Capítulo 5

Variáveis aleatórias

5.1 Introdução

Em experimentos aleatórios cujo espaço amostral contém alguns eventos de interesse é, em geral, mais fácil lidar como uma variável aleatória, isto é, é mais fácil sumarizar a informação do espaço amostral em valores associados a eventos. Por exemplo, em um estudo ecológico pode haver o interesse em determinar se certa espécie vegetal está ou não presente em n locais de um continente Atribuindo 1 à presença e 0 à ausência, o espaço amostral teria 2^n elementos. Não obstante, se a informação de interesse for o número de locais que contém a espécie, então poderia ser definida a variável X representando o número de locais onde a espécie está presente, captando assim a essência do problema.

Considere ainda o exemplo a seguir: Tome o ato de identificar o sexo de duas crias de uma égua como sendo um experimento aleatório. O espaço amostral associado é definido por $S = \{MM, MF, FM, FF\}$. Seja X a variável aleatória que representa o número de machos obtidos nas duas crias Tem-se então que: $X(MM) = 2$, $X(MF) = 1$, $X(FM) = 1$ e $X(FF) = 0$.

Ao especificar a quantidade X, definimos uma transformação a partir de cada elemento A pertencente ao espaço amostral S para um novo espaço amostral R, um conjunto de números reais (no último exemplo: 0, 1, 2). Essa função a partir do espaço amostral nos reais é o que chamamos de *variável aleatória*, como ilustra a Figura 5.1.

Probabilidades podem então ser associadas aos valores ou intervalo de valores de uma variável aleatória, constituindo assim a distribuição de probabilidades dessa variável. Muitas das técnicas estatísticas são baseadas em modelos de distribuição de probabilidades, os quais podem, obviamente, serem utilizados para calcular probabilidades de interesse. Um exemplo clássico

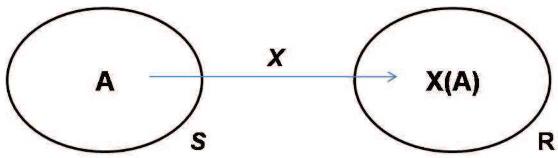

Figura 5.1: Ilustrando a definição de variável aleatória, domínimo (S) e contradomínio (R)

.

dessa aplicação é o cálculo do valor-p nos testes de hipóteses.

O uso de variáveis aleatórias equivale a descrever os resultados de um experimento aleatório por meio de valores numéricos ao invés de palavras, o que nos permite um melhor tratamento matemático.

Uma variável aleatória quantitativa pode ser *discreta* ou *contínua*.

5.2 Variável aleatória discreta

Uma variável aleatória X é considerada discreta se o conjunto de valores dessa variável, seu *espaço amostral*, for enumerável. Em geral, os valores assumidos são números inteiros, por exemplo: número de animais doentes, número de insetos praga por planta, tamanho da leitegada etc.

A distribuição de probabilidades de uma variável aleatória discreta X pode ser caracterizada pela sua *função de probabilidade* (f.p.), de modo que a probabilidade de X assumir um certo valor x é determinada pela f.p., denotada por $P_X(X = x)$ ou simplesmente $P_X(x)$.

A função P_X é dita f.p. de X se e somente se satisfizer:

1. $P_X(X = x) \geq 0 \;\forall\; x$;

2. $\sum_x P_X(X = x) = 1$.

Formalmente, denominamos *distribuição de probabilidades da v.a.d.* a coleção de pares $[x_i,\ P_X(x_i)]$, $i = 1, 2, ..., n$, que pode ser apresentada por meio de tabelas ou gráficos.

Revisitando o exemplo das duas crias de uma égua, poderíamos definir a seguinte distribuição de probabilidades (tabela 5.1) da variável discreta X, o número de machos.

Assim, podemos calcular probabilidades do tipo $P(X \geq 1)$, ao menos 1 macho:

Tabela 5.1: Distribuição de probabilidades de X, o número de machos em dois partos de uma égua.

X_i	0	1	2
$P(X_i)$	1/4	1/2	1/4

$$\begin{aligned} P(X \geq 1) &= P_X(X=1) + P_X(X=2) \\ &= 1/2 + 1/4 \\ &= 3/4 \end{aligned}$$

Calcule $P(X > 1)$.

Média e variância

O valor esperado ou média e a variância de uma variável aleatória discreta são obtidos fazendo

$$\begin{aligned} E(X) &= \textstyle\sum_x x P_X(X=x) \\ Var(X) &= E[X - E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Para a distribuição de probabilidades apresentada na tabela 5.1, temos:

$$\begin{aligned} E(X) &= \textstyle\sum_x x P_X(X=x) \\ &= x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3) \\ &= 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 \\ &= 1 \end{aligned}$$

e também,

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= [x_1^2 P(x_1) + x_2^2 P(x_2) + x_3^2 P(x_3)] - 1^2 \\ &= [0^2 \times 1/4 + 1^2 \times 1/2 + 2^2 \times 1/4] - 1 \\ &= 1/2 \end{aligned}$$

5.3 Variável aleatória contínua

Uma variável aleatória é do tipo contínua se puder assumir todo e qualquer valor em algum intervalo $[a, b]$. Em geral, tais valores são obtidos por um processo de medição, por exemplo: altura das plantas (em cm), peso dos animais (em kg), rendimento de grãos (em kg/ha) etc.

A distribuição de probabilidades de uma variável aleatória contínua X pode ser caracterizada pela sua *função densidade de probabilidade* (f.d.p.), denotada por f_X, de modo que probabilidades de X assumir valores num dado intervalo podem ser determinadas integrando f_X.

A função f_X com domínio real e contradomínio no intervalo $[0, \infty)$ é dita f.d.p. de X se e somente se satisfizer

1. $f_X(x) \geq 0 \ \forall \ x$;
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$;
3. $P_X(a \leq X \leq b) = \int_a^b f_X(x)dx$.

Exemplo: Seja a variável aleatória contínua X denotada pela função densidade de probabilidade a seguir, representada na figura 5.2:

$$f(x) = \begin{cases} 3x^2, & \text{se } 0 \leq x < 1 \\ 0, & \text{para outros valores de x} \end{cases}$$

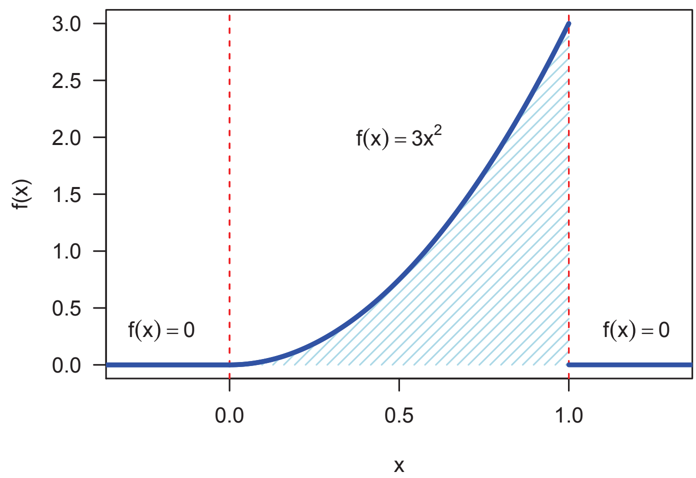

Figura 5.2: Função densidade de probabilidade de X.

Com $f(x)$ podemos calcular, por exemplo,

$$\begin{aligned} P(X > 0,5) &= \int_{0,5}^{1} f_X(x)dx \\ &= \int_{0,5}^{1} 3x^2 dx \\ &= 7/8 \end{aligned}$$

A figura 5.3 indica a área sob a curva referente a esta probabilidade.

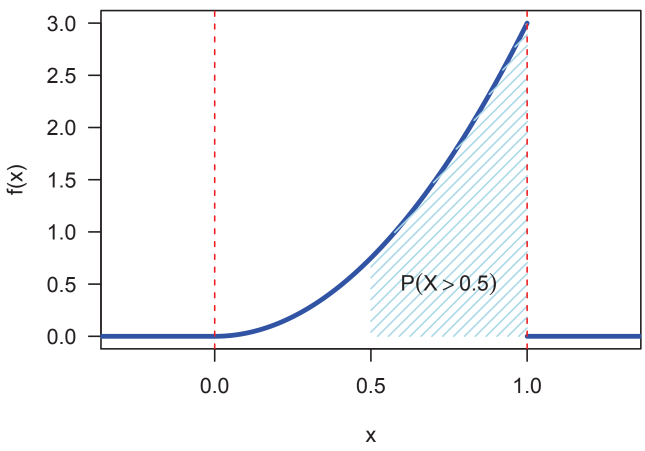

Figura 5.3: $P(X > 0,5)$.

Calcule $P(0,3 \leq X < 0.5)$.

Observações

- Se X é variável aleatória contínua no intervalo $[a, b]$, então $P_X(X = x) = 0 \ \forall \ x \in [a, b]$, isto é, probabilidades pontuais são consideradas nulas. Assim, por exemplo, $P(X \geq 2) = P(X > 2)$.

- Diferentemente do caso discreto, a f.d.p. não determina probabilidades diretamente, como ocorre com a f.p.

Média e variância

O valor esperado ou média e a variância de uma variável aleatória contínua são obtidos fazendo

$$\begin{array}{rcl} E(X) & = & \int_{-\infty}^{\infty} x f_X(x)dx \\ Var(X) & = & E(X^2) - [E(X)]^2 \\ & = & \left[\int_{-\infty}^{\infty} x^2 f_X(x)dx\right] - [E(X)]^2 \end{array}$$

Com $f(x)$ dada no exemplo anterior, temos que:

$$\begin{array}{rcl} E(X) & = & \int_{-\infty}^{\infty} x f_X(x)dx \\ & = & \int_0^1 x3x^2dx \\ & = & 3/4 \end{array}$$

e a variância de X:

$$\begin{array}{rcl} Var(X) & = & E(X^2) - [E(X)]^2 \\ & = & \left[\int_0^1 x^2 3x^2 dx\right] - [3/4]^2 \\ & = & 3/5 - 9/16 \\ & = & 3/80 \end{array}$$

5.4 Exercícios

1) Um pesquisador amostrou 10 plantas de alho da área útil de cada uma das parcelas de um experimento para realizar a contagem de plantas apresentando superbrotamento. Seja X uma v.a.d. denotada pelo número de plantas que apresentaram superbrotamento, conforme distribuição de probabilidades apresentada na tabela a seguir:

X_i	0	1	2	3	4	5	6	7	8	9	10
$P(X_i)$	0,20	0,30	0,20	0,15	0,07	0,05	0,02	0,01	0,00	0,00	0,00

Obter:

a. $P(X = 0)$

b. $P(X > 0)$

c. $P(X \leq 2)$

d. $P(1 \leq X < 5)$

e. $E(X)$

f. $E(X^2)$

g. $Var(X)$

h. $Mo(X)$, a moda de X

2) Suponha que $f(x) = 1,5x^2$ seja uma função densidade de probabilidade (f.d.p.) para $-1 < X < 1$. Determine:

a. $P(X > 0)$

b. $P(-0,5 \leq X \leq 0,5)$

c. O valor de k tal que $P(X < k) = 0,05$

3) A função densidade de probabilidade do peso líquido X, em libras, de um pacote de herbicida químico é $f(x) = 2$ para $49,75 \leq X \leq 50,25$ libras. A empresa empacotadora decide verificar a precisão com que os pacotes estão sendo distribuídos e, para isso, amostra 50 pacotes. Qual é a quantidade estimada de pacotes pesando mais que 50 libras?

4) O acréscimo anual na área atingida por certa praga numa certa região produtora de frutas pode ser modelado por uma variável aleatória contínua, medida em hectares, com densidade de probabilidade:

$$f(x) = \begin{cases} \frac{2}{3}x, & \text{se } 0 < x < 1 \text{ ha} \\ 1 - \frac{x}{3}, & \text{se } 1 \leq x < 3 \text{ ha} \\ 0, & \text{para outros valores de x} \end{cases}$$

Pede-se:

a. Qual é a probabilidade de a praga atingir mais que 2 hectares neste ano?

b. Qual é o tamanho da área à ser atingida com 50% de certeza?

c. Qual é o valor médio do acréscimo (X), em hectares?

d. Qual é o desvio padrão de X?

5) Admita que uma a v.a.c. Y que representa o diâmetro do colmo de uma certa variedade de cana-de-açúcar tem a seguinte função densidade de probabilidade (f.d.p.):

$$f(y) = \begin{cases} ky^2, & \text{se } 2 \leq y < 5 \\ k(8-y), & \text{se } 5 \leq y \leq 8 \\ 0, & \text{para outros valores de y} \end{cases}$$

Pede-se:

a. O valor da constante k para que $f(y)$ seja uma f.d.p.

b. $P(2 \leq Y < 4)$

c. $P(4 \leq Y < 6)$

d. $P(Y \geq 6)$

e. $E(Y)$

f. $E(Y^2)$

g. $Var(Y)$

Respostas

1) **a.** 0,20 **b.** 0,80 **c.** 0,70 **d.** 0,72 **e.** 1,87 **f.** 6,03 **g.** 2,5331 **h.** 1
2) **a.** 1/2 **b.** 1/8 **c.** $\approx -0,965$
3) 25 pacotes
4) **a.** 1/6 **b.** $\approx 1,268$ **c.** $\approx 1,33$ **d.** $\approx 0,39$
5) **a.** 6/261