Capítulo 6

Modelos básicos de distribuição de probabilidade

Muitas variáveis aleatórias, discretas e contínuas, podem ser descritas por modelos de probabilidade já conhecidos. Tais modelos permitem não só o cálculo de probabilidades, mas também a determinação de medidas como média e variância dessa variável. Estes modelos são funções dos valores assumidos pela variável aleatória e de determinado parâmetros (medidas populacionais) desta variável.

Estudaremos neste capítulo alguns dos principais modelos de distribuição de probabilidades de variáveis discretas e contínuas.

6.1 Modelos discretos

6.1.1 Distribuição Binomial

Considere uma variável aleatória discreta X que represente o número de sucessos (resultado de interesse) obtidos em N ensaios independentes cujos resultados são do tipo sucesso/fracasso (1 ou 0), também conhecidos como ensaios de Bernoulli. Então, diz-se que X segue distribuição binomial com a seguinte função de probabilidade

$$P_X(X = x|N, \pi) = \frac{N!}{x!(N-x)!}\pi^x(1-\pi)^{N-\pi}, \quad \begin{cases} x = 0, 1, 2, ..., N \\ 0 \leq \pi \leq 1 \end{cases}$$

Utiliza-se a notação $X \sim B(N, \pi)$, em que N e π são os parâmetros da distribuição, sendo N o número total de ensaios e π a probabilidade de sucesso, que é assumida ser a mesma em cada ensaio independente de Bernoulli.

Média e variância

$$
\begin{aligned}
E(X) &= N\pi \\
Var(X) &= N\pi(1-\pi)
\end{aligned}
$$

Aplicações

Em experimentos agrícolas envolvendo testes de produtos fitossanitários há o interesse não só de verificar se há parasitas presentes na planta após a aplicação do produto, mas também em saber quantos estão mortos e quantos ainda estão vivos. É o caso dos estudos de dose letal, frequentemente realizados em experimentos entomológicos cuja variável resposta é, comumente, um dado de proporção que pode ser modelado pela distribuição binomial.

Dados de proporção, isto é, números de sucesso e de fracassos de um total N de ensaios são muito comuns também em estudos de ecologia. Por exemplo, pesquisadores podem estar interessados em determinar a proporção de plantas de um determinado gênero em N diferentes locais de uma reserva de mata.

6.1.2 Distribuição Poisson

A distribuição Poisson tem apenas um parâmetro, λ, as vezes chamado de parâmetro de intensidade ou de taxa de ocorrência por unidade de medida. Uma variável aleatória discreta X tem distribuição Poisson com parâmetro λ, denotada por $X \sim Poisson(\lambda)$, se

$$
P_X(X=x|\lambda) = \frac{\exp(-\lambda)\lambda^x}{x!}, \quad \begin{cases} x = 0, 1, 2, ... \\ \lambda > 0 \end{cases}
$$

Média e variância

$$
E(X) = Var(X) = \lambda
$$

Aplicações

O modelo Poisson tem um importante papel na análise de dados em forma de contagens, e suas principais características são:

1. Proporciona, em geral, uma descrição satisfatória de dados cuja variância é proporcional à média.

2. Se eventos ocorrem independentemente e aleatoriamente no tempo, com taxa média de ocorrência constante (λ), o modelo determina o número de eventos em um intervalo de tempo especificado.

3. Aplica-se a dados de contagem, mas também pode ser usado na análise de dados contínuos que apresentam a característica descrita no ítem 1.

São exemplos de dados discretos que podem ser modelados pela distribuição Poisson: concentração de bactérias numa *placa de petri*, número de brotos por explante, número de plantas de uma determinada espécie por área, número de acidentes por dia, número de peças defeituosas produzidas por hora, entre outros.

Note que a unidade de medida (tempo, área, etc.) é contínua, mas a variável aleatória em si é discreta. Assim, uma variável Poisson surge, geralmente, de contagens realizadas no tempo ou no espaço.

6.2 Modelos contínuos

Nesta seção, estudaremos apenas a distribuição Normal, pelo fato de ser a mais importante das distribuições.

6.2.1 Distribuição Normal

Se a variável aleatória contínua X tem distribuição normal ou gaussiana com parâmetro de posição μ e de escala σ^2, denotada por $X \sim N(\mu, \sigma^2)$, então sua função densidade de probabilidade é dada por

$$f_X(x|\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I_{(-\infty,\infty)}(x), \quad \mu \in \Re, \sigma^2 > 0.$$

A figura 6.1 ilustra a densidade da curva normal, que é simétrica em torno da média (μ).

Média e variância

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

A distribuição normal apresenta as seguintes características:

- A variável pode assumir qualquer valor real

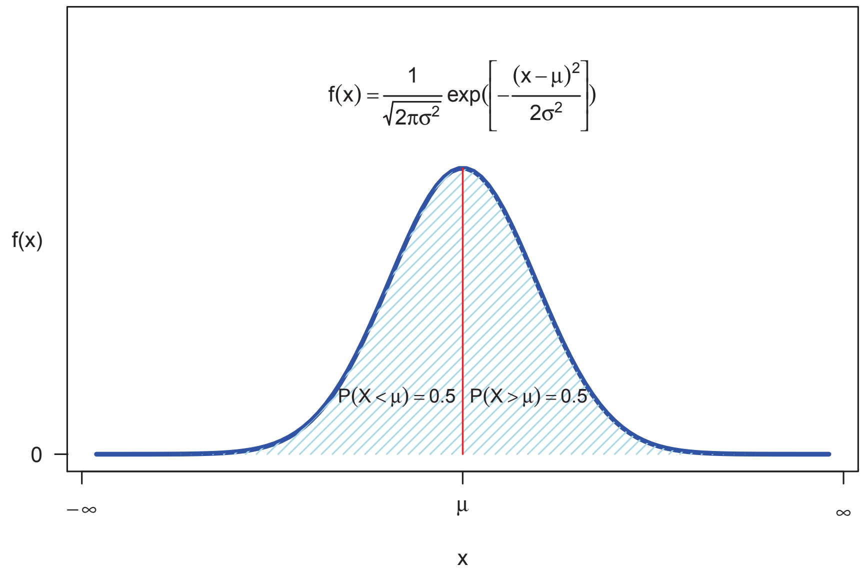

Figura 6.1: Densidade de probabilidade normal.

- A distribuição é em forma de sino, simétrica em torno da média (μ), como se observa na figura 6.1
- A área total sob a curva vale 1, pois esta é a probabilidade de a variável assumir qualquer valor real
- O formato da curva é determinado unicamente pelos dois parâmetros da distribuição, μ e σ^2. Alterando a média, alteramos a posição da distribuição, como mostra a figura 6.2A. Alterando a variância, alteramos a dispersão da distribuição, como mostra a figura 6.2B.

Distribuição Normal Padrão

A integral da curva normal não tem solução analítica. Assim sendo, os cálculos são, em geral, baseados na *distribuição normal padrão*, de acordo com a transformação

$$Z = \frac{X - \mu}{\sigma}$$

de modo que $Z \sim N(0, 1)$, isto é, Z tem média zero e variância unitária. Esse resultado é válido para quaisquer valores de μ e σ^2. Isso permite a construção de tabelas contendo probabilidades em função dos valores de Z, do tipo

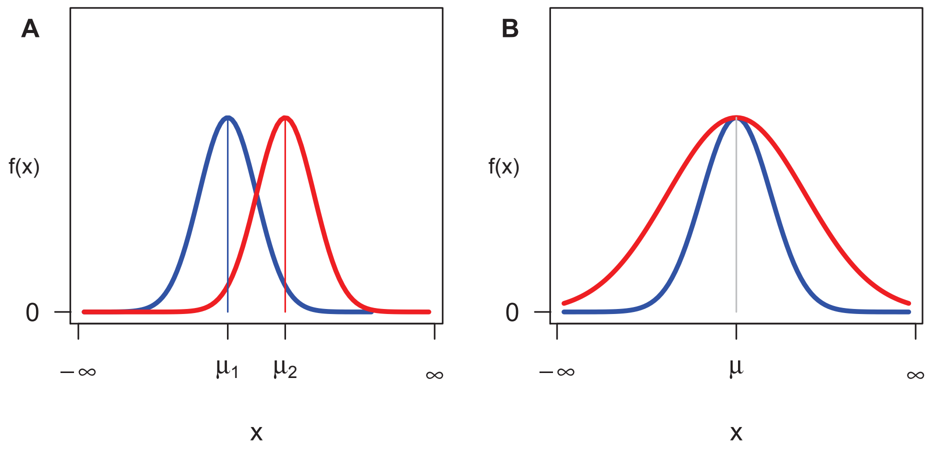

Figura 6.2: Densidades de probabilidade normal com (A) médias diferentes e mesma variância e (B) médias iguais e variâncias diferentes.

$$P(X \leq x) = P(Z \leq z)$$

A tabela da distribuição normal padrão que usaremos neste curso nos fornece a probabilidade da variável Z assumir um valor entre zero e um particular valor z_0, isto é,

$$P(0 \leq Z \leq z_0)$$

como mostrado na figura 6.3.

Aplicações

A distribuição normal é, certamente, a mais importante das distribuições probabilísticas, devido a quatro principais motivos:

1. Devido ao papel central que desempenha nas técnicas de inferência.

2. Outras importantes distribuições amostrais são derivadas da distribuição normal, tais como: qui-quadrado, F e t-Student.

3. Há ainda o *teorema central do limite* (TCL), que permite aproximações de vários outros modelos, contínuos e discretos, em grandes amostras. São exemplos clássicos as aproximações dos modelos binomial e Poisson.

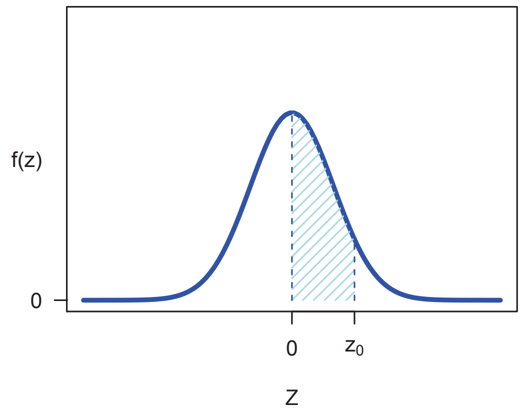

Figura 6.3: Densidade de probabilidade normal padrão e área sombreada indicando $P(0 \leq Z \leq z_0)$.

6.3 Alguns resultados úteis

Corolário 1. *Se $X \sim B(n, \pi)$, então assintoticamente ($n \to \infty$), tem-se que*

$$X \sim Poisson(\lambda = n\pi)$$

Corolário 2. *Seja $X \sim Binomial(n, \pi)$. Pelo teorema central do limite, para n grande e π não muito extremo (próximo de 0 ou 1), tem-se que*

$$X \sim Normal(\mu = n\pi, \sigma^2 = n\pi(1 - \pi))$$

Corolário 3. *Se $X \sim Normal(0, 1)$, então*

$$\mu + \sigma X \sim Normal(\mu, \sigma^2)$$

6.4 Exercícios

1) (Binomial) Determine a probabilidade de que, em 5 lançamentos de um dado, apareça a face 3...

a. ...duas vezes.

b. ...no máximo uma vez.

c. ...pelo menos duas vezes.

2) (Binomial) Considere a amostragem de 3 peças que saem de uma linha de produção. Estima-se que são produzidas 20% de peças defeituosas. Calcule as seguintes probabilidades:

a. De 2 das peças serem defeituosas.

b. De 2 das peças serem não defeituosas.

3) (Binomial) Uma firma determina o sexo de pintos de um dia com 95% de probabilidade de acerto.

a. Se comprarmos 5 pintos tidos como do sexo feminino, qual é a probabilidade de que pelo menos um seja macho?

b. Quantos pintos machos espera-se encontrar num lote de 500 pintinhos tidos como do sexo feminino.

4) (Poisson) Na pintura de paredes aparecem defeitos na proporção média de um defeito por metro quadrado. Qual é a probabilidade de aparecerem 3 defeitos numa parede de 2×2 m?

5) (Poisson) O departamento de trânsito registrou num certo ano, numa determinada via pública, 30 acidentes fatais. Qual é a probabilidade de que num determinado **mês** do próximo ano ocorram 3 acidentes fatais?

6) (Normal) A observação dos pesos (X) de um grande número de espigas de milho apresentou distribuição normal com média $\mu = 120$g e desvio padrão $\sigma = 10$g. Num programa de melhoramento genético, entre outras características, uma linhagem deve satisfazer a condição: $112 < X < 140$g. Num programa envolvendo 450 linhagens, qual deve ser o número provável de linhagens que atende à essa condição?

7) (Normal) Sabe-se que o peso médio, em arrobas, de abate de bovinos é normalmente distribuído com média 18 e variância 2,25. Um lote de 5000 cabeças, com essa característica, foi destinado ao frigorífico que abate só a partir de um peso mínimo W. Sabendo-se que foram abatidas 4200 cabeças, pede-se:

a. O número esperado de bovinos com peso entre 17 e 19 arrobas.

b. O valor de W.

8) As notas de uma prova são normalmente distribuídas com média 73 e variância 225. Os 15% melhores alunos recebem o conceito A e os 11,9% piores alunos recebem conceito R. Determine:

a. A nota mínima para receber conceito A.

b. A nota mínima para ser aprovado.

c. $P(X \geq 55,3)$.

9) Se $X \sim Normal(3,4)$, calcule o valor de k tal que $P(X \geq k) = 2P(X \leq k)$.

10) O diâmetro de um cabo elétrico é normalmente distribuído com média 0,8 mm e variância 0,0004 mm^2. Dentre uma amostra de 1000 cabos, quantos esperamos que tenha diâmetro...

a. ... maior ou igual a 0,81 mm?

b. ... entre 0,73 e 0,86 mm?

c. ... menor que 0,78 mm?

11) Um pesquisador decidiu que, para facilitar a classificação das aves em experimentos de nutrição, deve-se dividir as poedeiras no início da postura em três grupos **equiprováveis** de peso, a saber: poedeiras pesadas, poedeiras médias, poedeiras leves. Encontre os pesos correspondentes à cada classe, sabendo-se que o peso médio das aves nessa idade é 1,5 kg, e o desvio padrão é de 0,17 kg. Asmita a distribuição normal para a variável peso.

Respostas

1) **a**. $625/3888 \approx 0,161$ **b**. $3125/3888 \approx 0,803$ **c**. $763/3888 \approx 0,196$
2) **a**. 0,096 **b**. 0,384
3) **a**. 0,2263 **b**. 25
4) 0,1953
5) 0,2138
6) 345
7) **a**. 2486 **b**. 16,52
8) **a**. 88,66 **b**. 55,3 **c**. 0,881
9) 2,14
10) **a**. 308,5 $\approx$ 309 cabos **b**. 998,7 $\approx$ 999 cabos **c**. 158,7 $\approx$ 159 cabos
11) Leves: peso $<$ 1,43 kg; Médias: $1,43 \leq \text{peso} \leq 1,57$; Pesadas: peso $>$ 1,57