) = P(A)$$

e da mesma maneira o conhecimento de $B$ não afeta as chances de $A$ ocorrer.

# Exemplo

Considere o experimento caracterizado pelo lançamento simultâneo de uma moeda eqüilibrada e de um dado de seis faces também eqüilibrado.

Ou seja, o espaço amostral é:

$$\Omega = \{(\text{cara}, 1); (\text{cara}, 2), \ldots, (\text{cara}, 6); (\text{coroa}, 1), \ldots, (\text{coroa}, 6)\}$$

Intuitivamente, podemos dizer que o resultado do experimento com a moeda é independente daquele realizado com o dado.

Assim, a probabilidade de obtermos uma *cara* na moeda e simultaneamente o número 6 no dado pode ser obtida por:

$$P(\text{coroa}, 6) = P(\text{coroa}) \times P(6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

# Teorema de Bayes

Sejam $A_i,\ i = 1, 2, \ldots, n$, eventos exaustivos e disjuntos (ou seja, $\cup_i A_i = \Omega$ e $A_i \cap A_j = \phi,\ i \neq j$) com $P(A_i) > 0,\ \forall i$. Seja $B$ um evento tal que $P(B) > 0$.

Então:

$$P(A_j|B) = \frac{P(B|A_j) \times P(A_j)}{\sum_{i=1}^{n} P(B|A_i) P(A_i)}$$

# Exemplo

Consideremos o *Canal Binário Simétrico*

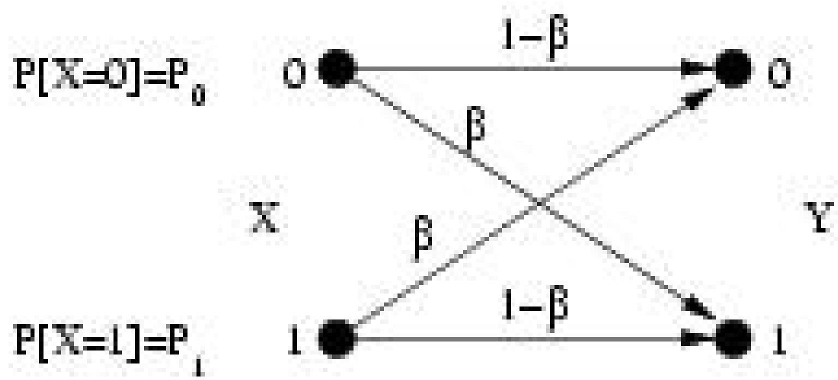


Figura 6: Canal binário simétrico com probabilidade de transição $\beta$.

Note que os símbolos de entrada são não-eqüiprováveis. Ou seja,

$$P(X=0)=P_0 \quad P(X=1)=1-P_0 \triangleq P_1$$

onde possivelmente $P_0 \neq 1/2$.

# Exemplo (cont.)

Calcule $P(X=1|Y=1)$.

*Solução*:

$$\begin{aligned} P(X=1|Y=1) &= \frac{P(X=1,Y=1)}{P(Y=1)} \\ &= \frac{P(Y=1|X=1)\ P(X=1)}{P(Y=1|X=0)\ P(X=0)+P(Y=1|X=1)\ P(X=1)} \\ &= \frac{P_1(1-\beta)}{P_0\beta+P_1(1-\beta)} \end{aligned}$$

Se $P_0 = P_1 = \frac{1}{2}$, teremos:

$$P(X=1|Y=1) = 1-\beta$$

# Outro Exemplo

Consideremos agora um experimento com um radar e definamos os seguintes eventos:

$$\begin{aligned} A &= \text{radar detecta um alvo} \\ B &= \text{o alvo de fato existe} \\ \overline{A} &= \text{o radar não detecta nada} \\ \overline{B} &= \text{não há alvo} \end{aligned}$$

Conhecemos os seguintes dados do problema: $P(A|B) = 0,95$, $P(\overline{A}|\overline{B}) = 0,95$ e $P(B) = 0,005$.

# Exemplo (cont.)

Calcule $P(B|A)$. O radar é eficiente?

*Solução*: Temos que

$$(\overline{B}|A) = 1 - P(B|A) = \text{probabilidade de alarme falso}$$

$$P(\overline{A}|B) = 1 - P(A|B) = 0,05 = \text{ probabilidade de falhar na detecção}$$

Assim,

$$\begin{aligned}
P(B|A) &= \frac{P(A|B)\ P(B)}{P(A|B)\ P(B) + P(A|\overline{B})\ P(\overline{B})} \\
&= \frac{0,95 \times 0,005}{0,95 \times 0,005 + [1 - P(\overline{A}|\overline{B})][1 - P(B)]} \\
&= \frac{0,95 \times 0,005}{0,95 \times 0,005 + 0,05 \times 0,995} = 0,087 = 8,7\%
\end{aligned}$$

# Esperança de F(X)

Seja $F$ uma função real cujo domínio inclui $X(\Omega)$. A esperança de F(X) é o número real

$$E\{F(X)\} = \overline{F(X)} = \sum_{x \in X} P_X(x)\, F(x)$$

## Casos Particulares:

1. *Média*:

$$F(X) = X \Rightarrow E\{X\} \stackrel{\Delta}{=} \overline{X} = \sum_{x \in X} x\, P_X(x)$$

2. *Variância*:

$$F(X) = (X - \overline{X})^2 \Rightarrow E\{(X - \overline{X})^2\} = \sum_{x \in X} (x - \overline{X})^2\, P_X(x) = \mathsf{Var}(X) \stackrel{\Delta}{=} \sigma_X^2$$

Analogamente, podemos ter a definição de esperança para duas variáveis aleatórias, digamos $X$ e $Y$, da seguinte forma:

$$E\{F(X,Y)\} = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y)\; F(x,y)$$

Para $F(X,Y) = XY$, temos a correlação cruzada.

# Esperança Condicionada

Podemos ter esperança de $F(X)$ condicionada à ocorrência do evento $Y = y$. Escrevemos então:

$$E\{F(X)|Y = y\} = \sum_x F(x) P_{X|Y}(x|y)$$

Sempre que necessário, usaremos $P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)}$.

Se $E\{F(X)|Y=y\}$ é conhecida $\forall y \in Y$, então

$$\begin{aligned}
E\{F(X)\} &= \sum_{y\in Y} E\{F(X)|Y=y\}\, P_Y(y) \\
&= \sum_{y\in Y}\sum_{x\in X} F(x)P_{X|Y}(x|y)P_Y(y) \\
&= \sum_{x\in X}\sum_{y\in Y} F(x)P_{XY}(x,y) \\
&= \sum_{x\in X} F(x) \sum_{y\in Y} P_{XY}(x,y) \\
&= \sum_{x\in X} F(x)P_X(x) \\
&= E\{F(X)\}
\end{aligned}$$

Note que na penúltima igualdade com somatória fizemos uso da relação

$$\sum_{y \in Y} P_{XY}(x,y) = P_X(x)$$

Quando obtemos $P_X(x)$ desta maneira, chamamos $P_X(x)$ de *Distribuição de Probabilidade Marginal*.

# Exemplo

Vamos supor que uma empresa de comunicação de dados tenha a opção de usar alguns tipos de canais, listados na Tabela 1.

Tabela 1: Dados para o problema dos Canais

| $Y$ | Tipo de Canal |
|---|---|
| 1 | satélite |
| 2 | cabo coaxial |
| 3 | enlace de microondas |
| 4 | fibra óptica |

Vamos admitir que a escolha do canal é baseada na disponibilidade, que é um fenômeno aleatório. Vamos supor que $P(Y = i) = 1/4$, $i = 1, 2, 3, 4$, onde $P(Y = i)$ é a probabilidade de o canal $i$ ser escolhido.

Seja $X$ o atraso médio da mensagem em milisegundos, que é diferente para cada canal, segundo as seguintes esperanças condicionadas:

$$\begin{aligned} E\{X|Y=1\} &= 500 \text{ ms} \\ E\{X|Y=2\} &= 300 \text{ ms} \\ E\{X|Y=3\} &= 200 \text{ ms} \\ E\{X|Y=4\} &= 100 \text{ ms} \end{aligned}$$

Calcule o atraso médio $E\{X\}$.

*Solução*:

Temos que

$$E\{X\} = \sum_{y=1}^{4} E\{X|Y = y\}P(y)$$

$$= 500 \times \frac{1}{4} + 300 \times \frac{1}{4} + 200 \times \frac{1}{4} + 100 \times \frac{1}{4} = 275 \text{ ms}$$

# Exemplos de Distribuição de Variáveis Aleatórias Discretas e suas Médias e Variâncias

1. Uniforme: $\Omega = \{1, 2, \ldots, n\}$, com $P(X = i) = 1/n$, $\forall i \in \Omega$.

   A média e a variância são dadas por:

$$E\{X\} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} (1+n) \frac{n}{2} = \frac{n+1}{2}$$

$$\mathsf{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (i - \frac{n+1}{2})^2 = \frac{n^2-1}{12}$$

2. Bernoulli: $\Omega = \{0, 1\}$, com $P(X = 0) = 1 - p$ e $P(X = 1) = p$.

   A média e a variância são dadas por:

$$E\{X\} = 0 \times (1 - p) + 1 \times p = p$$

$$\begin{aligned} \text{Var}(X) &= (0 - p)^2(1 - p) + (1 - p)^2 p = p^2 - p^3 + p - 2p^2 + p^3 = p - p^2 \\ &= p(1 - p) \end{aligned}$$

3. $\Omega = \{1, 2, \ldots, n\}$, com $P(X = i) = 1$ para algum $i$, e $P(X = j) = 0$ para $j \neq i$. A média e a variância são dadas por:

$$E\{X\} = i \times 1 = i$$

$$\text{Var}(X) = (i - i)^2 \times 1 = 0$$

# PROBLEMAS

1. Um experimento aleatório consiste em sortear uma bola de uma urna que contém quatro bolas vermelhas numeradas com 1, 2, 3 e 4, e três bolas pretas numeradas com 1, 2 e 3. Determine com precisão que resultados do experimento acima estão contidos nos seguintes eventos:

   (a) $E_1 =$ O número de bolas é par.
   (b) $E_2 =$ A cor da bola é vermelha e seu número é maior que 1.
   (c) $E_3 =$ O número da bola é menor que 3.
   (d) $E_4 = E_1 \cup E_3$.
   (e) $E_5 = E_1 \cup (E_2 \cap E_3)$.

2. Se todas as bolas no problema anterior são sorteadas da urna com a mesma probabilidade, encontre as probabilidades de $E_i$, $1 \leq i \leq 5$.

3. Em uma certa cidade, três marcas de carro $A$, $B$ e $C$ têm 20%, 30% e 50% da preferência, respectivamente. A probabilidade de que um carro necessite de manutenção durante seu primeiro ano desde a compra para as três marcas são 5%, 10% e 15%, respectivamente.

   (a) Qual a probabilidade de um carro nessa cidade necessitar de manutenção durante seu primeiro ano desde a compra?
   (b) Se um carro nesta cidade necessitar de manutenção durante seu primeiro ano desde a compra, qual a probabilidade de esse carro ser da marca $A$?

4. Em que condições dois eventos $A$ e $B$ disjuntos podem ser independentes?

5. Um fonte de informação produz 0 e 1 com probabilidades 0,3 e 0,7, respectivamente. A saída da fonte é transmitida por um canal cuja

probabilidade de erro (ou seja, a probabilidade de 1 se transformar em 0, ou vice-versa) é 0,2.

(a) Qual a probabilidade de na saída do canal ser observado um 1?
(b) Qual a probabilidade de a saída da fonte ter sido 1 dado que a saída observada do canal foi 1?

6. Uma moeda é lançada três vezes e a variável aleatória discreta $X$ denota o número total de *caras* que aparecem. A probabilidade de em um lançamento da moeda sair uma *cara* é denotada por $p$.

   (a) Quantos valores a variável aleatória $X$ pode ter?
   (b) Qual a distribuição de probabilidades de $X$?
   (c) Qual a probabilidade de $X$ ser maior que 1?

   **OBS.: Supor que os resultados dos três lançamentos da moeda sejam estatisticamente independentes.**

7. A moeda $A$ tem probabilidade de *cara* igual a 0,25 e de *coroa* 0,75. A moeda $B$ é equilibrada. Cada moeda é lançada 4 vezes. Considere a variável aleatória discreta $X$ que denota o número de *caras* resultantes da moeda $A$ e $Y$ que denota o número de *caras* resultantes da moeda $B$.

   (a) Qual a probabilidade de $X = Y = 2$?
   (b) Qual a probabilidade de $X = Y$?
   (c) Qual a probabilidade de $X > Y$?
   (d) Qual a probabilidade de $X + Y \leq 5$?

8. Considere a variável aleatória $X = \{1, 2, 3, 4\}$, com probabilidades $P_X(X = 1) = 0,4$, $P_X(X = 2) = 0,25$, $P_X(X = 3) = 0,2$ e $P_X(X = 4) = 0,15$. Calcule a esperança de $F(X)$ para as seguintes funções:

   (a) $F(X) = 1$.

(b) $F(X) = X$.
(c) $F(X) = X^2$.
(d) $F(X) = (X - \overline{X})^2$, onde $\overline{X}$ é a esperança obtida no item b.

9. Considerando o problema 5, calcule a esperança condicionada $E\{F(X)|Y = 1\}$, onde $F(X) = X$.