SME 0520 - Introdução à Estatística

Daiane de Souza

SME/ICMC/USP

Parte II - Março de 2025

Variáveis Aleatórias

Variáveis Aleatórias

Definição:

Seja E um experimento aleatório e Ω o espaço amostral associado a esse experimento. Uma função $X(\omega)$ que associa a cada elemento $\omega \in \Omega$ a um número real $x = x(\omega)$ é denominada **variável aleatória.**

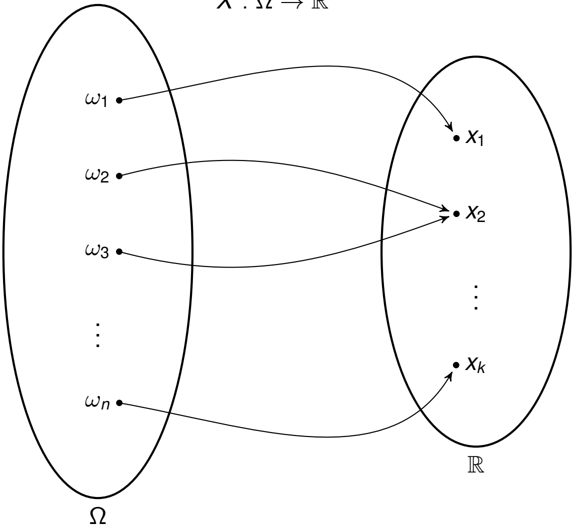
Definição:

Uma variável aleatória é uma função que mapeia o espaço amostral na reta real, sendo que cada elemento do espaço amostral é mapeado em um valor real.

Variáveis Aleatórias

Ilustração:

$X : \Omega \to \mathbb{R}$



Variáveis Aleatórias

Exemplo 2.1: Lançamento de uma moeda duas vezes. A variável aleatória X é o número de caras.

Variáveis Aleatórias

- **Exemplo:** Em uma linha do produção, peças são classificadas em defeituosas ou não-defeituosas. Podemos definir uma variável aleatória X como:

$$X = \begin{cases} 1, & \text{a peça é defeituosa} \\ 0, & \text{caso contrário} \end{cases}.$$

  - Nesse caso, $\Omega = \{\text{peça defeituosa; peça não-defeituosa}\}$.
  - Uma variável X desse tipo é chamada **variável aleatória de Bernoulli**.
  - A variável aleatória X assume um conjunto finito de valores.

Variáveis Aleatórias

Classificação de Variáveis Aleatórias

- Se a variável aleatória X assume valores em um **conjunto finito ou infinito enumerável** é chamada **variável aleatória discreta**.
    - **Exemplo:** X indica o número de residentes em um domicílio (X pode assumir valores em $\mathbb{N}$).
- Se a variável aleatória X assume valores em um **conjunto infinito não enumerável** é chamada **variável aleatória contínua**.
    - **Exemplo:** X indica o tempo de vida do componente eletrônico, em horas (X pode assumir valores em $\mathbb{R}^+$).

Definição: Variável Aleatória Discreta.

Seja X uma **variável aleatória discreta** que assume valores em R_x, $R_x = \{x_1, x_2, \ldots, x_n, \ldots\}$. A cada resultado x_i, associamos a um número

$$p_x(x_i) = P(X(\omega_i) = x_i),\ \omega_i \in \Omega \ \text{ e } \ x_i \in R_x,$$

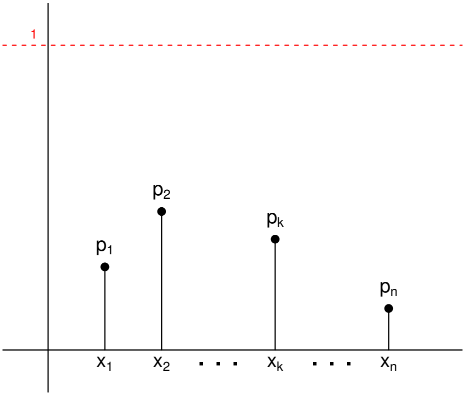
dito probabilidade de x_i.

A função $p_x(x)$ é definida como **função massa de probabilidade** de X (f.m.p. de X).

As probabilidades $p_X(x_i)$ devem satisfazer as seguintes condições:

i. $p_X(x_i) > 0$, para todo $x_i \in R_X$;
ii. $\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Ilustração da f.m.p.: Seja X uma variável aleatória discreta com $R_X = \{x_1, x_2, \ldots, x_n\}$ e $p_X(x_i) = p_i$.



Exemplo 2.2: Lançamento de uma moeda duas vezes e X é o número de caras.

Exemplo 2.3: Um carregamento de 8 computadores contém 3 defeituosos. Se uma empresa faz uma compra aleatória de dois desses computadores, apresente a função de probabilidade para o número de computadores com defeitos adquiridos.

Exemplo 2.4: A demanda diária de um item é uma variável aleatória discreta com função massa de probabilidade dada por:

$$P(D = d) = \frac{2^d k}{d!},\ d = 1, 2, 3, 4.$$

a) Determine a constante k.

b) Calcule $P(D > 2)$.

Introdução: Variável Aleatória Contínua.

Caracterização de variáveis cujos possíveis valores ocorrem aleatoriamente e pertencem a um intervalo dos números reais: renda, salário, tempo de uso de um equipamento, área atingida por uma praga agrícola, ...

Objetivos: De forma semelhante àquela desenvolvida para variáveis aleatória discretas, precisamos estabelecer, para as contínuas, a atribuição de probabilidades às suas diversas realizações que, neste caso, podem assumir um número infinito de valores diferentes.

Exemplo:

Estudos anteriores revelam a existência de um grande lençol de água no subsolo de uma região. No entanto, sua profundidade ainda não foi determinada, sabendo-se apenas que o lençol pode estar situado em qualquer ponto entre 20 e 100 metros.

Vamos supor que escolhemos, ao acaso, um ponto nessa região e dispomos de uma sonda que, ao fazer a perfuração, detecta com precisão a profundidade do reservatório de água. Denotamos por X a variável aleatória representando a **profundidade**.

Exemplo:

- Uma vez que não temos informações adicionais a respeito da profundidade do lençol, é razoável assumirmos que a sonda pode parar em qualquer ponto entre 20 e 100 metros;
- Assim, consideraremos todos os pontos como igualmente prováveis;
- Se utilizarmos a mesma ideia de atribuir a cada possível ponto uma probabilidade, teremos uma dificuldade extra, pois eles pertencem ao intervalo [20,100], em que existem infinitos números reais;
- Assim, se cada um deles tiver a mesma probabilidade maior que zero, a soma das probabilidades será igual a **infinito** e não 1, como requer a definição da **função de probabilidade**.

Exemplo:

- Em geral, em situações como esta, não é de interesse considerar um único valor para a variável aleatória, mas intervalos de valores na atribuição de probabilidades;
- Neste caso, sabemos que o espaço amostral corresponde ao intervalo [20,100] e as profundidades são igualmente prováveis;
- Suponha, por um momento, que dividimos o espaço amostral em 8 intervalos de comprimento 10.
- Logo, é razoável atribuir aos intervalos a probabilidade 1/8, correspondendo à relação entre o comprimento de cada um deles e o comprimento do espaço amostral (10 para 80).

Exemplo:

- Para construir um histograma, podemos supor que 1/8 é a frequência relativa da ocorrência de cada um dos intervalos;
- Note que, dada as características do problema, a divisão em 8 intervalos produziu o mesmo valor da densidade de 1/80 para todos eles;
- Se dividirmos o intervalo em 16 faixas iguais, temos que os intervalos terão todos a mesma probabilidade $1/16$.
- Apesar de termos diferentes intervalos, a densidade permanece com o mesmo valor, igual a 1/80.

Exemplo:

- Estamos agora em condições de caracterizar, completamente , a atribuição de probabilidades para o caso contínuo;
- Ela será definida pela área abaixo de uma função positiva, denominada **densidade de probabilidade**;
- Notemos que a densidade não é uma probabilidade, mas uma função matemática que nos auxilia na atribuição de probabilidade;
- Para a v.a. contínua X, a função densidade f é dada por

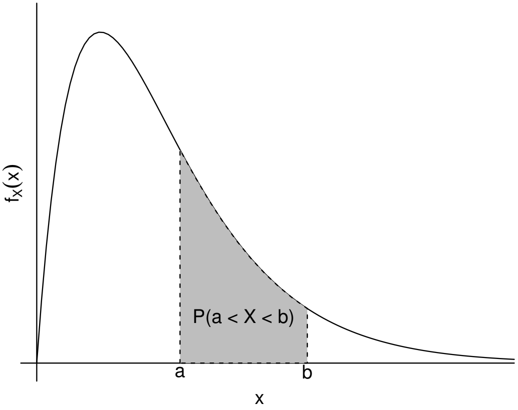
$$f_X(x) = \begin{cases} 1/80, & \text{para } 20 \leq x \leq 100 \\ 0, & \text{para } x < 20 \quad \text{ou} \quad x > 100. \end{cases}.$$

Definição: Variável Aleatória Contínua.

Seja X uma **variável aleatória contínua** que assume valores em R_x, $R_x \in \mathbb{R}$. A função $f_x(x)$ é a **função densidade de probabilidade** (f.d.p.) para X, se satisfaz as seguintes propriedades:

i. $f_x(x) \geq 0$, para todo $x \in R_x$;
ii. $\int_{R_x} f_x(x)dx = 1 \left(\int_{-\infty}^{+\infty} f_x(x)dx = 1\right)$.
iii. $P(a < X < b) = \int_a^b f_x(x)dx$.

Ilustração da f.d.p.:



- Se X é uma variável aleatória contínua assumindo valores em R_x, então para todo $a \in R_x$ temos:

  a) $P(X = a) = 0$;

  b) $P(X > a) = P(X \geq a)$;

  c) $P(X < a) = P(X \leq a)$;

  d) $P(X > a) = 1 - P(X \leq a) = 1 - P(X < a)$;

  e) $P(X < a) = 1 - P(X \geq a) = 1 - P(X > a)$.

Exemplo 2.5: O tempo de produção de um componente (em minutos) é uma variável aleatória X com f.d.p. dada por:

$$f_X(x) = \begin{cases} \dfrac{5-x}{4}, & \text{se } 2 < x < 4 \\ 0, & \text{caso contrário} \end{cases} .$$

a) Mostre que $f_X(x)$ é uma f.d.p..

b) Calcule a probabilidade de que o tempo de produção de um compenente escolhido ao acaso seja menos do que 3 minutos.

Exercício Proposto 2.1: Seja X uma variável aleatória contínua com f.d.p. dada por:

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

a) Verifique se $f_X(x)$ é uma f.d.p..

b) $P(X \leq \frac{1}{2})$.

c) $P(X \leq \frac{1}{2} | \frac{1}{3} \leq X \leq \frac{2}{3})$.

Resp.: b) $\frac{1}{4}$; c) $\frac{5}{12}$.

Função de Distribuição Acumulada

Definição: Função de Distribuição Acumulada (caso discreto).

Seja X uma **variável aleatória discreta** que assume valores em R_x e com f.m.p. $p_X(x) = P(X = x)$. Para qualquer $x \in \mathbb{R}$, a **função de distribuição acumulada (f.d.a.)** de X, denotada por $F_X(x)$, é definida como:

$$\begin{aligned} F_X(x) = P(X \leq x) &= \sum_{\substack{x_i \in R_x; \\ \forall\ x_i \leq x}} p_X(x_i) \\ &= \sum_{\substack{x_i \in R_x; \\ \forall\ x_i \leq x}} P(X = x_i). \end{aligned}$$

Exemplo 2.6: Considere o lançamento de uma moeda duas vezes e X é o número de caras. Obtenha a f.m.p. de X. Obtenha também a f.d.a. de X juntamente com a sua ilustração gráfica.

Definição: Função de Distribuição Acumulada (caso contínuo).

Seja X uma **variável aleatória contínua** que assume valores em R_x e com f.d.p. $f_x(x)$. Para qualquer $x \in \mathbb{R}$, a **função de distribuição acumulada (f.d.a.)** de X, denotada por $F_x(x)$, é definida como:

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f_x(t)dt, \text{ para } -\infty < x < +\infty.$$

- Como consequência imediata, podemos escrever os dois resultados:

$$P(a < X < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

e

$$f_X(x) = \frac{d}{dx}F_X(x),$$

se a derivada existir.

Exemplo 2.7: Para a função densidade dada por:

$$f_X(x) = \begin{cases} \dfrac{x^2}{3}, & \text{se } -1 < x < 2 \\ 0, & \text{caso contrário} \end{cases},$$

determine $F_X(x)$ e use-a para avaliar $P(0 \leq X < 1)$.

Propriedades de uma f.d.a.:

1. Para todo x, $0 \geq F(x) \leq 1$;
2. $F_x(x)$ é uma função monótona não decrescente;
3. $\lim_{x \to -\infty} F_x(x) = 0$ e $\lim_{x \to +\infty} F_x(x) = 1$;
4. Se $R_x = \{X_1, X_2, \ldots\}$, em que $X_1 < X_2 < \ldots$, então $f(X_i) = P(X = x_i) = F(X_i) - F(X_{i-1})$
5. Se a e b são tais que $a < b$, então
   (i) $P(X \leq a) = F_x(a)$;
   (ii) $P(X \geq a) = 1 - P(X < a) =$;
   (iii) $P(a < X \leq b) = F_x(b) - F_x(a)$;
   (iv) $P(a \leq X \leq b) = F_x(b) - F_x(a) + P(x = a)$;
   (v) $P(a < X < b) = F_x(b) - F_x(a) - P(X = b)$.

Exercício Proposto 2.2: Seja $F_x(x)$ dada por:

$$F_x(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1/8, & \text{se } 0 \leq x < 1 \\ 1/2, & \text{se } 1 \leq x < 2 \\ 5/8, & \text{se } 2 \leq x < 3 \\ 1, & \text{se } x \geq 3 \end{cases}.$$

Determinar:

a) $P(1 < X \leq 3)$.

b) $P(X > 2)$.

c) Encontre a $p_x(x)$.

Resp.: a) $\frac{1}{2}$; b) $\frac{3}{8}$; c) $\frac{1}{8}$ se $x = 0; 2$ e $\frac{3}{8}$ se $x = 1; 3$.

Exercício Proposto 2.3: Seja X uma variável aleatória contínua com f.d.p. dada por:

$$f_X(x) = \begin{cases} kx^2, & \text{se } 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases} .$$

a) Ache o valor de k.

b) Determine $F_X(x)$.

c) $P(\frac{1}{3} < X < \frac{1}{2})$.

Resp.: a) $k = 3$; b) x^3, se $0 < x \leq 1$; c) $\frac{19}{216}$.

Esperança Matemática de uma Variável Aleatória

Definição: Esperança Matemática.

Seja X uma variável aleatória com f.m.p. $p_X(x)$ (no caso discreto) ou f.d.p. $f_X(x)$ (no caso contínuo). Chamamos de esperança matemática ou valor médio de X a quantidade

$$\mu = E(X) = \sum_{x \in R_X} x p_X(x), \text{ no caso discreto}$$

ou

$$\mu = E(X) = \int_{R_X} x f_X(x), \text{ no caso contínuo,}$$

desde que o somatório e a integral existam.

Algumas propriedades:

- Considere $a, b \in \mathbb{R}$, constantes.

a) $E(aX) = aE(X)$;

b) Se $X = a$ (constante), então $E(X) = E(a) = a$;

c) $E\big(E(X)\big) = E(X)$;

d) $E(X \pm a) = E(X) \pm a$;

e) $E(aX \pm bY) = aE(X) \pm bE(Y)$.

Exemplo 2.8: Seja X uma variável aleatória com f.m.p. dada por:

$$p_X(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}},\ x = 0, 1, 2, 3.$$

Calcule o $E(X)$.

Exemplo 2.9: Considere a variável aleatória X com f.d.p. dada por:

$$f_X(x) = 2x,\ 0 \leq x \leq 1.$$

Calcule o $E(X)$.

Resultado: Seja X uma variável aleatória com f.m.p. $p_X(x)$ no caso discreto ou f.d.p. $f_X(x)$ no caso contínuo. Uma função de X, dita $g(X)$, é também uma variável aleatória e

$$E\big(g(X)\big) = \sum_{x \in R_X} g(x) p_X(x), \text{ no caso discreto}$$

ou

$$E\big(g(X)\big) = \int_{R_X} g(x) f_X(x), \text{ no caso contínuo.}$$

Variância de uma Variável Aleatória

Definição: Variância.

Seja X uma variável aleatória com f.m.p. $p_X(x)$ (no caso discreto) ou f.d.p. $f_X(x)$ (no caso contínuo), com média $\mu = E(X)$. Chamamos de variância da variável aleatória X o valor

$$\begin{aligned} \sigma^2 = Var(X) &= E\left((X - E(X))^2\right) \\ &= E\left((X - \mu)^2\right). \end{aligned}$$

Definição: Variância (continuação).

Ou seja:

$$\sigma^2 = Var(X) = \sum_{x \in R_x} (x-\mu)^2 p_X(x), \text{ no caso discreto}$$

ou

$$\sigma^2 = Var(X) = \int_{R_x} (x-\mu)^2 f_X(x), \text{ no caso contínuo.}$$

A raiz quadrada da variância é chamada de desvio-padrão de X, $\sigma = DP(X)$.

Resultado: Podemos escrever a variância da variável aleatória X por:

$$\sigma^2 = Var(X) = E(X^2) - \{E(X)\}^2.$$

Dem.:]

$$\begin{aligned} Var(X) &= E\big((X - E(X))^2\big) \\ &= E\big(X^2 - 2XE(X) + (E(X))^2\big) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

Algumas propriedades:

- Considere $a, b \in \mathbb{R}$, constantes.

a) $Var(aX) = a^2 Var(X)$;

b) Se $X = a$ (constante), então $Var(X) = Var(a) = 0$;

c) $Var(X \pm a) = Var(X)$;

d) $Var(aX \pm b) = a^2 Var(X)$;

e) Se X e Y são duas **variáveis aleatórias independentes**, então $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$.

Exemplo 2.10: Seja X uma variável aleatória com f.d.p. dada por:

$$f_X(x) = \frac{x^2}{3},\ -1 < x < 2.$$

Determine:

a) $E(X)$.

b) $Var(X)$.

c) $E(4X + 3)$.

Exercício Proposto 2.4: Seja X uma variável aleatória com f.m.p. dada por:

$$p_X(x) = \begin{cases} \frac{1}{2}, & \text{se } x = 0 \\ \frac{1}{4}, & \text{se } x = 1, 2 \\ 0, & \text{caso contrário} \end{cases} .$$

Considere a variável aleatória $g(X) = (X - a)^2$, $a = 0; \frac{1}{2}; 1$. Calcule:

a) $E(X)$ e $Var(X)$.

b) $E\big(g(X)\big)$ para cada a.

Resp.: a) $\mu = \frac{3}{4}$ e $\sigma^2 = \frac{11}{16}$; b) $E\big(g(X)\big) = a^2 - \frac{3}{2}a + \frac{5}{4}$ (avaliar para cada valor de a).

Exercício Proposto 2.5: A demanda semanal de certa bebida, em milhares de litros, em uma rede de lojas de conveniência é uma variável aleatória contínua $g(X) = X^2 + X - 2$, sendo que X tem f.d.p. dada por:

$$f_X(x) = \begin{cases} 2(x-1), & \text{se } 1 < x < 2 \\ 0, & \text{caso contrário} \end{cases}.$$

Determine o valor esperado da demanda semanal dessa bebida.

Resp.: $\frac{15}{6} = 2,5$.