Capítulo 3

Conceitos Básicos de Probabilidade

A probabilidade pode ser definida, basicamente, como a teoria matemática que é utilizada para estudar a incerteza oriunda de fenômenos aleatórios. Os fenômenos aleatórios são observados a partir da realização de um experimento não determinístico, também chamado de experimento probabilístico. Um experimento não determinístico é aquele em que o resultado do experimento é aleatório, ou seja, existe a incerteza do resultado. Os fenômenos aleatórios são situação ou acontecimento cujos resultados não podem ser determinados com certeza, como por exemplo o resultado do lançamento de um dado, o hábito de fumar de um estudante sorteado em sala, as condições climáticas do próximo domingo, a taxa de inflação do próximo mês ou o resultado de um exame de sangue.

Os resultados observados de um fenômeno aleatório são denominados Variáveis Aleatórias. Como o próprio nome diz, uma variável é uma quantidade ou qualidade observada em um conjunto de unidades amostrais em que seu resultado varia de unidade amostral para unidade amostral e é aleatória, pois, seu resultado é aleatório, ou seja, existe a incerteza do resultado que será observado. A seguir serão apresentadas dois tipos de variáveis aleatórias, que são utilizadas no estudo de probabilidade, as discretas e as contínuas.

3.1 Variáveis Aleatórias Discretas

Uma quantidade X é denominada *variável aleatória discreta* se assume valores num conjunto enumerável, ou seja, assume apenas valores inteiros, com certa probabilidade de ocorrência. O comportamento de uma variável aleatória discreta é estudado a partir da função discreta de probabilidade. Essa função atribui a probabilidade de ocorrência de cada valor da variável aleatória e é denominada *função de probabilidade* (f.p.), ela pode ser definida por,

$$P\left(X = x_i\right) = p\left(x_i\right) = p_i,\ i = 1, 2, \ldots \tag{3.1}$$

ou ainda,

$$\begin{array}{c|cccc} X & x_1 & x_2 & x_3 & \cdots \\ \hline p_i & p_1 & p_2 & p_3 & \cdots \end{array} \tag{3.2}$$

em que, $0 \leq p_i \leq 1$ e $\sum_i p_i = 1$.

Essa função matemática é muito útil, na prática, para resolver problemas envolvendo inferência estatística (assunto que será abordado no próximo capítulo). Dependendo do comportamento da variável aleatória discreta, existem modelos estatísticos (funções matemáticas) específicos para o estudo dessa variável. Esses modelos são denominados Modelos Discretos de Probabilidade e os principais serão abordados a seguir.

3.1.1 Modelo Uniforme Discreto

Seja X uma variável aleatória discreta cujos possíveis valores são representados por $x_1, x_2, x_3, \ldots, x_k$. A variável aleatória X vai seguir distribuição Uniforme Discreta; notação: $X \sim Ud\,(k)$; se a probabilidade de ocorrência de cada um desses k valores é a mesma, ou seja, $1/k$. Portanto, a função matemática que melhor descreve esse comportamento é,

$$P\,(X = x_i) = \frac{1}{k},\ i = 0, 1, 2, \ldots, k. \tag{3.3}$$

Exemplo: no lançamento de um dado, não viciado, com seis faces numeradas de um até seis, a probabilidade de sair qualquer número no intervalo de um a seis é a mesma. Portanto, a variável aleatória *X: número obtido no lançamento de um dado, não viciado, com seis faces numeradas de um até seis* segue distribuição uniforme discreta, ou seja, $X \sim Ud\,(6)$.

3.1.2 Modelo de Bernoulli

Uma variável aleatória X segue distribuição de Bernoulli; notação: $X \sim Br\,(p)$; se seus valores são dicotômicos, 0 ou 1, e representam a ocorrência de fracasso e sucesso, respectivamente. Com p representando a probabilidade de sucesso, $0 \leq p \leq 1$, sua função de probabilidade é dada por,

$$P\,(X = x) = p^x\,(1-p)^{1-x},\ x = 0, 1. \tag{3.4}$$

Exemplo: supor o interesse em estudar o comportamento de peças defeituosas e não defeituosas de um lote em uma linha de produção, assumindo que a variável aleatória *X: condição da peça no lote* assuma dois valores, são eles:

$$X = \begin{cases} 1 \text{ peça defeituosa} \\ 0 \text{ peça não defeituosa} \end{cases}, \tag{3.5}$$

logo, a variável aleatória *X: condição da peça no lote* segue distribuição de Bernoulli com probabilidade de sucesso p. Supor que a probabilidade de selecionar uma peça ao acaso desse lote e ela ser defeituosa seja de 12%, portanto, $X \sim Br\,(0,12)$.

3.1.3 Modelo de Poisson

Seja uma variável aleatória X que representa o número de ocorrências (contagem) de um evento de interesse num determinado intervalo de tempo, X segue distribuição de Poisson com parâmetro $\lambda > 0$; notação: $X \sim P\,(\lambda)$; se sua função de probabilidade é dada por,

$$P\,(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \ldots, \tag{3.6}$$

o parâmetro λ representa a frequência média ou esperada de ocorrências do evento de interesse num determinado intervalo de tempo. O parâmetro de um modelo de probabilidade é uma quantidade, que pode ser calculada (estimada) a partir de um conjunto de observações, que descreve o comportamento da função de probabilidade. Por exemplo, o parâmetro da distribuição uniforme discreta, definida em (3.3), é o k; o parâmetro da distribuição de Bernoulli, definida em (3.4), é o p; e o parâmetro da distribuição de Poisson, definida em (3.6), é o λ.

Exemplo: supor o interesse em estudar o comportamento de postura de ovos em uma granja, pode-se assumir que a variável aleatória *X: número de ovos postados por ave diariamente* (contagem) segue distribuição de Poisson. Supor que o número médio de ovos postos por ave seja de $3,2$ ovos por dia, portanto, $X \sim P\,(3,2)$.

3.2 Variáveis Aleatórias Contínuas

Uma quantidade X é denominada *variável aleatória contínua* se assume valores num conjunto não enumerável, ou seja, ela pode assumir qualquer valor pertencente a um intervalo dos números reais. O comportamento de uma variável aleatória contínua é estudado a partir da função densidade de probabilidade (f.d.p.). A função densidade de probabilidade é uma função matemática contínua $f(x)$ que descreve o comportamento de uma variável aleatória contínua X, para uma função matemática ser considerada f.d.p. ela deve satisfazer duas condições:

- $f(x) \geq 0$, para todo $x \in (-\infty, \infty)$.
- A área definida por $f(x)$ é igual a 1, ou seja,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1. \tag{3.7}$$

Assim como no casa discreto, essa função matemática é muito útil, na prática, para resolver problemas envolvendo inferência estatística (assunto que será abordado no próximo capítulo). Dependendo do comportamento da variável aleatória contínua, existem modelos estatísticos (funções matemáticas) específicos para o estudo dessa variável. Esses modelos são denominados Modelos Contínuos de Probabilidade e os principais serão abordados a seguir.

3.2.1 Modelo Uniforme Contínuo

Uma variável aleatória contínua X segue distribuição uniforme contínua no intervalo $[a, b]$, $a < b$; notação: $X \sim U[a, b]$, se sua função densidade de probabilidade é dada por,

$$f(x) = \frac{1}{b-a}, \ a \leq x \leq b. \tag{3.8}$$

O comportamento de uma variável aleatória contínua que segue distribuição uniforme contínua pode ser observado na Figura 3.1, ou seja, independente dos valores que a variável aleatória X assume sua f.d.p. vai ser sempre a mesma $(\frac{1}{b-a})$. Os parâmetros a e b, podem ser calculados (estimados) observando os respectivos valores mínimos e máximos de um conjunto de dados, ou seja, $a = min(X)$ e $b = max(X)$.

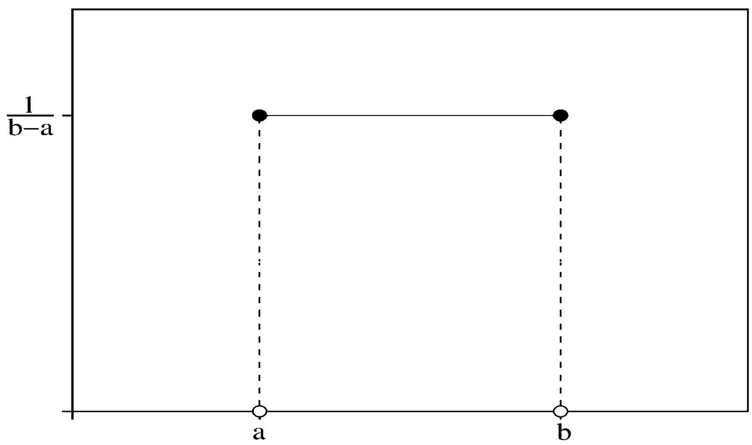


Figura 3.1: Comportamento de uma variável aleatória que segue distribuição uniforme contínua.

3.2.2 Modelo Exponencial

Uma variável aleatória contínua X, assumindo valores não negativos ($X \geq 0$), segue distribuição exponencial com parâmetro $\alpha > 0$; notação: $X \sim Exp\,(\alpha)$; se sua f.d.p. é dada por,

$$f\,(x) = \alpha e^{-\alpha x},\ x \geq 0. \tag{3.9}$$

O comportamento de uma variável aleatória contínua que segue distribuição exponencial pode ser observado na Figura 3.2, ou seja, sua f.d.p. tem um comportamento decrescente. O parâmetro α pode ser calculado (estimado) utilizando a média observada de um conjunto de dados, sua estimativa é dada por $\alpha = \frac{1}{\bar{x}}$, em que $\bar{x}$ é a média observada. Esse parâmetro define o grau de decréscimo da curva, quanto menor o valor de α maior o grau de decrescimento da curva, consequentemente, quanto maior seu valor menor o grau de decrescimento da curva (ver Figura 3.2).

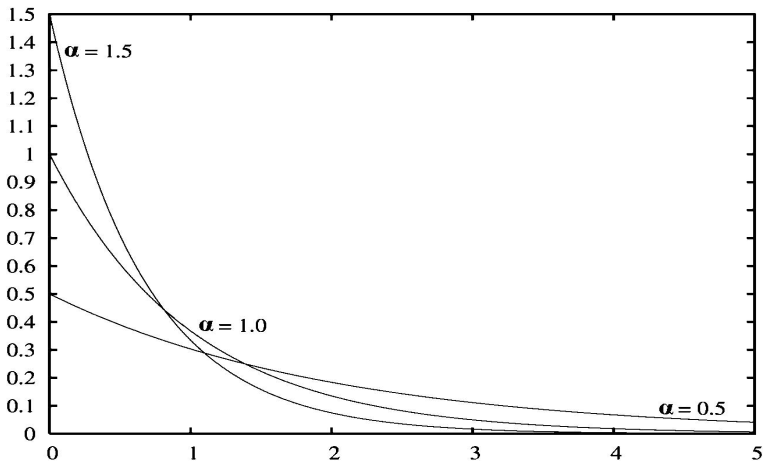


Figura 3.2: Comportamento de uma variável aleatória que segue distribuição exponencial.

3.2.3 Modelo Normal

Para muitas variáveis aleatórias contínuas, a distribuição de probabilidade é representada por uma curva específica, em formato de sino, chamada de curva normal ou curva gaussiana. É a distribuição de probabilidade mais utilizada na Estatística. Uma variável aleatória X segue distribuição normal com parâmetros de locação $-\infty \leq \mu \leq \infty$ e escala $\sigma^2 > 0$, se sua f.d.p. é dada por,

$$f\,(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\ \ -\infty < x < \infty, \tag{3.10}$$

em que, ∞ é a notação matemática usual para infinito, ou seja, a distribuição normal comporta qualquer valor positivo ou negativo, diferente da distribuição normal que só comporta valores positivos. Para indicar que uma variável aleatória segue distribuição normal com parâmetros de locação μ e escala σ^2, usa-se a notação $X \sim N\,(\mu, \sigma^2)$. O comportamento de uma variável aleatória contínua que segue distribuição normal pode ser observado na Figura 3.3.

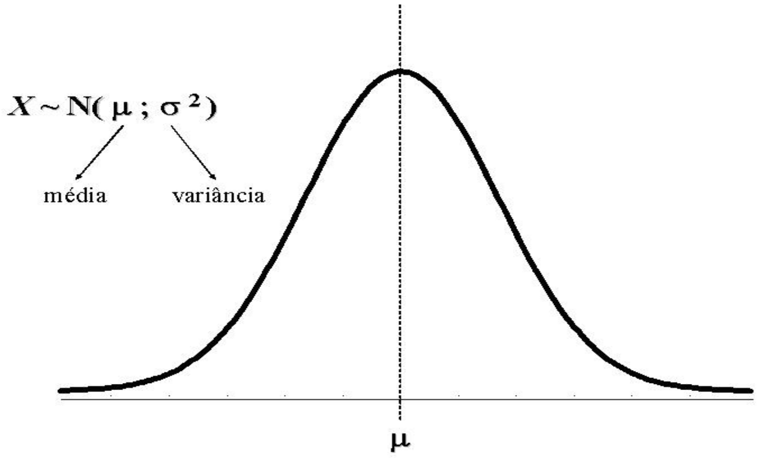
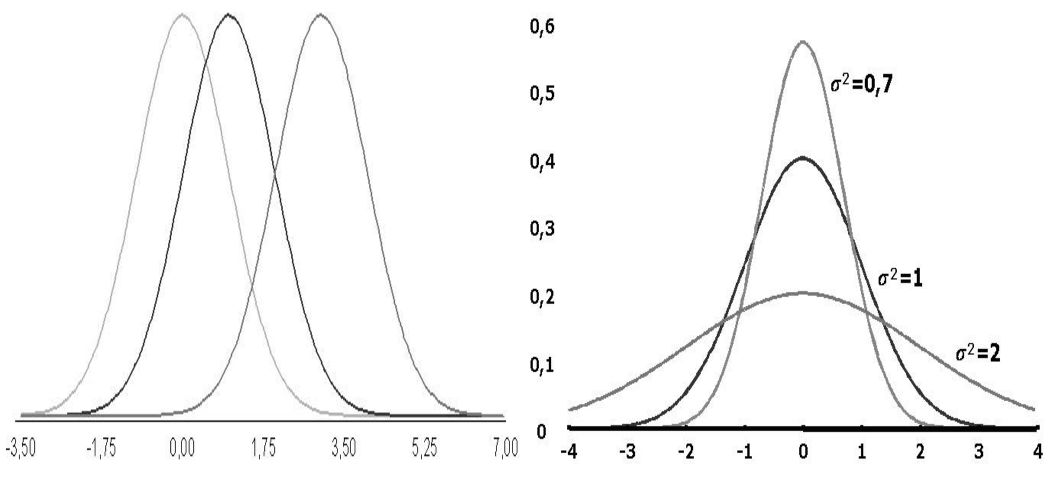


Figura 3.3: Comportamento de uma variável aleatória que segue distribuição normal.

Observando a Figura 3.3, pode-se verificar algumas propriedades da distribuição normal, percebe-se que a função densidade $f(x)$ é simétrica em relação a μ; quando $f(x)$ tende a zero x tende a $\pm\infty$ e o valor máximo de $f(x)$ se dá para $x = \mu$. Pode-se verificar que os parâmetros μ e σ^2 representam, respectivamente, a média e a variância da distribuição, ou seja, eles podem ser respectivamente calculados (estimados) utilizando a média e a variância observada de um conjunto de dados.

Na Figura 3.4, tem-se o comportamento da distribuição normal para alguns valores de μ e σ^2. Percebe-se que o valor de μ define a locação (posição) da curva (Figura 3.4(a)), quanto menor o valor de μ mais a esquerda a curva vai se posicionar, consequentemente, quanto maior o valor de μ mais a direita a curva vai se posicionar; o valor de σ^2 define a escala da curva (Figura 3.4(b)), quanto menor o valor de σ^2 mais estreita é a curva, consequentemente, quanto maior o valor de σ^2 menos estreita é a curva.

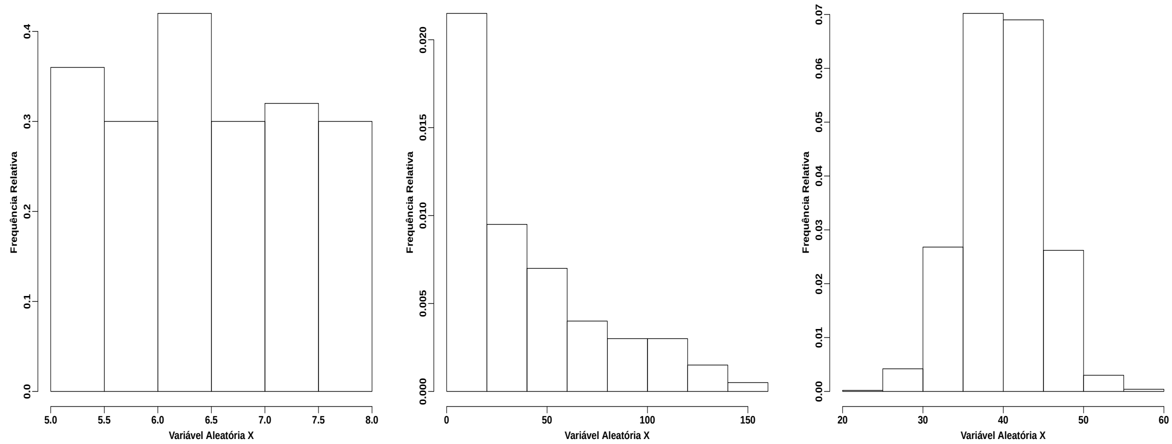


(a) (b)

Figura 3.4: Exemplos do comportamento da distribuição normal para diferentes valores de μ e σ^2.

3.2.4 Aplicação Dos Modelos De Probabilidade: Ajuste De Dados Aos Modelos De Probabilidade Contínuos

Uma ferramenta bastante útil na descrição de variáveis aleatórias contínuas é o histograma, os primeiros histogramas foram utilizados em 1781 na área da economia. O histograma consiste em retângulos contíguos com base nas faixas de valores da variável e com área igual à frequência relativa da respectiva faixa. Desta forma, a altura de cada retângulo é denominada *densidade de frequência* ou simplesmente *densidade* e é definida pelo quociente da área pela amplitude da faixa. É importante ressaltar que, é possível utilizar a frequência absoluta ou a frequência relativa (porcentagem) na construção do histograma. Aqui será utilizado a frequência relativa, pois há uma relação direta entre a frequência relativa e a função densidade de probabilidade, por exemplo, na Figura 3.5 tem-se o comportamento do histograma para algumas variáveis aleatórias contínuas que seguem os modelos de probabilidade contínuos vistos nas três seções anteriores desse texto.



(a) Modelo Uniforme Contínuo (b) Modelo Exponencial (c) Modelo Normal

Figura 3.5: Exemplos do comportamento do histograma para três modelos contínuos de probabilidade.