

INFERÊNCIA ESTATÍSTICA

TRADUÇÃO DA 2ª EDIÇÃO
NORTE-AMERICANA

GEORGE CASELLA

ROGER L. BERGER

Famílias comuns de distribuições

“Como todos esses aspectos incomuns afetam você, Watson?”

“O seu efeito cumulativo certamente é considerável e, mesmo assim, cada um deles é bastante possível em si mesmo.”

SHERLOCK HOLMES E DR. WATSON (*A aventura do Abbey Grange*)

3.1 INTRODUÇÃO

Distribuições estatísticas são utilizadas para modelar populações; deste modo, geralmente lidamos com uma *família* de distribuições em vez de uma única. Esta família é indexada por um ou mais parâmetros, o que nos permite variar certas características da distribuição, ao mesmo tempo em que permanece com uma forma funcional. Por exemplo, podemos especificar que a distribuição normal é uma opção razoável para modelar uma população em particular, mas não podemos especificar precisamente a média. Então, lidamos com uma família paramétrica de distribuições normais com média μ , onde μ é um parâmetro não especificado, $-\infty < \mu < \infty$.

Neste capítulo catalogamos muitas das distribuições estatísticas mais comuns, algumas das quais encontramos anteriormente. Para cada distribuição daremos sua média e variância e muitas outras medidas úteis ou descritivas, que podem ajudar no entendimento deste tópico. Também indicaremos algumas aplicações típicas dessas distribuições e inter-relações interessantes e úteis. Alguns destes fatos são resumidos em tabelas no final do livro. Este capítulo não pretende, de modo algum, ser abrangente em sua abordagem das distribuições estatísticas. Esta tarefa foi realizada por Johnson e Kotz (1969 – 1972) em seu trabalho composto de vários volumes, *Distribuições e estatística*, e nos volumes atualizados por Johnson, Kotz e Balakrishnan (1994, 1995) e Johnson, Kotz e Kemp (1992).

3.2 DISTRIBUIÇÕES DISCRETAS

Uma variável aleatória X é considerada como tendo uma distribuição discreta se o conjunto de valores de X , o espaço amostral, for contável. Na maioria das situações, a variável aleatória tem como resultados números inteiros.

DISTRIBUIÇÃO UNIFORME DISCRETA

Uma variável aleatória X tem *distribuição uniforme discreta* $(1, N)$ se

$$P(X = x|N) = \frac{1}{N}, \quad x = 1, 2, \dots, N, \quad (3.2.1)$$

onde N é um número inteiro especificado. Esta distribuição coloca massa igual em cada um dos resultados $1, 2, \dots, N$.

Uma observação sobre notação: quando estamos lidando com distribuições paramétricas, como quase sempre será o caso, a distribuição é dependente de valores dos parâmetros. A fim de enfatizar este fato e manter o controle dos parâmetros, os escrevemos na fp precedida por um "I" (dado que). Esta convenção também será utilizada com fdas, fdps, expectâncias e outros lugares onde possa ser necessário manter o controle dos parâmetros. Quando não existe possibilidade de confusão, os parâmetros podem ser omitidos, a fim de não complicar muito a notação.

Para calcular a média e a variância de X , lembre-se das identidades (prováveis pela indução)

$$\sum_{i=1}^k i = \frac{k(k+1)}{2} \quad \text{e} \quad \sum_{i=1}^k i^2 = \frac{k(k+1)(2k+1)}{6}.$$

Então, temos

$$EX = \sum_{x=1}^N xP(X = x|N) = \sum_{x=1}^N x \frac{1}{N} = \frac{N+1}{2}$$

e

$$EX^2 = \sum_{x=1}^N x^2 \frac{1}{N} = \frac{(N+1)(2N+1)}{6},$$

e, assim,

$$\begin{aligned} \text{Var } X &= EX^2 - (EX)^2 \\ &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{(N+1)(N-1)}{12}. \end{aligned}$$

Esta distribuição pode ser generalizada de modo que o espaço amostral seja qualquer amplitude de números inteiros, $N_0, N_0 + 1, \dots, N_1$, com fp $P(X = x|N_0, N_1) = 1/(N_1 - N_0 + 1)$.

DISTRIBUIÇÃO HIPERGEOMÉTRICA

A distribuição hipergeométrica tem muitas aplicações em amostragem de população finita e é mais bem compreendida por meio do exemplo clássico do modelo de urna.

Suponha que temos uma grande urna preenchida com N bolas que são idênticas em todos os aspectos, exceto pelo fato de que M são vermelhas e $N - M$ são verdes. Abrimos a urna com os olhos vendados e selecionamos K aleatoriamente (as K bolas são coletadas todas de uma vez; um caso de amostragem sem reposição). Qual é a probabilidade de que exatamente x das bolas sejam vermelhas?

O número total de amostras de tamanho K , que podem ser retiradas das N bolas, é $\binom{N}{K}$, como foi discutido na Seção 1.2.4. É exigido que x das bolas sejam vermelhas, e isso pode ser obtido de $\binom{M}{x}$ maneiras, deixando $\binom{N-M}{K-x}$ maneiras de completar a amostra com $K - x$ bolas verdes. Assim, se considerarmos que X denota o número de bolas vermelhas em uma amostra de tamanho K , então X tem uma *distribuição hipergeométrica* dada por

$$P(X = x|N, M, K) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}, \quad x = 0, 1, \dots, K. \quad (3.2.2)$$

Observe que existe, implicitamente em (3.2.2), outra suposição sobre a variabilidade de X . Coeficientes binomiais da forma $\binom{n}{r}$ são definidos se $n \geq r$ e, assim, o conjunto de valores de X é também restringido pelo par de desigualdades

$$M \geq x \quad \text{e} \quad N - M \geq K - x,$$

que pode ser combinado como

$$M - (N - K) \leq x \leq M.$$

Em muitos casos K é pequeno em comparação a M e N , de modo que o intervalo $0 \leq x \leq K$ estará contido no intervalo acima e, portanto, será apropriado. Geralmente, é bastante difícil lidar com a fórmula para a função de probabilidade hipergeométrica. Na verdade, nem mesmo é comum verificar que

$$\sum_{x=0}^K P(X = x) = \sum_{x=0}^K \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}} = 1.$$

A distribuição hipergeométrica ilustra o fato de que, estatisticamente, lidar com populações finitas (N finito) é uma tarefa difícil.

A média da distribuição hipergeométrica é dada por

$$EX = \sum_{x=0}^K x \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}} = \sum_{x=1}^K x \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}. \quad (\text{somando } 0 \text{ em } x=0)$$

Para avaliar esta expressão, utilizamos as identidades (já encontradas na Seção 2.3)

$$\begin{aligned} x \binom{M}{x} &= M \binom{M-1}{x-1}, \\ \binom{N}{K} &= \frac{N}{K} \binom{N-1}{K-1}, \end{aligned}$$

e obtemos

$$EX = \sum_{x=1}^K \frac{M \binom{M-1}{x-1} \binom{N-M}{K-x}}{\frac{N}{K} \binom{N-1}{K-1}} = \frac{KM}{N} \sum_{x=1}^K \frac{\binom{M-1}{x-1} \binom{N-M}{K-x}}{\binom{N-1}{K-1}}.$$

Agora, podemos reconhecer a segunda soma como a soma das probabilidades para outra distribuição hipergeométrica com base nos valores de parâmetros $N-1$, $M-1$ e $K-1$. Isto pode ser visto claramente definindo $y = x-1$ e escrevendo

$$\begin{aligned} \sum_{x=1}^K \frac{\binom{M-1}{x-1} \binom{N-M}{K-x}}{\binom{N-1}{K-1}} &= \sum_{y=0}^{K-1} \frac{\binom{M-1}{y} \binom{(N-1)-(M-1)}{K-1-y}}{\binom{N-1}{K-1}} \\ &= \sum_{y=0}^{K-1} P(Y = y | N-1, M-1, K-1) = 1, \end{aligned}$$

onde Y é uma variável aleatória hipergeométrica com parâmetros $N-1$, $M-1$ e $K-1$. Portanto, para a distribuição hipergeométrica,

$$EX = \frac{KM}{N}.$$

Um cálculo similar, porém mais demorado, estabelecerá que

$$\text{Var } X = \frac{KM}{N} \left(\frac{(N-M)(N-K)}{N(N-1)} \right).$$

Observe as manipulações utilizadas aqui para calcular EX . A soma foi transformada em outra distribuição hipergeométrica com diferentes valores de parâmetros e, reconhecendo este fato, estamos em condições de somar a série.

EXEMPLO 3.2.1 Amostragem por aceitação

A distribuição hipergeométrica tem aplicação na amostragem por aceitação, como este exemplo ilustrará. Suponha que uma loja varejista compre mercadorias em lotes e que cada item possa ser considerado aceitável ou com defeito. Seja

$N = \#$ de itens em um lote,

$M = \#$ de produtos com defeito em um lote.

Então, calculamos a probabilidade de que uma amostra de tamanho K contenha x mercadorias com defeito. Para sermos específicos, suponha que um lote de 25 peças de máquinas sejam entregues e que uma peça é considerada aceitável somente se estiver dentro do limite de tolerância. Seleccionamos 10 peças e descobrimos que nove não apresentam nenhum defeito (todas as nove estão dentro do limite de tolerância). Qual é a probabilidade deste evento se houver 6 peças com defeito no lote de 25 peças? Aplicando a distribuição hipergeométrica com $N = 25$, $M = 6$, $K = 10$, temos

$$P(X=0) = \frac{\binom{6}{0} \binom{19}{10}}{\binom{25}{10}} = 0,028,$$

mostrando que nosso evento observado é bastante improvável se houver 6 (ou mais!) peças com defeito no lote.

DISTRIBUIÇÃO BINOMIAL

A distribuição binomial, uma das mais úteis distribuições discretas, é baseada na ideia de uma *Prova de Bernoulli* (em homenagem a James Bernoulli, um dos criadores da teoria da probabilidade), um experimento com dois, e somente dois, resultados possíveis. Uma variável aleatória X tem uma *distribuição de Bernoulli*(p) se

$$X = \begin{cases} 1 & \text{com probabilidade } p \\ 0 & \text{com probabilidade } 1-p, \end{cases} \quad 0 \leq p \leq 1. \quad (3.2.3)$$

O valor $X = 1$ geralmente é identificado como um "sucesso" e p é chamado de probabilidade de sucesso. O valor $X = 0$ é identificado como "fracasso". A média e a variância de uma variável aleatória de Bernoulli(p) são facilmente determinadas como sendo

$$\begin{aligned} EX &= 1p + 0(1-p) = p, \\ \text{Var } X &= (1-p)^2 p + (0-p)^2 (1-p) = p(1-p). \end{aligned}$$

Muitos experimentos podem ser modelados como uma sequência de provas de Bernoulli, sendo a mais simples o lançamento repetido de uma moeda; $p =$ probabilidade de sair cara, $X = 1$ se a moeda mostrar cara. Outros exemplos incluem jogos de apostas (por exemplo, em uma roleta, seja $X = 1$ se os dados caírem no vermelho, portanto, $p =$ probabilidade de ocorrer vermelho), pesquisas eleitorais ($X = 1$ se o candidato A receber um voto), e a incidência de uma doença ($p =$ probabilidade de que uma pessoa, aleatoriamente, seja infectada).

Se n idênticas provas de Bernoulli forem realizadas, defina os eventos

$$A_i = \{X = 1 \text{ na } i\text{-ésima tentativa}\}, \quad i = 1, 2, \dots, n.$$

Se assumirmos que os eventos A_1, \dots, A_n são uma sequência de eventos independentes (como no caso do lançamento da moeda), então é fácil derivar a distribuição do número total de sucesso em n provas. Defina uma variável aleatória Y por

$$Y = \text{o número total de sucessos em } n \text{ tentativas.}$$

O evento $\{Y = y\}$ ocorrerá somente se, de todos os eventos A_1, \dots, A_n , exatamente y deles ocorrer e necessariamente $n - y$ deles não ocorrer. Um resultado em particular (uma determinada ordenação de ocorrências e não ocorrências) das n provas de Bernoulli pode ser $A_1 \cap A_2 \cap A_3^c \cap \dots \cap A_{n-1} \cap A_n^c$. Isto tem a probabilidade de ocorrência

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3^c \cap \dots \cap A_{n-1} \cap A_n^c) &= pp(1-p) \cdot \dots \cdot p(1-p) \\ &= p^y(1-p)^{n-y}, \end{aligned}$$

onde utilizamos a independência dos A_i s neste cálculo. Observe que o cálculo não depende de *qual* conjunto de y A_i s ocorre, mas somente de que *algum* conjunto de y A_i s ocorra. Além disso, o evento $\{Y = y\}$ ocorrerá, não importando qual conjunto de y A_i s ocorra. Considerando tudo isto, vemos que uma sequência em particular de n provas com exatamente y sucessos tem probabilidade $p^y(1-p)^{n-y}$ de ocorrer. Uma vez que existem $\binom{n}{y}$ dessas sequências (o número de ordenações de y 1s e $n - y$ 0s), temos

$$P(Y = y | n, p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \dots, n,$$

e Y é chamado de uma *variável aleatória binomial* (n, p) .

A variável aleatória Y pode ser, de forma alternativa e equivalente, definida da seguinte maneira: em uma sequência de n provas de Bernoulli idênticas e independentes, cada uma delas com probabilidade de sucesso p , defina as variáveis aleatórias X_1, \dots, X_n por

$$X_i = \begin{cases} 1 & \text{com probabilidade } p \\ 0 & \text{com probabilidade } 1-p. \end{cases}$$

A variável aleatória

$$Y = \sum_{i=1}^n X_i$$

tem a distribuição binomial (n, p) .

O fato de que $\sum_{y=0}^n P(Y = y) = 1$ resulta do seguinte teorema geral.

TEOREMA 3.2.2

Teorema binomial

Para quaisquer números reais x e y e inteiros $n \geq 0$,

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i}. \quad (3.2.4)$$

PROVA:

Escreva

$$(x + y)^n = (x + y)(x + y) \cdot \dots \cdot (x + y)$$

e considere como o lado direito deve ser calculado. A partir de cada fator $(x + y)$ escolhemos um x ou y , e multiplicamos, conjuntamente, as n opções. Para cada $i = 0, 1, \dots, n$, o número de termos nos quais x aparece exatamente i vezes é $\binom{n}{i}$. Portanto, este termo não está na forma $\binom{n}{i} x^i y^{n-i}$, e o resultado é o que segue. ■

Se assumirmos que $x = p$ e $y = 1 - p$ em (3.2.4), obtemos

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

e verificamos que cada termo na soma é uma probabilidade binomial. Como outro caso especial, consideramos que $x = y = 1$ no Teorema 3.2.2 e obtemos a identidade

$$2^n = \sum_{i=0}^n \binom{n}{i}.$$

A média e a variância da distribuição binomial já foram derivadas nos Exemplos 2.2.3 e 2.3.5, de modo que não as repetiremos aqui. Para concluir, as estabelecemos. Se $X \sim \text{binomial}(n, p)$, então

$$EX = np, \quad \text{Var } X = np(1-p).$$

A fgm da distribuição binomial foi calculada no Exemplo 2.3.9. Isto é

$$M_X(t) = [pe^t + (1-p)]^n.$$

EXEMPLO 3.2.3 Probabilidades com dados

Suponha que estejamos interessados em descobrir a probabilidade de obter pelo menos um 6 em quatro lançamentos de um dado equilibrado. Este experimento pode ser modelado como uma seqüência de quatro provas de Bernoulli com probabilidade de sucesso $p = \frac{1}{6} = P$ (o dado mostre 6). Defina a variável aleatória X por

$X =$ número total de 6s em quatro lançamentos.

Então $X \sim \text{binomial}(4, \frac{1}{6})$ e

$$\begin{aligned} P(\text{em pelo menos um 6}) &= P(X > 0) = 1 - P(X = 0) \\ &= 1 - \binom{4}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^4 \\ &= 1 - \left(\frac{5}{6}\right)^4 \\ &= 0,518. \end{aligned}$$

Agora, consideremos outro jogo; jogue um par de dados 24 vezes e questione qual é a probabilidade de sair, pelo menos, um duplo 6. Isto, mais uma vez, pode ser modelado pela distribuição binomial com probabilidade de sucesso p , onde

$$p = P(\text{lançamento de um duplo 6}) = \frac{1}{36}.$$

Assim, se $Y =$ número de ocorrências de duplo 6 em 24 lançamentos, $Y \sim \text{binomial}(24, \frac{1}{36})$ e

$$\begin{aligned} P(\text{lançamento de um duplo 6}) &= P(Y > 0) \\ &= 1 - P(Y = 0) \\ &= 1 - \binom{24}{0} \left(\frac{1}{36}\right)^0 \left(\frac{35}{36}\right)^{24} \\ &= 1 - \left(\frac{35}{36}\right)^{24} \\ &= 0,491. \end{aligned}$$

Este é o cálculo que originalmente era feito no século XVIII, por Pascal, a pedido do apostador de Meré, que considerou que ambos os eventos tinham a mesma probabilidade. (Ele começou a acreditar que estava errado e passou a perder dinheiro na segunda aposta.)

DISTRIBUIÇÃO DE POISSON

Esta é uma distribuição discreta amplamente aplicada e pode servir como modelo para uma série de diferentes tipos de experimentos. Por exemplo, se estivermos modelando um fenômeno no qual esperamos por uma ocorrência (assim como esperamos um ônibus, ou clientes cheguem em um banco), o número de ocorrências em um determinado intervalo de tempo pode, algumas vezes, ser modelado pela distribuição de Poisson. Uma das suposições básicas sobre a qual esta distribuição é desenvolvida é a de que, para pequenos intervalos de tempo, a probabilidade de uma chegada é proporcional ao tempo de espera. Esta suposição torna este um modelo razoável para situações como aquelas indicadas anteriormente. Por exemplo, faz sentido assumir que quanto mais longa a espera, mais provável é que o cliente entre no banco. Veja a seção Itens Diversos para verificar uma abordagem mais formal a este respeito.

Outra área de aplicação é a das distribuições espaciais, onde, por exemplo, a distribuição de Poisson pode ser utilizada para modelar a distribuição de uma bomba que atinge uma área ou a distribuição de peixes em um lago.

A distribuição de Poisson tem um único parâmetro λ , às vezes chamado de parâmetro de intensidade. Uma variável aleatória X , assumindo valores nos números inteiros não negativos, tem uma *Distribuição de Poisson*(λ) se

$$P(X = x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots \quad (3.2.5)$$

Para verificar que $\sum_{x=0}^{\infty} P(X = x|\lambda) = 1$, lembre-se da expansão em série, de Taylor, de e^y ,

$$e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!}.$$

Assim,

$$\sum_{x=0}^{\infty} P(X = x|\lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

A média de X é facilmente verificada como sendo

$$\begin{aligned} EX &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \quad (\text{substituir } y = x - 1) \\ &= \lambda \end{aligned}$$

Um cálculo similar mostrará que

$$\text{Var } X = \lambda,$$

e, portanto, o parâmetro λ é a média e a variância da distribuição de Poisson.

A fgm também pode ser obtida por um cálculo direto, novamente seguindo a série de Taylor, de e^y . Temos

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

(Veja o Exercício 2.33 e o Exemplo 2.3.13.)

EXEMPLO 3.2.4 Tempo de espera

Como exemplo de uma aplicação de “espera para a ocorrência”, considere um operador de telefone que, em média, atende cinco chamadas a cada três minutos. Qual é a probabilidade de que não ocorrerão chamadas no minuto seguinte? E, pelo menos, duas chamadas?

Se considerarmos que X = número de chamadas em um minuto, então X tem uma distribuição de Poisson com $EX = \lambda = \frac{5}{3}$. Desse modo

$$\begin{aligned} P(\text{nenhuma chamada no minuto seguinte}) &= P(X = 0) \\ &= \frac{e^{-5/3} \left(\frac{5}{3}\right)^0}{0!} \\ &= e^{-5/3} = 0,189; \end{aligned}$$

$$\begin{aligned} P(\text{pelo menos duas chamadas no minuto seguinte}) &= P(X \geq 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - 0,189 - \frac{e^{-5/3} \left(\frac{5}{3}\right)^1}{1!} \\ &= 0,496. \end{aligned}$$

O cálculo das probabilidades, de Poisson, pode ser feito rapidamente notando a seguinte relação de recursão:

$$P(X = x) = \frac{\lambda}{x} P(X = x - 1), \quad x = 1, 2, \dots \quad (3.2.6)$$

Esta relação é facilmente provada escrevendo a fp da distribuição de Poisson. Relações similares são mantidas para outras distribuições discretas. Por exemplo, se $Y \sim \text{binomial}(n, p)$, então

$$P(Y = y) = \frac{(n - y + 1)p}{y(1 - p)} P(Y = y - 1). \quad (3.2.7)$$

As relações de recursão (3.2.6) e (3.2.7) podem ser utilizadas para estabelecer a aproximação de Poisson para a binomial, que já vimos na Seção 2.3, onde a aproximação foi justificada utilizando fgms. Defina que $\lambda = np$ e, se p for pequeno, podemos escrever

$$\frac{n - y + 1}{y} \frac{p}{1 - p} = \frac{np - p(y - 1)}{y - py} \approx \frac{\lambda}{y}$$

uma vez que, para p pequeno, os termos $p(y - 1)$ e py podem ser ignorados. Deste modo, para este nível de aproximação, (3.2.7) se torna

$$P(Y = y) = \frac{\lambda}{y} P(Y = y - 1), \quad (3.2.8)$$

que é a relação de recursão de Poisson. Para completar a aproximação, precisamos apenas estabelecer que $P(X = 0) \approx P(Y = 0)$, uma vez que todas as outras probabilidades seguirão a partir de (3.2.8). Agora

$$P(Y = 0) = (1 - p)^n = \left(1 - \frac{np}{n}\right)^n = \left(1 - \frac{\lambda}{n}\right)^n$$

definindo que $np = \lambda$. Lembre-se, da Seção 2.3, que para λ , $\lim_{n \rightarrow \infty} (1 - (\lambda/n))^n = e^{-\lambda}$ fixo, de modo que para n grande temos a aproximação

$$P(Y = 0) = \left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} = P(X = 0),$$

completando a aproximação da Poisson para a Binomial.

A aproximação é válida quando n é grande e p é pequeno, que é exatamente quando é mais útil, pois nos libera de calcular coeficientes binomiais e potências para grandes n .

EXEMPLO 3.2.5 Aproximação de Poisson

Um tipógrafo, em média, comete um erro a cada 500 palavras impressas. Uma página típica contém 300 palavras. Qual é a probabilidade de que não ocorram mais que dois erros em cinco páginas?

Se assumirmos que a impressão de uma palavra é uma prova de Bernoulli com probabilidade de sucesso $p = \frac{1}{500}$ (observe que estamos identificando um erro como sendo um "sucesso") e que as provas são independentes, então $X =$ número de erros em cinco páginas (1500 palavras) é binomial $(1500, \frac{1}{500})$. Portanto,

$$\begin{aligned} P(\text{não mais que dois erros}) &= P(X \leq 2) \\ &= \sum_{x=0}^2 \binom{1500}{x} \left(\frac{1}{500}\right)^x \left(\frac{499}{500}\right)^{1500-x} \\ &= 0,4230, \end{aligned}$$

que é um cálculo bastante complicado. Se utilizarmos a aproximação de Poisson com $\lambda = 1500\left(\frac{1}{500}\right) = 3$, temos

$$P(X \leq 2) \approx e^{-3} \left(1 + 3 + \frac{3^2}{2}\right) = 0,4232.$$

DISTRIBUIÇÃO BINOMIAL NEGATIVA

A distribuição binomial conta o número de sucessos em um número fixo de provas de Bernoulli. Suponha que, em vez disto, contamos o número de provas de Bernoulli necessários para obter um número fixo de sucessos. Esta última formulação leva à distribuição binomial negativa.

Em uma sequência de provas de Bernoulli independentes (p), seja a variável aleatória X denotando a prova na qual o r -ésimo sucesso ocorre, onde r é um número inteiro fixo. Então

$$P(X = x|r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots, \quad (3.2.9)$$

e dizemos que X tem uma *distribuição binomial*(r, p) *negativa*.

A derivação de (3.2.9) segue rapidamente a partir da distribuição binomial. O evento $\{X = x\}$ pode ocorrer somente se houver exatamente $r-1$ sucessos nas primeiras $x-1$ provas, e um sucesso na x -ésima prova. A probabilidade de $r-1$ sucessos em $x-1$ provas é a probabilidade binomial $\binom{x-1}{r-1} p^{r-1} (1-p)^{x-r}$, e com probabilidade p existe um sucesso na x -ésima prova. Multiplicar essas probabilidades resulta em (3.2.9).

A distribuição binomial negativa, algumas vezes, é definida em termos da variável aleatória $Y =$ número de fracassos antes do r -ésimo sucesso. Esta formulação é estatisticamente equivalente à que é dada acima, em termos de $X =$ prova na qual ocorre o r -ésimo sucesso,

uma vez que $Y = X - r$. Utilizando a relação entre Y e X , a forma alternativa da distribuição binomial negativa é

$$P(Y = y) = \binom{r + y - 1}{y} p^r (1 - p)^y, \quad y = 0, 1, \dots \quad (3.2.10)$$

A menos que seja notado de outra forma, quando nos referirmos à distribuição binomial (r, p) utilizaremos esta fp.

A distribuição binomial negativa recebe seu nome por causa da relação

$$\binom{r + y - 1}{y} = (-1)^y \binom{-r}{y} = (-1)^y \frac{(-r)(-r-1)(-r-2)\cdots(-r-y+1)}{(y)(y-1)(y-2)\cdots(2)(1)},$$

que é, na verdade, a equação de definição para coeficientes binomiais com números inteiros negativos (veja Feller, 1968, para uma abordagem completa). Substituir em (3.2.10) gera

$$P(Y = y) = (-1)^y \binom{-r}{y} p^r (1 - p)^y,$$

que tem uma notável semelhança com a distribuição binomial.

O fato de que $\sum_{y=0}^{\infty} P(Y = y) = 1$ não é fácil de ser verificado, mas segue a partir de uma extensão do Teorema Binomial, uma extensão que inclui expoentes negativos. Aqui, não nos aprofundaremos neste assunto. Uma excelente exposição sobre coeficientes binomiais pode ser encontrada em Feller (1968).

A média e a variância de Y podem ser calculadas utilizando-se técnicas similares às empregadas para a distribuição binomial:

$$\begin{aligned} EY &= \sum_{y=0}^{\infty} y \binom{r + y - 1}{y} p^r (1 - p)^y \\ &= \sum_{y=1}^{\infty} \frac{(r + y - 1)!}{(y-1)!(r-1)!} p^r (1 - p)^y \\ &= \sum_{y=1}^{\infty} r \binom{r + y - 1}{y-1} p^r (1 - p)^y. \end{aligned}$$

Agora, escrevemos $z = y - 1$, e a soma passa a ser

$$\begin{aligned} EY &= \sum_{z=0}^{\infty} r \binom{r + z}{z} p^r (1 - p)^{z+1} \\ &= r \frac{(1-p)}{p} \sum_{z=0}^{\infty} \binom{(r+1) + z - 1}{z} p^{r+1} (1-p)^z \quad (\text{o somando é fp binomial negativo}) \\ &= r \frac{(1-p)}{p}. \end{aligned}$$

Uma vez que a soma envolve todos os valores de uma distribuição binomial negativa $(r+1, p)$, ela é igual a 1. Um cálculo similar mostrará

$$\text{Var } Y = \frac{r(1-p)}{p^2}.$$

Existe uma reparametrização interessante e, às vezes, útil da distribuição binomial negativa, em termos de sua média. Se definirmos o parâmetro $\mu = r(1-p)/p$, então $EY = \mu$ e um pouco de álgebra mostrarão que

$$\text{Var } Y = \mu + \frac{1}{r} \mu^2.$$

A variância é uma função quadrática da média. Esta relação pode ser útil tanto para a análise de dados como para considerações teóricas (Morris, 1982).

A família binomial negativa de distribuições inclui a distribuição de Poisson como um caso limite. Se $r \rightarrow \infty$ e $p \rightarrow 1$, de modo que $r(1-p) \rightarrow \lambda, 0 < \lambda < \infty$, então

$$EY = \frac{r(1-p)}{p} \rightarrow \lambda,$$

$$\text{Var } Y = \frac{r(1-p)}{p^2} \rightarrow \lambda,$$

que concorda com a média de Poisson e a variância. Para demonstrar que a binomial negativa $(r, p) \rightarrow \text{Poisson}(\lambda)$, podemos mostrar que todas as probabilidades convergem. O fato de que as fgms convergem nos leva a esperar por isto (veja o Exercício 3.15).

EXEMPLO 3.2.6 Amostragem binomial inversa

Uma técnica conhecida como amostragem binomial inversa é útil na amostragem de populações biológicas. Se a proporção de indivíduos que apresentam uma determinada característica é p e fazemos a amostragem até observarmos r desses indivíduos, então, o número de indivíduos amostrados é uma variável aleatória binomial negativa.

Por exemplo, suponhamos que em uma população de moscas-da-fruta estamos interessados na proporção que tem asas vestigiais, e decidimos fazer a amostragem até encontrarmos 100 dessas moscas. A probabilidade de que teremos de examinar pelo menos N moscas é (utilizando (3.2.9))

$$P(X \geq N) = \sum_{x=N}^{\infty} \binom{x-1}{99} p^{100} (1-p)^{x-100}$$

$$= 1 - \sum_{x=100}^{N-1} \binom{x-1}{99} p^{100} (1-p)^{x-100}.$$

Para determinados p e N , podemos avaliar esta expressão a fim de determinar para quantas moscas-da-fruta provavelmente estaremos olhando. (Embora a avaliação seja problemática, o uso de uma relação de recursão agilizará este processo.)

O Exemplo 3.2.6 mostra que a distribuição binomial negativa pode, assim como a de Poisson, ser utilizada para modelar fenômenos nos quais estamos esperando por uma ocorrência. No caso da binomial negativa, estamos esperando por um número específico de sucessos.

DISTRIBUIÇÃO GEOMÉTRICA

A distribuição geométrica é a mais simples das distribuições de tempo de espera e é um caso especial da distribuição binomial negativa. Se definirmos $r = 1$ em (3.2.9) teremos

$$P(X = x|p) = p(1-p)^{x-1}, \quad x = 1, 2, \dots,$$

o que define a fp de uma *variável aleatória geométrica* X com probabilidade de sucesso p . X pode ser interpretado como a prova na qual ocorre o primeiro sucesso, de modo que estamos "esperando por um sucesso". O fato de que $\sum_{x=1}^{\infty} P(X = x) = 1$ segue a partir das propriedades da série geométrica. Para qualquer número a com $|a| < 1$,

$$\sum_{x=1}^{\infty} a^{x-1} = \frac{1}{1-a},$$

a qual já encontramos no Exemplo 1.5.4.

A média e a variância de X podem ser calculadas utilizando as fórmulas binomiais negativas e escrevendo $X = Y + 1$ para obter

$$EX = EY + 1 = \frac{1}{p} \quad \text{e} \quad \text{Var } X = \frac{1-p}{p^2}.$$

A distribuição geométrica tem uma propriedade interessante, conhecida como a propriedade “sem memória”. Para números inteiros $s > t$, este é o caso em que

$$P(X > s | X > t) = P(X > s - t); \quad (3.2.11)$$

isto é, a distribuição geométrica “se esquece” do que ocorreu. A probabilidade de ocorrer outros fracassos $s - t$, depois de já terem sido observados t fracassos, é a mesma que a probabilidade de observar $s - t$ fracassos no início da sequência. Em outras palavras, a probabilidade de se obter uma sequência de fracassos depende somente do tamanho da sequência, não de sua posição.

Para estabelecer (3.2.11), primeiro notamos que para qualquer número inteiro n ,

$$\begin{aligned} P(X > n) &= P(\text{sem sucesso em } n \text{ tentativas}) \\ &= (1-p)^n, \end{aligned} \quad (3.2.12)$$

e, assim,

$$\begin{aligned} P(X > s | X > t) &= \frac{P(X > s \text{ e } X > t)}{P(X > t)} \\ &= \frac{P(X > s)}{P(X > t)} \\ &= (1-p)^{s-t} \\ &= P(X > s - t). \end{aligned}$$

EXEMPLO 3.2.7 Tempos de falha

Algumas vezes, a distribuição geométrica é utilizada para modelar “tempo de vida” ou “tempo até a ocorrência da falha” de componentes. Por exemplo, se a probabilidade é de 0,001 de que uma lâmpada irá queimar em determinado dia, então a probabilidade de que ela irá durar pelo menos 30 dias é

$$P(X > 30) = \sum_{x=31}^{\infty} 0,001(1-0,001)^{x-1} = (0,999)^{30} = 0,970.$$

A propriedade sem memória da distribuição geométrica descreve uma propriedade de “ausência de envelhecimento” muito especial, que indica que a distribuição geométrica não é aplicável à modelagem de tempo de vida para o qual se espera que a probabilidade de falha aumente com o decorrer do tempo. Existem outras distribuições utilizadas para modelar vários tipos de envelhecimento; veja, por exemplo, Barlow e Proschan (1975).

3.3 DISTRIBUIÇÕES CONTÍNUAS

Nesta seção, discutiremos algumas das famílias mais comuns de distribuições contínuas, aquelas com nomes bem conhecidos. As distribuições mencionadas aqui não constituem, de forma alguma, todas as utilizadas em estatística. Na verdade, como foi visto na Seção 1.6, qualquer função integrável, não negativa, pode ser transformada em uma fdp.

DISTRIBUIÇÃO UNIFORME

A *distribuição uniforme* contínua é definida pela massa uniformemente espalhada sobre um intervalo $[a, b]$. Sua fdp é dada por

$$f(x|a, b) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{caso contrário.} \end{cases} \quad (3.3.1)$$

É fácil verificar que $\int_a^b f(x) dx = 1$. Também temos

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2};$$

$$\text{Var } X = \int_a^b \frac{\left(x - \frac{b+a}{2}\right)^2}{b-a} dx = \frac{(b-a)^2}{12}.$$

DISTRIBUIÇÃO GAMA

A família de distribuições gama é a família flexível de distribuições flexível em $[0, \infty)$ e pode ser derivada pela construção discutida na Seção 1.6. Se α é uma constante positiva, a integral

$$\int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

é finita. Se α é um número inteiro positivo, a integral pode ser expressa de forma fechada; do contrário, isto não é possível. Em qualquer caso, seu valor define a *função gama*,

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt. \quad (3.3.2)$$

A função gama satisfaz muitas relações úteis, em particular,

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0, \quad (3.3.3)$$

que pode ser verificada por meio da integração por partes. Combinando (3.3.3) com o fato facilmente verificado de que $\Gamma(1) = 1$, temos, para qualquer número inteiro $n > 0$,

$$\Gamma(n) = (n-1)!. \quad (3.3.4)$$

(Outro caso útil especial, que será visto em (3.3.15), é que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.)

As expressões (3.3.3) e (3.3.4) fornecem relações de recursão que facilitam os problemas para calcular valores da função gama. A relação de recursão nos permite calcular qualquer valor da função gama conhecendo somente os valores de $\Gamma(c)$, $0 < c \leq 1$.

Uma vez que o integrando em (3.3.2) é positivo, imediatamente segue que

$$f(t) = \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)}, \quad 0 < t < \infty, \quad (3.3.5)$$

é uma fdp. A família gama completa, no entanto, tem dois parâmetros e pode ser derivada mudando as variáveis para obter a fdp da variável aleatória $X = \beta T$ em (3.3.5), onde β é uma constante positiva. Ao fazermos isto, obtemos a *família gama* (α, β) ,

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0. \quad (3.3.6)$$

O α é conhecido como o parâmetro de forma, uma vez que é o que mais influencia o pico da distribuição, ao passo que β é chamado de parâmetro de escala, porque a maior parte de sua influência ocorre na dispersão da distribuição.

A média da distribuição $\text{gama}(\alpha, \beta)$ é

$$EX = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x x^{\alpha-1} e^{-x/\beta} dx. \quad (3.3.7)$$

Para avaliar (3.3.7), observe que o integrando é o núcleo de uma fdp $\text{gama}(\alpha + 1, \beta)$. A partir de (3.3.6) sabemos que, para qualquer $\alpha, \beta > 0$,

$$\int_0^\infty x^{\alpha-1} e^{-x/\beta} dx = \Gamma(\alpha)\beta^\alpha, \quad (3.3.8)$$

portanto, temos

$$\begin{aligned} EX &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^\alpha e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha + 1)\beta^{\alpha+1} \\ &= \frac{\alpha\Gamma(\alpha)\beta}{\Gamma(\alpha)} \quad (\text{a partir de (3.3.3)}) \\ &= \alpha\beta. \end{aligned}$$

Note que para avaliar EX novamente utilizamos a técnica de reconhecer a integral como o núcleo de outra fdp. (Já utilizamos esta técnica para calcular a fgm gama no Exemplo 2.3.8 e, em um caso discreto, para fazer cálculos binomiais, nos Exemplos 2.2.3 e 2.3.5.)

A variância da distribuição $\text{gama}(\alpha, \beta)$ é calculada de maneira análoga àquela utilizada para a média. Em particular, ao calcular EX^2 lidamos com o núcleo de uma distribuição $\text{gama}(\alpha + 2, \beta)$. O resultado é

$$\text{Var } X = \alpha\beta^2.$$

No Exemplo 2.3.8 calculamos a fgm de uma distribuição $\text{gama}(\alpha, \beta)$. Ela é dada por

$$M_X(t) = \left(\frac{1}{1 - \beta t} \right)^\alpha, \quad t < \frac{1}{\beta}.$$

EXEMPLO 3.3.1 Relação Gama-Poisson

Existe uma interessante relação entre as distribuições gama e Poisson . Se X é uma variável aleatória $\text{gama}(\alpha, \beta)$, onde α é um número inteiro, então, para qualquer x ,

$$P(X \leq x) = P(Y \geq \alpha), \quad (3.3.9)$$

onde $Y \sim \text{Poisson}(x/\beta)$. A Equação (3.3.9) pode ser estabelecida por sucessivas integrações por partes, como segue. Como α é um número inteiro, escrevemos $\Gamma(\alpha) = (\alpha - 1)!$ para obter

$$\begin{aligned} P(X \leq x) &= \frac{1}{(\alpha - 1)!\beta^\alpha} \int_0^x t^{\alpha-1} e^{-t/\beta} dt \\ &= \frac{1}{(\alpha - 1)!\beta^\alpha} \left[-t^{\alpha-1} \beta e^{-t/\beta} \Big|_0^x + \int_0^x (\alpha - 1)t^{\alpha-2} \beta e^{-t/\beta} dt \right], \end{aligned}$$

onde utilizamos a integração por substituição de partes $u = t^{\alpha-1}$, $dv = e^{-t/\beta} dt$. Continuando nossa avaliação, temos

$$\begin{aligned} P(X \leq x) &= \frac{-1}{(\alpha - 1)!\beta^{\alpha-1}} x^{\alpha-1} e^{-x/\beta} + \frac{1}{(\alpha - 2)!\beta^{\alpha-1}} \int_0^x t^{\alpha-2} e^{-t/\beta} dt \\ &= \frac{1}{(\alpha - 2)!\beta^{\alpha-1}} \int_0^x t^{\alpha-2} e^{-t/\beta} dt - P(Y = \alpha - 1), \end{aligned}$$

onde $Y \sim \text{Poisson}(x/\beta)$. Continuando desta maneira, podemos estabelecer (3.3.9). (Veja o Exercício 3.19.)

Existe uma série de casos especiais importantes da distribuição gama. Se definirmos $\alpha = p/2$, onde p é um número inteiro, e $\beta = 2$, então, a fdp gama passa a ser

$$f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 < x < \infty, \quad (3.3.10)$$

que é a fdp *qui-quadrada com p graus de liberdade*. A média, a variância e a fgm da distribuição qui-quadrada podem ser calculadas utilizando-se as fórmulas gama previamente derivadas.

A distribuição qui-quadrada representa um importante papel na inferência estatística, especialmente ao fazer a amostragem de uma distribuição normal. Este tópico será abordado no Capítulo 5.

Outro importante caso especial da distribuição gama é obtido quando definimos $\alpha = 1$. Então, temos

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 < x < \infty, \quad (3.3.11)$$

a fdp *exponencial* com parâmetro de escala β . Sua média e sua variância foram calculadas nos Exemplos 2.2.2 e 2.3.3.

A distribuição exponencial pode ser utilizada para modelar tempo de vida, de modo análogo ao uso da distribuição geométrica no caso discreto. Na verdade, a distribuição exponencial compartilha a propriedade "sem memória" da distribuição geométrica. Se $X \sim \text{exponencial}(\beta)$, isto é, com a fdp dada por (3.3.11), então para $s > t \geq 0$,

$$P(X > s | X > t) = P(X > s - t),$$

uma vez que

$$\begin{aligned} P(X > s | X > t) &= \frac{P(X > s, X > t)}{P(X > t)} \\ &= \frac{P(X > s)}{P(X > t)} && \text{(uma vez que } s > t) \\ &= \frac{\int_s^{\infty} \frac{1}{\beta} e^{-x/\beta} dx}{\int_t^{\infty} \frac{1}{\beta} e^{-x/\beta} dx} \\ &= \frac{e^{-s/\beta}}{e^{-t/\beta}} \\ &= e^{-(s-t)/\beta} \\ &= P(X > s - t). \end{aligned}$$

Outra distribuição relacionada às famílias exponencial e gama é a *distribuição de Weibull*. Se $X \sim \text{exponencial}(\beta)$, então $Y = X^{1/\gamma}$ tem uma distribuição de Weibull(γ, β),

$$f_Y(y|\gamma, \beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}, \quad 0 < y < \infty, \quad \gamma > 0, \quad \beta > 0. \quad (3.3.12)$$

Claramente, poderíamos ter começado com a distribuição de Weibull e, então, derivado a exponencial como um caso especial ($\gamma = 1$). É uma questão de gosto. A distribuição de Weibull representa um papel extremamente importante na análise dos dados referentes ao tempo para falha (veja Kalbfleisch e Prentice, 1980, para conhecer uma abordagem abrangente deste tópico). A distribuição de Weibull, em particular, é muito útil para a modelagem de *funções risco* (veja os Exercícios 3.25 e 3.26).

DISTRIBUIÇÃO NORMAL

A distribuição normal (algumas vezes chamada de *distribuição Gaussiana*) representa um papel central em um grande conjunto de estatísticas. Existem três principais razões para isto. Primeiro, a distribuição normal e as associadas a ela são muito tratáveis analiticamente (embora isto talvez não seja percebido à primeira vista). Segundo, a distribuição normal tem o familiar formato de um sino, cuja simetria a transforma em uma escolha atrativa para muitos modelos de populações. Embora existam muitas outras distribuições que também têm formato de sino, a maior parte delas não tem a tratabilidade analítica da distribuição normal. Terceiro, existe o Teorema do Limite Central (veja o Capítulo 5 para saber mais detalhes), o que mostra que, em condições moderadas, a distribuição normal pode ser utilizada para aproximar uma grande variedade de distribuições em grandes amostras.

A distribuição normal tem dois parâmetros, geralmente denotados por μ e σ^2 , que são sua média e sua variância. A fdp da *distribuição normal* com média μ e variância σ^2 (geralmente denotada por $n(\mu, \sigma^2)$) é dada por

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty. \quad (3.3.13)$$

Se $X \sim n(\mu, \sigma^2)$, então a variável aleatória $Z = (X - \mu)/\sigma$ tem uma distribuição $n(0,1)$, também conhecida como a *normal padrão*. Isto é facilmente estabelecido escrevendo

$$\begin{aligned} P(Z \leq z) &= P\left(\frac{X - \mu}{\sigma} \leq z\right) \\ &= P(X \leq z\sigma + \mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z\sigma + \mu} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt, \quad \left(\text{substituir } t = \frac{x - \mu}{\sigma}\right) \end{aligned}$$

mostrando que $P(Z \leq z)$ é a fda normal padrão.

Portanto, segue que todas as probabilidades normais podem ser calculadas em termos da normal padrão. Além disso, cálculos de valores esperados podem ser simplificados desenvolvendo os detalhes no caso $n(0,1)$, transformando, então, o resultado para o caso $n(\mu, \sigma^2)$. Por exemplo, se $Z \sim n(0,1)$,

$$EZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = -\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Big|_{-\infty}^{\infty} = 0,$$

e, portanto, se $X \sim n(\mu, \sigma^2)$, segue que, a partir do Teorema 2.2.5,

$$EX = E(\mu + \sigma Z) = \mu + \sigma EZ = \mu.$$

De modo similar, temos que $\text{Var } Z = 1$ e, a partir do Teorema 2.3.4, $\text{Var } X = \sigma^2$.

Ainda não estabelecemos que a integral de (3.3.13) é igual a 1 ao longo de toda a reta real. Aplicando a transformação de padronização, precisamos somente mostrar que

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1.$$

Observe que o integrando acima é simétrico em torno de 0, implicando que a integral sobre $(-\infty, 0)$ é igual à integral sobre $(0, \infty)$. Portanto, reduzimos o problema mostrando

$$\int_0^{\infty} e^{-z^2/2} dz = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}. \quad (3.3.14)$$

A função $e^{-z^2/2}$ não tem uma antiderivada que possa ser escrita explicitamente em termos de funções elementares (isto é, em forma fechada), assim, não podemos realizar a integração diretamente. Na verdade, este é um exemplo de uma integração que, se você não souber como fazer, perderá muito tempo sem "ir a lugar nenhum". Uma vez que ambos os lados de (3.3.14) são positivos, a igualdade se manterá se estabelecermos que os quadrados são iguais. Definimos o quadrado da integral em (3.3.14) para obter

$$\begin{aligned} \left(\int_0^\infty e^{-z^2/2} dz \right)^2 &= \left(\int_0^\infty e^{-t^2/2} dt \right) \left(\int_0^\infty e^{-u^2/2} du \right) \\ &= \int_0^\infty \int_0^\infty e^{-(t^2+u^2)/2} dt du. \end{aligned}$$

As variáveis de integração são apenas variáveis aparentes, por isso é permitido mudar seus nomes. Agora, convertemos para coordenadas polares. Defina

$$t = r \cos \theta \quad e \quad u = r \sin \theta.$$

Então $t^2 + u^2 = r^2$ e $dt du = r d\theta dr$, e os limites de integração passam a ser $0 < r < \infty$, $0 < \theta < \pi/2$ (o limite superior em θ é $\pi/2$ porque t e u têm a restrição de serem positivos). Agora, temos

$$\begin{aligned} \int_0^\infty \int_0^\infty e^{-(t^2+u^2)/2} dt du &= \int_0^\infty \int_0^{\pi/2} r e^{-r^2/2} d\theta dr \\ &= \frac{\pi}{2} \int_0^\infty r e^{-r^2/2} dr \\ &= \frac{\pi}{2} \left[-e^{-r^2/2} \Big|_0^\infty \right] \\ &= \frac{\pi}{2}, \end{aligned}$$

que estabelece (3.3.14).

Esta integral está estritamente relacionada com a função gama; de fato, fazendo a substituição $w = \frac{1}{2}z^2$ em (3.3.14), verificamos que esta integral é essencialmente $\Gamma(\frac{1}{2})$. Se tivermos o cuidado de obter as constantes corretas, veremos que (3.3.14) implica

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty w^{-1/2} e^{-w} dw = \sqrt{\pi}. \tag{3.3.15}$$

A distribuição normal é, de certo modo, especial, no sentido de que seus dois parâmetros, μ (a média) e σ^2 (a variância), nos fornecem informações completas sobre o formato e a localização exatos da distribuição. Esta propriedade, de que a distribuição é determinada por μ e σ^2 , não é única para a fdp normal, mas compartilhada por um grupo de fdps chamado de famílias de locação-escala, que serão discutidas na Seção 3.5.

O cálculo direto mostra que a fdp normal (3.3.13) tem seu máximo em $x = \mu$ e a pontos de inflexão (onde a curva se modifica de côncava para convexa) em $\mu \pm \sigma$. Além disso, o conteúdo de probabilidade dentro de 1, 2 ou 3 desvios padrão da média é

$$P(|X - \mu| \leq \sigma) = P(|Z| \leq 1) = 0,6826,$$

$$P(|X - \mu| \leq 2\sigma) = P(|Z| \leq 2) = 0,9544,$$

$$P(|X - \mu| \leq 3\sigma) = P(|Z| \leq 3) = 0,9974,$$

onde $X \sim n(\mu, \sigma^2)$, $Z \sim n(0, 1)$, e os valores numéricos podem ser obtidos a partir de muitos pacotes de programas de computador ou de tabelas. Geralmente, os valores de dois dígitos relatados são 0,68, 0,95 e 0,99, respectivamente. Embora não representem os valores arredondados, eles são comumente utilizados. A Figura 3.3.1 mostra a fdp normal juntamente com esses recursos principais.

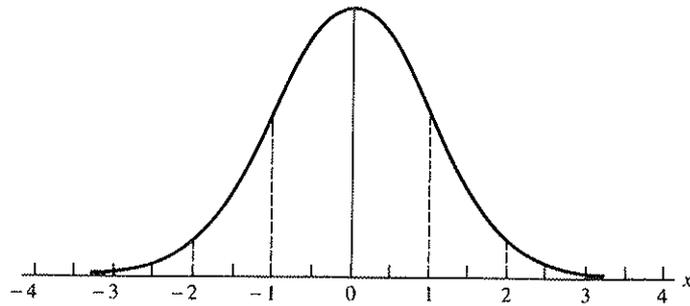


Fig. 3.3.1 Densidade normal padrão

Entre os muitos usos da distribuição normal, um importante é como uma aproximação de outras distribuições (o que é parcialmente justificado pelo Teorema do Limite Central). Por exemplo, se $X \sim \text{binomial}(n, p)$, então $EX = np$ e $\text{Var } X = np(1-p)$, e sob condições adequadas, a distribuição de X pode ser aproximada por aquela de uma variável aleatória normal com média $\mu = np$ e variância $\sigma^2 = np(1-p)$. As "condições adequadas" são que n deverá ser grande e p não deverá ser extremo (próximo de 0 ou 1). Queremos que n seja grande, de modo que existam valores (discretos) suficientes de X para fazer uma aproximação por uma distribuição contínua razoável, e p deverá estar "no meio", de modo que a binomial é praticamente simétrica, assim como a normal. Como acontece com a maioria das aproximações, não existem regras absolutas, e cada aplicação deverá ser verificada para decidir se a aproximação é suficientemente boa para seu uso pretendido. Uma regra de conservação a ser seguida é que a aproximação será boa, se $\min(np, n(1-p)) \geq 5$.

EXEMPLO 3.3.2 Aproximação normal

Seja $X \sim \text{binomial}(25, 0,6)$. Podemos aproximar X com uma variável aleatória normal, Y com média $\mu = 25(0,6) = 15$ e desvio padrão $\sigma = ((25)(0,6)(0,4))^{1/2} = 2,45$. Deste modo

$$P(X \leq 13) \approx P(Y \leq 13) = P\left(Z \leq \frac{13 - 15}{2,45}\right) = P(Z \leq -0,82) = 0,206,$$

ao passo que o cálculo binomial exato resulta em

$$P(X \leq 13) = \sum_{x=0}^{13} \binom{25}{x} (0,6)^x (0,4)^{25-x} = 0,267,$$

mostrando que a aproximação normal é boa, mas não chega a ser ótima. No entanto, a aproximação pode ser bastante melhorada por uma "correção de continuidade". Para ver como isto funciona, observe a Figura 3.3.2, que mostra a fp binomial $(25, 0,6)$ e a fdp $n(15, (2,45)^2)$. Obtemos a fp binomial utilizando barras com largura 1, com altura igual à probabilidade. Portanto, as áreas das barras dão as probabilidades binomiais. Na aproximação, observe como a área da aproximação normal é menor que a área binomial (a normal inclui tudo à esquerda da linha em 13, ao passo que a binomial inclui toda a barra, de 13 até 13,5). A correção de continuidade acrescenta novamente esta área adicionando $\frac{1}{2}$ ao ponto de corte. Assim, em vez de aproximar $P(X \leq 13)$, aproximamos a expressão equivalente (por causa da singularidade) $P(X \leq 13,5)$ e obtemos

$$P(X \leq 13) = P(X \leq 13,5) \approx P(Y \leq 13,5) = P(Z \leq -0,61) = 0,271,$$

uma aproximação muito melhor. Em geral, a aproximação normal com a correção de continuidade é muito superior à aproximação sem a correção de continuidade.

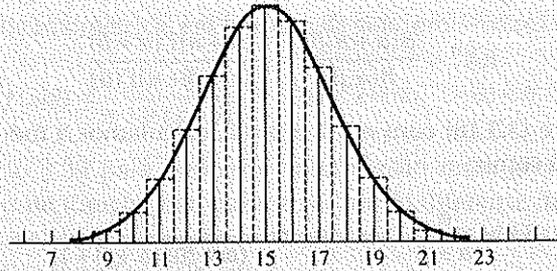


Fig. 3.3.2. Aproximação normal(15, (2,45)²) à binomial(25, 0,6)

Também fazemos a correção na extremidade inferior. Se $X \sim \text{binomial}(n, p)$ e $Y \sim n(np, np(1-p))$, então aproximamos

$$P(X \leq x) \approx P(Y \leq x + 1/2),$$

$$P(X \geq x) \approx P(Y \geq x - 1/2).$$

DISTRIBUIÇÃO BETA

A família beta de distribuições é uma família contínua com (0,1) indexada por dois parâmetros. A fdp beta(α, β) é

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0, \quad (3.3.16)$$

onde $B(\alpha, \beta)$ denota a função beta,

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx.$$

A função beta está relacionada à função gama por meio da seguinte identidade:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \quad (3.3.17)$$

A Equação (3.3.17) é muito útil ao lidar com a função beta, permitindo tirar proveito das propriedades da função gama. Na verdade, nunca lidaremos diretamente com a função beta, mas, em vez disso, utilizaremos (3.3.17) para todas as nossas avaliações.

A distribuição beta é uma das poucas distribuições "chamadas" comuns que atribuem probabilidade 1 para um intervalo finito, aqui assumido como sendo (0,1). Desta maneira, a beta geralmente é utilizada para modelar proporções, o que naturalmente está entre 0 e 1. Veremos ilustrações relacionadas a isto no Capítulo 4.

O cálculo de momentos da distribuição beta é bastante simples, devido à forma particular da fdp. Para $n > -\alpha$ temos

$$\begin{aligned} EX^n &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^n x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{(\alpha+n)-1} (1-x)^{\beta-1} dx. \end{aligned}$$

Agora, reconhecemos a integranda como o núcleo de uma fdp beta($\alpha + n, \beta$); portanto,

$$EX^n = \frac{B(\alpha + n, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + n)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)\Gamma(\alpha)}. \quad (3.3.18)$$

Utilizando (3.3.3) e (3.3.18) com $n = 1$ e $n = 2$, calculamos a média e a variância da distribuição beta(α, β) como

$$EX = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad \text{Var } X = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Como os parâmetros α e β variam, a distribuição beta assume muitas formas, como mostra a Figura 3.3.3. A fdp pode estar estritamente aumentando ($\alpha > 1, \beta = 1$), estritamente diminuindo ($\alpha = 1, \beta > 1$), em formato de U ($\alpha < 1, \beta < 1$), ou pode ser unimodal ($\alpha > 1, \beta > 1$). O caso $\alpha = \beta$ gera uma fdp simétrica sobre $\frac{1}{2}$ com média $\frac{1}{2}$ (necessariamente) e variância $(4(2\alpha + 1))^{-1}$. A fdp se torna mais concentrada à medida que α aumenta, mas permanece simétrica, como mostra a Figura 3.3.4. Por fim, se $\alpha = \beta = 1$, a distribuição beta se reduz à uniforme(0,1), mostrando que esta pode ser considerada como um membro da família beta. A distribuição beta também está relacionada, por meio de uma transformação, à distribuição F , uma distribuição que representa um papel extremamente importante na análise estatística (veja a Seção 5.3).

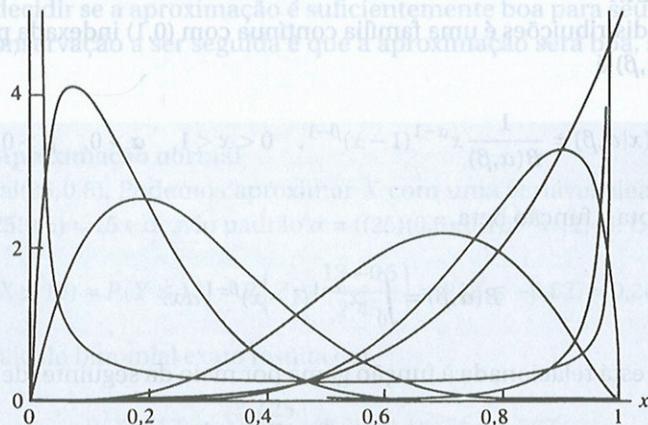


Fig. 3.3.3 Densidades beta

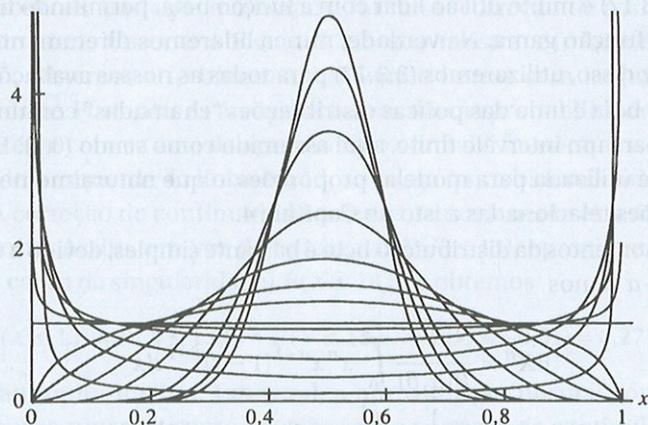


Fig. 3.3.4 Densidades beta simétricas

DISTRIBUIÇÃO DE CAUCHY

Esta é uma distribuição simétrica, em formato de sino, em $(-\infty, \infty)$ com fdp

$$f(x|\theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty. \quad (3.3.19)$$

(Veja o Exercício 3.39 para conhecer uma versão mais geral da fdp de Cauchy.) Visivelmente, a fdp de Cauchy não parece ser muito diferente da distribuição normal. Contudo, na realidade, existe uma grande diferença. Como já vimos no Capítulo 2, a média da distribuição de Cauchy não existe, isto é,

$$E|X| = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{|x|}{1 + (x - \theta)^2} dx = \infty. \quad (3.3.20)$$

É fácil verificar que (3.3.19) define uma fdp apropriada para todo θ . Lembre-se de que $\frac{d}{dt} \arctg(t) = (1 + t^2)^{-1}$; portanto,

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} dx = \frac{1}{\pi} \arctg(x - \theta) \Big|_{-\infty}^{\infty} = 1,$$

uma vez que $\arctg(\pm\infty) = \pm\pi/2$.

Como $E|X| = \infty$, segue que não existe nenhum momento da distribuição de Cauchy ou, em outras palavras, todos os momentos absolutos são iguais a ∞ . Em particular, a fgm não existe.

O parâmetro θ em (3.3.19) mede o centro da distribuição; ela é a mediana. Se X tem uma distribuição de Cauchy com parâmetro θ , então, a partir do Exercício 3.37 segue que $P(X \geq \theta) = \frac{1}{2}$, mostrando que θ é a mediana da distribuição. A Figura 3.3.5 mostra uma distribuição de Cauchy(0) juntamente com um $n(0,1)$, onde vemos a similaridade no formato, exceto pelas caudas muito mais espessas, de Cauchy.

A distribuição de Cauchy representa um papel especial na teoria da estatística. Ela demonstra um caso extremo, em relação ao qual conjecturas podem ser testadas. Mas não cometa o erro de considerar a distribuição de Cauchy como sendo somente um caso patológico, porque ela tem um jeito de se transformar quando menos se espera. Por exemplo, é prática comum para experimentadores calcular razões de observações, isto é, razões de variáveis aleatórias. (Em medidas de crescimento, é comum combinar peso e altura em uma única medição do peso em relação à altura, isto é, peso / altura.) Um fato surpreendente é que a razão de duas normais padrão tem uma distribuição de Cauchy (veja o Exemplo 4.3.6). Assumir proporções pode levar a distribuições com comportamentos anormais.

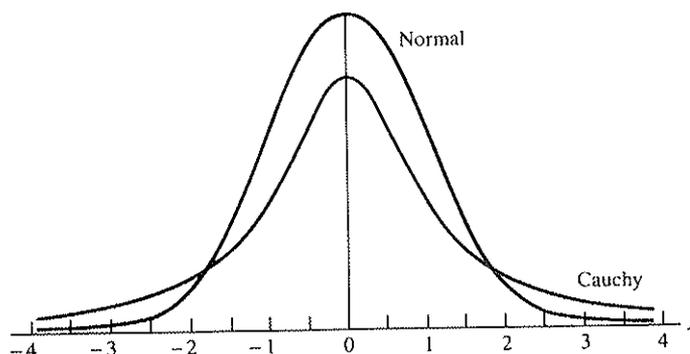


Fig. 3.3.5 Densidade padrão normal e densidade de Cauchy

DISTRIBUIÇÃO LOGNORMAL

Se X é uma variável aleatória cujo logaritmo é normalmente distribuído (isto é, $\log X \sim n(\mu, \sigma^2)$), então X tem uma distribuição lognormal. A fdp de X pode ser obtida pela transformação direta da fdp normal utilizando o Teorema 2.1.5, gerando

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2 / (2\sigma^2)}, \quad 0 < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad (3.3.21)$$

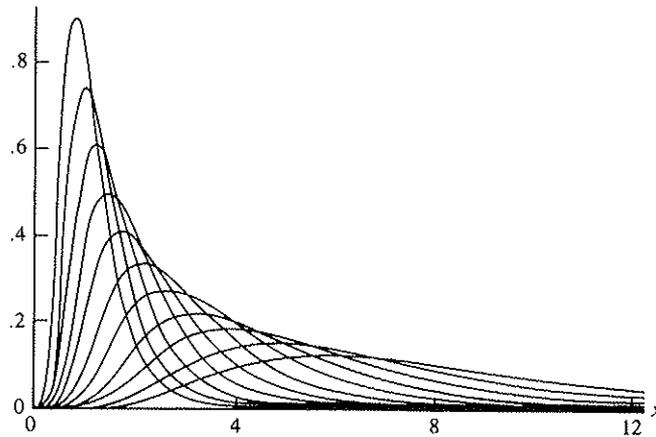
para a fdp lognormal. Os momentos de X podem ser calculados diretamente, utilizando (3.3.21) ou explorando a relação com a normal e escrevendo

$$\begin{aligned} EX &= Ee^{\log X} \\ &= Ee^Y && (Y = \log X \sim n(\mu, \sigma^2)) \\ &= e^{\mu + (\sigma^2/2)}. \end{aligned}$$

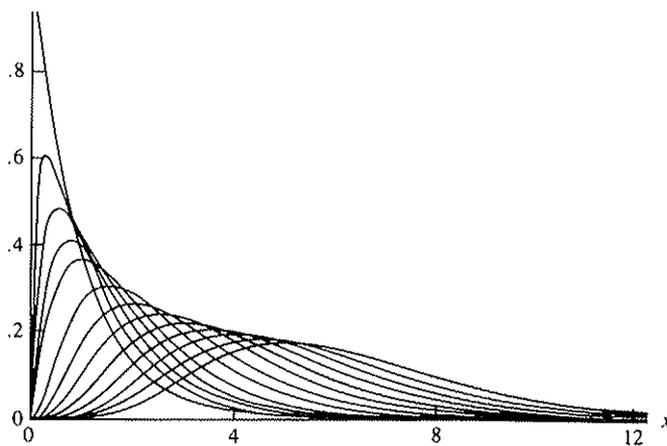
A última igualdade é obtida reconhecendo a fgm da distribuição normal (considerando $t = 1$; veja o Exercício 2.33). Podemos utilizar uma técnica similar para calcular EX^2 e obter

$$\text{Var } X = e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2}.$$

A distribuição lognormal é similar em aparência à distribuição gama, como mostra a Figura 3.3.6. A distribuição é muito popular na modelagem de aplicações, quando a



a.



b.

Fig. 3.3.6 (a) Algumas densidades lognormais; (b) Algumas densidades gama

variável de interesse é assimétrica à direita. Por exemplo, rendimentos são necessariamente assimétricos à direita, e a modelagem com uma lognormal permite o uso de estatística da teoria normal em $\log(\text{rendimento})$, uma circunstância muito conveniente.

DISTRIBUIÇÃO EXPONENCIAL DUPLA

Esta é formada por reflexão da distribuição exponencial em torno de sua média. A fdp é dada por

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0. \quad (3.3.22)$$

A exponencial dupla apresenta uma distribuição simétrica, com extremidades "gordas" (muito mais que o normal), mas ainda retém todos os seus momentos. É possível calcular diretamente

$$EX = \mu \quad \text{e} \quad \text{Var } X = 2\sigma^2.$$

A distribuição exponencial dupla não tem formato de sino. De fato, ela tem um pico (ou, mais formalmente, um ponto de não diferenciabilidade) em $x = \mu$. Quando lidamos com esta distribuição analiticamente, é importante lembrar este ponto. Os sinais de valor absoluto também podem ser problemáticos quando se realizam integrações, e é melhor dividir a integral em regiões ao seu redor $x = \mu$:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} \frac{x}{2\sigma} e^{-|x-\mu|/\sigma} dx \\ &= \int_{-\infty}^{\mu} \frac{x}{2\sigma} e^{(x-\mu)/\sigma} dx + \int_{\mu}^{\infty} \frac{x}{2\sigma} e^{-(x-\mu)/\sigma} dx. \end{aligned} \quad (3.3.23)$$

Observe que podemos remover os sinais de valor absoluto nas duas regiões de integração. (Esta estratégia é útil, em geral, para lidar com integrais contendo valores absolutos; divide-se a região de integração, de modo que os valores absolutos possam ser removidos.) A avaliação de (3.3.23) pode ser concluída fazendo-se a integração em partes, em cada integral.

Existem muitas outras distribuições contínuas, que têm uso em diferentes aplicações estatísticas, muitas das quais serão mostradas ao longo do restante deste livro. O trabalho abrangente, realizado por Johnson e coautores, mencionado no início deste capítulo, é uma referência valiosa para as distribuições estatísticas mais úteis.

3.4 FAMÍLIAS EXPONENCIAIS

Uma família de fdps ou fps é chamada de *exponencial*, se puder ser expressa como

$$f(x|\theta) = h(x)c(\theta) \exp\left(\sum_{i=1}^k w_i(\theta) t_i(x)\right). \quad (3.4.1)$$

Aqui $h(x) \geq 0$ e $t_1(x), \dots, t_k(x)$ são funções com valores reais, de observação x (elas não dependem de θ), e $c(\theta) \geq 0$ e $w_1(\theta), \dots, w_k(\theta)$ são funções com valores reais do parâmetro que possivelmente tem seu valor definido pelo vetor θ (elas não podem depender de x). Muitas famílias comuns apresentadas na seção anterior são exponenciais. Estas incluem as famílias contínuas — normal, gama e beta, e as discretas — binomial, de Poisson e binomial negativa.

Para verificar que a família de fdps ou fps é uma família exponencial, devemos identificar as funções $h(x)$, $c(\theta)$, $w_i(\theta)$ e $t_i(x)$ e mostrar que ela tem a forma (3.4.1). O exemplo a seguir ilustra isto.

EXEMPLO 3.4.1 Família exponencial binomial

Seja n um número inteiro positivo e considere a família binomial (n, p) com $0 < p < 1$. Então, a fp para esta família, para $x = 0, \dots, n$ e $0 < p < 1$, é

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{x} (1-p)^n \left(\frac{p}{1-p}\right)^x \\ &= \binom{n}{x} (1-p)^n \exp\left(\log\left(\frac{p}{1-p}\right)x\right). \end{aligned} \quad (3.4.2)$$

Definimos

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, \dots, n \\ 0 & \text{do contrário,} \end{cases} \quad c(p) = (1-p)^n, \quad 0 < p < 1,$$

$$w_1(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1, \quad \text{e} \quad t_1(x) = x.$$

Então, temos

$$f(x|p) = h(x)c(p) \exp[w_1(p)t_1(x)], \quad (3.4.3)$$

que é a forma (3.4.1) com $k = 1$. Em particular, observe que $h(x) > 0$ somente se $x = 0, \dots, n$, e $c(p)$ é definido somente se $0 < p < 1$. Isto é importante, uma vez que (3.4.3) deve corresponder a (3.4.2) para *todos* os valores de x , e é uma família exponencial somente se $0 < p < 1$ (de modo que as funções do parâmetro são definidas somente aqui). Além disso, os valores do parâmetro $p = 0$ e 1 algumas vezes são incluídos no modelo binomial, mas não os incluímos aqui por causa do conjunto de x valores para o qual $f(x|p) > 0$ é diferente para $p = 0$ e 1 do que para outros p valores.

A forma específica de (3.4.1) resulta em famílias exponenciais que têm muitas propriedades matemáticas úteis. Mas o mais importante para um modelo estatístico é que a forma de (3.4.1) resulta em muitas propriedades estatísticas úteis. A seguir, ilustramos um atalho para o cálculo de momentos de uma família exponencial.

TEOREMA 3.4.2

Se X é uma variável aleatória com fdp ou fp da forma (3.4.1), então

$$E\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log c(\theta); \quad (3.4.4)$$

$$\text{Var}\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log c(\theta) - E\left(\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)\right). \quad (3.4.5)$$

Embora essas equações possam parecer complicadas demais, quando aplicadas a casos específicos elas podem funcionar muito bem. Sua vantagem é que podemos substituir a integração ou soma pela diferenciação, que geralmente é mais direta.

EXEMPLO 3.4.3 Média e variância da binomial

A partir do Exemplo 3.4.1 temos

$$\begin{aligned}\frac{d}{dp} w_1(p) &= \frac{d}{dp} \log \frac{p}{1-p} = \frac{1}{p(1-p)} \\ \frac{d}{dp} \log c(p) &= \frac{d}{dp} n \log(1-p) = \frac{-n}{1-p}\end{aligned}$$

e, deste modo, a partir do Teorema 3.4.2 temos

$$E\left(\frac{1}{p(1-p)} X\right) = \frac{n}{1-p}$$

E, rearranjando, obtemos $E(X) = np$. A identidade de variância funciona de maneira similar.

A prova do Teorema 3.4.2 é uma excursão de cálculo, deixada como Exercício 3.31. Veja também o Exercício 3.32 para conhecer um caso especial.

Agora, podemos verificar outro exemplo e alguns outros recursos das famílias exponenciais.

EXEMPLO 3.4.4 Família exponencial normal

Seja $f(x|\mu, \sigma^2)$ a família $n(\mu, \sigma^2)$ de fdps, onde $\theta = (\mu, \sigma)$, $-\infty < \mu < \infty$, $\sigma > 0$. Então,

$$\begin{aligned}f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right).\end{aligned}\tag{3.4.6}$$

Definimos

$$\begin{aligned}h(x) &= 1 \text{ para todo } x; \\ c(\theta) &= c(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right), \quad -\infty < \mu < \infty, \sigma > 0;\end{aligned}$$

$$\begin{aligned}w_1(\mu, \sigma) &= \frac{1}{\sigma^2}, \quad \sigma > 0; & w_2(\mu, \sigma) &= \frac{\mu}{\sigma^2}, \quad \sigma > 0; \\ t_1(x) &= -x^2/2; & e \quad t_2(x) &= x.\end{aligned}$$

Então,

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp\{w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)\},$$

que é a forma (3.4.1) com $k = 2$. Observe novamente que as funções de parâmetro são definidas somente sobre o domínio do parâmetro.

Em geral, o conjunto de x valores para os quais $f(x|\theta) > 0$ não pode depender de θ em uma família exponencial. A definição completa da fdp ou da fp deve ser incorporada na forma (3.4.1). Isto é mais facilmente obtido incorporando-se a amplitude de x na expressão para $f(x|\theta)$ por meio do uso de uma função indicadora.

DEFINIÇÃO 3.4.5

A função indicadora de um conjunto A , mais frequentemente denotada por $I_A(x)$, é a função

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Uma notação alternativa é $I(x \in A)$.

Portanto, a fdp da normal do Exemplo 3.4.4 seria escrita como

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp[w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)]I_{(-\infty, \infty)}(x).$$

Como a função indicadora é uma função somente de x , ela pode ser incorporada à função $h(x)$, mostrando que esta fdp é da forma (3.4.1).

A partir de (3.4.1), como o fator $\exp(\cdot)$ sempre é positivo, pode ser visto que para qualquer $\theta \in \Theta$, isto é, para qualquer θ para o qual $c(\theta) > 0$, $\{x: f(x|\theta) > 0\} = \{x: h(x) > 0\}$, e este conjunto não depende de θ . Assim, por exemplo, o conjunto de fdps dado por $f(x|\theta) = \theta^{-1} \exp(1 - (x/\theta))$, $0 < \theta < x < \infty$, não é uma família exponencial, mesmo se pudermos escrever $\theta^{-1} \exp(1 - (x/\theta)) = h(x)c(\theta) \exp(w(\theta)t(x))$, onde $h(x) = e^1$, $c(\theta) = \theta^{-1}$, $w(\theta) = \theta^{-1}$, e $t(x) = -x$. Escrever a fdp com funções indicadoras torna isto muito claro. Temos

$$f(x|\theta) = \theta^{-1} \exp\left(1 - \left(\frac{x}{\theta}\right)\right) I_{[\theta, \infty)}(x).$$

A função indicadora não pode ser incorporada em nenhuma das funções de (3.4.1), uma vez que não é uma função somente de x , nem uma função somente de θ , e não pode ser expressa como uma exponencial. Assim, esta não é uma família exponencial.

Algumas vezes, uma família exponencial é reparametrizada como

$$f(x|\eta) = h(x)c^*(\eta) \exp\left(\sum_{i=1}^k \eta_i t_i(x)\right). \quad (3.4.7)$$

Aqui, as funções $h(x)$ e $t_i(x)$ são as mesmas que na parametrização original (3.4.1). O conjunto $\mathcal{H} = \{\eta = (\eta_1, \dots, \eta_k): \int_{-\infty}^{\infty} h(x) \exp(\sum_{i=1}^k \eta_i t_i(x)) dx < \infty\}$ é chamado de *espaço paramétrico natural* para a família. (A integral é substituída por uma soma dos valores de x para a qual $h(x) > 0$ se X for discreta.) Para os valores de $\eta \in \mathcal{H}$, devemos ter $c^*(\eta) = [\int_{-\infty}^{\infty} h(x) \exp(\sum_{i=1}^k \eta_i t_i(x)) dx]^{-1}$ para garantir que a integral da fdp é igual a 1. Uma vez que a $f(x|\theta)$ original em (3.4.1) é uma fdp ou fp, o conjunto $\{\eta = (w_1(\theta), \dots, w_k(\theta)): \theta \in \Theta\}$ deve ser um subconjunto do espaço do parâmetro natural. Mas também podem existir outros valores de $\eta \in \mathcal{H}$. A parametrização natural e o espaço do parâmetro natural têm propriedades matemáticas muito úteis. Por exemplo, \mathcal{H} é convexa.

EXEMPLO 3.4.6 Continuação do Exemplo 3.4.4

A fim de determinar o espaço paramétrico natural para a família de distribuições normais, substituímos $w_i(\mu, \sigma)$ por η_i em (3.4.6) para obter

$$f(x|\eta_1, \eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp\left(-\frac{\eta_2^2}{2\eta_1}\right) \exp\left(-\frac{\eta_1 x^2}{2} + \eta_2 x\right). \quad (3.4.8)$$

A integral será finita se, e somente se, o coeficiente em x^2 for negativo. Isto significa que η_1 deve ser positivo. Se $\eta_1 > 0$, a integral será finita, independentemente do valor de η_2 . Portanto, o espaço do parâmetro natural é $\{(\eta_1, \eta_2): \eta_1 > 0, -\infty < \eta_2 < \infty\}$. Identificando (3.4.8) com (3.4.6), observamos que $\eta_2 = \mu/\sigma^2$ e $\eta_1 = 1/\sigma^2$. Embora parâmetros naturais

ofereçam uma formulação matemática conveniente, algumas vezes eles não dispõem de interpretações simples, como a média e a variância.

Na representação (3.4.1), geralmente é o caso em que a dimensão do vetor θ é igual a k , o número de termos na soma, no expoente. Mas não é preciso que seja assim, e é possível que a dimensão do vetor θ seja igual a $d < k$. Esta família exponencial é chamada de *família exponencial curva*.

DEFINIÇÃO 3.4.7

Uma *família exponencial curva* é um grupo de densidades da forma (3.4.1) para a qual a dimensão do vetor θ é igual a $d < k$. Se $d = k$, é uma *família exponencial completa*. (Veja também Itens Diversos 3.8.3.)

EXEMPLO 3.4.8 Uma família exponencial curva

A família normal do Exemplo 3.4.4 é uma exponencial completa. No entanto, se assumirmos que $\sigma^2 = \mu^2$, a família se torna curva. (Tal modelo pode ser utilizado na análise de variância; veja os Exercícios 11.1 e 11.2.) Temos então

$$\begin{aligned} f(x|\mu) &= \frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{(x-\mu)^2}{2\mu^2}\right) \\ &= \frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{1}{2}\right) \exp\left(-\frac{x^2}{2\mu^2} + \frac{x}{\mu}\right). \end{aligned} \quad (3.4.9)$$

Para a família normal, a família exponencial completa teria espaço paramétrico $(\mu, \sigma^2) = \mathbb{R} \times (0, \infty)$, enquanto o espaço paramétrico da família curva $(\mu, \sigma^2) = (\mu, \mu^2)$ é uma parábola.

Famílias exponenciais curvas são úteis de muitas maneiras. O exemplo a seguir ilustra um uso simples.

EXEMPLO 3.4.9 Aproximações normais

No Capítulo 5 veremos que se X_1, \dots, X_n é uma amostra de uma população de Poisson(λ), então a distribuição de $\bar{X} = \sum_i X_i/n$ é aproximadamente

$$\bar{X} \sim n(\lambda, \lambda/n),$$

uma família exponencial curva.

A aproximação $n(\lambda, \lambda/n)$ é justificada pelo Teorema do Limite Central, ou CLT (Teorema 5.5.14). Na verdade, podemos perceber que a maioria das aproximações com base no CLT resultará em uma família normal curva. Verificamos a aproximação binomial normal (Exemplo 3.3.2): Se X_1, \dots, X_n forem provas de Bernoulli(p) iid (independentes e identicamente distribuídas), então

$$\bar{X} \sim n(p, p(1-p)/n),$$

aproximadamente. Para observar outra ilustração, veja o Exemplo 5.5.16.

Apesar do fato de o espaço paramétrico ser um espaço de baixa dimensionalidade, ele tem alguma influência nas propriedades da família; veremos que famílias curvas ainda apresentam muitas das propriedades das famílias complexas. Em particular, o Teorema 3.4.2

aplica-se às famílias exponenciais curvas. Além disso, famílias exponenciais completas e curvas têm outras propriedades estatísticas, que serão discutidas ao longo do restante deste livro. Por exemplo, suponha que temos um grande número de valores de dados a partir de uma população que tem uma fdp ou fp da forma (3.4.1). Então, somente k números ($k =$ número de termos na soma em (3.4.1)) que podem ser calculados a partir dos dados resumem todas as informações sobre θ , que está nos dados. Esta propriedade de "redução de dados" é tratada mais detalhadamente no Capítulo 6 (Teorema 6.2.10), no qual discutimos sobre estatísticas suficientes.

Para obter mais que uma introdução às famílias exponenciais, veja Lehmann (1986, Seção 2.7) ou Lehmann e Casella (1998, Seção 1.5 e Nota 1.10.6). Uma introdução completa, em um nível um pouco mais avançado, é fornecida na clássica monografia desenvolvida por Brown (1986).

3.5 FAMÍLIAS DE LOCAÇÃO E DE ESCALA

Nas Seções 3.3 e 3.4 discutimos diversas famílias comuns de distribuições contínuas. Nesta, abordamos três técnicas para a construção de famílias de distribuições. As famílias resultantes têm prontas interpretações físicas que as tornam úteis para a modelagem, assim como têm propriedades matemáticas convenientes.

Os três tipos de famílias são chamadas de locação, de escala e de locação-escala. Cada uma das famílias é construída pela especificação de uma única fdp, seja $f(x)$, chamada de *fdp padrão* para a família. Então, todos os outros fdps na família são gerados pela transformação da fdp padrão de um modo prescrito. Começamos com um teorema simples sobre fdps.

TEOREMA 3.5.1

Seja $f(x)$ qualquer fdp e que μ e $\sigma > 0$ sejam quaisquer constantes dadas. Então, a função

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

é uma fdp.

PROVA:

Para verificar que a transformação produziu uma fdp legítima, precisamos verificar se $(1/\sigma)f((x - \mu)/\sigma)$, como uma função de x , é uma fdp para cada valor de μ e σ que podemos substituir na fórmula. Isto é, devemos checar se $(1/\sigma)f((x - \mu)/\sigma)$ é não negativa e se a integral é igual a 1. Uma vez que $f(x)$ é uma fdp, $f(x) \geq 0$ para todos os valores de x . De modo que, $(1/\sigma)f((x - \mu)/\sigma) \geq 0$ para todos os valores de x, μ e σ . Em seguida, observamos que

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) dx = \int_{-\infty}^{\infty} f(y) dy \quad \left(\text{substituir } y = \frac{x - \mu}{\sigma}\right)$$

$$= 1, \quad \text{(uma vez que } f(y) \text{ é uma fdp)}$$

como deveria ser verificado. ■

Agora, recorremos à primeira de nossas construções, a das famílias de locação.

DEFINIÇÃO 3.5.2

Seja $f(x)$ qualquer fdp. Então, a família de fdps $f(x - \mu)$, indexada pelo parâmetro μ , $-\infty < \mu < \infty$, é chamada de *família de locação com fdp padrão* $f(x)$ e μ é chamada de *parâmetro de locação* para a família.

Para verificar o efeito da introdução de um parâmetro de locação μ , considere a Figura 3.5.1. Em $x = \mu$, $f(x - \mu) = f(0)$; em $x = \mu + 1$, $f(x - \mu) = f(1)$; e, em geral, em $x = \mu + a$,

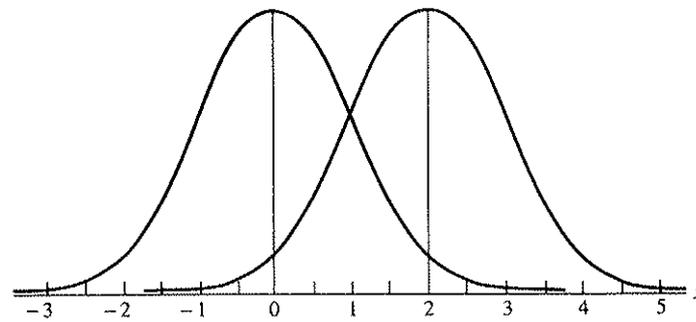


Fig. 3.5.1 Dois membros da mesma família de localização: médias em 0 e 2

$f(x - \mu) = f(x)$. Naturalmente, $f(x - \mu)$ para $\mu = 0$ é apenas $f(x)$. Assim, o parâmetro de localização μ simplesmente muda a fdp $f(x)$, de modo que o formato do gráfico permanece inalterado, mas o ponto no gráfico que era $x = 0$ para $f(x)$ é $x = \mu$ para $f(x - \mu)$. Fica claro, a partir da Figura 3.5.1, que a área sob o gráfico de $f(x)$ entre $x = -1$ e $x = 2$ é igual à área sob o gráfico de $f(x - \mu)$ entre $x = \mu - 1$ e $x = \mu + 2$. Portanto X é uma variável aleatória com fdp $f(x - \mu)$; podemos escrever

$$P(-1 \leq X \leq 2|0) = P(\mu - 1 \leq X \leq \mu + 2|\mu),$$

onde a variável aleatória X tem fdp $f(x - 0) = f(x)$ à esquerda da igualdade e fdp $f(x - \mu)$ à direita.

Várias das famílias apresentadas na Seção 3.3 são, ou têm como subfamílias, famílias de locação. Por exemplo, se $\sigma > 0$ é um número específico conhecido e definimos

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}, \quad -\infty < x < \infty,$$

então, a família de locação com fdp padrão $f(x)$ é o conjunto de distribuições normais com média desconhecida μ e variância conhecida σ^2 . Para verificar isto, observe se substituir x por $x - \mu$ na fórmula acima gera fdps da forma definida em (3.3.13). De maneira similar, as famílias de Cauchy e a exponencial dupla, com σ como um valor específico e μ como um parâmetro, são exemplos de famílias de locação. Mas o apontado pela definição 3.5.2 é que podemos começar com *qualquer* fdp $f(x)$ e gerar uma família de fdps introduzindo um parâmetro de locação.

Se X é uma variável aleatória com fdp $f(x - \mu)$, então X pode ser representado como $X = Z + \mu$, onde Z é uma variável aleatória com fdp $f(z)$. Esta representação é uma consequência do Teorema 3.5.6 (com $\sigma = 1$), o que será comprovado posteriormente. A consideração desta representação indica quando uma família de locação pode ser um modelo apropriado para uma variável observada X . Descreveremos duas destas situações.

Primeiro, suponhamos que um experimento seja projetado para medir alguma constante física μ , digamos, a temperatura de uma solução. Mas existe algum erro de medição envolvido na observação, de modo que o real valor observado X é $Z + \mu$, onde Z é o erro de medição. X será maior que μ se $Z > 0$ para esta observação, e será menor que μ se $Z < 0$. A distribuição do erro de medição aleatório pode ser bem conhecida a partir da experiência anterior na utilização deste dispositivo de medição para medir outras soluções. Se esta distribuição tem fdp $f(z)$, então a fdp do valor observado X é $f(x - \mu)$.

Como outro exemplo, suponha que a distribuição dos tempos de reação de motoristas em um teste de coordenação motora é conhecida a partir de um experimento realizado anteriormente. Denote o tempo de reação para um motorista escolhido aleatoriamente, utilizando a variável aleatória Z . Seja a fdp de Z descrevendo a distribuição conhecida $f(z)$. Agora, considere a "aplicação de um tratamento" à população. Por exemplo, considere o

que aconteceria se todos os motoristas bebessem três copos de cerveja. Podemos assumir que o tempo de reação de todos eles muda por meio de alguma quantidade desconhecida μ . (Este modelo muito simples, em que o tempo de reação de todos eles muda a mesma quantidade μ , provavelmente não é o melhor. Por exemplo, sabe-se que o efeito do álcool depende do peso, portanto, pessoas mais pesadas provavelmente serão menos afetadas pela cerveja consumida.) Como cientistas de mente aberta, podemos, até mesmo, permitir a possibilidade de que $\mu < 0$, isto é, que o tempo de reação diminuiu. Então, se observarmos o tempo de reação de um motorista que foi selecionado aleatoriamente, após o "tratamento", o tempo de reação seria $X = Z + \mu$ e a família de possíveis distribuições para X seria dada por $f(x - \mu)$.

Se o conjunto de x para o qual $f(x) > 0$ não é a reta real completa, então, o conjunto de x para o qual $f(x - \mu) > 0$ dependerá de μ . O Exemplo 3.5.3 ilustra isto.

EXEMPLO 3.5.3 Família de localização exponencial

Seja $f(x) = e^{-x}$, $x \geq 0$ e $f(x) = 0$, $x < 0$. Para formar uma família de localização, substituímos x por $x - \mu$ para obter

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x - \mu \geq 0 \\ 0 & x - \mu < 0 \end{cases}$$

$$= \begin{cases} e^{-(x-\mu)} & x \geq \mu \\ 0 & x < \mu. \end{cases}$$

Gráficos de $f(x|\mu)$ para diversos valores de μ são mostrados na Figura 3.5.2. Como na Figura 3.5.1, o gráfico foi modificado. Agora, a parte positiva do gráfico começa em μ em vez de em 0. Se X mede o tempo, então μ está restrito a ser não negativo, de modo que X será positivo, com probabilidade 1 para cada valor de μ . Neste tipo de modelo, onde μ denota um limite sobre o conjunto de valores de X , μ algumas vezes é chamado de *parâmetro limite*.

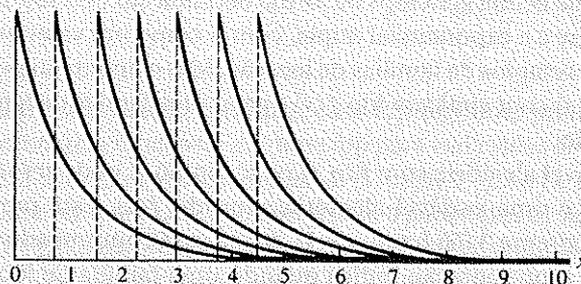


Fig. 3.5.2 Densidades de localização exponenciais

Os dois outros tipos de famílias a serem discutidos nesta seção são os de escala e os de localização-escala.

DEFINIÇÃO 3.5.4

Seja $f(x)$ qualquer fdp. Então, para qualquer $\sigma > 0$, a família de fdps $(1/\sigma)f(x/\sigma)$, indexada pelo parâmetro σ , é chamada de *família de escala com fdp padrão $f(x)$* e σ é chamada de *parâmetro de escala* da família.

O efeito de introduzir o parâmetro de escala σ é o de alongar ($\sigma > 1$) ou contrair ($\sigma < 1$) o gráfico de $f(x)$ enquanto ainda mantém o mesmo formato básico do gráfico. Isto é ilustrado na Figura 3.5.3. Na maioria das vezes, quando os parâmetros de escala são utilizados, $f(x)$ é

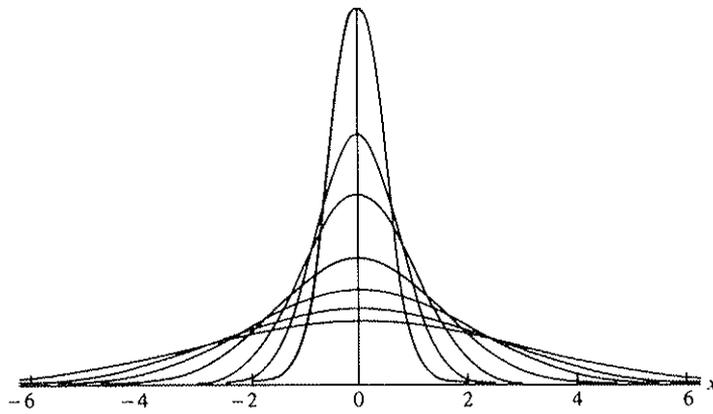


Fig. 3.5.3 Membros da mesma família de escala

simétrico em 0 ou positivo somente para $x > 0$. Nesses casos, o alongamento ou é simétrico em 0 ou somente na direção positiva. Mas, na definição, qualquer fdp pode ser utilizada como o padrão.

Diversas das famílias introduzidas na Seção 3.3 ou são de escala ou as têm como subfamílias. Estas são a família gama se α for um valor fixo e β , o parâmetro de escala; são a família normal, se $\mu = 0$ e σ for o parâmetro de escala; são a família exponencial e a família exponencial dupla, se $\mu = 0$ e σ forem o parâmetro de escala. Em cada caso, a fdp padrão é a fdp obtida ao se definir o parâmetro de escala igual a 1. Então, todos os outros membros da família podem ser mostrados como na Definição 3.5.4.

DEFINIÇÃO 3.5.5

Seja $f(x)$ qualquer fdp. Então, para qualquer μ , $-\infty < \mu < \infty$, e qualquer $\sigma > 0$, a família de fdps $(1/\sigma)f((x-\mu)/\sigma)$, indexada pelo parâmetro (μ, σ) , é chamada de *família de locação-escala com fdp padrão $f(x)$* ; μ é chamado de *parâmetro de locação* e σ de *parâmetro de escala*.

O efeito de introduzir os parâmetros de locação e de escala é o de alongar ($\sigma > 1$) ou contrair ($\sigma < 1$) o gráfico com o parâmetro de escala e, então, modificá-lo, de modo que o ponto que estava acima de 0 fique agora acima de μ . A Figura 3.5.4 ilustra esta transformação de $f(x)$. As famílias exponenciais normal e dupla são exemplos de famílias de locação-escala. O Exercício 3.39 apresenta a família de Cauchy como uma família de locação-escala.

O seguinte teorema relata a transformação da fdp $f(x)$ que define uma família de locação-escala para a transformação de uma variável aleatória Z com fdp $f(z)$. Como

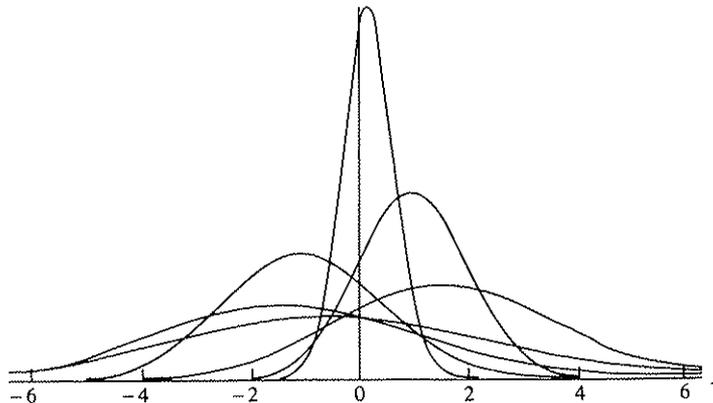


Fig. 3.5.4 Membros da mesma família de escala-localização

foi mencionado na discussão sobre famílias de locação, a representação em termos de Z é uma ferramenta matemática útil e pode nos ajudar a entender quando uma família de locação-escala pode ser apropriada em um contexto de modelagem. Definir $\sigma = 1$ no Teorema 3.5.6 gera um resultado (somente) para as famílias de locação, e definir $\mu = 0$ gera um resultado (somente) para as famílias de escala.

TEOREMA 3.5.6

Seja $f(\cdot)$ qualquer fdp e seja μ um número real qualquer, e σ um número real qualquer positivo. Então X é uma variável aleatória com fdp $(1/\sigma)f((x-\mu)/\sigma)$ se, e somente se, existir uma variável aleatória Z com fdp $f(z)$ e $X = \sigma Z + \mu$.

PROVA:

Para provar a parte referente a "se", definimos $g(z) = \sigma z + \mu$. Então $X = g(Z)$, g é uma função monótona, $g^{-1}(x) = (x - \mu)/\sigma$, e $|(d/dx)g^{-1}(x)| = 1/\sigma$. Portanto, segundo o Teorema 2.1.5, a fdp de X é

$$f_X(x) = f_Z(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right| = f\left(\frac{x-\mu}{\sigma}\right) \frac{1}{\sigma}.$$

Para provar a parte referente a "somente se", definimos que $g(x) = (x - \mu)/\sigma$ e que $Z = g(X)$. Novamente, o Teorema 2.1.5 se aplica: $g^{-1}(z) = \sigma z + \mu$, $|(d/dz)g^{-1}(z)| = \sigma$, e a fdp de Z é

$$f_Z(z) = f_X(g^{-1}(z)) \left| \frac{d}{dz} g^{-1}(z) \right| = \frac{1}{\sigma} f\left(\frac{(\sigma z + \mu) - \mu}{\sigma}\right) \sigma = f(z).$$

Além disso,

$$\sigma Z + \mu = \sigma g(X) + \mu = \sigma \left(\frac{X - \mu}{\sigma}\right) + \mu = X. \quad \blacksquare$$

Um fato importante que se pode extrair do Teorema 3.5.6 é que a variável aleatória $Z = (X - \mu)/\sigma$ tem fdp

$$f_Z(z) = \frac{1}{1} f\left(\frac{z-0}{1}\right) = f(z).$$

Isto é, a distribuição de Z é aquele membro da família de locação-escala correspondente a $\mu = 0, \sigma = 1$. Isto já foi comprovado para o caso especial da família normal, na Seção 3.3.

Geralmente, os cálculos podem ser desenvolvidos para a variável aleatória "padrão" Z com fdp $f(z)$ e, então, o resultado correspondente para a variável aleatória X com fdp $(1/\sigma)f((x-\mu)/\sigma)$ pode ser derivado facilmente. Um exemplo é dado a seguir, que é uma generalização de um cálculo realizado na Seção 3.3 para a família normal.

TEOREMA 3.5.7

Seja Z uma variável aleatória com fdp $f(z)$. Suponha que EZ e $\text{Var } Z$ existam. Se X é uma variável aleatória com fdp $(1/\sigma)f((x-\mu)/\sigma)$, então

$$EX = \sigma EZ + \mu \quad \text{e} \quad \text{Var } X = \sigma^2 \text{Var } Z.$$

Em particular, se $EZ = 0$ e $\text{Var } Z = 1$, então $EX = \mu$ e $\text{Var } X = \sigma^2$.

PROVA:

De acordo com o Teorema 3.5.6, existe uma variável aleatória Z^* com fdp $f(z)$ e $X = \sigma Z^* + \mu$. Portanto $EX = \sigma EZ^* + \mu = \sigma EZ + \mu$ e $\text{Var } X = \sigma^2 \text{Var } Z^* = \sigma^2 \text{Var } Z$. \blacksquare

Para qualquer família de locação-escala com média e variância finitas, a fdp padrão $f(z)$ pode ser escolhida de tal maneira que $EZ = 0$ e $\text{Var } Z = 1$. (A prova de que esta escolha pode ser feita é deixada como o Exercício 3.40.) Isto resulta na conveniente interpretação de μ e σ^2 como a média e a variância de X , respectivamente. Este é o caso para a definição usual da família normal, como foi dado na Seção 3.3. Contudo, esta não é a escolha para a definição usual da família exponencial dupla, como foi dado na Seção 3.3. Neste caso, $\text{Var } Z = 2$.

As probabilidades para que qualquer membro de uma família de locação-escala podem ser calculadas em termos da variável padrão Z porque

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right).$$

Deste modo, se $P(Z \leq z)$ for tabulado ou facilmente calculável para a variável padrão Z , então as probabilidades para X podem ser obtidas. Os cálculos das probabilidades normais utilizando a tabela normal padrão são exemplos disto.

3.6 DESIGUALDADES E IDENTIDADES

A teoria estatística está, literalmente, repleta de desigualdades e identidades — são tantas, que livros inteiros são dedicados a este tópico. O importante trabalho de Marshall e Olkin (1979) contém muitas desigualdades utilizando o conceito de majorização. Outro trabalho, mais antigo, realizado por Hardy, Littlewood e Polya (1952), é um compêndio de desigualdades clássicas. Nesta seção e na Seção 4.7 mesclaremos algumas desigualdades clássicas e novas, para dar uma ideia dos tipos de resultados que existem. Esta seção é dedicada às identidades e desigualdades que surgem das preocupações probabilísticas, ao passo que as da Seção 4.7 dependem mais das propriedades básicas de números e funções.

3.6.1 DESIGUALDADES DE PROBABILIDADE

A mais famosa e, talvez, a mais útil desigualdade de probabilidade é a Desigualdade de Chebychev. Sua utilidade deve-se a sua ampla aplicabilidade. Assim como acontece com muitos resultados importantes, sua prova é quase trivial.

TEOREMA 3.6.1

Desigualdade de Chebychev

Seja um X uma variável aleatória e $g(x)$ uma função não negativa. Então, para qualquer $r > 0$,

$$P(g(X) \geq r) \leq \frac{Eg(X)}{r}.$$

PROVA:

$$\begin{aligned} Eg(X) &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \\ &\geq \int_{\{x: g(x) \geq r\}} g(x) f_X(x) dx && (g \text{ não é negativa}) \\ &\geq r \int_{\{x: g(x) \geq r\}} f_X(x) dx \\ &= rP(g(X) \geq r). && (\text{definição}) \end{aligned}$$

Rearranjar agora produz a desigualdade desejada. ■

EXEMPLO 3.6.2 Ilustrando Chebychev

O mais amplo uso da Desigualdade de Chebychev envolve médias e variâncias. Seja $g(x) = (x - \mu)^2 / \sigma^2$, onde $\mu = EX$ e $\sigma^2 = \text{Var } X$. Por conveniência, escrevemos $t = t^2$. Então

$$P\left(\frac{(X - \mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2} E \frac{(X - \mu)^2}{\sigma^2} = \frac{1}{t^2}.$$

Aplicando alguma álgebra óbvia, obtemos a desigualdade

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$

e seu complemento

$$P(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2},$$

que dá um limite universal ao desvio $|X - \mu|$ em termos de σ . Por exemplo, assumindo que $t = 2$, obtemos

$$P(|X - \mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0,25.$$

Deste modo, existem pelo menos 75% de chances de que uma variável aleatória estará a menos de 2σ de sua média (não importa qual seja a distribuição de X).

Embora a Desigualdade de Chebychev seja amplamente aplicável, ela é necessariamente conservadora. (Veja, por exemplo, o Exercício 3.46 e Itens Diversos 3.8.2.) Em particular, geralmente, podemos obter limites mais estritos para algumas distribuições específicas.

EXEMPLO 3.6.3 Uma desigualdade de probabilidade normal

Se Z é normal padrão, então

$$P(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}, \quad \text{para todo } t > 0. \quad (3.6.1)$$

Compare isto com a Desigualdade de Chebychev. Para $t = 2$, Chebychev resulta $P(|Z| \geq t) \leq 0,25$, mas $\sqrt{(2/\pi)}e^{-2/2} = 0,054$, uma grande melhoria.

Para provar (3.6.1), escrevemos

$$\begin{aligned} P(Z \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2} dx \\ &\leq \frac{1}{\sqrt{2\pi}} \int_t^{\infty} \frac{x}{t} e^{-x^2/2} dx && \text{(uma vez que } x/t > 1 \text{ para } x > t) \\ &= \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t} \end{aligned}$$

e utilizamos o fato de que $P(|Z| \geq t) = 2P(Z \geq t)$. Um limite inferior em $P(|Z| \geq t)$ pode ser estabelecido de maneira similar (veja o Exercício 3.47).

Existem muitas outras desigualdades de probabilidade, e quase todas elas são similares, em essência, à de Chebychev. Por exemplo, veremos (Exercício 3.45) que

$$P(X \geq a) \leq e^{-at} M_X(t),$$

mas, naturalmente, esta desigualdade exige a existência da fgm. Existem outras desigualdades, mais estritas do que a Chebychev, mas exigindo mais suposições (como é detalhado em Itens Diversos 3.8.2).

3.6.2 IDENTIDADES

Nesta seção, apresentamos uma amostra de várias identidades que podem ser úteis em estabelecer teoremas, mas também em facilitar cálculos numéricos. Uma classe completa de identidades pode ser considerada como sendo "relações de recursão", algumas das quais já vimos. Lembre-se de que se X é $\text{Poisson}(\lambda)$, então

$$P(X = x + 1) = \frac{\lambda}{x + 1} P(X = x), \quad (3.6.2)$$

permitindo-nos calcular probabilidades de Poisson recursivamente, começando a partir de $P(X = 0) = e^{-\lambda}$. Relações como (3.6.2) existem para quase todas as distribuições discretas (veja o Exercício 3.48). Algumas vezes, elas existem de forma um pouco diferente para as distribuições contínuas.

TEOREMA 3.6.4

Seja $X_{\alpha, \beta}$ denotando uma variável aleatória gama(α, β) com fdp $f(x|\alpha, \beta)$, onde $\alpha > 1$. Então, para quaisquer constantes a e b ,

$$P(a < X_{\alpha, \beta} < b) = \beta (f(a|\alpha, \beta) - f(b|\alpha, \beta)) + P(a < X_{\alpha-1, \beta} < b). \quad (3.6.3)$$

PROVA:

Por definição,

$$\begin{aligned} P(a < X_{\alpha, \beta} < b) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_a^b x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \left[-x^{\alpha-1} \beta e^{-x/\beta} \Big|_a^b + \int_a^b (\alpha-1)x^{\alpha-2} \beta e^{-x/\beta} dx \right], \end{aligned}$$

onde fizemos uma integração por partes com $u = x^{\alpha-1}$ e $dv = e^{-x/\beta} dx$. Continuando, temos

$$P(a < X_{\alpha, \beta} < b) = \beta (f(a|\alpha, \beta) - f(b|\alpha, \beta)) + \frac{(\alpha-1)}{\Gamma(\alpha)\beta^{\alpha-1}} \int_a^b x^{\alpha-2} e^{-x/\beta} dx.$$

Utilizando o fato de que $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, vemos que o último termo é $P(a < X_{\alpha-1, \beta} < b)$. ■

Se α é um número inteiro, o uso repetido de (3.6.3) eventualmente levará a uma integral que pode ser avaliada analiticamente (quando $\alpha = 1$, a distribuição exponencial). Portanto, podemos facilmente calcular essas probabilidades gama.

Existe uma classe inteira de identidades que dependem da integração por partes. A primeira delas é atribuída a Charles Stein, que a utilizou em seu trabalho sobre estimação de médias normais multivariadas (Stein, 1973, 1981).

LEMA 3.6.5

O lema de Stein

Seja $X \sim n(\theta, \sigma^2)$ e g uma função diferenciável que satisfaz $E|g'(X)| < \infty$. Então,

$$E[g(X)(X - \theta)] = \sigma^2 E g'(X).$$

PROVA:

O lado esquerdo é

$$E[g(X)(X - \theta)] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g(x)(x - \theta) e^{-(x-\theta)^2/(2\sigma^2)} dx.$$

Utilizamos a integração por partes com $u = g(x)$ e $dv = (x - \theta) e^{-(x-\theta)^2/(2\sigma^2)} dx$ para obter

$$E[g(X)(X - \theta)] = \frac{1}{\sqrt{2\pi}\sigma} \left[-\sigma^2 g(x) e^{-(x-\theta)^2/(2\sigma^2)} \Big|_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} g'(x) e^{-(x-\theta)^2/(2\sigma^2)} dx \right].$$

A condição em g' é suficiente para garantir que o primeiro termo é 0 e que o restante, no lado direito, é $\sigma^2 E g'(X)$. ■

EXEMPLO 3.6.6 Momentos normais de ordem superior

O Lema de Stein facilita muito o cálculo de momentos de ordem mais elevada. Por exemplo, se $X \sim n(\theta, \sigma^2)$, então

$$\begin{aligned} EX^3 &= EX^2(X - \theta + \theta) \\ &= EX^2(X - \theta) + \theta EX^2 \\ &= 2\sigma^2 EX + \theta EX^2 && (g(x) = x^2, g'(x) = 2x) \\ &= 2\sigma^2\theta + \theta(\sigma^2 + \theta^2) \\ &= 3\theta\sigma^2 + \theta^3. \end{aligned}$$

Identidades similares às com integração por partes existem para muitas distribuições (veja o Exercício 3.49 e Hudson, 1978). É possível obter identidades úteis explorando propriedades de uma distribuição específica, como mostra o teorema a seguir.

TEOREMA 3.6.7

Seja χ_p^2 denotando uma variável aleatória qui-quadrada, com p graus de liberdade. Para qualquer função $h(x)$,

$$Eh(\chi_p^2) = pE\left(\frac{h(\chi_{p+2}^2)}{\chi_{p+2}^2}\right) \quad (3.6.4)$$

desde que as expectâncias existam.

PROVA:

A frase “desde que as expectâncias existam” é uma maneira “preguiçosa” de evitar a especificação de condições em h . Em geral, funções razoáveis irão satisfazer (3.6.4). Temos

$$\begin{aligned} Eh(\chi_p^2) &= \frac{1}{\Gamma(p/2)2^{p/2}} \int_0^\infty h(x)x^{(p/2)-1} e^{-x/2} dx \\ &= \frac{1}{\Gamma(p/2)2^{p/2}} \int_0^\infty \left(\frac{h(x)}{x}\right) x^{((p+2)/2)-1} e^{-x/2} dx, \end{aligned}$$

onde multiplicamos a integranda por x/x . Agora, escrevemos

$$\Gamma\left(\frac{p}{2}\right)2^{p/2} = \frac{\Gamma((p+2)/2)2^{(p+2)/2}}{p},$$

deste modo, temos

$$\begin{aligned} Eh(\chi_p^2) &= \frac{p}{\Gamma((p+2)/2)2^{(p+2)/2}} \int_0^\infty \left(\frac{h(x)}{x}\right) x^{((p+2)/2)-1} e^{-x/2} dx \\ &= pE\left(\frac{h(\chi_{p+2}^2)}{\chi_{p+2}^2}\right). \quad \blacksquare \end{aligned}$$

Alguns cálculos de momentos são muito fáceis com (3.6.4). Por exemplo, a média de um χ_p^2 é

$$E\chi_p^2 = pE\left(\frac{\chi_{p+2}^2}{\chi_{p+2}^2}\right) = pE(1) = p,$$

e o segundo momento é

$$E(\chi_p^2)^2 = pE\left(\frac{(\chi_{p+2}^2)^2}{\chi_{p+2}^2}\right) = pE(\chi_{p+2}^2) = p(p+2).$$

Assim $\text{Var } \chi_p^2 = p(p+2) - p^2 = 2p$.

Encerramos nossa seção sobre identidades com alguns análogos discretos das identidades anteriores. Uma versão geral das duas identidades, no Teorema 3.6.8, se deve a Hwang (1982).

TEOREMA 3.6.8

Hwang

Seja $g(x)$ uma função com $-\infty < Eg(X) < \infty$ e $-\infty < g(-1) < \infty$. Então:

a. Se $X \sim \text{Poisson}(\lambda)$,

$$E(\lambda g(X)) = E(Xg(X-1)). \quad (3.6.5)$$

b. Se $X \sim \text{binomial negativa}(r, p)$,

$$E((1-p)g(X)) = E\left(\frac{X}{r+X-1}g(X-1)\right). \quad (3.6.6)$$

PROVA:

Comprovaremos a parte (a), deixando a parte (b) para o Exercício 3.50. Temos

$$\begin{aligned} E(\lambda g(X)) &= \sum_{x=0}^{\infty} \lambda g(x) \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} g(x) \frac{e^{-\lambda} \lambda^{x+1}}{x!} \frac{(x+1)}{(x+1)} \\ &= \sum_{x=0}^{\infty} (x+1)g(x) \frac{e^{-\lambda} \lambda^{x+1}}{(x+1)!}. \end{aligned}$$

Agora, transformamos o índice da somatória, escrevendo $y = x + 1$. À medida que x vai de 0 a ∞ , y vai de 1 até ∞ . Portanto,

$$\begin{aligned} E(\lambda g(X)) &= \sum_{y=1}^{\infty} yg(y-1) \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \sum_{y=0}^{\infty} yg(y-1) \frac{e^{-\lambda} \lambda^y}{y!} \quad (\text{o termo adicionado é } 0) \\ &= E(Xg(X-1)), \end{aligned}$$

uma vez que esta última somatória é uma expectativa de Poisson(λ). ■

Hwang (1982) utilizou sua identidade de maneira similar a Stein, demonstrando resultados sobre estimadores multivariados. A identidade tem outras aplicações, em particular nos cálculos de momentos.

EXEMPLO 3.6.9 Momentos de Poisson de ordem superior

Para $X \sim \text{Poisson}(\lambda)$, assumimos que $g(x) = x^2$ e utilizamos (3.6.5):

$$E(\lambda X^2) = E(X(X-1)^2) = E(X^3 - 2X^2 + X).$$

Portanto, o terceiro momento de uma distribuição de Poisson(λ) é

$$\begin{aligned} EX^3 &= \lambda EX^2 + 2EX^2 - EX \\ &= \lambda(\lambda + \lambda^2) + 2(\lambda + \lambda^2) - \lambda \\ &= \lambda^3 + 3\lambda^2 + \lambda. \end{aligned}$$

Para a binomial negativa, a média pode ser calculada assumindo que $g(x) = r + x$ em (3.6.6):

$$E((1-p)(r+X)) = E\left(\frac{X}{r+X-1}(r+X-1)\right) = EX,$$

deste modo, rearranjando, obtemos

$$(EX)((1-p)-1) = -r(1-p)$$

ou

$$EX = \frac{r(1-p)}{p}.$$

Outros momentos podem ser calculados de maneira similar.

3.7 EXERCÍCIOS

- 3.1 Encontre as expressões para EX e $\text{Var } X$ se X for uma variável aleatória com a distribuição uniforme geral discreta (N_0, N_1) que atribui igual probabilidade a cada um dos valores N_0, N_0+1, \dots, N_1 . Aqui $N_0 \leq N_1$ e são números inteiros.
- 3.2 Um fabricante recebe um lote de 100 peças de um fornecedor. O lote será considerado inaceitável se mais que cinco das peças apresentarem defeito. O fabricante selecionará aleatoriamente K peças do lote para a realização de inspeção, e o lote será aceito se não forem encontradas peças com defeito na amostra.
 - a. Qual deverá ser o tamanho de K para garantir que a probabilidade de o fabricante aceitar um lote inaceitável seja menor que 0,10?
 - b. Suponha que o fabricante decida aceitar o lote se houver, no máximo, uma peça com defeito na amostra. Qual deverá ser o tamanho de K para assegurar que a probabilidade de o fabricante aceitar um lote inaceitável seja menor que 0,10?
- 3.3 O fluxo do tráfego em determinadas esquinas, algumas vezes, pode ser modelado como uma sequência de provas de Bernoulli, assumindo que a probabilidade de um carro passar durante qualquer segundo específico é uma constante p e que não existe interação entre a passagem de carros em diferentes segundos. Se tratarmos os segundos como unidades de tempo indivisíveis (provas), o modelo de Bernoulli se aplica. Suponha que um pedestre possa atravessar a rua somente se nenhum carro passar nos 3 segundos seguintes. Descubra a probabilidade de que o pedestre tenha de esperar por exatamente 4 segundos antes de começar a atravessar a rua.
- 3.4 Um homem que tem n chaves quer abrir sua porta e experimenta as chaves aleatoriamente. Apenas uma chave abrirá a porta. Descubra o número médio de tentativas, se
 - a. as chaves erradas não forem excluídas das escolhas seguintes.
 - b. as chaves erradas forem excluídas.
- 3.5 Sabe-se que um medicamento padrão tem efeito em 80% dos casos em que é utilizado. Um novo medicamento é testado em 100 pacientes e descobre-se que ele foi efetivo em 85 casos. O novo medicamento é superior? (*Dica:* avalie a probabilidade de observar 85 ou mais casos de sucesso, assumindo que o medicamento novo e o antigo são igualmente efetivos.)

- 3.6 Espera-se que um grande número de insetos seja atraído para uma determinada variedade de rosas. Um inseticida comercial é anunciado como sendo 99% efetivo. Suponha que 2 mil insetos infestem um jardim de rosas onde o inseticida foi aplicado e seja $X =$ o número de insetos que sobrevivem ao inseticida.
- Que distribuição de probabilidade pode fornecer um modelo razoável para este experimento?
 - Escreva, mas não avalie, uma expressão para a probabilidade de que menos de 100 insetos sobreviva, utilizando o modelo na parte (a).
 - Desenvolva uma aproximação para probabilidade, na parte (b).
- 3.7 Considere que o número de gotas de chocolate em um determinado tipo de biscoito tem uma distribuição de Poisson. Queremos que a probabilidade de que um biscoito escolhido aleatoriamente tenha pelo menos duas gotas de chocolate seja maior que 0,99. Encontre o menor valor da média da distribuição que assegura esta probabilidade.
- 3.8 Dois cinemas competem para conseguir um público de mil espectadores. Assuma que cada espectador escolhe entre os cinemas independentemente e com "indiferença". Seja N denotando o número de poltronas em cada cinema.
- Utilizando um modelo binomial, encontre um expressão para N que possa garantir que a probabilidade de perder um espectador (porque a sala está cheia) é menor que 1%.
 - Utilize a aproximação normal para obter um valor numérico para N .
- 3.9 Frequentemente, histórias que são relatadas como surpreendentes coincidências, que acontecem "uma vez em um milhão", são na verdade, sob um exame mais detalhado, eventos não raros e que até mesmo se espera que aconteçam.
- Há alguns anos, uma escola primária, no estado de Nova York, relatou que sua classe de alunos do jardim de infância tinha cinco duplas de alunos gêmeos. Isso, naturalmente, foi divulgado para todo o Estado, com uma observação feita pelo diretor da escola de que esta era uma "impossibilidade estatística".
- Esta observação é verdadeira? Ou seria este um exemplo do que Diaconis e Mosteller (1989) chamaram de a "lei dos números verdadeiramente grandes"? Vamos fazer alguns cálculos.
- A probabilidade de um nascimento de gêmeos é de aproximadamente $1/90$, e podemos assumir que uma escola primária terá aproximadamente 60 crianças entrando para o jardim da infância (três classes com 20 alunos cada uma). Explique como nosso evento "estatisticamente impossível" pode ser considerado como a probabilidade de 5 ou mais sucessos a partir de uma binomial($60, 1/90$). Esta probabilidade é rara o suficiente para ser considerada uma notícia?
 - Mesmo se a probabilidade na parte (a) for suficientemente rara para ser considerada uma notícia, é preciso levar em conta que isso poderia ter acontecido em qualquer escola da cidade e em qualquer cidade do estado, e, ainda assim, teria sido relatada exatamente da mesma maneira. (A "lei dos números verdadeiramente grandes" está começando a entrar em ação.) O estado de Nova York tem 62 municípios, e é razoável assumir que cada município tenha cinco escolas primárias. O evento ainda pode ser qualificado como uma "impossibilidade estatística" ou está se tornando algo que se pode esperar que ocorra?
 - Se a probabilidade na parte (b) ainda parece ser pequena, considere também que este evento poderia ter acontecido em qualquer um dos 50 estados, durante qualquer um dos últimos 10 anos e, ainda assim, teria recebido o mesmo destaque da mídia.
- Além de Diaconis e Mosteller (1989), veja Hanley (1992) para conhecer mais sobre coincidências.
- 3.10 Shuster (1991) descreve uma série de cálculos de probabilidade que fez para um caso judicial envolvendo a venda de cocaína. Um departamento de polícia da Flórida apreendeu 496 *supostos* pacotes de cocaína, dos quais quatro foram selecionados aleatoriamente e testados, e ficou comprovado que realmente se tratava de cocaína. A polícia então escolheu mais dois pacotes aleatoriamente, e um policial fingindo ser traficante de drogas vendeu os pacotes para o réu. Estes dois últimos pacotes foram perdidos antes que pudessem ser testados para verificar se realmente continham cocaína.

- a. Se os 496 pacotes originais eram compostos de N pacotes de cocaína e $M = 496 - N$ de uma substância que não era cocaína, mostre que a probabilidade de selecionar 4 pacotes de cocaína e então 2 com uma substância diferente da cocaína, que é a probabilidade de que o réu seja inocente em relação à compra da cocaína, é

$$\frac{\binom{N}{4} \binom{M}{2}}{\binom{N+M}{4} \binom{N+M-4}{2}}$$

- b. Maximizar (em M e N) a probabilidade na parte (a) aumenta a "probabilidade de inocência" do réu. Mostre que esta probabilidade é de 0,022, e é atingida em $M = 165$ e $N = 331$.
- 3.11 A distribuição hipergeométrica pode ser aproximada pela binomial ou pela distribuição de Poisson. (Naturalmente, ela pode ser aproximada por outras distribuições, mas neste exercício nos concentraremos somente nestas duas.) Seja X com distribuição hipergeométrica

$$P(X = x|N, M, K) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}, \quad x = 0, 1, \dots, K.$$

- a. Mostre que quando $N \rightarrow \infty, M \rightarrow \infty$ e $M/N \rightarrow p$,

$$P(X = x|N, M, K) \rightarrow \binom{K}{x} p^x (1-p)^{K-x}, \quad x = 0, 1, \dots, K.$$

(A Fórmula de Stirling, do Exercício 1.23, pode ser útil.)

- b. Utilize o fato de que a binomial pode ser aproximada pela de Poisson para mostrar que se $N \rightarrow \infty, M \rightarrow \infty, K \rightarrow \infty, M/N \rightarrow 0$ e $KM/N \rightarrow \lambda$, então

$$P(X = x|N, M, K) \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots$$

- c. Verifique a aproximação na parte (b) diretamente, sem utilizar a aproximação de Poisson para a binomial. (O Lema 2.3.14 é útil.)

- 3.12 Suponha que X tenha uma distribuição binomial(n, p) e que Y tenha uma distribuição binomial negativa(r, p). Mostre que $F_X(r-1) = 1 - F_Y(n-r)$.

- 3.13 Uma distribuição discreta *truncada* é aquela na qual uma determinada classe não pode ser observada e é eliminada do espaço amostral. Em particular, se X tem amplitude $0, 1, 2, \dots$ e a classe 0 não pode ser observada (como geralmente é o caso), a variável aleatória *truncada* no 0 X_T tem fp

$$P(X_T = x) = \frac{P(X = x)}{P(X > 0)}, \quad x = 1, 2, \dots$$

Encontre a fp, a média e a variância da variável aleatória *truncada* no 0, começando a partir de

- a. $X \sim \text{Poisson}(\lambda)$.
 b. $X \sim \text{binomial negativa}(r, p)$, como em (3.2.10).
- 3.14 Começando da binomial negativa truncada no 0 (consulte o Exercício 3.13), se considerarmos que $r \rightarrow 0$, obtemos uma distribuição interessante, a *distribuição da série logarítmica*. Uma variável aleatória X tem uma distribuição de série logarítmica com parâmetro p se

$$P(X = x) = \frac{-(1-p)^x}{x \log p}, \quad x = 1, 2, \dots, \quad 0 < p < 1.$$

- a. Verifique se isto define uma legítima função de probabilidade.
 b. Encontre a média e a variância de X . (A distribuição de série logarítmica mostrou ser útil na modelagem de abundância de espécie. Veja Stuart e Ord, 1987, para uma discussão mais detalhada desta distribuição.)

3.15 Na Seção 3.2 foi afirmado que a distribuição de Poisson(λ) é o limite da distribuição binomial negativa(r, p) considerando que $r \rightarrow \infty$, $p \rightarrow 1$ e $r(1-p) \rightarrow \lambda$. Mostre que sob estas condições a fgm da binomial negativa converge para a de Poisson.

3.16 Verifique essas duas identidades em relação à função gama, que foram dadas no livro:

a. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

b. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

3.17 Estabeleça uma fórmula similar a (3.3.18) para a distribuição gama. Se $X \sim \text{gama}(\alpha, \beta)$, então, para qualquer constante positiva v ,

$$EX^v = \frac{\beta^v \Gamma(v + \alpha)}{\Gamma(\alpha)}.$$

3.18 Existe uma interessante relação entre a binomial negativa e as variáveis aleatórias gama, o que, às vezes, pode oferecer uma aproximação útil. Seja Y uma variável aleatória binomial negativa com parâmetros r e p , onde p é a probabilidade de sucesso. Mostre que como $p \rightarrow 0$, a fgm da variável aleatória pY converge para a de uma distribuição gama com parâmetros r e 1.

3.19 Mostre que

$$\int_x^\infty \frac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-z} dz = \sum_{y=0}^{\alpha-1} \frac{x^y e^{-x}}{y!}, \quad \alpha = 1, 2, 3, \dots$$

(Dica: utilize a integração por partes.) Expresse esta fórmula como uma relação probabilística entre as variáveis aleatórias de Poisson e gama.

3.20 Seja a variável aleatória X com fdp

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty.$$

a. Encontre a média e a variância de X . (Esta distribuição às vezes é chamada de *normal dobrada*.)

b. Se X tem a distribuição normal dobrada, encontre a transformação $g(X) = Y$ e os valores de α e β de modo que $Y \sim \text{gama}(\alpha, \beta)$.

3.21 Escreva a integral que definiria a fgm da fdp

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

A integral é finita? (Você espera que ela seja?)

3.22 Para cada uma das seguintes distribuições, verifique as fórmulas para EX e $\text{Var } X$ dadas no livro.

a. Verifique $\text{Var } X$ se X tem uma distribuição de Poisson(λ). (Dica: Calcule $EX(X-1) = EX^2 - EX$.)

b. Verifique $\text{Var } X$ se X tem uma distribuição binomial negativa(r, p).

c. Verifique $\text{Var } X$ se X tem uma distribuição gama(α, β).

d. Verifique EX e $\text{Var } X$ se X tem uma distribuição beta(α, β).

e. Verifique EX e $\text{Var } X$ se X tem uma distribuição exponencial dupla(μ, σ).

3.23 A distribuição de Pareto, com parâmetros α e β , tem fdp

$$f(x) = \frac{\beta \alpha^\beta}{x^{\beta+1}}, \quad \alpha < x < \infty, \quad \alpha > 0, \quad \beta > 0.$$

a. Verifique se $f(x)$ é uma fdp.

b. Derive a média e a variância desta distribuição.

c. Prove que a variância não existe se $\beta \leq 2$.

3.24 Muitas distribuições "nomeadas" são casos especiais das distribuições mais comuns já discutidas. Para cada uma das seguintes distribuições nomeadas derive a forma da fdp, verifique se esta é uma fdp e calcule a média e a variância.

a. Se $X \sim \text{exponencial}(\beta)$, então $Y = X^{1/\gamma}$ tem a distribuição de Weibull(γ, β), onde $\gamma > 0$ é uma constante.

- b. Se $X \sim \text{exponencial}(\beta)$, então $Y = (2X/\beta)^{1/2}$ tem a *distribuição de Rayleigh*.
 c. Se $X \sim \text{gama}(a, b)$, então $Y = 1/X$ tem a *distribuição G(gama invertida)(a, b)*. (Esta distribuição é útil na estimação Bayesiana de variâncias; veja o Exercício 7.23.)
 d. Se $X \sim \text{gama}(\frac{3}{2}, \beta)$, então $Y = (X/\beta)^{1/2}$ tem a *distribuição de Maxwell*.
 e. Se $X \sim \text{exponencial}(1)$, então $Y = \alpha - \gamma \log X$ tem a *distribuição de Gumbel*(α, γ), onde $-\infty < \alpha < \infty$ e $\gamma > 0$. (A distribuição de Gumbel também é conhecida como *distribuição do valor extremo*.)

3.25 Suponha que a variável aleatória T seja a duração da vida útil de um objeto (possivelmente, a vida útil de um componente elétrico ou de um determinado indivíduo recebendo um tratamento em particular). A *função risco* $h_T(t)$ associada com a variável aleatória T é definida por

$$h_T(t) = \lim_{\delta \rightarrow 0} \frac{P(t \leq T < t + \delta | T \geq t)}{\delta}.$$

Portanto, podemos interpretar $h_T(t)$ como o índice de mudança da probabilidade de que o objeto sobreviva um pouco mais que o tempo t , considerando que o objeto sobreviva ao tempo t . Mostre que se T é uma variável aleatória contínua, então

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{d}{dt} \log(1 - F_T(t)).$$

3.26 Verifique se as seguintes fdps têm as funções risco indicadas (veja o Exercício 3.25).

- a. Se $T \sim \text{exponencial}(\beta)$, então $h_T(t) = 1/\beta$.
 b. Se $T \sim \text{Weibull}(\gamma, \beta)$, então $h_T(t) = (\gamma/\beta)t^{\gamma-1}$.
 c. Se $T \sim \text{logística}(\mu, \beta)$, isto é,

$$F_T(t) = \frac{1}{1 + e^{-(t-\mu)/\beta}},$$

então $h_T(t) = (1/\beta)F_T(t)$.

3.27 Para cada uma das seguintes famílias, mostre se todas as fdps na família são unimodais (veja o Exercício 2.27).

- a. uniforme(a, b)
 b. gama(α, β)
 c. $n(\mu, \sigma^2)$
 d. beta(α, β)

3.28 Mostre que cada uma das seguintes famílias é uma família exponencial.

- a. família normal com o parâmetro μ ou σ conhecido
 b. família gama com o parâmetro α ou β conhecido, ou ambos desconhecidos
 c. família beta, com o parâmetro α ou β conhecido, ou ambos desconhecidos
 d. família de Poisson
 e. família binomial negativa com r conhecido, $0 < p < 1$

3.29 Para cada família no Exercício 3.28, descreva o espaço paramétrico natural.

3.30 Utilize as identidades do Teorema 3.4.2 para

- a. calcular a variância de uma variável aleatória binomial.
 b. calcular a média e a variância de uma variável aleatória Poisson(λ).

3.31 Neste exercício, provaremos o Teorema 3.4.2.

- a. Comece a partir da igualdade

$$\int h(x)c(\theta) \exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right) dx = 1,$$

diferencie ambos os lados e, então, rearranje os termos para estabelecer (3.4.4). (O fato de que $\frac{d}{dx} \log g(x) = g'(x)/g(x)$ será útil.)

- b. Diferencie a igualdade acima uma segunda vez; então, rearranje para estabelecer (3.4.5). (O fato de que $\frac{d^2}{dx^2} \log g(x) = (g''(x)/g(x)) - (g'(x)/g(x))^2$ será útil.)

- 3.32 a. Se uma família exponencial pode ser escrita na forma (3.4.7), mostre que as identidades do Teorema 3.4.2 simplificam para

$$E(t_j(X)) = -\frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta}),$$

$$\text{Var}(t_j(X)) = -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\boldsymbol{\eta}).$$

- b. Utilize esta identidade para calcular a média e a variância de uma variável aleatória gama(a, b).

- 3.33 Para cada uma das seguintes famílias:

- i. Verifique se esta é uma família exponencial.
- ii. Descreva a curva na qual está o vetor do $\boldsymbol{\theta}$ parâmetro.
- iii. Esboce um gráfico do espaço paramétrico curvo.
 - a. $n(\boldsymbol{\theta}, \theta)$
 - b. $n(\theta, a\theta^2)$, a conhecido
 - c. gama($\alpha, 1/\alpha$)
 - d. $f(x|\theta) = C \exp\{-(x-\theta)^4\}$, C uma constante de normalização

- 3.34 No Exemplo 3.4.9 vimos que as aproximações normais podem resultar em famílias exponenciais curvas. Para cada uma das seguintes aproximações normais:

- i. Descreva a curva na qual está o parâmetro $\boldsymbol{\theta}$.
- ii. Esboce um gráfico do espaço paramétrico curvo.
 - a. Aproximação de Poisson: $\tilde{X} \sim n(\lambda, \lambda/n)$
 - b. Aproximação binomial: $\tilde{X} \sim n(p, p(1-p)/n)$
 - c. Aproximação binomial negativa: $\tilde{X} \sim n(r(1-p)/p, r(1-p)/np^2)$

- 3.35 a. A família normal que se aproxima de uma Poisson também pode ser parametrizada como $n(e^\theta, e^\theta)$, onde $-\infty < \theta < \infty$. Esboce um gráfico do espaço paramétrico e compare com a aproximação no Exercício 3.34(a).

- b. Suponha que $X \sim \text{gama}(\alpha, \beta)$ e assumo que $EX = \mu$. Esboce um gráfico do espaço paramétrico.

- c. Suponha que $X_i \sim \text{gama}(\alpha_i, \beta_i)$, $i = 1, 2, \dots, n$, e assumo que $EX_i = \mu$. Descreva o espaço paramétrico $(\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n)$.

- 3.36 Considere a fdp $f(x) = \frac{63}{4}(x^6 - x^8)$, $-1 < x < 1$. Represente o gráfico $(1/\sigma)f((x-\mu)/\sigma)$ para cada um dos seguintes itens nos mesmos eixos.

- a. $\mu = 0, \sigma = 1$
- b. $\mu = 3, \sigma = 1$
- c. $\mu = 3, \sigma = 2$

- 3.37 Mostre que se $f(x)$ é uma fdp simétrica em torno de 0, então μ é a mediana da fdp de localização-escala $(1/\sigma)f((x-\mu)/\sigma)$, $-\infty < x < \infty$.

- 3.38 Seja Z uma variável aleatória com fdp $f(z)$. Defina que z_α seja um número que satisfaz esta relação:

$$\alpha = P(Z > z_\alpha) = \int_{z_\alpha}^{\infty} f(z) dz.$$

Mostre que se X é uma variável aleatória com fdp $(1/\sigma)f((x-\mu)/\sigma)$ e $x_\alpha = \sigma z_\alpha + \mu$, então $P(X > x_\alpha) = \alpha$. (Portanto, se uma tabela de z_α valores estivesse disponível, então os valores de x_α poderiam facilmente ser calculados para qualquer membro da família de localização-escala.)

- 3.39 Considere a família de Cauchy, definida na Seção 3.3. Esta família pode ser ampliada para uma família de localização-escala, gerando fdps da forma

$$f(x|\mu, \sigma) = \frac{1}{\sigma\pi \left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}, \quad -\infty < x < \infty.$$

A média e a variância não existem para a distribuição de Cauchy. Assim, os parâmetros μ e σ^2 não são a média e a variância. Mas elas têm significado importante. Mostre que se X é uma variável aleatória com uma distribuição de Cauchy com parâmetros μ e σ , então:

- μ é a mediana da distribuição de X , isto é, $P(X \geq \mu) = P(X \leq \mu) = \frac{1}{2}$.
- $\mu + \sigma$ e $\mu - \sigma$ são os quartis da distribuição de X , isto é, $P(X \geq \mu + \sigma) = P(X \leq \mu - \sigma) = \frac{1}{4}$.
(Dica: prove isto primeiro para $\mu = 0$ e $\sigma = 1$ e então utilize o Exercício 3.38.)

- 3.40 Seja $f(x)$ qualquer fdp com média μ e variância σ^2 . Mostre como criar uma família de locação-escala com base em $f(x)$ de modo que a fdp padrão da família, digamos $f^*(x)$, tenha média θ e variância 1.
- 3.41 Uma família de fdas $\{F(x|\theta), \theta \in \Theta\}$ é *estocasticamente crescente em θ* se $\theta_1 > \theta_2 \Rightarrow F(x|\theta_1)$ for estocasticamente maior que $F(x|\theta_2)$. (Veja o Exercício 1.49 para ver a definição de estocasticamente maior.)
- Mostre que a família $n(\mu, \sigma^2)$ é estocasticamente crescente em μ para σ^2 fixo.
 - Mostre que a família $\text{gama}(\alpha, \beta)$ de (3.3.6) é estocasticamente crescente em β (parâmetro de escala) para α fixo (parâmetro de formato).
- 3.42 Consulte o Exercício 3.41 para a definição de uma família estocasticamente crescente.
- Mostre que uma família de locação é estocasticamente crescente em seu parâmetro de locação.
 - Mostre que uma família de escala é estocasticamente crescente em seu parâmetro de escala se o espaço amostral for $(0, \infty)$.
- 3.43 Uma família de fdas $\{F(x|\theta), \theta \in \Theta\}$ é *estocasticamente decrescente em θ* se $\theta_1 > \theta_2 \Rightarrow F(x|\theta_2)$ for estocasticamente maior que $F(x|\theta_1)$. (Veja os Exercícios 3.41 e 3.42.)
- Prove que se $X \sim F_X(x|\theta)$, onde o espaço amostral de X é $(0, \infty)$ e $F_X(x|\theta)$ é estocasticamente crescente em θ , então $F_Y(y|\theta)$ é estocasticamente decrescente em θ , onde $Y = 1/X$.
 - Prove que se $X \sim F_X(x|\theta)$, onde $F_X(x|\theta)$ é estocasticamente crescente em θ e $\theta > 0$, então $F_X(x|\frac{1}{\theta})$ é estocasticamente decrescente em θ .
- 3.44 Para qualquer variável aleatória X para a qual EX^2 e $E|X|$ existem, mostre que $P(|X| \geq b)$ não excede nem EX^2/b^2 nem $E|X|/b$, onde b é uma constante positiva. Se $f(x) = e^{-x}$ para $x > 0$, mostre que um limite é melhor quando $b = 3$, e o outro, quando $b = \sqrt{2}$. (Note a Desigualdade de Markov, em Itens Diversos 3.8.2.)
- 3.45 Seja X uma variável aleatória com função de geração de momentos $M_X(t)$, $-h < t < h$.
- Prove que $P(X \geq a) \leq e^{-at} M_X(t)$, $0 < t < h$. (Uma prova similar àquela utilizada para a Desigualdade de Chebychev irá funcionar.)
 - De modo similar, prove que $P(X \leq a) \leq e^{-at} M_X(t)$, $-h < t < 0$.
 - Um caso especial da parte (a) é que $P(X \geq 0) \leq Ee^{tX}$ para todo $t \geq 0$ para o qual a fgm é definida. Quais são as condições gerais de uma função $h(t, x)$ de modo que $P(X \geq 0) \leq Eh(t, X)$ para todo $t \geq 0$ para o qual $Eh(t, X)$ existe? (Na parte (a), $h(t, x) = e^{tx}$.)
- 3.46 Calcule $P(|X - \mu_X| \geq k\sigma_X)$ para $X \sim \text{uniforme}(0, 1)$ e $X \sim \text{exponencial}(\lambda)$, e compare suas respostas com o limite da Desigualdade de Chebychev.
- 3.47 Se Z é uma variável aleatória normal padrão, prove este complemento da desigualdade no Exemplo 3.6.3:

$$P(|Z| \geq t) \geq \sqrt{\frac{2}{\pi}} \frac{t}{1+t^2} e^{-t^2/2}.$$

- 3.48 Derive as relações de recursão, similares àsquelas dadas em (3.6.2), para as distribuições binomial, binomial negativa e hipergeométrica.
- 3.49 Prove os seguintes análogos do Lema de Stein, assumindo condições apropriadas na função g .
- Se $X \sim \text{gama}(\alpha, \beta)$, então

$$E(g(X)(X - \alpha\beta)) = \beta E(Xg'(X)).$$

- Se $X \sim \text{beta}(\alpha, \beta)$, então

$$E\left[g(X)\left(\beta - (\alpha - 1)\frac{(1-X)}{X}\right)\right] = E((1-X)g'(X)).$$

- 3.50 Prove a identidade para a distribuição binomial negativa dada no Teorema 3.6.8, parte (b).

3.8 ITENS DIVERSOS

3.8.1 OS POSTULADOS DE POISSON

A distribuição de Poisson pode ser derivada a partir de um conjunto de suposições básicas, algumas vezes chamadas de postulados de Poisson. Estas suposições estão relacionadas às propriedades físicas do processo em consideração. Embora, genericamente falando, as suposições não sejam muito fáceis de serem verificadas, elas fornecem ao experimentador um conjunto de diretrizes para avaliar se os postulados de Poisson fornecerão um modelo razoável. Para uma abordagem mais completa dos postulados de Poisson, veja o clássico livro de Feller (1968) ou Barr e Zehna (1983).

TEOREMA 3.8.1

Para cada $t \geq 0$, seja N_t uma variável aleatória com valores inteiros, com as seguintes propriedades. (Pense em N_t como denotando o número de chegadas no período de tempo de 0 a t .)

- i. $N_0 = 0$ (inicia sem nenhuma chegada)
- ii. $s < t \Rightarrow N_s$ e $N_t - N_s$ são independentes. (chegadas em períodos de tempo disjuntos são independentes)
- iii. N_s e $N_{t+s} - N_t$ são identicamente distribuídos. (o número de chegadas depende somente do comprimento do período)
- iv. $\lim_{t \rightarrow 0} \frac{P(N_t = 1)}{t} = \lambda$ (a probabilidade de chegada é proporcional ao comprimento do período se o comprimento do período for pequeno)
- v. $\lim_{t \rightarrow 0} \frac{P(N_t > 1)}{t} = 0$ (nenhuma chegada simultânea)

Se i-v se mantiver, então, para qualquer número inteiro n ,

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

isto é, $N_t \sim \text{Poisson}(\lambda t)$.

Os postulados também podem ser interpretados como descrição do comportamento de objetos espacialmente (por exemplo, o movimento de insetos), dando a aplicação de Poisson em distribuições espaciais.

3.8.2 ALÉM DE CHEBYCHEV

Ghosh e Meeden (1977) discutem o fato de que a Desigualdade de Chebychev é muito conservadora, e quase nunca é atingida. Se escrevermos \bar{X}_n para a média das variáveis aleatórias X_1, X_2, \dots, X_n , então a Desigualdade de Chebychev define que

$$P(|\bar{X}_n - \mu| \geq k\sigma) \leq \frac{1}{nk^2}.$$

Elas provam o seguinte teorema.

TEOREMA 3.8.2

Se $0 < \sigma < \infty$, então

- a. Se $n = 1$, a desigualdade é atingível para $k \geq 1$ e inatingível para $0 < k < 1$.
- b. Se $n = 2$, a desigualdade é atingível se, e somente se, $k = 1$.
- c. Se $n \geq 3$, a desigualdade não é atingível.

São dados exemplos para os casos em que a desigualdade é atingida. A maior parte de seus argumentos técnicos é baseada na seguinte desigualdade, conhecida como Desigualdade de Markov.

LEMA 3.8.3**Desigualdade de Markov**

Se $P(Y \geq 0) = 1$ e $P(Y = 0) < 1$, então, para qualquer $r > 0$,

$$P(Y \geq r) \leq \frac{EY}{r}$$

com igualdade se, e somente se, $P(Y = r) = p = 1 - P(Y = 0)$, $0 < p \leq 1$.

A Desigualdade de Markov pode então ser aplicada à quantidade

$$Y = \frac{(\bar{X}_n - \mu)^2}{\sigma^2}$$

para obter os resultados acima.

Uma razão por que a Desigualdade de Chebychev é tão vaga é que ela não impõe restrições à distribuição básica de Y . Com a restrição adicional da *unimodalidade*, podemos obter limites mais estritos, e as desigualdades de Gauss e Vysochanskii-Petunin. (Veja Pukelsheim, 1994, para ver detalhes e provas dessas desigualdades, baseadas em cálculo elementar.)

TEOREMA 3.8.4**Desigualdade de Gauss**

Seja $X \sim f$, onde f é unimodal com moda ν , e defina $\tau^2 = E(X - \nu)^2$. Então

$$P(|X - \nu| > \varepsilon) \leq \begin{cases} \frac{4\tau^2}{9\varepsilon^2} & \text{para todo } \varepsilon \geq \sqrt{4/3}\tau \\ 1 - \frac{\varepsilon}{\sqrt{3}\tau} & \text{para todo } \varepsilon \leq \sqrt{4/3}\tau. \end{cases}$$

Embora este seja um limite mais estrito do que o de Chebychev, a dependência da moda limita sua utilidade. A extensão de Vysochanskii-Petunin remove esta limitação.

TEOREMA 3.8.5**Desigualdade de Vysochanskii-Petunin**

Seja $X \sim f$, onde f é unimodal, e defina $\xi^2 = E(X - \alpha)^2$ para um ponto arbitrário α . Então

$$P(|X - \alpha| > \varepsilon) \leq \begin{cases} \frac{4\xi^2}{9\varepsilon^2} & \text{para todo } \varepsilon \geq \sqrt{8/3}\xi \\ \frac{4\xi^2}{9\varepsilon^2} - \frac{1}{3} & \text{para todo } \varepsilon \leq \sqrt{8/3}\xi. \end{cases}$$

Pukelsheim indica que assumir $\alpha = \mu = E(X)$ e $\varepsilon = 3\sigma$, onde $\sigma^2 = \text{Var}(X)$, gera

$$P(|X - \mu| > 3\sigma) \leq \frac{4}{81} < 0,05,$$

a chamada *regra dos três sigmas*, e que a probabilidade é menor que 5% de que X esteja a mais do que três desvios padrão da média da população.

3.8.3 MAIS SOBRE FAMÍLIAS EXPONENCIAIS

A distribuição lognormal está na família exponencial? A densidade dada em (3.3.21) pode ser colocada na forma especificada por (3.4.1). Assim, colocamos a lognormal na família exponencial.

De acordo com Brown (1986, Seção 1.1), para definir uma família exponencial de distribuições, começamos com uma função não negativa $v(x)$ e definimos o conjunto \mathcal{N} por

$$\mathcal{N} = \left\{ \theta : \int_{\mathcal{X}} e^{\theta x} v(x) dx < \infty \right\}.$$

Se assumirmos que $\lambda(\theta) = \int_{\mathcal{X}} e^{\theta x} v(x) dx$, o conjunto de densidades de probabilidade definido por

$$f(x|\theta) = \frac{e^{\theta x} v(x)}{\lambda(\theta)}, \quad x \in \mathcal{X}, \quad \theta \in \mathcal{N},$$

é uma *família exponencial*. A função geradora de momentos de $f(x|\theta)$ é

$$M_X(t) = \int_{\mathcal{X}} e^{tx} f(x|\theta) dx = \frac{\lambda(t+\theta)}{\lambda(\theta)}$$

e, assim, existe por construção. Se o espaço paramétrico Θ for igual ao conjunto \mathcal{N} , a família exponencial é chamada *completa*. Casos onde Θ é um conjunto dimensional inferior de \mathcal{N} dão origem a famílias exponenciais curvas.

Retornando à distribuição lognormal, sabemos que ela não tem uma fgm; deste modo, ela não pode satisfazer a definição de Brown de uma família exponencial. Contudo, a lognormal satisfaz as identidades de expectâncias do Teorema 3.4.2 e tira proveito das propriedades de suficiência detalhadas na Seção 6.2.1 (Teorema 6.2.10). Para nossos propósitos, estas são as principais propriedades de que precisamos e as principais razões para identificar um membro da família exponencial. Propriedades mais avançadas, que não investigaremos aqui, podem precisar da existência da fgm.