# Probabilidade (PPGECD000000001)

# Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD)

Sessão 5

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

# Variável aleatória

O conceito de variável aleatória (v.a.) é um mecanismo que permite relacionar qualquer resultado de um experimento aleatório com uma medida numérica.

## Definição 1

*Sejam* $(\Omega, \mathfrak{F})$ *e* $(\Omega', \mathfrak{F}')$ *espaços mensuráveis. Uma aplicação* $X : \Omega \to \Omega'$ *se diz* $\mathfrak{F} - \mathfrak{F}'-$*mensurável se para cada* $A \in \mathfrak{F}'$*, a imagem inversa*

$$X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathfrak{F}.$$

## Definição 2

*Se* $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathfrak{B})$ *então falamos de funções reais mensuráveis. Se o par* $(\Omega', \mathfrak{F}') = (\overline{\mathbb{R}}, \overline{\mathfrak{B}})$*, em que os reais estendidos são definidos por* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ *e a* $\sigma$*-álgebra associada a* $\overline{\mathbb{R}}$ *é*

$$\overline{\mathfrak{B}} := \{B, B \cup \{\infty\}, B \cup \{-\infty, \infty\} : B \in \mathfrak{B}\},$$

*em que* $\mathfrak{B}$ *é a* $\sigma$*-álgebra de Borel nos reais, então falamos de funções numéricas mensuráveis.*

## Definição 3 (Variável aleatória)

*Se $(\Omega, \mathfrak{F}, P)$ é um espaço de probabilidade e $(\Omega', \mathfrak{F}')$ é um espaço mensurável, então uma função $X : \Omega \to \Omega'$ chama-se $(\Omega', \mathfrak{F}')-$variável aleatória (v.a.) se $X$ é uma função $\mathfrak{F} - \mathfrak{F}'$-mensurável.*

## Proposição 1

*Seja $(\Omega, \mathfrak{F})$ um espaço mensurável (por exemplo, um espaço de probabilidade) e seja $X$ uma função numérica (assume valores reais). As seguintes afirmações são equivalentes:*

1. *$X$ é $\mathfrak{F} - \mathfrak{B}$-mensurável*
2. *Para cada número real c, o conjunto $\{\omega \in \Omega : X(\omega) > c\} \in \mathfrak{F}$.*
3. *Para cada número real c, o conjunto $\{\omega \in \Omega : X(\omega) \geq c\} \in \mathfrak{F}$.*
4. *Para cada número real c, o conjunto $\{\omega \in \Omega : X(\omega) < c\} \in \mathfrak{F}$.*
5. *Para cada número real c, o conjunto $\{\omega \in \Omega : X(\omega) \leq c\} \in \mathfrak{F}$.*
   *Além disso, quaisquer das afirmações acima implicam em:*
6. *Para cada número real c, o conjunto $\{\omega \in \Omega : X(\omega) = c\} \in \mathfrak{F}$.*

## Demonstração.

A prova da proposição anterior é relativamente simples.Lembre sempre de olhar para as imagens inversas dos conjuntos e mostre que 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 4) $\Rightarrow$ 5) $\Rightarrow$ 6) $\Rightarrow$ 1). $\square$

Resolva como exercício.

## Exemplo 1

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade em que $\mathcal{F} = \mathcal{B}$ e seja $A \in \mathfrak{F}$ fixo. A função indicadora de $A$, $I_A$ definida por

$$I_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A \\ 0, & \text{se } \omega \notin A \end{cases}$$

é uma função $\mathfrak{F} - \mathfrak{B}$ mensurável, i.e., $I_A(\omega)$ é uma variável aleatória. De fato, para $a \in \mathbb{R}$

$$I_A^{-1}((a, \infty)) = \{\omega : I_A(\omega) > a\} = \begin{cases} \Omega, & \text{se } a < 0, \\ A, & \text{se } 0 \leq a \leq 1, \\ \emptyset, & \text{se } a > 1. \end{cases}$$

# Variáveis aleatória real

## Motivação

Suponha que uma moeda é lançada cinco vezes. Qual é o número de caras? Esta quantidade é o que tradicionalmente tem sido chamada de *variável aleatória*. Intuitivamente, é uma variável porque seus valores variam, dependendo da sequência de lançamentos da moeda realizada; o adjetivo "aleatória" é usado para enfatizar que o seu valor é de certo modo incerto. Formalmente, contudo, uma variável aleatória não é nem "aleatória" nem é uma variável.

## Definição 4

*Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Uma função $X : \Omega \to R$ é chamada de variável aleatória se para todo evento Boreliano $B$, $X^{-1}(B) \in \mathcal{A}$.*

Por definição, temos que $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ é o conjunto de elementos do espaço amostral cuja imagem segundo $X$ está em $B$.

O próximo teorema prova que para determinar se uma dada função real $X$ é uma variável aleatória só precisamos checar que a imagem inversa de intervalos da forma $(-\infty, x]$ pertence à $\sigma$-álgebra $\mathcal{A}$.

## Teorema 1

*Seja* $(\Omega, \mathcal{A})$ *um espaço mensurável. Uma função real* $X : \Omega \to R$ *é uma variável aleatória se e somente se*

$$X^{-1}((-\infty, \lambda]) = \{w : X(w) \leq \lambda\} \in \mathcal{A}, \forall \lambda \in R.$$

Considere os seguintes Lemas para a sua demonstração

## Lema 1

*Seja* $\mathcal{B}$ *a* $\sigma$*-álgebra de Borel, então* $X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\}$ *é uma* $\sigma$*-álgebra de eventos de* $\Omega$*.*

## Prova do lema 1.

Considere os três postulados para uma $\sigma$-álgebra:

(i) $\Omega \in X^{-1}(\mathcal{B})$.
Como $R \in \mathcal{B}$, nós temos $X^{-1}(R) = \Omega \in X^{-1}(\mathcal{B})$.

(ii) Se $A \in X^{-1}(\mathcal{B})$, então $A^c \in X^{-1}(\mathcal{B})$.
Suponha que $A \in X^{-1}(\mathcal{B})$, então existe $A' \in \mathcal{B}$ tal que $A = X^{-1}(A')$. Como $\mathcal{B}$ é uma $\sigma$-álgebra, temos que $(A')^c \in \mathcal{B}$. Logo, $X^{-1}((A')^c) \in X^{-1}(\mathcal{B})$. Como $X^{-1}((A')^c) = (X^{-1}(A'))^c$, temos que $A^c \in X^{-1}(\mathcal{B})$.

(iii) Se $A_1, A_2, \ldots \in X^{-1}(\mathcal{B})$, então $\cup_{i=1}^{\infty} A_i \in X^{-1}(\mathcal{B})$.
Suponha que $A_1, A_2, \ldots \in X^{-1}(\mathcal{B})$, então existem $A_1', A_2', \ldots \in \mathcal{B}$ tais que $A_i = X^{-1}(A_i')$ para $i \geq 1$. Como $\mathcal{B}$ é uma $\sigma$-álgebra, temos que $\cup_{i=1}^{\infty} A_i' \in \mathcal{B}$. Logo, $X^{-1}(\cup_{i=1}^{\infty} A_i') \in X^{-1}(\mathcal{B})$. Como $\cup_{i=1}^{\infty} X^{-1}(A_i') = X^{-1}(\cup_{i=1}^{\infty} A_i')$, temos que $\cup_{i=1}^{\infty} A_i \in X^{-1}(\mathcal{B})$.

$\square$

Dado qualquer classe de conjuntos $\mathcal{C}$, denotamos por $\sigma(\mathcal{C})$ a menor $\sigma$-álgebra contendo $\mathcal{C}$. Desta forma se $\mathcal{B}' = \{(-\infty, \lambda] : \lambda \in R\}$, então $\mathcal{B} = \sigma(\mathcal{B}')$. O próximo lema prova um resultado semelhante ao do lema anterior, porém mais forte.

## Lema 2

$X^{-1}(\mathcal{B}) = \sigma(X^{-1}(\mathcal{B}'))$, *isto é, a imagem inversa de eventos Borelianos é igual a menor* $\sigma$*-álgebra contendo as imagens inversas dos intervalos da forma* $(\infty, \lambda]$, *onde* $\lambda \in R$.

## Prova do lema 2.

De acordo com Lema 1, $X^{-1}(\mathcal{B})$ é uma $\sigma$-álgebra. Como $\mathcal{B}' \subseteq \mathcal{B}$, temos que $X^{-1}(\mathcal{B}') \subseteq X^{-1}(\mathcal{B})$. Então, por definição de menor $\sigma$-álgebra, temos que

$$\sigma(X^{-1}(\mathcal{B}')) \subseteq X^{-1}(\mathcal{B}).$$

Para provar igualdade, definimos

$$\mathcal{F} = \{B' \subseteq R : X^{-1}(B') \in \sigma(X^{-1}(\mathcal{B}'))\}.$$

É fácil provar que $\mathcal{F}$ é uma $\sigma$-álgebra; nós omitimos os detalhes. Por definição, temos que $X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{B}'))$ e $\mathcal{B}' \subseteq \mathcal{F}$. Como $\mathcal{F}$ é uma $\sigma$-álgebra, $\mathcal{B} = \sigma(\mathcal{B}') \subseteq \mathcal{F}$. Portanto,

$$X^{-1}(\mathcal{B}) \subseteq X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{B}')).$$

□

## Prova do Teorema 1.

Agora nós podemos provar o teorema 1. Suponha que $X^{-1}(\mathcal{B}') \subseteq \mathcal{A}$. Por definição de menor $\sigma$-álgebra,

$$\sigma(X^{-1}(\mathcal{B}')) \subseteq \mathcal{A}.$$

Então, pelo Lema 2, $X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$, o que implica que $X$ é uma variável aleatória. $\square$

De forma gera temos a seguinte definição

## Definição 5

*(Evento aleatório): Seja* $X$ *uma variável aleatória definida sobre o espaço de probabilidade* $(\Omega, \mathfrak{F}, P)$ *e com valores num espaço mensurável* $(\Omega', \mathfrak{F}')$. *Definimos o evento aleatório como o conjunto*

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}, \quad \textit{para todo } B \in \mathfrak{F}'.$$

em particular, a definição é válida para variáveis aleatórias reais, i.e., quando $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathfrak{B})$ .

## Exemplo 2

Seja $(\Omega, \mathfrak{F}, P)$ um espaço de probabilidade com $\mathfrak{F} = \mathcal{P}(\Omega)$. Neste caso, qualquer função $X$ a valores reais é uma variável aleatória. Prove!

## Exemplo 3

Um dado é lançado ao acaso uma vez. Neste caso o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Se observamos o número da face obtida no lançamento podemos definir a variável aleatória $X$ como:

$$\begin{array}{cccc} X: & \Omega & \to & \mathbb{R} \\ & \omega & \mapsto & X(\omega) = \omega, \end{array}$$

a qual atribui a cada elemento $\omega$ do espaço amostral $\Omega$ um número real $X(\omega)$ da seguinte forma

$$\begin{array}{cccc} X: & \Omega & \to & \mathbb{R} \\ & 1 & \mapsto & X(1) = 1 \\ & 2 & \mapsto & X(2) = 2 \\ & 3 & \mapsto & X(3) = 3 \\ & 4 & \mapsto & X(4) = 4 \\ & 5 & \mapsto & X(5) = 5 \\ & 6 & \mapsto & X(6) = 6 \end{array}$$

Logo, dizemos que a variável aleatória assume valores em $X = \{1, 2, 3, 4, 5, 6\}$.

## Probabilidade Induzida

Dada uma variável aleatória $X$, pode-se definir uma probabilidade induzida $P_X$ no espaço mensurável $(R, \mathcal{B})$ da seguinte maneira: para todo $A \in \mathcal{B}$, definimos $P_X(A) = P(X^{-1}(A))$. Por definição de variável aleatória, tem-se que $X^{-1}(A) \in \mathcal{A}$, então $P_X$ está bem definida. Resta provar que $P_X$ satisfaz os axiomas K1, K2, e K4$'$ de probabilidade:

De forma geral, temos a seguinte definição

## Definição 6

*(Distribuição da variável aleatória). Seja $X : (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ uma aplicação mensurável e seja $P$ uma medida de probabilidade definida sobre $\Omega$. Então a função*

$$P_X(B) = P(\{X \in B\}) = P\left(X^{-1}(B)\right) = P\left(\omega \in \Omega : X(\omega) \in B\right),$$

*em que $B \in \mathfrak{F}'$, define uma medida de probabilidade sobre $\Omega'$. A medida $P_X$ chama-se* **medida transportada ou induzida por** *$X$ ou, simplesmente, a distribuição da variável aleatória $X$, já o conjunto $\{X \in B\}$ é o evento aleatório.*

## Axiomas

K1. $P_X(A) = P(X^{-1}(A)) \geq 0.$

K2. $P_X(R) = P(X^{-1}(R)) = P(\Omega) = 1.$

K4.′ Suponha que $A_1, A_2, \ldots$ são eventos Borelianos disjuntos. Então,

$$\begin{aligned} P_X(\cup_i A_i) &= P(X^{-1}(\cup_i A_i)) = P(\cup_i X^{-1}(A_i)) \\ &= \sum_i P(X^{-1}(A_i)) = \sum_i P_X(A_i). \end{aligned}$$

## Exemplo 4

Seja $\Omega = \{1, 2, 3\}$, $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ e $P$ dada por: $P(\emptyset) = 0$, $P(\{1\}) = 1/5$, $P(\{2, 3\}) = 4/5$ e $P(\Omega) = 1$. Se $\Omega' = \{a, b\}$, $\mathfrak{F}' = \mathcal{P}(\Omega')$ e $X : \Omega \to \Omega'$ dada por

$$\begin{cases} a, & \text{se } \omega = 1, \\ b, & \text{se } \omega = 2 \text{ ou } \omega = 3, \end{cases}$$

então, a distribuição $P_X$ de $X$ é dada por: $P_X(\emptyset) = 0$, $P_X(\{a\}) = P(\{1\}) = 1/5$, $P_X(\{b\}) = P(\{2, 3\}) = 4/5$ e $P(\Omega') = 1$.

# Função de Distribuição Acumulada

Para uma variável aleatória $X$, uma maneira simples e básica de descrever a probabilidade induzida $P_X$ é utilizando sua *função de distribuição acumulada*.

## Definição 7

*A função de distribuição acumulada de uma variável aleatória $X$, representada por $F_X$, é definida por*

$$F_X(x) = P_X((-\infty, x]), \forall x \in R.$$

## Propriedades

A função de distribuição acumulada $F_X$ satisfaz as seguintes propriedades:

F1. Se $x \leq y$, então $F_X(x) \leq F_X(y)$.

De fato

$$\begin{aligned}
& x \leq y \Rightarrow (-\infty, x] \subseteq (-\infty, y] \\
& \Rightarrow P_X((-\infty, x]) \leq P_X((-\infty, y]) \\
& \Rightarrow F_X(x) \leq F_X(y).
\end{aligned}$$

F2. Se $x_n \downarrow x$, então $F_X(x_n) \downarrow F_X(x)$.

De fato. Se $x_n \downarrow x$, então os eventos $(-\infty, x_n]$ são decrescentes e $\cap_n(-\infty, x_n] = (-\infty, x]$. Logo, pela continuidade da medida de probabilidade, tem-se que $P_X((-\infty, x_n]) \downarrow P((-\infty, x])$, ou seja, $F_X(x_n) \downarrow F_X(x)$.

F3. Se $x_n \downarrow -\infty$, então $F_X(x_n) \downarrow 0$, e se $x_n \uparrow \infty$, então $F_X(x_n) \uparrow 1$.

De fato. Se $x_n \downarrow -\infty$, então os eventos $(-\infty, x_n]$ são decrescentes e $\cap_n(-\infty, x_n] = \emptyset$. Logo, pela continuidade da medida de probabilidade, tem-se que $P_X((-\infty, x_n]) \downarrow P(\emptyset)$, ou seja, $F_X(x_n) \downarrow 0$. Similarmente, se $x_n \uparrow \infty$, então os eventos $(-\infty, x_n]$ são crescentes e $\cup_n(-\infty, x_n] = \mathbb{R}$. Logo, pela continuidade da medida de probabilidade, tem-se que $P_X((-\infty, x_n]) \uparrow P(\Omega)$, ou seja, $F_X(x_n) \uparrow 1$.

## Teorema 2

*Uma função real $G$ satisfaz F1–F3 se e somente se $G$ é uma distribuição de probabilidade acumulada.*

## Demonstração.

Aprova de que se $G$ for uma distribuição de probabilidade acumulada, então $G$ satisfaz F1-F3 foi dada acima. A prova de que toda função real que satisfaz F1-F3 é uma função de probabilidade acumulada é complexa envolvendo o Teorema da Extensão de Carathéodory. Nós apresentamos aqui um esquema de como a prova é feita. Primeiro define-se $P_X((-\infty, x]) = F_X(x)$, $P_X((x, \infty)) = 1 - F_X(x)$, e $P_X((a, b]) = F_X(b) - F_X(a)$. Com esta definição, considera-se a álgebra formada por união finita de intervalos e prova-se que $P_X$ é $\sigma$-aditiva nesta álgebra. Finalmente, aplica-se o Teorema da Extensão de Carathéodory para provar que $P_X$ pode ser estendida para todo evento Boreliano. □

## Nota 1

Uma função de distribuição pode corresponder a várias variáveis aleatórias no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Por exemplo, se $X$ tem uma distribuição normal com parâmetros 0 e 1, então por simetria é fácil ver que $-X$ também distribuição normal com parâmetros 0 e 1. Consequentemente, $F_X = F_{-X}$. No entanto, $P(X = -X) = P(X = 0) = 0$.

Condição F2 significa que toda função distribuição de probabilidade acumulada $F_X$ é continua à direita. Ainda mais, como $F_X$ é não-decrescente e possui valores entre 0 e 1, pode-se provar que ela tem um número enumerável de descontinuidades do tipo salto. Pela continuidade à direita , o salto no ponto $x$ é igual a

$$\begin{aligned} F_X(x) - F_X(x^-) &= F_X(x) - \lim_{n\to\infty} F(x - \frac{1}{n}) \\ &= P_X((-\infty, x]) - \lim_{n\to\infty} P_X((-\infty, x - \frac{1}{n}]) \\ &= \lim_{n\to\infty} P_X((x - \frac{1}{n}, x]). \end{aligned}$$

Como a sequência de eventos $(x - \frac{1}{n}, x]$ é decrescente e $\cap_n (x - \frac{1}{n}, x] = \{x\}$. Temos que $\{x\}$ é Boreliano e

$$P_X(x) = F_X(x) - F_X(x^-).$$

**Teorema 3**

*Seja $D$ o conjunto de pontos de descontinuidade da função de distribuição $F$. Então, $D$ é enumerável.*

## Demonstração.

Pela monotonicidade, temos que para todo $x \in \mathbb{R}$, $F(x^-) \leq F(x) \leq F(x^+)$. Logo, $x \in D$ se, e somente se, $F(x^+) > F(x^-)$. Para $n = 1, 2, 3, \ldots$ seja

$$A_n = \{x : F(x^+) - F(x^-) > \frac{1}{n}\}.$$

Então, $D = \cup_{n=1}^{\infty} A_n$. Vamos verificar que todo $A_n$ contém menos que $n$ pontos e, portanto, é finito. Dessa forma, $D$ será enumerável.
Por absurdo, suponha que exista $A_n$ que contém $n$ pontos. Assim, $A_n = \{x_1, x_2, \ldots, x_n\}$, onde $x_1 < x_2 < \cdots x_n$ e

$$\begin{aligned} &0 \leq F(x_1^-) \leq F(x_1^+) \leq F(x_2^-) \leq F(x_2^+) \leq \cdots \\ &\cdots \leq F(x_n^-) \leq F(x_n^+) \leq 1. \end{aligned}$$

Então, temos $\sum_{k=1}^{n} [F(x_k^+) - F(x_k^-)] \leq 1$. Mas por definição do conjunto $A_n$, temos que $F(x_i^+) - F(x_i^-) > \frac{1}{n}$ para todo $x_i \in A_n$. Portanto, $\sum_{k=1}^{n} [F(x_k^+) - F(x_k^-)] > n \times \frac{1}{n} = 1$, absurdo. Logo, $A_n$ contém menos que $n$ pontos. $\square$

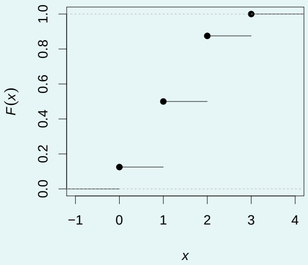
## Exemplo 5

Suponhamos que se atira ao acaso uma moeda corrente três vezes consecutivas. Neste caso $\Omega = \{Cara, Coroa\} \times \{Cara, Coroa\} \times \{Cara, Coroa\}$. Definimos

$$X = \text{“número de caras obtidas”}.$$

Note que, $X = \{0, 1, 2, 3\}$. Então, a função de distribuição $F_X$ da variável aleatória $X$ está dada por

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ \frac{1}{8}, & \text{se } 0 \leq x < 1, \\ \frac{1}{2}, & \text{se } 1 \leq x < 2, \\ \frac{7}{8}, & \text{se } 2 \leq x < 3, \\ 1, & \text{se } x \geq 3. \end{cases}$$



## Exercício 1

Da definição de função de distribuição podem ser obtidas as seguintes propriedades:

1. $F(x^-) = \lim_{0<h\to 0} F(x-h) = P(X < x)$.
2. $P(a \leq X \leq b) = F(b) - F(a^-)$.
3. $P(a < X \leq b) = F(b) - F(a)$.
4. $P(a \leq X < b) = F(b^-) - F(a^-)$.
5. $P(a < X < b) = F(b^-) - F(a)$.
6. $P(X = a) = F(a) - F(a^-)$ ( deduz-se então que $F$ é contínua em $a$, se e somente se, $P(X = a) = 0$ ).
7. Se $P(a < X < b) = 0$ então $F$ é constante sobre o intervalo $(a, b)$.

# Tipos de Variável Aleatória

## Definição

Existem três tipos de variáveis aleatórias:

- **Discreta.** Uma variável aleatória $X$ é *discreta* se assume um número enumerável de valores, ou seja, se existe um conjunto enumerável $\{x_1, x_2, \ldots\} \subseteq R$ tal que $X(w) \in \{x_1, x_2, \ldots\}, \forall w \in \Omega$. A função $p(x_i)$ definida por $p(x_i) = P_X(\{x_i\}), i = 1, 2, \ldots$ e $p(x) = 0$ para $x \notin \{x_1, x_2, \ldots\}$, é chamada de *função probabilidade* de $X$. Note que neste caso, temos

$$F_X(x) = \sum_{i:x_i \leq x} p(x_i).$$

- **Contínua.** Uma variável aleatória $X$ é *contínua* se existe uma função $f_X(x) \geq 0$ que é Riemman integrável tal que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \forall x \in R.$$

  Neste caso, a função $f_X$ é chamada de *função densidade de probabilidade* de $X$.

- **Singular.** Uma variável aleatória $X$ é *singular* se $F_X$ é uma função contínua cujos pontos de crescimento formam um conjunto de comprimento (medida de Lebesgue) nulo.

Mais adiante veremos que pode-se decompor qualquer função de distribuição de probabilidade acumulada $F_X$ na soma de no máximo três funções de distribuição de probabilidade acumuladas, sendo uma discreta, uma contínua e outra singular.

## Nota 2

Para o caso de funções limitadas, se $f$ é uma função definida num intervalo $[a, b] \subset \mathbb{R}$ e $\mathcal{P} : a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ é uma partição arbitrária de $[a, b]$. Dizemos que a função $f$ é Riemann integrável no intervalo $[a, b]$, se existir (e for finito) o limite seguinte: $\lim_{|\mathcal{P}|\to 0} \sum_{i=1}^{n} f(\epsilon_i)\Delta x_i$, independentemente da partição $\mathcal{P}$ do intervalo $[a, b]$, ou de como os pontos $\epsilon_i$ pertencentes aos subintervalos $[x_{i-1}, x_i]$ são escolhidos, em que $|\mathcal{P}|$ é o comprimento do maior intervalo contido na partição $\mathcal{P}$ e $\Delta x_i = x_i - x_{i-1}$.

# Variável Aleatória Discreta

- Se uma variável aleatória é discreta, então pode-se definir uma função de probabilidade $p$ de modo que $p(x_i) = P_X(\{x_i\}), i = 1, 2, \ldots$, onde $X \subseteq \{x_1, x_2, \ldots\}$ e $p(x) = 0$ para $x \notin \{x_1, x_2, \ldots\}$. Toda função de probabilidade é uma função dos reais $R$ e assume valores entre 0 e 1, sendo positiva para um número enumerável de pontos e satisfaz a seguinte propriedade $\sum_i p(x_i) = 1$.
- Reciprocamente, dada uma função $p : R \to [0, 1]$, onde $p$ é positiva para um número enumerável de pontos $\{x_1, x_2, \ldots\}$ e satisfaz $\sum_i p(x_i) = 1$, uma função $P$ definida nos eventos Borelianos de modo que $P(A) = \sum_{x_i \in A} p(x_i), \forall A \in \mathcal{B}$ é uma medida de probabilidade em $(R, \mathcal{B})$ (verifique os axiomas de Kolmogorov!). Logo, a distribuição de uma variável aleatória discreta $X$ pode ser determinada tanto pela função de distribuição acumulada $F_X$ ou pela sua função de probabilidade $p$.

## Exemplo 6

Consideremos a variável aleatória $X$ que assume os valores 1, 2 e 3, com probabilidades 0,3, 0,5 e 0,2 respectivamente. Então a função de probabilidade de $X$ é

$$p(x) = \begin{cases} 0,3 & \text{se } x = 1, \\ 0,5 & \text{se } x = 2, \\ 0,2 & \text{se } x = 3, \\ 0 & \text{caso contrário.} \end{cases}$$

cujo gráfico é indicado na Figura 1.

# Variável Aleatória Contínua

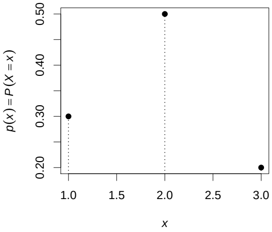



Figura: Representação gráfica da função de probabilidade $p(x)$.

## Variável Aleatória Contínua

Se uma variável aleatória é (absolutamente) contínua, então existe uma função $f_X(x) \geq 0$ tal que $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. Deste modo, $F_X$ é contínua e $f_X(x) = F'_X(x)$, exceto num conjunto de medida de Lebesgue nula. Uma função $f(x) \geq 0$ é densidade de alguma variável aleatória se e somente se, $\int_{-\infty}^{\infty} f(x)dx = 1$, já que neste caso é fácil provar que a função $F$ definida por $\int_{-\infty}^{x} f(t)dt$ satisfaz as condições F1, F2, e F3, e portanto, pelo Teorema 2 $F$ é uma função de distribuição acumulada. Logo, a distribuição de uma variável aleatória contínua $X$ pode ser determinada tanto pela função de distribuição acumulada $F_X$ ou pela sua função de densidade $f_X$.

Uma variável aleatória $X$ tem densidade se $F_X$ é a integral (de Lebesgue) de sua derivada; sendo neste caso a derivada de $F_X$ uma função densidade para $X$. Este fato pode ser provado utilizando argumentos de Teoria da Medida. Sem recorrer a argumentos envolvendo Teoria da Medida, em quase todos os casos encontrados na prática, uma variável aleatória $X$ tem densidade se $F_X$ é (i) contínua e (ii) derivável por partes, ou seja, se $F_X$ é derivável no interior de um número finito ou enumerável de intervalos fechados cuja união é a reta $R$.

No gráfico da Figura 2 podemos observar a função de distribuição representada como a área embaixo da curva da função de densidade (painel esquerdo) e como uma função nos valores $x$ (painel direito).

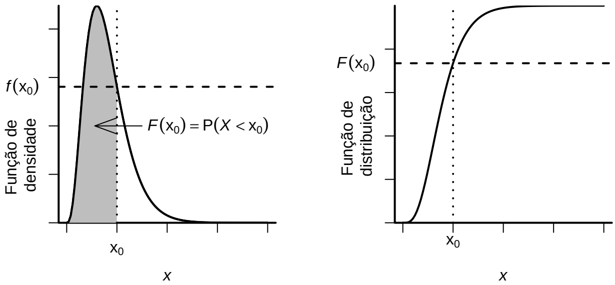



Figura: Representação gráfica de $f(x)$ e $F(x)$.

Das propriedades da $F(x)$ pode-se ver que

$$2(\Delta x)f(x) \approx F(x+\Delta x) - F(x-\Delta x) = P(x-\Delta x < X \leq x+\Delta x)$$

isto é, a probabilidade de que $X$ esteja num intervalo de comprimento pequeno ao redor de $x$ é igual a $f(x)$ pelo comprimento do intervalo.

## Exemplo 7

Por exemplo, considere

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } 0 \leq x < 1, \\ 1 & \text{se } x \geq 1. \end{cases}$$

Então $X$ tem densidade pois $F_X$ é contínua e derivável em todos os pontos da reta exceto em $\{0, 1\}$.

## Exemplo 8

Seja $X$ uma variável aleatória contínua cuja função de distribuição é dada por:

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } 0 \leq x < 1, \\ 1 & \text{se } x \geq 1. \end{cases} \qquad f(x) = \begin{cases} 1 & \text{se} 0 \leq x \leq 1, \\ 0 & \text{caso contrário} \end{cases}$$

é uma função de densidade para $X$.

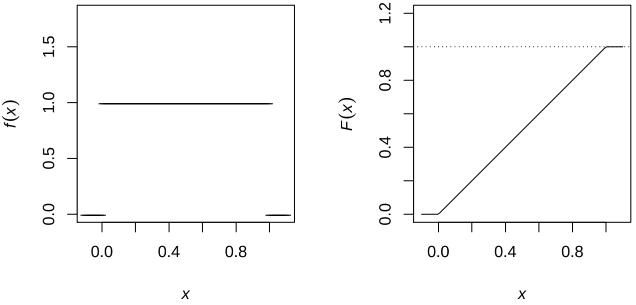



Figura: Representação gráfica de $f(x)$ e $F(x)$.

## Variável Aleatória Singular

Vamos analisar um exemplo de uma função de distribuição de uma variável aleatória singular conhecida como *função de Cantor.* Esta função é contínua, derivável em todo ponto exceto em um conjunto de medida de Lebesgue nula, mas não é absolutamente contínua. Seja $F(x) = 0$ se $x < 0$ e $F(x) = 1$ se $x > 1$. Continuemos por etapas:

Etapa 1: Seja $F(x) = \frac{1}{2}$ para $x \in (1/3, 2/3)$. Então, o valor de $F$ neste intervalo é igual a média dos valores de $F$ nos intervalos vizinhos em que $F$ já está definida: $(-\infty, 0)$ e $(1, \infty)$. $F$ continua sem definição em dois intervalos: $[0, 1/3]$ e $[2/3, 1]$ de comprimento total 2/3.

Etapa $n+1$: No terço central de cada um dos $2^n$ intervalos restantes após a etapa $n$, seja $F(x)$ igual à média dos valores nos dois intervalos vizinhos onde $F$ já está definida. Por exemplo, na etapa 2 defina $F(x) = 1/4$ para $x \in (1/9, 2/9)$ e $F(x) = 3/4$ para $x \in \cup(7/9, 8/9)$. Restarão então $2^{n+1}$ intervalos (o dobro do número restante após a etapa $n$), de comprimento total $(2/3)^{n+1}$, em que $F$ ainda não estará definida.

Então definimos $F$ por indução em um número enumerável de intervalos abertos, cujo complementar (ou seja, o conjunto onde $F$ ainda não está definida) é o conjunto de Cantor, um conjunto de comprimento 0.

Podemos estender a definição de $F$ até o conjunto de Cantor $C$ por continuidade: se $x \in C$, a diferença entre os valores de $F$ nos dois intervalos vizinhos após a etapa $n$ é $1/2^n$.

Note que $F$ é monótona não decrescente em $C^c$. Se $a_n$ é o valor de $F$ no intervalo vizinho esquerdo após a etapa $n$, e $b_n$ é o valor no intervalo vizinho direito após a etapa $n$, então, $a_n \uparrow$, $b_n \downarrow$ e $b_n - a_n \downarrow 0$.

Seja $F(x)$ o limite comum de $a_n$ e $b_n$. Deste modo $F$ está definida em toda reta e é de fato uma função de distribuição (verifique!).

Seja $X$ uma variável aleatória cuja função de distribuição é $F$, a função de Cantor. Então $X$ não é discreta e nem contínua pois $X$ não tem densidade $F'(x) = 0$ em $C^c$ e $\int_{-\infty}^{x} F'(t)dt = 0$, ou seja, $F$ não é a integral de sua derivada, ou melhor, não é absolutamente contínua.

Como $F$ é contínua e $F'(x) = 0$ para $x \in C^c$ e $C$ tem comprimento nulo, temos que $X$ é uma variável aleatória singular.

## Decomposição de uma Variável Aleatória

Seja $X$ uma variável aleatória qualquer e seja $F$ sua função de distribuição. Se $J = \{x_1, x_2, \ldots\}$ é o conjunto dos pontos de salto de $F$ (se $F$ for contínua $J = \emptyset$), indiquemos com $p_i$ o salto no ponto $x_i$, ou seja,

$$p_i = F(x_i) - F(x_i^-).$$

Definimos $F_d(x) = \sum_{i:x_i \leq x} p_i$. $F_d$ é uma função degrau não-decrescente: a *parte discreta* de $F$. Como uma função monótona possui derivada em quase toda parte, seja

$$f(x) = \begin{cases} F'(x) & \text{se } F \text{ é diferenciável em } x, \\ 0 & \text{se } F \text{ não é diferenciável em } x. \end{cases}$$

Seja $F_{ac}(x) = \int_{-\infty}^{x} f(t)dt$. $F_{ac}$ é não-decrescente, pois a integral indefinida de uma função nao-negativa ($f \geq 0$ porque $F$ é não-decrescente). A sua derivada é igual a $f$ em quase toda parte, de modo que $F_{ac}$ é absolutamente contínua: $F_{ac}$ é a parte *absolutamente contínua* de $F$.
Seja $F_s(x) = F(x) - F_d(x) - F_{ac}(x)$. $F_s$ é contínua pois é a diferença de duas funções contínuas. A derivada de $F_s$ é igual a zero em quase toda parte, porque $F$ e $F_{ac}$ têm a mesma derivada $f$, e $F_d$ possui derivada zero em quase toda parte. Pode-se provar que $F_s$ também é não-decrescente, mas está fora do escopo deste curso. $F_s$ é a *parte singular* de $F$.
Esta discussão nos dá um método de decompor $F$ em suas partes discreta, absolutamente contínua e singular.

## Exemplo 9

Suponha que $X \sim U[0,1]$ e $Y = min(X, 1/2)$. Note que

$$F_Y(x) = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } 0 \leq x < 1/2, \\ 1 & \text{se } x \geq 1/2. \end{cases}$$

$F_Y$ tem apenas um salto em $x = 1/2$ e $p_1 = 1/2$. Logo, $F_d(x) = 0$ se $x < 1/2$ e $F_d(x) = 1/2$ se $x \geq 1/2$. Diferenciando $F_Y$, temos

$$F'_Y(x) = \begin{cases} 0 & \text{se } x < 0 \text{ ou } x > 1/2, \\ 1 & \text{se } 0 < x < 1/2. \end{cases}$$

Logo, por definição,

$$f(x) = \begin{cases} 0 & \text{se } x \leq 0 \text{ ou } x \geq 1/2, \\ 1 & \text{se } 0 < x < 1/2. \end{cases}$$

Portanto,

$$F_{ac}(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } 0 \leq x \leq 1/2, \\ 1/2 & \text{se } x > 1/2. \end{cases}$$

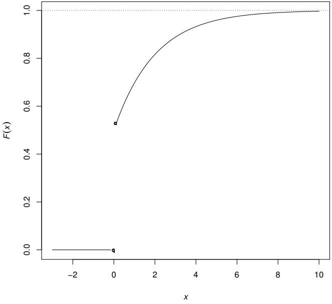
Como $F_d + F_{ac} = F_Y$, temos que $F_s(x) = 0, \forall x \in \mathbb{R}$ e não há parte singular. *Uma variável aleatória que possui apenas partes discreta e absolutamente contínua é conhecida como uma variável aleatória mista.* Na prática, é pouco provável que surja uma variável aleatória singular. Portanto, quase todas as variáveis aleatórias são discretas, contínuas ou mistas.

## Exemplo 10

Seja $X$ a variável aleatória com função de distribuição dada por:

$$F(x) = \begin{cases} 1 - \frac{1}{2}e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{caso contrário,} \end{cases}$$

em que $\lambda > 0$ é uma constante. A variável aleatória $X$ não é discreta nem contínua pois apresenta um ponto de descontinuidade em $x = 0$. Ou seja é Singular. Para $\lambda = 1/2$ o gráfico de $F(x)$ é apresentado na seguinte figura.



# Principais Distribuições Discretas

## Aleatória - Uniforme discreta

$X$ tem uma distribuição *aleatória* com parâmetro $n$, onde $n$ é um número inteiro, se $X(w) \in \{x_1, x_2, \ldots, x_n\}$ e $p(x_i) = \frac{1}{n}$, para $i \in \{1, \ldots, n\}$.

- Utilizando a propriedade de aditividade da probabilidade, é fácil ver que para qualquer evento $A \subseteq \{x_1, x_2, \ldots, x_n\}$, temos que $P(X \in A) = \frac{||A||}{n}$.
- Aplicações: modelar mecanismos de jogos (dados e moedas balanceados, cartas bem embaralhadas).

# Principais Distribuições Discretas

## Bernoulli

$X$ tem uma distribuição *Bernoulli* com parâmetro $p$, onde $0 \leq p \leq 1$, se $X(w) \in \{x_0, x_1\}$ e $p(x_1) = p = 1 - p(x_0)$.

- Aplicações: modelar a probabilidade de sucesso em uma única realização de um experimento. Em geral, qualquer variável aleatória dicotômica pode ser modelada por uma distribuição Bernoulli.

# Principais Distribuições Discretas

## Binomial

$X$ tem uma distribuição *Binomial* com parâmetros $n$ e $p$, onde $n$ é um número inteiro e $0 \leq p \leq 1$, se $X(w) \in \{0, 1, \ldots, n\}$ e $p(k) = \binom{n}{k} p^k (1-p)^{1-k}$, para $k \in \{0, 1, \ldots, n\}$. Note que utilizando o Teorema Binomial, temos que

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

Logo, esta é uma legítima função probabilidade de massa.

# Principais Distribuições Discretas

## Binomial (cont.)

- Pode-se provar que a soma de $n$ variáveis aleatórias Bernoulli com parâmetro $p$ independentes tem uma distribuição $Binomial(n, p)$.
- Aplicações: modelar a quantidade de erros em um texto de $n$ símbolos quando os erros entre símbolos são assumidos independentes e a probabilidade de erro em um símbolo do texto é igual a $p$; modelar o número de caras em $n$ lançamentos de uma moeda que possui probabilidade $p$ de cair cara em cada lançamento. Se $p = 1/2$, temos um modelo para o número de 1's em uma sequência binária de comprimento $n$ escolhida aleatoriamente ou o número de caras em $n$ lançamentos de uma moeda justa.

# Principais Distribuições Discretas

## Geométrica

$X$ tem uma distribuição *Geométrica* com parâmetro $\beta$, onde $0 \leq \beta < 1$, se $X(w) \in \{0, 1, \ldots\}$ e $p(k) = (1-\beta)\beta^k$, para $k \in \{0, 1, \ldots\}$.
Utilizando o resultado de uma soma infinita de uma Progressão Geométrica, temos que

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} (1-\beta)\beta^k = (1-\beta)\sum_{k=0}^{\infty} \beta^k = 1.$$

Logo, esta é uma legítima função probabilidade de massa.

- Aplicações: modelar o tempo de espera medido em unidades de tempo inteira até a chegada do próximo consumidor em uma fila, até a próxima emissão de um fóton, ou até a primeira ocorrência de cara numa sequência de lançamentos de uma moeda.

# Principais Distribuições Discretas

## Binomial Negativa ou Pascal

É uma generalização da distribuição geométrica. Suponha que ao invés de estarmos interessados no tempo de espera até a primeira ocorrência de um evento, estejamos interessados em calcular o tempo de espera até a $r$-ésima ocorrência de um evento. Seja $Y$ o tempo de espera necessário a fim de que um evento $A$ possa ocorrer exatamente $r$ vezes. Temos que $Y = k$ se, e somente se, $A$ ocorrer na $(k+1)$-ésima repetição e $A$ tiver ocorrido $r-1$ vezes nas $k$ repetições anteriores. Assumindo independência entre os experimentos, esta probabilidade é igual $p\binom{k}{r-1}p^{r-1}(1-p)^{k-r+1}$. Portanto,

$$P(Y = k) = \binom{k}{r-1}p^r(1-p)^{k-r+1}, \text{ onde } k \geq r-1.$$

Note que se $r = 1$, temos que $Y$ tem uma distribuição geométrica com parâmetro $\beta = 1-p$. No caso geral, dizemos que $Y$ tem uma distribuição *Binomial Negativa ou Pascal*.

# Principais Distribuições Discretas

## Relação entre as Distribuições Binomial e Binomial Negativa

Se $X \sim Binomial(n, p)$ e $Y \sim Pascal(r, p)$, então temos que $\{X \geq r\} = Y + 1 \leq n$, ou seja, o número de sucessos em $n$ repetições de um experimento é maior ou igual a $r$ se, e somente se, o tempo de espera para o $r$-ésimo sucesso for menor ou igual a $n - 1$. Portanto,

$$P(X \geq r) = P(Y \leq n - 1).$$

# Principais Distribuições Discretas

## Relação entre as Distribuições Binomial e Binomial Negativa

Observe que estas duas distribuições tratam de experimentos repetidos.

- A distribuição binomial surge quando lidamos com um número fixo de experimentos e estamos interessados no número de sucessos que venham a ocorrer.
- A distribuição binomial negativa é encontrada quando fixamos o número de sucessos e então registramos o tempo de espera necessário.

# Principais Distribuições Discretas

## Zeta ou Zipf

$X$ tem uma distribuição Zeta ou Zipf com parâmetro $\alpha$, onde $\alpha > 1$, se $X(w) \in \{1, 2, \ldots\}$ e

$$p(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}, k = 1, 2, \ldots,$$

onde $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ é conhecida como a função Zeta de Riemann.

- A função de probabilidade Zeta ou Zipf é um exemplo de uma distribuição de cauda pesada; observe que ela tem decaimento polinomial.
- Aplicações: número de consumidores afetados por um blackout, tamanhos de arquivos solicitados em transferência via Web e atraso de pacotes na internet.

# Principais Distribuições Discretas

## Hipergeométrica

A distribuição hipergeométrica descreve o número de sucessos em uma sequência de $n$ amostras de uma população finita sem reposição.

Considere que tem-se uma carga com $N$ objetos dos quais $D$ têm defeito. A distribuição hipergeométrica descreve a probabilidade de que em uma amostra de $n$ objetos distintos escolhidos da carga aleatoriamente exatamente $k$ objetos sejam defeituosos.

Formalmente, se uma variável aleatória $X$ segue uma distribuição hipergeométrica com parâmetros $N$, $D$, e $n$, então a probabilidade de termos exatamente $k$ sucessos é dada por

$$p(k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}.$$

Esta probabilidade é positiva se: $N - D \geq n - k$, ou seja $k \geq \max(0, D + n - N)$, e $k \leq \min(n, D)$.

# Principais Distribuições Discretas

## Hipergeométrica (cont.)

- Quando a população é grande quando comparada ao tamanho da amostra (ou seja, $N$ for muito maior que $n$) a distribuição hipergeométrica é aproximada razoavelmente bem por uma distribuição binomial com parâmetros $n$ (tamanho da amostra) e $p = D/N$ (probabilidade de sucesso em um único ensaio).
- Note que se a amostragem é feita com reposição, temos que a distribuição correta é dada por uma $Binomial(n, D/N)$.

# Principais Distribuições Discretas

## Poisson

$X$ tem uma distribuição *Poisson* com parâmetro $\lambda$, onde $\lambda \geq 0$, se $X(w) \in \{0, 1, \ldots\}$ e $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, para $k \in \{0, 1, \ldots\}$.
Por definição, temos que para todo $x$ real,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Utilizando este fato, temos que

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

- Aplicações: modelar a contagem do número de ocorrências de eventos aleatórios em um certo tempo $T$, como número de fótons emitidos por uma fonte de luz de intensidade $I$ fótons/seg em $T$ segundos ($\lambda = IT$), número de clientes chegando em uma fila no tempo $T$ ($\lambda = CT$), número de ocorrências de eventos raros no tempo $T$ ($\lambda = CT$).

# Principais Distribuições Discretas

## Poisson como um Limite de Eventos Raros de Binomial

Suponhamos que chamadas telefônicas cheguem em uma grande central, e que em um período particular de três horas (180 minutos), um total de 270 chamadas tenham sido recebidas, ou seja, 1,5 chamadas por minuto. Suponhamos que queiramos calcular a probabilidade de serem recebidas $k$ chamadas durante os próximos três minutos.

Ao considerar o fenômeno da chegada de chamadas, poderemos chegar à conclusão de que, a qualquer instante, uma chamada telefônica é tão provável de ocorrer como em qualquer outro instante. Como em qualquer intervalo de tempo, temos um número infinito de pontos, vamos fazer uma série de aproximações para este cálculo.

# Principais Distribuições Discretas

## Poisson como um Limite de Eventos Raros de Binomial

Para começar, pode-se dividir o intervalo de 3 minutos em nove intervalos de 20 segundos cada um. Poderemos então tratar cada um desses nove intervalos como um ensaio de Bernoulli, durante o qual observaremos uma chamada (sucesso) ou nenhuma chamada (falha), com probabilidade de sucesso igual a $p = 1,5 \times \frac{20}{60} = 0,5$. Desse modo, poderemos ser tentados a afirmar que a probabilidade de 2 chamadas é igual a $\binom{9}{2}(0,5)^9 = \frac{9}{128}$. Porém, este cálculo ignora a possibilidade de que mais de uma chamada possa ocorrer em um único intervalo. Então, queremos aumentar o número $n$ de subintervalos de tempo de modo que cada subintervalo corresponde a $\frac{180}{n}$ segundos e então a probabilidade de ocorrência de uma chamada em um subintervalo é igual a $p = 1,5 \times \frac{180}{60n}$. Desta maneira temos que $np = 4,5$ permanece constante ao crescermos o número de subintervalos. Utilizando novamente o modelo binomial, temos que a probabilidade de ocorrerem $k$ chamadas é dada por: $\binom{n}{k}(\frac{4,5}{n})^k(1 - \frac{4,5}{n})^{n-k}$. Queremos saber então o que acontece com esta probabilidade quando $n \to \infty$. A resposta como veremos a seguir é que esta distribuição tende a distribuição de Poisson e este resultado é conhecido como *limite de eventos raros*.

# Principais Distribuições Discretas

## Poisson como um Limite de Eventos Raros de Binomial

Consideremos a expressão geral da probabilidade binomial,

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$
$$= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}.$$

Como queremos estudar o caso em que $np$ é constante, façamos $np = \alpha$, ou seja, $p = \alpha/n$ e $1 - p = \frac{n-\alpha}{n}$. Então,

$$p(k) = \frac{n(n-1)\cdots(n-k+1)}{k!} (\frac{\alpha}{n})^k (\frac{n-\alpha}{n})^{n-k}$$
$$= \frac{\alpha^k}{k!} [(1)(1 - \frac{1}{n})\cdots(1 - \frac{k-1}{n})][1 - \frac{\alpha}{n}]^{n-k}$$

## Poisson como um Limite de Eventos Raros de Binomial

Fazendo $n \to \infty$, temos que os termos da forma $(1 - \frac{j}{n})$, para $1 \leq j \leq k - 1$, tendem para 1 e como existe um número fixo $k$ deles, o seu produto também tende a 1. O mesmo ocorre com $(1 - \frac{\alpha}{n})^{-k}$. Finalmente, por definição do número $e$, temos que $(1 - \frac{\alpha}{n})^n \to e^{-\alpha}$ quando $n \to \infty$. Portanto,

$$\lim_n p(k) = e^{-\alpha} \frac{\alpha^k}{k!},$$

ou seja obtemos a expressão de Poisson.

Mais geralmente, pode-se provar o seguinte teorema:

# Principais Distribuições Discretas

## Poisson como um Limite de Eventos Raros de Binomial

### Teorema 4

*Se* $\lim_{n\to\infty} np_n = \alpha > 0$*, então*

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\alpha}\frac{\alpha^k}{k!}.$$

### Demonstração.

Nós utilizamos os seguintes fatos:

1. $\lim_{n\to\infty} \binom{n}{k} = \lim_{n\to\infty} \frac{n^k}{k!}$.
2. $\lim_{n\to\infty} np_n^2 = 0$.
3. $(1-x)^n \leq e^{-nx}$, para $x \geq 0$.
4. $(1-x)^n \geq e^{-nx-nx^2}$, para $0 \leq x \leq \frac{1}{2}$.

Usando fatos 2, 3, e 4, nós obtemos $\lim_{n\to\infty}(1-p_n)^{n-k} = \lim_{n\to\infty} e^{-(n-k)p_n}$. Logo, usando fato 1,

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \lim_{n\to\infty} \frac{(np_n)^k}{k!} e^{-(n-k)p_n} = e^{-\alpha}\frac{\alpha^k}{k!}.$$

# Principais Distribuições Contínuas

## Uniforme

$X$ tem uma distribuição *uniforme* com parâmetros $a$ e $b$, onde $a$ e $b$ são números reais e $a < b$, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{1}{b-a} U(x-a)U(b-x).$$

- Este modelo é frequentemente usado impropriamente para representar "completa ignorância" sobre valores de um parâmetro aleatório sobre o qual apenas sabe-se estar no intervalo finito $[a, b]$.
- Também é frequentemente utilizada para modelar a fase de osciladores e fase de sinais recebidos em comunicações incoerentes.

# Principais Distribuições Contínuas

## Exponencial

$X$ tem uma distribuição *Exponencial* com parâmetro $\lambda$, onde $\lambda > 0$ é um número real, se a função densidade de $X$ é igual a

$$f_X(x) = \lambda e^{-\lambda x} U(x).$$

- Aplicações: modelar tempo de vida de componentes que falham sem efeito de idade; tempo de espera entre sucessivas chegadas de fótons, emissões de elétrons de um cátodo, ou chegadas de consumidores; e duração de chamadas telefônicas.

# Principais Distribuições Contínuas

## Qui-quadrado

$X$ tem uma distribuição *Qui-quadrado* com parâmetro $n$, onde $n$ é número natural, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)} U(x),$$

onde $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$ para $p > 0$ é a função gama. $n$ é conhecido como número de *graus de liberdade* da distribuição qui-quadrado.

- Pode-se provar que a soma dos quadrados de $n$ variáveis aleatórias independentes com distribuição normal padrão possui uma distribuição qui-quadrado com $n$ graus de liberdade.
- A distribuição Qui-quadrado tem inúmeras aplicações em inferência estatística. Por exemplo, em testes qui-quadrados e na estimação de variâncias.

# Principais Distribuições Contínuas

## Gama

$X$ tem uma distribuição *Gama* com parâmetros $\alpha$ e $\beta$, onde $\alpha > 0$ e $\beta > 0$ são números reais, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} U(x).$$

- Pode-se provar que a soma de $\alpha$ variáveis aleatórias exponenciais independentes com média $1/\beta$ tem uma distribuição Gama. É fácil ver que se $\alpha = 1$, temos uma distribuição exponencial com parâmetro $\beta$, e se $\alpha = n/2$ e $\beta = 1/2$ temos uma distribuição Qui-quadrado com $n$ graus de liberdade.

# Principais Distribuições Contínuas

## Beta

Dizemos que $X$ tem uma distribuição *Beta* com parâmetros $\alpha$ e $\beta$, onde $\alpha > 0$ e $\beta > 0$ são números reais, se a função densidade de $X$ é igual a

$$\begin{aligned} f_X(x) &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du} U(x)U(1-x) \\ &= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} U(x)U(1-x), \end{aligned}$$

onde $B(\alpha, \beta)$, para $\alpha > 0, \beta > 0$, é a função beta que é o fator de normalização que garante que $f_X$ é uma densidade.

- Distribuições Beta são usadas exaustivamente em Estatística Bayesiana, pois elas são uma família de distribuições *a priori* conjugadas para distribuições binomiais e geométricas.
- A distribuição beta pode ser utilizada para modelar eventos que tem restrição de estar em um intervalo finito.

# Principais Distribuições Contínuas

## t de Student

$X$ tem uma distribuição *t de Student* com parâmetro $n$, onde $n$ é número natural, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{\Gamma[(n+1)/2]}{\Gamma[n/2]\sqrt{\pi n}}\left(1 + \frac{x^2}{n}\right)^{\frac{-(n+1)}{2}},$$

onde $n$ é conhecido como número de *graus de liberdade* da distribuição t de Student.

- Pode-se provar que se $Z$ tem uma distribuição normal padrão, $V$ tem uma distribuição qui-quadrado com $n$ graus de liberdade e $Z$ e $V$ forem independentes, então $\frac{Z}{\sqrt{V/n}}$ tem uma distribuição t de Student com $n$ graus de liberdade.
- A distribuição t de Student é bastante utilizada em inferência estatística. Por exemplo, pode-se utilizá-la para calcular intervalos de confiança para a média de uma amostra quando a variância da população não é conhecida.

# Principais Distribuições Contínuas

## Pareto

$X$ tem uma distribuição *Pareto* com parâmetros $\alpha$ e $\tau$, onde $\alpha$ e $\tau$ são números reais positivos, se a função densidade de $X$ é igual a

$$f_X(x) = \alpha\tau^{\alpha} x^{-\alpha-1} U(x-\tau).$$

- A distribuição de Pareto é o exemplo mais fundamental de uma distribuição contínua de cauda pesada. Ela pode ser utilizada para modelar distribuição de riquezas; atrasos em transmissão de pacotes; e duração sessões de Internet.

# Principais Distribuições Contínuas

## Normal ou Gaussiana

$X$ tem uma distribuição *Normal (ou Gaussiana)* com parâmetros $m$ e $\sigma$, onde $m$ e $\sigma > 0$ são números reais, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-m)^2}{2\sigma^2}}.$$

- Historicamente, esta distribuição foi chamada de “normal” porque ela era amplamente aplicada em fenômenos biológicos e sociais que era sempre tida como a distribuição antecipada ou normal.
- Se $m = 0$ e $\sigma = 1$, diz-se que $X$ tem uma distribuição *normal padrão* ou *normal reduzida*.
- Aplicações: modelar ruído térmico em resistores e em outros sistemas físicos que possuem um componente dissipativo; ruídos de baixa-frequência como os em encontrados em amplificadores de baixa frequência; e variabilidade em parâmetros de componentes manufaturados e de organismos biológicos (por exemplo, altura, peso, inteligência).

# Principais Distribuições Contínuas

## Cauchy

$X$ tem uma distribuição *Cauchy* com parâmetro $a > 0$, se a função densidade de $X$ é igual a

$$f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2}.$$

- A razão entre duas variáveis aleatórias com distribuição Normal padrão independentes tem uma distribuição Cauchy com parâmetro 1.