vermelhas (e três brancas) em cada uma das seguintes condições:

(a) A urna tem quatro vermelhas e seis brancas e as retiradas são feitas sem reposição;

(b) A urna tem oito vermelhas e doze brancas e as retiradas são feitas sem reposição;

(c) A urna tem dezesseis vermelhas e 24 brancas e as retiradas são feitas sem reposição;

(d) A urna tem quatro vermelhas e seis brancas e as retiradas são feitas com reposição;

(e) A urna tem oito vermelhas e doze brancas e as retiradas são feitas com reposição; e

(f) A urna tem dezesseis vermelhas e 24 brancas e as retiradas são feitas com reposição.

Interprete os resultados obtidos em (a)-(f) em vista da possível aproximação de uma v.a. hipergeométrica por uma v.a. binomial quando o número de Experimentos de Bernoulli é suficientemente grande. Justifique essa aproximação com uma interpretação coerente dos Experimentos de Bernoulli, dependência das retiradas sem reposição quando m/N *for pequeno.*

Exercício 4.14 *Entre as quatorze e dezessete horas, em dias úteis, passam por um pedágio noventa carros por hora (em média). Calcule (para dias úteis):*

(a) Probabilidade de que passem cinco carros entre 15:06 e 15:08;

(b) Esperança e variância do número de carros entre 15:06 e 15:08;

(c) Probabilidade de que passem cinco carros entre 16:00 e 16:02;

(d) Esperança e variância do número de carros entre 16:00 e 16:02;

(e) Probabilidade de que passem no máximo quatro carros entre 14:10 e 14:15;

e

(f) Probabilidade de que passem mais do que oito carros entre 14:14 e 14:18.

O que se pode dizer sobre a probabilidade de que passem cinco carros entre 5:06 e 5:08, baseado apenas no enunciado? Justifique.

Exercício 4.15 *Bactérias de certa espécie aparecem na água à razão de 0,8 por* cm^3*. Calcule a probabilidade de que, em cinco* cm^3 *de água, tenhamos:*

(a) nenhuma bactéria;

(b) pelo menos duas bactérias; e

(c) pelo menos treze bactérias.

Exercício 4.16 *A taxa semanal de suicídios em um certo país escandinavo é de um para cada* 250000 *habitantes. Considere uma cidade de* 500000 *habitantes e calcule a probabilidade de que ocorram seis ou mais suicídios em uma determinada semana. Indique qual seria a probabilidade de que, no inverno, haveria tal evento. Você utilizar-se-ia de um modelo deste tipo para estudo de sarampo ou paralisia infantil? Justifique.*

Exercício 4.17 *Compare cada uma das seguintes distribuições:* $Bin(10; 0,05)$, $Bin(10; 0,10)$, $Bin(10; 0,20)$, $Bin(20; 0,05)$, $Bin(20; 0,10)$, $Bin(20; 0,20)$, $Bin(25; 0,10)$ *e* $Bin(25; 0,20)$ *respectivamente com as distribuições de Poisson de médias iguais. Interprete os resultados em face do que foi visto no texto.*

Exercício 4.18 *Ache a(s) moda(s) de uma v.a. distribuída como Poisson, sem consultas ao texto.*

Exercício 4.19
(a) Demonstre a proposição 4.8.3;
(b) Proponha uma fórmula para calcular o k-ésimo momento a partir da fgp; e

(c) Analise por que a fgp e a fgm nos dão ambas os momentos de ordem k. Tente encontrar uma relação analítica entre $M_X^{(k)}(\cdot)$ e $\phi_X^{(k)}(\cdot)$ que justifique essa coincidência.

Dica: *Use argumentos análogos aos utilizados na demonstração da proposição 4.8.2.*

Capítulo 5

Principais Modelos Contínuos

5.1 Introdução

Ao contrário dos modelos discretos, não existe, ao nível do curso proposto, a possibilidade de um tratamento geral das variáveis aleatórias contínuas. Como elemento comum deste capítulo, podemos apenas destacar a distribuição normal. Esta distribuição que, via Teorema Central do Limite[1], pode ser considerada a responsável pela existência da Estatística como um ramo independente da metodologia científica.

Esta caracterização, no entanto, não diz respeito, somente, às variáveis aletórias contínuas e sim a uma classe geral de variáveis aleatórias, definida não pelo tipo (discreta, contínua, mista etc.) mas por suas características de momentos e caudas.

Como conseqüência destas observações, este capítulo será provavelmente de leitura menos agradável do que o anterior, sendo suas seções conectadas de forma apenas secundária.

5.2 Distribuição Uniforme

Em analogia à definição clássica de probabilidade (de uma forma ingênua, como deve ser quase toda analogia entre discreto e contínuo), a primeira distribuição contínua que se pode imaginar é aquela que dá *probabilidades iguais* a eventos de *mesma natureza*.

No caso discreto, a noção de *mesma natureza* é de fácil acesso pois faz-se equivalente à própria definição de **eventos equiprováveis**. No caso

[1] a ser apresentado e discutido em detalhes no Capítulo 6

contínuo, essa equivalência não é tão trivial. Uma forma de *estendê-la* para o caso contínuo é optar por outorgar probabilidades para os intervalos proporcionalmente ao seus respectivos comprimentos:

$$\mathbb{P}(X \in I) \propto i_2 - i_1,$$

onde $I = (i_1, i_2)$.

A primeira observação a ser feita sobre tal variável X é a de que intervalos abertos, fechados ou semi-abertos devem ter probabilidades iguais desde que tenham comprimentos iguais. Uma outra observação, importantíssima, é a de que o suporte de uma tal variável aleatória tem que ser necessariamente finito: em geral, será um intervalo em $\mathbb{R}$. Mais tarde, será visto que, sem perda de generalidade, tal variável pode ser definida com suporte em $(0, 1)$.

Definição 5.2.1 (Distribuição Uniforme-padrão)

Seja X uma variável aleatória cuja função de densidade de probabilidade é dada por:

$$f_X(x) = \mathbf{1}_{(0,1)}(x).$$

X é dita variável aleatória uniforme-padrão, com alcunha uniforme em (0,1) e símbolo $X \sim U(0, 1)$.

Note que:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(u)du = \int_0^x du = x, \end{aligned}$$

se $0 \leq x \leq 1$.

Esta forma da função de distribuição acumulada nos leva a uma generalização natural da v.a. uniforme, dada pela definição 5.2.2

Definição 5.2.2 (Distribuição Uniforme)

Seja Y uma variável aleatória cuja função de distribuição acumulada:

$$F_Y(y) = \frac{y-a}{b-a}\mathbf{1}_{(a,b)}(y),$$

onde $a < b$, $a, b \in \mathbb{R}$. Y é dita variável aleatória uniforme em (a, b), com símbolo $Y \sim U(a, b)$.

Note-se que, se $a \leq y \leq b$, a probabilidade de $Y \in (-\infty, y)$ se resume à probabilidade de $Y \in (a, y)$, sendo proporcional ao comprimento, $y - a$, desse intervalo.

O termo **uniforme-padrão** se justifica pela proposição 5.2.1.

Proposição 5.2.1 (Transformação Linear da Uniforme)

Seja X uma variável aleatória $U(0, 1)$. Defina uma outra v.a. $Y = (b - a)X + a$, onde $a < b$, $a, b \in \mathbb{R}$. Então, Y também tem distribuição uniforme, no intervalo (a, b).

A demonstração deste fato é bem simples, sendo deixada ao leitor (ver exercício 5.1).

O mais importante na proposição 5.2.1 é o fato de que uma transformação linear em uma v.a. uniforme mantém a característica principal (uniformidade) da distribuição, isto é, podemos dizer que a *classe* de v.a.'s uniformes é fechada com relação a transformações lineares.

Uma outra característica importantíssima da variável aleatória uniforme que, por si só, justifica seu estudo é a geração de valores pseudo-aleatórios.

A simulação tem sido, desde o advento dos computadores digitais, uma importante ferramenta de pesquisa: faz-se fundamental em problemas inferenciais complexos ou em análise de sistemas avançados. Não é, nem pretende ser, substituto para resultados teóricos. Pretende, na realidade, indicar possíveis caminhos para as soluções teóricas quando essas são por demais caras ou inviáveis.

A geração de valores pseudo-aleatórios pode ser resumida, de maneira simples, pelo seguinte exemplo:

Exemplo 5.2.1 (Joaquim Apaixonado)

Suponha que Joaquim proponha à Maria um jogo. Como ele está por ela apaixonado, gostaria de facilitar sua vida e dar-lhe chances amplas de vitória: a cada lançamento, Maria tem $0,75$ *de probabilidade de vitória contra apenas* $0,25$ *de derrota.*

Caso este jogo fosse jogado com moedas equilibradas, seriam necessários dois lançamentos, sendo Joaquim o vencedor com um entre os quatro resultados possíveis e Maria vencedora com os outros três. No entanto, visto que Maria não aceitaria tal artifício, Joaquim cria um jogo no computador, de forma que apareçam J's ou M's a cada rodada, indicando o vencedor.

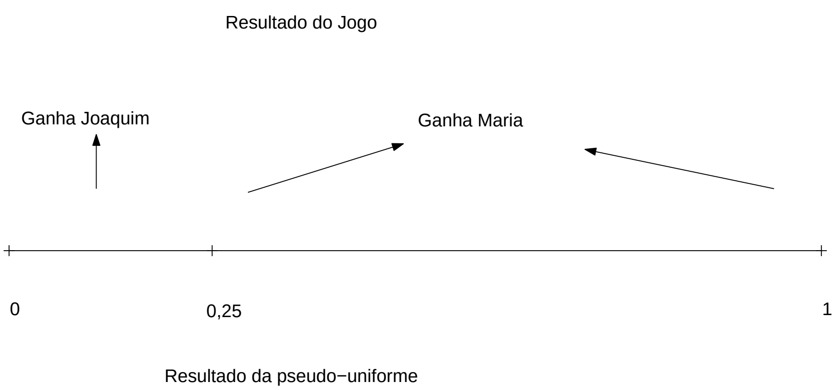
Para tanto, ele precisa programar o computador para gerar números, X*, de acordo com uma v.a. Bernoulli de parâmetro* $0,75$*. Uma maneira de fazê-lo seria gerar uma uniforme-padrão,* U*, e considerar* J *caso esta uniforme seja menor do que* $0,25$ *ou* M*, caso contrário, como ilustrado na figura 5.1.*

Três questões cruciais podem ser levantadas no exemplo 5.2.1:

- O que é um número aleatório?
- Quão fácil é gerar uma $U(0,1)$?
- É possível gerar qualquer v.a. a partir de uma $U(0,1)$?

A primeira das três perguntas, sendo a mais fundamental, é também a mais profunda e merece especial consideração. Uma das conseqüências mais interessantes da noção de probabilidade, no caso a de probabilidade condicional, é que, quando $B \cap A = A$, $\mathbb{P}(B|A) = 1$, isto é, o acréscimo de informação sobre B (ocorrência de A) é suficiente para tornar um evento aleatório em um fato *determinado*.

Figura 5.1: Simulação Primitiva



O caso $B \cap A = A$ é claramente um extremo mas, em todas as situações, com exceção de A e B independentes, a informação de ocorrência muda a probabilidade de B. Em linhas gerais, isto pode ser interpretado como probabilidade sendo uma forma de *ignorância quantificada* sobre o fenômeno (quando adquirimos mais conhecimento, na ocorrência de A, nossa função de probabilidade muda).

Portanto, a verdadeira natureza de uma seqüência ser aleatória se torna irrelevante e uma seqüência intrinsicamente determinística pode ser utilizada como se o fosse, desde que observada a ignorância sobre sua composição futura.

Seqüências que possuam esta características são, portanto, chamadas de pseudo-aleatórias: não são verdadeiramente aleatórias mas, para os propósitos em questão, comportam-se como se o fossem. Esta natureza dúbia traz consigo outra grande qualidade: repetibilidade e reprodutibilidade de um experimento. Pensemos que, se a seqüência utilizada num experimento fosse verdadeiramente aleatória, ser-nos-ia impossível repetí-lo ou a um outro pesquisador, reproduzí-lo.

A segunda questão pode ser respondida de maneira análoga à própria exposição motivadora de uma v.a. uniforme: gerar números de uma distribuição equilibrada é muito mais simples do que fazê-lo diretamente para uma outra particular distribuição. Isso é verdade por que é muito mais fácil controlar um processo homogêneo do que um heterogêneo. O que se afirma aqui não é que seja fácil gerar pseudo-observações de uma uniforme em $(0, 1)$: apenas que é mais fácil fazê-lo do que para a grande maioria das outras distribuições.

A facilidade de se gerarem pseudo-valores unformes seria praticamente inútil, não fosse pelo seguinte resultado fundamental, expresso no teorema 5.2.1.

Teorema 5.2.1 (Teorema da Inversão)

Sejam F uma função de distribuição acumulada qualquer e U uma v.a. com distribuição uniforme em $(0, 1)$. Defina-se F^-, a função inversa generalizada de F, por:

$$F^-(t) = \inf_{x \in \mathbb{R}} \{F(x) > t\},$$

para $t \in [0, 1]$.

Então, a variável aleatória $Y = F^-(U)$ tem distribuição F.

O teorema 5.2.1 é, provavelmente, a mais poderosa ferramenta teórica de simulação, pois, simplesmente, significa que, para qualquer variável aleatória imaginável, existe pelo menos uma maneira de simulá-la. Basta-nos gerar uma uniforme em $(0, 1)$ e aplicar F^-. Obviamente que, na prática, problemas surgirão, pelas dificuldades em inverter F, tanto por inexistência de soluções fechadas como pelo custo computacional de fazê-lo numericamente. Mesmo assim, isso não tira o tremendo potencial de tornar a uniforme em $(0, 1)$ como fundamental para a simulação estocástica.

De posse desse teorema, o problema de João no exemplo 5.2.1 pode ser solucionado facilmente. Basta, para tanto, que ele seja capaz de gerar variáveis unformes (o que quase todo pacote ou linguagem computacional faz) e utilize-se do teorema 5.2.1. A função de distribuição acumulada de quem ganha uma rodada no jogo por João idealizado seria:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0,25 & 0 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$$

Portanto,

$$F^{-}(t) = \begin{cases} 0 & t \leq 0,25 \\ 1 & t > 0,25. \end{cases}$$

Isto nos diz que basta escrever $[0,1] = A_0 \cup A_1$, $A_0 = [0;0,25)$ e $A_1 = [0,25;1]$ e assinalar respectivos valores 0 e 1 caso a observação pseudo-uniforme esteja em A_0 ou A_1. De forma análoga, podemos gerar pseudo-observações, sendo as dificuldades apenas de natureza aplicada.

A prova do teorema 5.2.1 se simplifica bastante se supusermos que F é absolutamente contínua. A demonstração para o caso geral é apenas técnica, não trazendo nova interpretação além da obtida quando F é absolutamente contínua, e será omitida.

Demonstração:

Suponha que F seja absolutamente contínua. Então, F^{-} pode ser escrita como $F^{-}(t) = \{x \in \mathbb{R} \ : \ F(x) = t\}$. Portanto:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(F^{-}(U) \leq y) \\ &= \mathbb{P}(U \leq F(y)) = F(y). \end{aligned}$$

■

5.3 Distribuição Normal

A distribuição normal, como já comentado na seção 5.1, é a mais importante da Estatística. Como será visto mais à frente, dá-se a ela o nome de *normal* por motivos bastantes fortes. Historicamente, foi redescoberta pelo grande matemático alemão, Carl Friederich Gauss, em suas pesquisas sobre erros em observações astronômicas e estimação por quadrados minímos, sendo, por muitos autores, chamada de distribuição *Gaussiana*[2].

[2] Essa nomenclatura sobrevive apesar de a distribuição normal ter sido originariamente descoberta por de Moivre, em suas pesquisas nos teoremas-limite. Veja o capítulo 6.

Definição 5.3.1 (Distribuição Normal-padrão)

X é dita uma variável aleatória normal-padrão, $N(0,1)$, se sua função-densidade for escrita como:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

Como visto na seção 3.4, para que X seja uma variável aleatória, é necessário que f_X seja não-negativa e tenha integral um. A primeira das condições é facilmente confirmada. Quanto à segunda:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx = 1, \tag{5.1}$$

não há forma direta de fazê-lo. A maneira de confirmar (5.1), que gera interpretações interessantes em cursos futuros de Estatística, é demonstrar que, em vez de a integral ser igual a um, o seu quadrado o é.

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} f_X(x)dx\right)^2 &= \left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx\right)^2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left\{-\frac{x^2+y^2}{2}\right\} dxdy \\
&\quad \text{(usando } x = r \times sen(\theta) \text{ e } y = r \times cos(\theta)\text{)} \\
&= \int_{-\pi}^{\pi}\int_{0}^{+\infty} \frac{1}{2\pi} \exp\left\{-\frac{r^2}{2}\right\} r dr d\theta \\
&= \int_{0}^{+\infty} r \exp\left\{-\frac{r^2}{2}\right\} dr \\
&= \exp\left\{-\frac{r^2}{2}\right\}\Bigg| = 0 - (-1) = 1.
\end{aligned}$$

Portanto, X, definida por 5.3.1, é uma variável aleatória, absolutamente contínua, gozando das propriedades detalhadas na seção 3.4. Podemos, então, expor abaixo, na proposição 5.3.1, algumas das principais propriedades de uma v.a. normal-padrão.

Proposição 5.3.1 (Propriedades de uma Normal-padrão)

Seja X uma v.a. normal-padrão. Então,

$$\mathbb{E}(X) = 0;$$
$$Var(X) = 1;$$
$$\mu_k = \begin{cases} 0 & \forall k \text{ ímpar} \\ (k-1)(k-3)\ldots 3.1 & \forall k \text{ par; e} \end{cases}$$

Demonstração:

O fato de a esperança de X ser nula é conseqüência imediata de f_X ser simétrica (com relação a 0). Quanto à variância, note, primeiramente, que;

$$Var(X) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}(X)\right)^2 = \mathbb{E}\left(X^2\right).$$

Portanto,

$$\begin{aligned} Var(X) &= \int_{-\infty}^{+\infty} x^2 f_X(x)dx \\ &= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}dx \\ &= -\frac{1}{\sqrt{2\pi}}x \exp\left\{-\frac{x^2}{2}\right\}\Bigg|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \left(-\frac{1}{\sqrt{2\pi}}\right) \exp\left\{-\frac{x^2}{2}\right\} \\ &= 0 + 1 \quad = \quad 1. \end{aligned}$$

A demonstração de cada momento central é realizada da mesma forma que o da esperança e é deixada para o leitor (veja o exercício 5.2). ■

O fato de X, definida por 5.3.1, ser chamada de normal-padrão tem razão análoga ao da variável aleatória uniforme-padrão, dada pela proposição 5.3.2. Antes, porém, é necessário definir o que é uma variável aleatória com distribuição normal, no caso geral.

Definição 5.3.2 (Variável Aleatória Normal)

Seja Y uma variável aleatória com densidade de probabilidade f_Y. Diz-se que Y tem distribuição normal com parâmetros μ ($\in \mathbb{R}$) e σ^2 (> 0), $Y \sim N\left(\mu, \sigma^2\right)$, se:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}.$$

Proposição 5.3.2 (Estabilidade da Normal)

Seja X uma variável aleatória distribuída segundo a definição 5.3.1. Suponha que Y seja uma outra variável aleatória, definida por $Y = \sigma X + \mu$, onde σ é uma constante diferente de zero e $\mu \in \mathbb{R}$. Então, Y é distribuída como uma v.a. $N\left(\mu, \sigma^2\right)$.

A interpretação da proposição 5.3.2 é análoga ao da proposição 5.2.1, isto é, a classe de variáveis aleatórias normais é invariante com relação a mudanças de escala e posição. Portanto, faz-se consistente a noção de uma variável aleatória padrão. No caso da normal, como se vê na proposição 5.3.3, sua esperança e variância são definidas respectivamente por μ e σ^2, de forma exclusiva. O padrão da classe normal fica, então, dado pela posição (média zero) e dispersão (variância um).

Proposição 5.3.3 (Propriedades de N $\left(\mu, \sigma^2\right)$)

Seja Y uma variável aleatória normal $N(\mu, \sigma^2)$. Então,

$$\begin{aligned}
&\mathbb{E}(Y) = \mu;\\
&Var(Y) = \sigma^2;\\
&\mu_k = \textstyle\sum_{r=0}^{k} \binom{r}{k} \mu^r \sigma^{k-r} \mu_{k-r}^{(0)},
\end{aligned}$$

onde $\mu_j^{(0)}$ é o momento de ordem j da normal-padrão, $j = 0, 1, \ldots, k$, isto é,

$$\mu_j^{(0)} = \begin{cases} 0 & \forall j \text{ ímpar} \\ (j-1)(j-3)\ldots3.1 & \forall j \text{ par; e} \end{cases}$$

A demonstração é muito simples: basta-nos escrever $Y = \sigma X + \mu$, onde X é $N(0,1)$. Veja o exercício 5.2.

O leitor já deve ter notado que o estudo da classe de variáveis aleatórias normais pode ser feita exclusivamente pelo estudo da v.a. normal-padrão e, conseqüentemente, por sua densidade. Por sua importância, a densidade da normal-padrão recebe o símbolo ϕ e a respectiva função de distribuição acumulada, Φ.

Até agora, discutiram-se vários pontos favoráveis da distribuição normal: sua importância assintótica (Teorema Central do Limite); sua invariância com relação a mudanças de escala e posição, a relação biunívoca e exclusiva entre os seus parâmetros e os dois primeiros momentos.

No entanto, o estudo da distribuição normal tem um problema fundamental: a função Φ não possui fórmula fechada. Para se trabalhar com a distribuição, faz-se necessária a utilização de valores tabelados (veja as Tabelas C.5-C.7 nos apêndices) ou aproximações polinomiais. Por sua fundamental importância em diversas áreas do conhecimento, aproximações de Φ e sua inversa existem em abundância, sejam sob formas de tabelas, expressões polinomiais ou algoritmos em diversas linguagens.

Definição 5.3.3 (Função distribuição da normal-padrão)

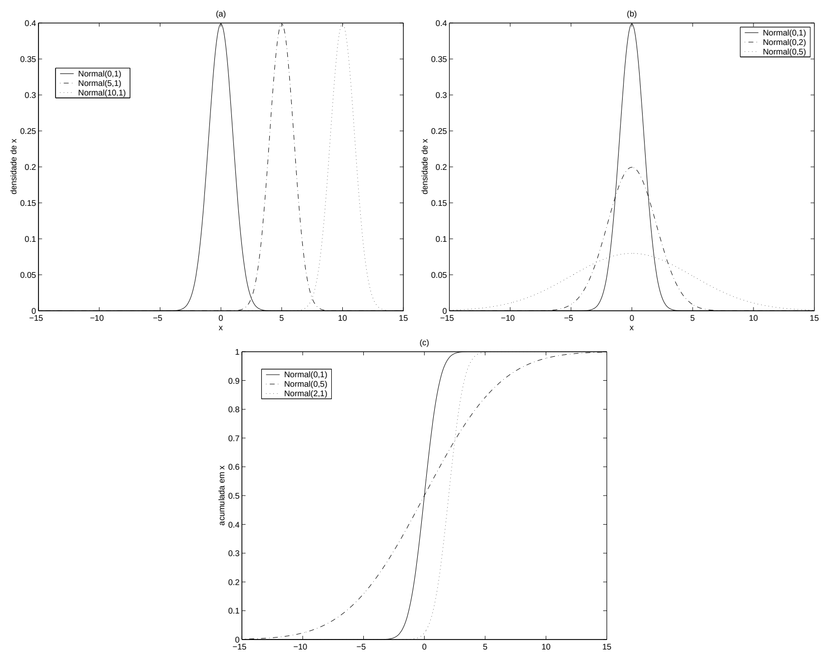
Seja X uma v.a. $N(0,1)$. Definem-se ϕ e Φ, ambas funções de $\mathbb{R}$ em $\mathbb{R}$, respectivamente por:

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}; \text{ e} \\ \Phi(x) &= \int_{-\infty}^{x} \phi(u)du. \end{aligned}$$

A função Φ é chamada de uma *função especial*: funções assim denominadas são encontradas em diversos campos do conhecimento. Têm algumas características qualitativas em comum: não possuem, em geral, soluções fechadas, aparecem como passos intermediários ou mesmo soluções de diversos problemas importantes. Nas seções 5.5 e 5.6, estaremos usando outras destas funções: gama e beta. No capítulo 4, vimos um caso particular de outra função desta classe: os coeficientes hipergeométricos.

Note que a utilidade de Φ reside principalmente na proposição 5.3.4, que nos permite calcular probabilidades para qualquer variável normal em função de Φ.

Figura 5.2: Família de Distribuições Normal



Proposição 5.3.4 (Cálculo de Probabilidades para Normais)

Seja Y uma v.a. $N\left(\mu, \sigma^2\right)$. Então, para qualquer $y \in \mathbb{R}$,

$$\mathbb{P}(Y \leq y) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

A demonstração da proposição 5.3.4 segue imediatamente da proposição 5.3.2 e é deixada para o leitor (veja o exercício 5.4). Devemos notar que a proposição 5.3.4 pode ser utilizada como justificativa de uma v.a. $N(0,1)$ ser chamada de normal-padrão, pois todo cálculo a ser realizado com qualquer variável da classe normal pode ser reduzido a um cálculo com uma normal-padrão, pela simples mudança de escala e posição (ou seja, pela padronização do evento em questão). Operacionalmente, isto significa que, para trabalharmos com normais, basta-nos um tabelamento (de Φ e de sua inversa), não nos sendo necessária a obtenção de outras tabelas.

5.4 Distribuição Exponencial

Uma das distribuições de maior utilização na prática é a distribuição exponencial (ou exponencial negativa). Há várias maneiras (cada uma delas dotada de uma interpretação igualmente interessante) de definir uma variável aleatória exponencial. Iremos, portanto, defini-la através de sua função de densidade e as outras possíveis interpretações serão apresentadas como propriedades.

Definição 5.4.1 (Variável Aleatória Exp(λ))

Seja X uma variável aleatória com função de densidade:

$$f_X(x) = \lambda \exp\{-\lambda x\} \mathbf{1}_{[0,\infty)}(x).$$

X é dita ter uma distribuição exponencial de parâmetro λ ou $X \sim Exp(\lambda)$.

Primeiramente, devemos confirmar que f_X é uma densidade fidedigna. Ela é claramente não-negativa. Quanto a sua norma,

$$\begin{aligned}\int_{-\infty}^{+\infty} f_X(x)dx &= \int_0^{\infty} \lambda \exp\{-\lambda u\}\, du \\ &= \exp\{-\lambda u\}]_0^{+\infty} = 1.\end{aligned}$$

Portanto, f_X define uma variável aleatória absolutamente contínua. Esta variável tem diversas propriedades interessantes. Como veremos a seguir, várias características dessa distribuição são facilmente calculadas, o que a torna extremamente atraente para aplicações.

Uma propriedade muito interessante diz respeito à função de distribuição acumulada de uma v.a. exponencial.

Proposição 5.4.1 (Função-distribuição de uma Exp(λ))

Seja X uma variável aleatória com distribuição Exp(λ). Sua função-distribuição, F, é dada por:

$$F(x) = 1 - \exp\{-\lambda x\}.$$

A proposição 5.4.1 tem como principal ponto de interesse o fato de que a *forma geral* da função-distribuição é a mesma da função de densidade de uma v.a. exponencial, muito interessante para cálculos.

Note que a forma exponencial negativa desta distribuição nos permite *intuitivamente prever* que a distribuição exponencial não terá problemas de momentos, como exposto em 5.4.2.

Proposição 5.4.2 (Momentos da Distribuição Exponencial)

Seja X uma variável aleatória com distribuição exponencial de parâmetro $\lambda > 0$. Então, X tem todos os momentos positivos finitos (na realidade, tem todos os momentos finitos de ordem

$s > -1$), sendo seus principais momentos recuperáveis de sua fgm:

$$\nu(t) = \frac{\lambda}{t - \lambda},$$

para todo $t < \lambda$.

Em particular, sua esperança e variância são, respectivamente:

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad e \quad Var(X) = \frac{1}{\lambda^2}.$$

Demonstração:

Note que:

$$\mathbb{E}(X^s) = \int_0^{+\infty} x^s \lambda \exp\{-\lambda x\} dx$$

não terá problemas para qualquer $s \in \mathbb{R}$ no seu limite superior mas, quando $s \leq -1$, divergirá no seu limite inferior.

$$\begin{aligned}
\nu(t) &= \mathbb{E}(\exp\{tX\}) \\
&= \int_0^{+\infty} \exp\{tu\} \lambda \exp\{-\lambda u\} du \\
&= \int_0^{+\infty} \lambda \exp\{(t-\lambda) u\} du \\
&= \frac{\lambda}{t-\lambda} \int_0^{+\infty} (t-\lambda) \exp\{(t-\lambda) u\} du \\
&= \frac{\lambda}{t-\lambda},
\end{aligned}$$

para $t < \lambda$.

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^{+\infty} x\lambda \exp\{-\lambda x\} dx \\
&= -\, x \exp\{-\lambda x\}]_0^{+\infty} + \\
&+ \frac{1}{\lambda} \int_0^{+\infty} \lambda \exp\{-\lambda x\} dx \\
&= \frac{1}{\lambda}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \int_0^{+\infty} x^2\lambda \exp\{-\lambda x\} dx \\
&= -\, x^2 \exp\{-\lambda x\}\Big]_0^{+\infty} + \\
&+ \frac{2}{\lambda} \int_0^{+\infty} x\lambda \exp\{-\lambda x\} dx \\
&= \frac{2}{\lambda^2}.
\end{aligned}$$

Logo,

$$Var(X) = \frac{1}{\lambda^2}.$$

■

A distribuição exponencial tem aplicação rotineira em áreas onde o interesse resida em tempo de vida. Esse tempo de vida pode ser relacionado a funcionamento de equipamentos (Controle de Qualidade) ou, no que se tem desenvolvido ultimamente, para **Análise de Sobrevivência** (sobrevida de portadores de certas doenças ou pessoas expostas a certos agentes de risco). Por esse motivo, algumas vezes, é de interesse exposicional trabalhar com a função-sobrevivência (definição 5.4.2) em vez de apresentar-se a função-distribuição .

Definição 5.4.2 (Função-sobrevivência)

Seja X uma variável aleatória com função de distribuição acumulada F_X. A função-sobrevivência de X, S_X, é dada por

$$S_X = 1 - F_X.$$

Note que S_X é definida para qualquer tipo de variável aleatória. No caso de variáveis não-negativas (como a exponencial), ela toma contornos de *sobrevivência* (ou *durabilidade*), repectivamente, de um determinado ser ou equipamento. A função-sobrevivência tem duas possíveis interpretações, de forma análoga à própria medida de probabilidade. Por definição, ela é simplesmente uma probabilidade, isto é, atende a todos os axiomas definidores de uma tal função de eventos.

Podemos pensá-la, no entanto, sob dois pontos de vista diversos: como medida individual de crença, $S_X(x)$ representa as *chances* de que um determinado indivíduo (equipamento) sobreviva (dure) mais do que x unidades de tempo. Ela pode ser também interpretada como a freqüência relativa de sobrevivência (vida útil) de um ser (equipamento), em uma população homogênea, assumindo, portanto, contornos de uma medida global de comportamento. Em qualquer um dos casos, ela é uma medida relevante, de fácil entendimento, uso geral e, no caso da distribuição exponencial, de facílimo cálculo.

Proposição 5.4.3 (Função-sobrevivência da Exp(λ))

Seja X uma variável aleatória com distribuição Exp(λ). Sua função-sobrevivência, S_X, é dada por:

$$S_X(x) = \exp\{-\lambda x\}.$$

O primeiro comentário de interesse sobre a proposição 5.4.3 é que, a menos de uma constante, λ, a função-sobrevivência num determinado ponto x é exatamente igual à densidade neste mesmo ponto. Em termos práticos, isto nos indica que os cálculos são muito facilitados. Uma propriedade, exposta na proposição 5.4.4, é a de que, não só a função-sobrevivência de uma v.a. exponencial é proporcional a sua densidade mas que a distribuição exponencial é a única distribuição absolutamente contínua sem memória.

Definição 5.4.3 (Variável Aleatória Sem Memória)

Seja X uma v.a. não-negativa (com probabilidade um). Então, X é dita ser uma v.a. sem memória se:

$$\mathbb{P}(X > s + t / X > t) = \mathbb{P}(X > s), \quad \forall s, t \geq 0. \tag{5.2}$$

Primeiramente, note que a equação 5.2 equivale a:

$$S_{X|X>t}(s+t) = S_X(s) \tag{5.3}$$

e (5.3) mostra que a função-sobrevivência *condicionada* ao evento $X > t$ é idêntica à função-sobrevivência primitiva. Isto quer dizer que o conhecimento que $X > t$ não modifica a estrutura de sobrevivência (durabilidade) do ser (equipamento). Isto tem um lado bom (cálculos condicionados equivalentes aos usuais) mas pode parecer um pouco exagerado: estamos dizendo que as pessoas não *envelhecem* ou que equipamentos não se estragam. Antes de mais nada, esta propriedade de *falta de memória* só é válida para a distribuição exponencial, pela proposição 5.4.4, sendo portanto facultativa a utilização de tal *restrição* nas análises. Incrivelmente, modelos exponenciais são muito bem sucedidos, demonstrando que a *falta de memória* não é muitas vezes tão absurda quanto parece. Nossa interpretação da adequação destes modelos deve ser análoga à da análise de qualquer modelo proposto para qualquer fenômeno natural: um modelo será sempre apenas uma aproximação da realidade. Não se pode, portanto, exigir de um modelo que ele esteja *correto* ou argumentar que ele esteja *errado*. No caso da *falta de memória* (obviamente *errada* para quase todos os casos), imagina-se que grande parte dos estudos realizados seja linear o suficiente para que a *falta de memória* não seja um comprometimento extremo.

Proposição 5.4.4 (Falta de Memória da Exponencial)

A distribuição exponencial (def. 5.4.1) é a única distribuição contínua com a propriedade de falta de memória (def. 5.4.3).

Uma das maneiras de definir uma distribuição exponencial, de especial interesse em processos estocásticos, é apresentada na proposição 5.4.5.

Proposição 5.4.5 (Ocorrências de um Proc. Poisson)

Seja $\{N_t\}_{t\geq 0}$ um processo de Poisson homogêneo de intensidade $\lambda \in \mathbb{R}^+$. Cada tempo entre duas ocorrências sucessivas deste processo tem distribuição exponencial, também de parâmetro λ.

Demonstração:

Sejam $N(t)$ o estado do processo de Poisson no tempo $t > 0$ e T o tempo para a primeira ocorrência de E.

$$\begin{aligned}
\mathbb{P}\left(T \leq t\right) &= \mathbb{P}\left(N(t) \geq 1\right) \\
&= \sum_{j=1}^{\infty} \mathbb{P}\left(N(t) = j\right) \quad = \quad 1 - \mathbb{P}\left(N(t) = 0\right) \\
&= 1 - \exp\{\lambda t\},
\end{aligned}$$

ou seja, T tem distribuição exponencial de parâmetro λ. ■

Além de conduzir-nos a uma ligação da distribuição exponencial com um dos processos estocásticos mais fundamentais e, conseqüentemente, com a distribuição de mesmo nome, a proposição 5.4.5 nos fornece uma interpretação interessantíssima, mesmo que óbvia, sobre a natureza do parâmetro λ, sendo essa a razão por que, ao contrário de alguns autores, preferimos, neste livro, trabalhar com exponenciais de esperança $1/\lambda$ ao invés de fazê-lo com exponenciais de esperança λ.

Na seção 5.5, veremos que a exponencial nada mais do que um caso particular de uma família extremamente importante para a teoria estatística.

5.5 Distribuição Gama

Ao final da seção 5.4, vimos que o tempo de espera para a primeira realização de um processo de Poisson de intensidade λ tem distribuição exponencial com o mesmo parâmetro λ. Uma pergunta natural é como seria distribuído o tempo até a n-ésima ocorrência, para $n \in \mathbb{N}$, T_n. A distribuição de T_n é chamada de distribuição de Erlang com parâmetros n e λ.

Definição 5.5.1 (Distribuição de Erlang)

A variável X tem distribuição de Erlang com parâmetros n e λ, $X \sim E(n, \lambda)$, se sua função de densidade, f_X é dada por:

$$f_X(x) = \frac{1}{(n-1)!}\lambda^n x^{n-1} \exp\{-\lambda x\}\mathbf{1}_{\mathbb{R}^+}(x).$$

A proposição 5.5.1 mostra, ao mesmo tempo que T_n tem distribuição $E(n, \lambda)$ e que a soma de n variáveis independentes e identicamente distribuídas como exponenciais de parâmetro λ tem distribuição também Erlang de parâmetros n e λ .

Proposição 5.5.1 (Tempo da n-ésima Ocorr. - Proc. Poisson) *Seja T_n o tempo para a n-ésima ocorrência num Processo de Poisson de intensidade λ. Então, T_n tem distribuição Erlang com parâmetros n e λ, isto é, com densidade:*

$$f_{T_n}(t) = \frac{1}{(n-1)!}\lambda^n t^{n-1} \exp\{-\lambda t\}\mathbb{I}(t > 0).$$

Demonstração:

Basta-nos achar a densidade de T_n, mostrando ser exatamente igual à definida em 5.5.1. Esta demonstração segue, em parte, a utilizada para a proposição 5.4.5.

$$
\begin{aligned}
\mathbb{P}\left(T_n \le t\right) &= \mathbb{P}\left(N(t) \ge n\right) \\
&= \sum_{j=n}^{\infty} \mathbb{P}\left(N(t) = j\right) \\
&= \sum_{j=n}^{\infty} \frac{exp\{-\lambda t\}\,(\lambda t)^j}{j!}.
\end{aligned}
$$

Portanto,

$$
F_{T_n}(t) = \left(1 - \sum_{j=0}^{n-1} \frac{exp\{-\lambda t\}\,(\lambda t)^j}{j!}\right) \mathbf{1}_{\mathbb{R}^+}(t)
$$

e sua densidade pode ser escrita, para $t > 0$ como:

$$
\begin{aligned}
f_{T_n}(t) &= \frac{d}{dt}\left(1 - \sum_{j=0}^{n-1} \frac{exp\{-\lambda t\}\,(\lambda t)^j}{j!}\right) \\
&= -\sum_{j=0}^{n-1} \frac{d}{dt}\frac{exp\{-\lambda t\}\,(\lambda t)^j}{j!} \\
&= -\sum_{j=1}^{n-1} \frac{exp\{-\lambda t\} j \lambda^j t^{j-1}}{j!} - \sum_{j=0}^{n-1} \frac{(-\lambda)\,exp\{-\lambda t\}\,(\lambda t)^j}{j!} \\
&\ \text{(trocando os somatórios e fazendo } k = j-1\text{)} \\
&= \sum_{j=0}^{n-1} \frac{exp\{-\lambda t\}\lambda^{j+1} t^j}{j!} - \sum_{k=0}^{n-2} \frac{exp\{-\lambda t\}\lambda^{k+1} t^k}{k!} \\
&= \frac{exp\{-\lambda t\}\lambda^n t^{n-1}}{(n-1)!}.
\end{aligned}
$$

■

Novamente, a família definida em 5.5.1 nada mais é do que uma subfamília de uma das classes mais importantes de variáveis aleatórias: as distribuições da família Gama. Antes de mais nada, definamos a função gama, uma das conhecidas funções especiais.

Definição 5.5.2 (Função Gama)

A função Gama, $\Gamma : \mathbb{R}^+ \to \mathbb{R}$, *é definida da seguinte forma:*

$$\Gamma(t) = \int_0^{+\infty} \exp\{-u\} u^{t-1} du.$$

Alguns de seus principais valores e propriedades são encontrados na proposição 5.5.2. Não há solução fechada para o cálculo da função Gama. Têm-se, no entanto, algumas propriedades que nos garantem o cálculo de Γ. O item (iv) da proposição 5.5.2 garante que nos basta calcular numericamente os valores de $\Gamma(t)$, para $t \in (0,1]$ para que possamos achar o valor de $\Gamma(t)$, para qualquer t real positivo [3] .

Proposição 5.5.2 (Propriedades da Função Gama)

Seja Γ *como na definição 5.5.2. Então:*

$$\begin{array}{lcl}
\textit{(i)}\ \Gamma(1/2) & = & \sqrt{\pi}; \\
\textit{(ii)}\ \Gamma(1) & = & 1; \\
\textit{(iii)}\ \Gamma(t) & = & (t-1)\Gamma(t-1) \quad \forall t > 1 \\
\textit{(iv)}\ \Gamma(t) & = & (t-1)(t-2)\ldots r\Gamma(r) \quad \forall t = n + r, \quad n \in \mathbb{N}, \ r > 0; \textit{ e} \\
\textit{(v)}\ \Gamma(n) & = & (n-1)! \quad \forall n \in \mathbb{N}^*.
\end{array}$$

Demonstração:

(i) Note que

$$\begin{aligned}
\Gamma(1/2) &= \int_0^{+\infty} \exp\{-u\} u^{-1/2} du \\
&\quad \big(\text{fazendo-se } v = u^2\big) \\
&= \int_{-\infty}^{+\infty} \exp\{-v^2\} dv \quad = \quad \sqrt{\pi},
\end{aligned}$$

[3] Por convenção, iremos utilizar $\Gamma(0) = 1$.

sendo a última integral simplesmente $\sqrt{\pi}$ vezes a integral de uma normal-padrão.

(ii) Note que

$$\Gamma(1) \quad = \quad \int_0^{+\infty} \exp\{-u\} du \quad = \quad 1,$$

sendo esse integrando a densidade de uma Exp(1), tendo, portanto, norma 1.

(iii)

$$\begin{aligned}
\Gamma(t) &= \int_0^{+\infty} \exp\{-u\} u^{t-1} du \\
&\quad \text{(integrando-se por partes)} \\
&= -\exp\{-u\} u^{t-1}\Big]_0^{\infty} + \\
&+ (t-1)\int_0^{+\infty} \exp\{-u\} u^{t-2} du \\
&= (t-1)\Gamma(t-1).
\end{aligned}$$

A propriedade (iv) se prova por indução, a partir de (iii), da mesma forma que (v) se utiliza de (iv) e (i) (veja o exercício 5.7). ■

A proposição 5.5.2 nos motiva a definir uma família mais geral de variáveis aleatórias, que contenha a família Erlang como caso particular. Para isso, duas mudanças são necessárias. A primeira, muito simples, é a de permitir-se que r, nas v.a.'s Erlang possa ser qualquer valor estritamente positivo, em vez de um número natural. A segunda, que se explica melhor pela figura 5.3, tem, como função fornecer à função gama (para cada t fixo), uma noção de escala (da mesma forma que o parâmetro da exponencial o faz). Note no gráfico (a) que a forma da densidade não muda quando variamos λ entre 1, 2 e 5; o mesmo ocorre no gráfico (b). Entretanto, em 5.3(c), nota-se que, quando variamos r (em 1, 2 ou 5), temos densidades completamente diferentes. Note também como se comportam as funções de distribuição acumuladas nesses casos. Portanto, o primeiro parâmetro da Erlang (e de sua generalização) nos fornece a forma da distribuição, enquanto o segundo nos fornece a escala com que estamos trabalhando.

O leitor curioso já deve ter notado que, desses parâmetros, apenas o primeiro existe originalmente na função matemática gama e que isto ocorre pois, para funções tabeladas, não faz sentido trabalhar-se com simples mudanças de **escala**. Para ilustrar o que se dizendo, tome uma exponencial, X, de média 1. Defina $Y = aX$, $a > 0$. Então, Y também terá distribuição exponencial, com média a em lugar de 1. Este resultado está generalizado na proposição 5.5.3.

Antes dissso, porém, note-se esta igualdade fundamental para a distribuição gama:

$$\begin{aligned} 1 &= \int_0^\infty \frac{1}{\Gamma(r)} x^{r-1} \lambda^r exp\{-\lambda x\} \\ &= \int_0^\infty \frac{1}{\Gamma(r)} (ax)^{r-1} (\lambda/a)^r \, exp\{(-\lambda/a)(ax)\}. \end{aligned} \tag{5.4}$$

A identidade (5.4) nos indica que mudanças de escala não violam a relação entre a densidade da variável e a função gama. Portanto, faz sentido definir-se uma variável aleatória com distribuição gama com dois parâmetros.

Definição 5.5.3 (Distribuição Gama)

A variável aleatória X, cuja densidade é dada por

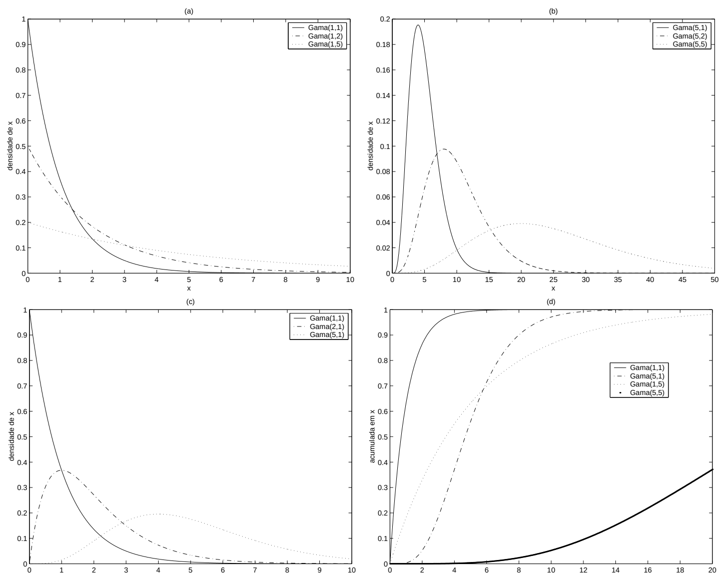
$$f_X(x) = \frac{1}{\Gamma(r)} x^{r-1} \lambda^r exp\{-\lambda x\} \mathbf{1}_{\mathbb{R}^+}(x),$$

é chamada de variável gama com parâmetros r e λ ou, simplesmente, $X \sim Gama(r, \lambda)$.

Proposição 5.5.3 (Invariância da Família Gama)

Seja X uma variável aleatória com distribuição $Gama(r, \lambda)$. Então, $Y = aX$, $a > 0$, também terá distribuição gama, com respectivos parâmetros r e λ/a.

Figura 5.3: Família de Distribuições Gama



Demonstração:

Basta-nos calcular a função de distribuição de Y e demonstrar que seu integrando tem forma análoga à densidade de X, dada na definição 5.5.3.

$$\begin{aligned}
\mathbb{P}\left(Y \leq y\right) &= \mathbb{P}\left(aX \leq y\right) = \mathbb{P}\left(X \leq y/a\right) \\
&= \int_0^{y/a} \frac{1}{\Gamma(r)} x^{r-1} \lambda^r exp\{-\lambda x\} dx \\
&\quad \text{(fazendo } v = ax\text{)} \\
&= \int_0^{y} \frac{1}{\Gamma(r)} (v/a)^{r-1} \lambda^r exp\left\{-\frac{\lambda}{a} v\right\} \frac{dv}{a}
\end{aligned}$$

$$= \int_0^y \frac{1}{\Gamma(r)} v^{r-1} \left(\frac{\lambda}{a}\right)^r exp\left\{-\left(\frac{\lambda}{a}\right) v\right\} dv.$$

. ■

O que a proposição 5.5.3 nos diz é que a família de v.a.'s gamas é invariante com relação a mudanças de escala. Isto é importantíssimo pois aponta mais uma vez que o caráter da distribuição gama é diretamente ligado ao parâmetro de forma, original da função matemática gama, sendo o segundo apenas um recurso para mudanças de escala.

Esse fato é confirmado na proposição 5.5.4, em que se discutem algumas propriedades dos momentos da distribuição gama.

Proposição 5.5.4 (Momentos da Distribuição Gama)

Seja X uma variável aleatória com distribuição $Gama(r, \lambda)$. Então, X tem todos os momentos positivos finitos (na realidade, tem todos os momentos finitos de ordem $s > -r$). Em particular,

$$\mathbb{E}(X) = \frac{r}{\lambda} \text{ e } Var(X) = \frac{r}{\lambda^2}.$$

Sua função geradora de momentos é dada por:

$$\nu(t) = \left(\frac{\lambda}{t-\lambda}\right)^r,$$

para $t < \lambda$.

Demonstração:

Note que:

$$\mathbb{E}(X^s) = \int_0^{+\infty} \frac{1}{\Gamma(r)} x^{s+r-1} \lambda^r \exp\{-\lambda x\} dx$$

não terá problemas para qualquer $s \in \mathbb{R}$ no seu limite superior mas, quando $s + r \leq 0$, divergirá no seu limite inferior.

$$\begin{aligned}
\nu(t) &= \mathbb{E}\left(\exp\{tX\}\right) \\
&= \int_0^{+\infty} \exp\{tu\} \frac{1}{\Gamma(r)} x^{r-1} \lambda^r \exp\{-\lambda u\} du \\
&= \int_0^{+\infty} \frac{1}{\Gamma(r)} x^{r-1} \lambda^r \exp\{(t-\lambda)\, u\} du \\
&= \left(\frac{\lambda}{t-\lambda}\right)^r \int_0^{+\infty} \frac{1}{\Gamma(r)} x^{r-1} (t-\lambda)^r \exp\{(t-\lambda)\, u\} du \\
&= \left(\frac{\lambda}{t-\lambda}\right)^r,
\end{aligned}$$

para $t < \lambda$.

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^{+\infty} u \frac{1}{\Gamma(r)} u^{r-1} \lambda^r \exp\{-\lambda u\} du \\
&= \frac{\Gamma(r+1)}{\lambda\Gamma(r)} \int_0^{+\infty} \frac{1}{\Gamma(r+1)} u^r \lambda^{r+1} \exp\{-\lambda u\} du \\
&= \frac{r\Gamma(r)}{\lambda\Gamma(r)} = \frac{r}{\lambda}
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \int_0^{+\infty} u^2 \frac{1}{\Gamma(r)} u^{r-1} \lambda^r \exp\{-\lambda u\} du \\
&= \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)} \int_0^{+\infty} \frac{1}{\Gamma(r+2)} u^{r+1} \lambda^{r+2} \exp\{-\lambda u\} du \\
&= \frac{(r+1)\, r\Gamma(r)}{\lambda^2\Gamma(r)} = \frac{(r+1)\, r}{\lambda^2}.
\end{aligned}$$

Logo,

$$\begin{aligned}
Var(X) &= \frac{(r+1)\, r}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 \\
&= \frac{r}{\lambda^2}. \qquad \blacksquare
\end{aligned}$$

5.6 Distribuição Beta

Uma outra distribuição de grande interesse em diversos campos de aplicação da Estatística é a Beta. Em muitos problemas, interessa-nos aplicar um modelo contínuo com suporte finito. Raramente, no entanto, a distribuição uniforme será adequada. Uma das opções é a utilização da distribuição, análoga à função especial de mesmo nome.

Definição 5.6.1 (Função Beta)

Sejam $\alpha > 0$ e $\beta > 0$ números reais. A função Beta, escrita como $B(\alpha, \beta)$, $B : \mathbb{R}_+^2 \to \mathbb{R}_+$, é definida pela seguinte integral:

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

Proposição 5.6.1 (Relação entre as Funções Beta e Gama)

Sejam $\alpha > 0$ e $\beta > 0$. Então,

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

Demonstração:

Esta demonstração tem basicamente dois passos: escreve-se o produto das duas gamas e transformam-se ambas as variáveis de integração, chegando-se à identidade proposta.

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_0^{+\infty} \exp\{-u\} u^{\alpha-1} du \int_0^{+} \infty \exp\{-v\} u^{\beta-1} dv \\
&= \int_0^{+\infty} \int_0^{+\infty} u^{\alpha-1} v^{\beta-1} \exp\{-(u+v)\} du dv \\
& \quad (u, v) \to (x, y), \text{ onde } x = u + v \text{ e } y = \frac{u}{u+v}
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{+\infty}\int_0^1 (xy)^{\alpha-1}(x(1-y))^{\beta-1}\exp\{-(x)\}x\,dy\,dx \\
&= \int_0^{+\infty} x^{\alpha+\beta-1}\exp\{-(x)\}dx\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy \\
&= \Gamma(\alpha+\beta)B(\alpha,\beta).\blacksquare
\end{aligned}
$$

Portanto, a proposição nos diz que, para calcular valores da função Beta, basta saber os valores da função Gama, acabando com a possível necessidade de novas tabelas para a Beta. Note que a riqueza da família de distribuições de probabilidade Gama estava concentrada nos pequenos valores pois os valores caudais sempre se comportavam de acordo com o expoente negativo.

O caso da Beta é bem diferente: o suporte é limitado a $(0,1)$ mas o estudo assintótico da respectiva função de densidade não só é interessante como fundamental para a interpretação de algumas das suas principais propriedades teóricas (e, conseqüentemente, para suas aplicações).

Definição 5.6.2 (Distribuição Beta)

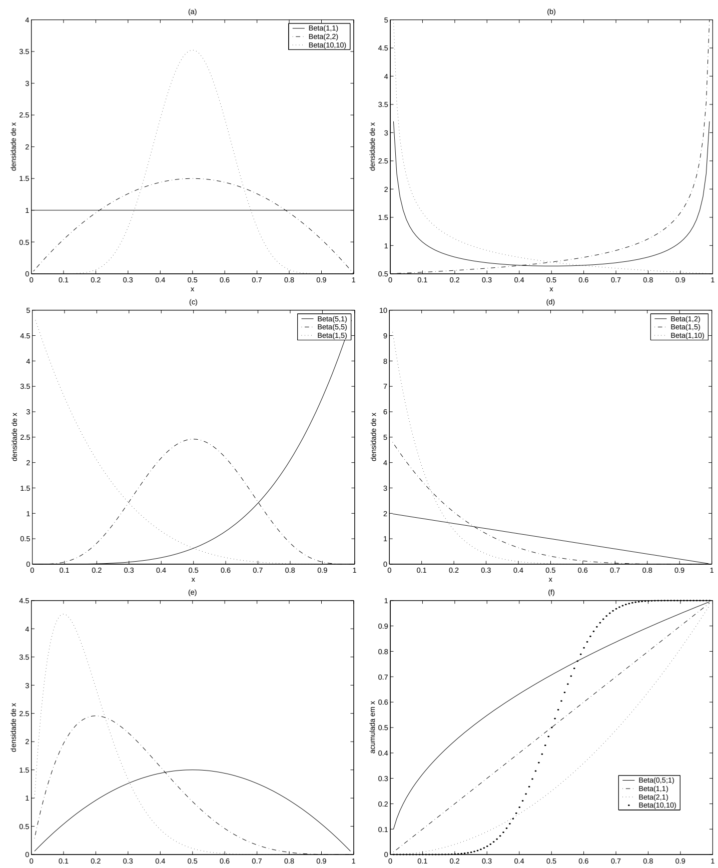
Uma variável aleatória, X, é dita distribuída como uma $Beta(\alpha,\beta)$ se sua densidade tiver a seguinte forma:

$$f_X(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{(0,1)}(x),$$

onde $\alpha,\beta > 0$.

O fato de a família de distribuições Beta poder representar tantos *comportamentos diferentes* para variáveis aleatórias de suporte limitado a torna muito atrativa em diversos problemas reais. Outras das suas características, que excedem os objetivos deste livro, a colocam como uma das principais distribuições na **Estatística Bayesiana**. Para um estudo mais aprofundado dessa matéria, o leitor interessado pode consultar [26].

Figura 5.4: Família de Distribuições Beta



Nos gráficos da figura 5.4, nota-se que elementos da família de distribuições Beta têm um entre os seguintes possíveis comportamentos modais :

- quando $\alpha = \beta > 1$, a distribuição é unimodal, sendo este ponto sempre $1/2$;
- quando $1 < \alpha < \beta$, a distribuição é unimodal, sendo sua moda estritamente menor do que $1/2$;
- quando $1 < \beta < \alpha$, a distribuição é unimodal, sendo sua moda estritamente maior do que $1/2$; e
- quando $\min(\alpha, \beta) \leq 1$, o comportamento modal não representa um verdadeiro valor-padrão da respectiva distribuição.

As demonstrações formais destas propriedades são deixadas para o leitor (veja exercício 5.12).

Portanto, a figura 5.4 ilustra que valores pequenos de α ou β (menores do que 1) geram assíntotas ilimitadas em 0 ou 1, respectivamente, sendo uma de suas conseqüências a inexistência de moda para distribuições baseadas nesses parâmetros. Problemas sérios para simulação também ocorrem para as distribuições cujo $\min(\alpha, \beta) < 1$. Quando $\min(\alpha, \beta) \geq 1$, nota-se que não existem problemas assintóticos e os parâmetros definem as condições de assimetria e o valor máximo da densidade.

Proposição 5.6.2 (Momentos da Distribuição Beta)

Seja X uma variável aleatória com distribuição Beta(α, β), $\min(\alpha, \beta) > 0$. *Então, suas esperança e variância são, respectivamente,*

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta}$$

e

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

Demonstração:

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^1 x\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}\int_0^1 \frac{1}{B(\alpha+1,\beta)}x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\times\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\times 1 \\
&= \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \int_0^1 x^2\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)}\int_0^1 \frac{1}{B(\alpha+2,\beta)}x^{\alpha+1}(1-x)^{\beta-1}dx \\
&= \frac{(\alpha+1)\,\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)\,(\alpha+\beta)\,\Gamma(\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)}.
\end{aligned}$$

Conseqüentemente,

$$\begin{aligned}
Var(X) &= \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)}-\left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{1}{(\alpha+\beta+1)\,(\alpha+\beta)^2}\left[(\alpha+\beta)\,(\alpha+1)\,\alpha+\right. \\
&- \left.(\alpha+\beta+1)\left(\alpha^2\right)\right] \\
&= \frac{1}{(\alpha+\beta+1)\,(\alpha+\beta)^2}\left[(\alpha+\beta)\,(\alpha+1)\,\alpha-\alpha^2\right]
\end{aligned}$$

$$= \frac{\beta\alpha}{(\alpha+\beta+1)(\alpha+\beta)^2}. \qquad \blacksquare$$

Basicamente, a proposição 5.6.2 nos diz que a assimetria de uma distribuição Beta, que ocorre quando $\alpha \neq \beta$ tem conseqüências também no seu valor esperado: quanto maior o valor relativo de α comparado a β, mais próximo de 1 estará a esperança. Por outro lado, a assimetria da densidade de uma variável aleatória $Beta(\alpha, \beta)$ não influencia a sua variância, cujo valor é invariante à permutação de α e β. Além disso, quanto maiores os valores de α e β, menor será a respectiva variância.

Exemplo 5.6.1 *Suponha $\alpha = \beta = 1$. A distribuição Beta, neste caso, é uma uniforme em $(0,1)$, cujas esperança e variância são, respectivamente, $1/2$ e $1/12$, confirmando-se os expostos na proposição 5.6.2.*

Por outro lado, suponha que $\alpha = \beta = 100$. A esperança também será $1/2$ mas a variância será diferente: $0,00124$.

Quaisquer que sejam os valores de α e β, desde que iguais, a simetria da respectiva densidade (e, conseqüentemente, a proposição 5.6.2) nos garante que a esperança será $1/2$. O que muda nessas distribuições é a variância: quanto $\alpha = \beta$ cresce, concentra-se a densidade em torno do valor central (que também é modal) e, portanto, decresce a variância.

5.7 Distribuição Cauchy

Exemplo 5.7.1 (Em Que Direção o Louco Armado Atira)

Suponha que você esteja passeando alegremente em uma praça e mais um desajustado armado comece a atirar, a esmo. Suponhamos também que o atirador tenha a possibilidade de atirar em qualquer direção. Nosso interesse está em saber qual o ângulo (dado um referencial) do próximo tiro.

É razoável supor que os tiros serão dados realmente a esmo, ou seja, que o seu ângulo seja distribuído uniformemente no círculo unitário. Qual seria a distribuição da tangente de tal variável aleatória? Seja θ o ângulo de tiro e X sua tangente.

$$\begin{aligned}\mathbb{P}\left(X \leq x\right) &= \mathbb{P}\left(tg(\theta) \leq x\right) \\ &= \mathbb{P}\left(\theta \leq arctg(x)\right) = arctg\,(x) - \frac{\pi}{2},\end{aligned}$$

que é a função de distribuição acumulada de X. Portanto, para conhecer sua densidade, basta-nos derivar F_X:

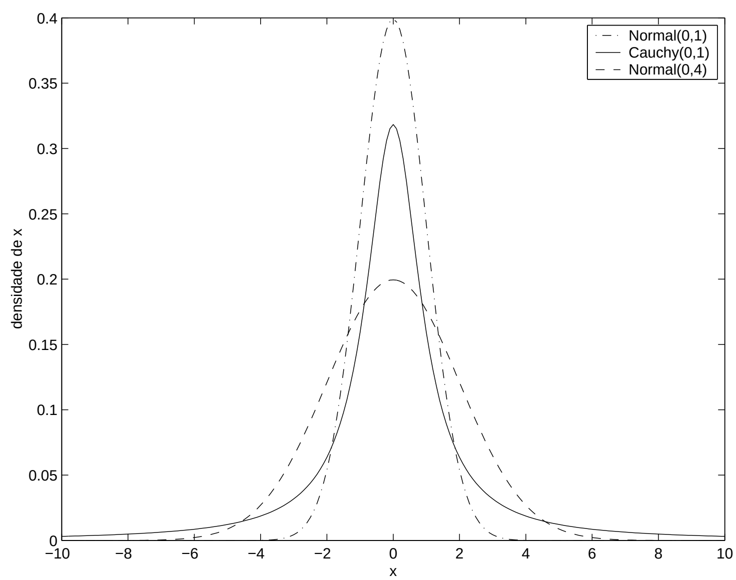
$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}. \tag{5.5}$$

Se fizermos o gráfico de f_X, definida por (5.5), notamos que, aparentemente, existe um comportamento semelhante ao da densidade da normal-padrão, isto é, tem-se uma função simétrica, com moda e mediana em 0 e caudas estritamente decrescentes. No entanto, a primeira diferença fundamental é a de que a densidade definida por (5.5), ao contrário da densidade normal-padrão, que tem decaimento exponencial, tem apenas decaimento quadrático. Esse fato terá conseqüências interessantíssimas.

Na figura 5.5, vemos que a concentração da distrição Cauchy, cuja densidade é dada por (5.5), em torno de sua mediana é menor do que a equivalente para uma densidade de uma normal-padrão. Portanto, o fato de a Cauchy ter caudas mais pesadas pode parecer natural, visto que ambas as integrais têm mesmo valor, 1. No entanto, note-se, no mesmo gráfico, que mesmo que tomemos uma densidade normal de menor concentração, no caso uma normal de média 0 e variância 4, cuja concentração em torno da mediana é menor do que a da Cauchy, as caudas desta distribuição continuam bem mais pesadas do que as daquela.

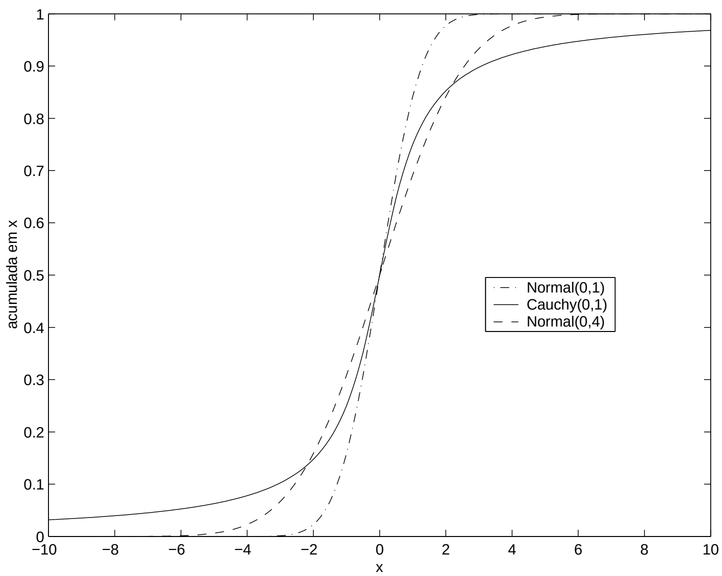
Essa diferença, que se mostrará fundamental, drasticamente ilustrada na proposição 5.7.1, é também clara na figura 5.6, em que se apresentam as respectivas distribuições acumuladas de uma normal-padrão, Cauchy com densidade dada por (5.5) e de uma $N(0,4)$. Note-se que as semelhanças entre as distribuições normal e de Cauchy só existem fracamente nos centros dos suportes. De resto, ambas as distribuições normais se mostram extremamente concentradas em suas probabilidades enquanto a Cauchy permite quase 10% de probabilidade para valores com mais de dez unidades longe da mediana.

Figura 5.5: Densidade da Cauchy e da Normal



Uma semelhança entre as famílias de distribuições normal e Cauchy, no entanto, é a existência de uma distribuição *central*, que estaremos chamando analogamente de distribuição Cauchy-padrão.

Figura 5.6: Distribuições Acumuladas da Cauchy e da Normal



Definição 5.7.1 (Distribuição Cauchy-padrão)

Seja X uma variável aleatória com densidade de probabilidade dada por:

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

X é dita ter distribuição Cauchy-padrão, ou $X \sim C(0,1)$.

Antes de entendermos o que o termo padrão quer dizer, devemos ressaltar que o resultado apresentado na proposição 5.7.1 nos revela que a idéia de padronização para as distribuições, seja ela qual for, tem que ser diferente da utilizada para a normal ou uniforme.

Proposição 5.7.1 (Momentos da Distribuição Cauchy-padrão)

Seja X uma variável aleatória com distribuição Cauchy-padrão. Então, X não tem esperança definida.

Demonstração:

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{+\infty} u\frac{1}{\pi}\frac{1}{1+u^2}du \\ &= \frac{1}{\pi}\, arctg(u)]_{-\infty}^{+\infty},\end{aligned}$$

que, claramente, não pode ser finita. ∎

O que o resultado exposto na proposição 5.7.1 nos revela é que faltam à distribuição Cauchy algumas das características mais utilizadas na pesquisa estatística: os momentos. Se, por um lado, isto pode ser desalentador, pois indicaria que essa distribuição seria *inútil* pois *contradiz* o que de *melhor* podemos ter, devemos pensar que o fato de a densidade de uma v.a. Cauchy poder ser confundida com uma v.a. normal, em uma análise superficial (como, em geral, podemos classificar qualquer análise de uma função contínua, feita por meio de uma amostra finita, por mais sofisticadas as técnicas de análise) , a proposição 5.7.1 deve servir de alerta que aparentes *ligeiras* modificações em relação a nossas hipóteses de modelo podem levar a resultados desastrosos, como será visto para o caso da Cauchy na **Lei Forte dos Grandes Números**. .

A pergunta que deve estar incomodando o leitor é a seguinte: por que chamamos a distribuição definida por 5.7.1 de Cauchy-padrão se todas as padronizações anteriormente aplicadas eram baseadas nos dois primeiros momentos, inexistentes nesse caso? Na realidade, a técnica de padronização faz uso de quaisquer medidas de locação e dispersão à disposição. Nos casos anteriores, os adequados eram: a esperança, para locação; e o desvio-padrão, para dispersão. Note que 0 para a esperança e 1 para a variância eram os únicos valores que verdadeiramente definiriam um padrão para a família de distribuições normais. Portanto, o fato de se utilizarem a esperança e desvio-padrão nada tem de arbitrário para essa distribuição mas, por outro, não tem que ser universal para as famílias de distribuições. No caso da Cauchy, o valor padrão de locação, 0, será representado pela mediana, enquanto o valor padrão de dispersão, 1, será representado pelo desvio interquartílico.

Definição 5.7.2 (Família de Distribuições Cauchy)

X é dita ter uma distribuição Cauchy com parâmetros m e a, $X \sim C(m,a)$, se sua densidade puder ser escrita como:

$$f_X(x) = \frac{a}{\pi}\frac{1}{a^2 + (x-m)^2},$$

onde $m \in \mathbb{R}$ e $a > 0$.

Proposição 5.7.2 (Invariância da Distribuição Cauchy)

Seja $X \sim C(0,1)$. Então $Y = aX + m$ tem também distribuição Cauchy, com parâmetros m e a. Além disso, m e $2a$ são, respectivamente, a mediana e desvio interquartílico de Y.

Demonstração:

Se $X \sim C(0,1)$, então:

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

Seja $Y = aX + m$.

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(aX + m \leq y) \\
&= \mathbb{P}\left(X \leq \frac{y-m}{a}\right) \\
&= \int_{-\infty}^{\frac{y-m}{a}} \frac{1}{\pi}\frac{1}{1+x^2}dx
\end{aligned}$$

fazendo-se $u = ax + m$

$$= \int_{-\infty}^{y} \frac{1}{\pi} \frac{a^2}{a^2 + (x-m)^2} \frac{du}{a},$$

sendo portanto a densidade de Y a de uma $C(m, a)$.

A demonstração do valor mediano é obtida diretamente da simetria da densidade em relação a m e é deixada para o leitor, juntamente com o do desvio interquartílico (veja o exercício 5.13). ∎

5.8 Distribuições Relacionadas à Exponencial

O leitor deve ter notado que as famílias de distribuições Cauchy e Normal têm diferenças extremas assim como o têm distribuições dentro da família Gama e Beta. Já foi ressaltado o ponto de que as distribuições exponenciais são casos particulares de distribuições Gamas, quando o parâmetro de forma é fixado. Portanto, a distribuição Gama pode ser pensada como uma *generalização* da distribuição exponencial. Essa generalização é feita do ponto de vista da forma da densidade associada à variável aleatória.

No entanto, se quiséssemos comparar uma exponencial com uma normal ou até mesmo com uma Cauchy, como fazê-lo? Uma primeira providência seria estender o domínio da distribuição exponencial para todos os reais, de forma simétrica. A distribuição resultante de tal operação é adequadamente chamada de **exponencial dupla**.

Definição 5.8.1 (Distribuição Exponencial Dupla)

Seja X uma variável aleatória com a seguinte densidade de probabilidade:

$$f_X(x) = \frac{\lambda}{2} \exp(-\lambda |x|).$$

X é dita uma exponencial dupla de parâmetro λ, $X \sim EE(\lambda)$.

Proposição 5.8.1 (Esperança e Variância da exponencial Dupla)

Seja X uma variável aleatória com distribuição exponencial dupla, de parâmetro λ. Então,

$$\mathbb{E}(X) = 0 \quad e \; Var(X) = \frac{2}{\lambda^2}.$$

Demonstração:

Note, para a esperança de X, simplesmente que a densidade de X é simétrica. Quanto a sua variância,

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \int_{-\infty}^{\infty} x^2 \frac{\lambda}{2} \exp\left(-\lambda|x|\right) dx \\
&= \int_{0}^{\infty} x^2 \lambda \exp\left(-\lambda x\right) dx \\
&= E(Y^2) \text{ para } Y \sim Exp(\lambda) \\
&= \frac{2}{\lambda^2}.
\end{aligned}$$

■

Antes de mais nada, note que a variância de uma exponencial dupla é exatamente duas vezes maior do que a variância de uma exponencial simples, com parâmetro igual. Esse fato é intuitivo, sabendo-se que o suporte da exponencial dupla é o *dobro* do suporte da exponencial e nenhuma modificação funcional foi realizada na densidade. Os gráficos da figura 5.7 nos mostram algumas comparações entre a exponencial dupla e uma normal-padrão. Note que o parâmetro da exponencial dupla foi escolhido de forma a termos sua variância 1, igual à da normal-padrão.

O primeiro gráfico, que compara as respectivas densidades da exponencial dupla e da normal-padrão tem um ponto muito interessante. Notemos que a densidade da exponencial dupla se mostra muito mais concentrada em torno de 0 do que a da normal. Como se explica então que ambas tenham variância 1? Para resolver esse aparente problema, notemos no segundo gráfico da figura 5.7 que as caudas da exponencial dupla são bem mais pesadas do que as da normal[4]. No terceiro gráfico, temos a razão entre as densidades da exponencial dupla e normal-padrão: vê-se, claramente, que a distribuição exponencial concentra sua densidade em torno de 0 mais do que o faz a densidade da normal-padrão mas suas caudas se tornam bem mais pesadas do que as equivalentes da normal-padrão. Note também um grande senão da distribuição exponencial dupla: a ausência de derivada em zero, com o claro *bico.*

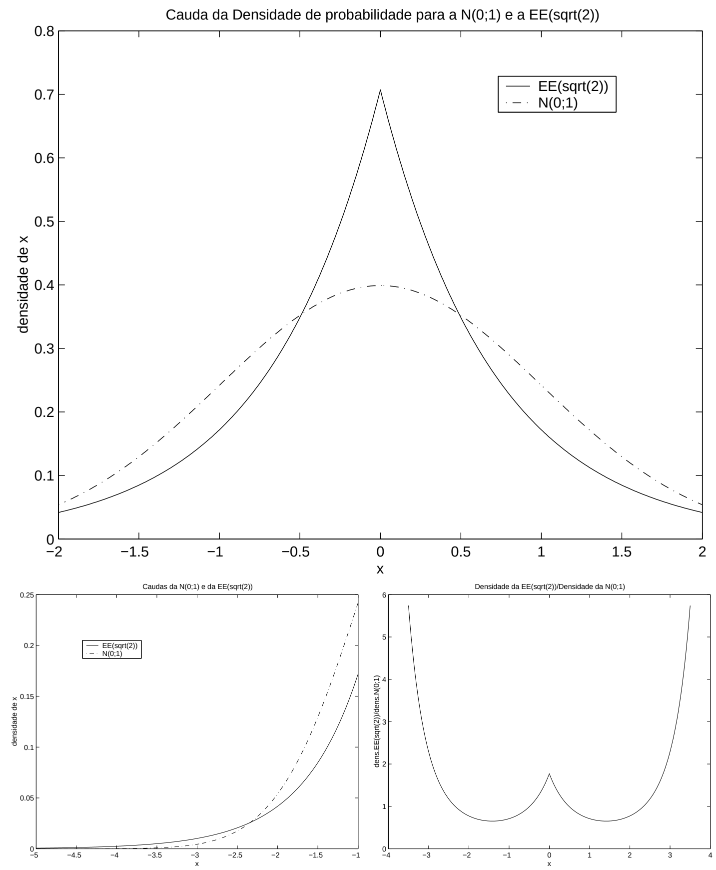
Sobre o comportamento heterogêneo da razão entre as densidades, faz-se interessante uma pequena conta. Suponhemos $X \sim EE(0, \sqrt{2})$ e Y normal-padrão .

$$
\begin{aligned}
\frac{f_X(t)}{f_Y(t)} &= \frac{\sqrt{2}\exp\left(-\sqrt{2}|t|\right)/2}{\exp\left(-t^2/2\right)/\sqrt{2\pi}} \\
&= \sqrt{\pi}\exp\left(\frac{1}{2}\left(t^2 - 2\sqrt{2}|t|\right)\right) \\
&= \sqrt{\pi}\exp\left(\frac{1}{2}\left(t - \sqrt{2}\right)^2 + 1\right)
\end{aligned}
$$

Conseqüentemente, para pequenos valores de $|t|$, a razão entre as densidades se mostra próxima a $\sqrt{\pi} \approx 1,7725$ mas, para valores de $|t|$ maiores do que $\sqrt{2} \approx 1,4142$, a razão entre as densidades da exponencial dupla e da normal é exponencial, isto é, com as caudas daquela exponencialmente maiores do que as desta.

[4]Lembre-se que, apesar de apenas as caudas esquerdas estarem retratadas, ambas as distribuições são simétricas

Figura 5.7: Distribuições Exponencial Dupla e Normal-padrão



Já vimos uma generalização da exponencial em sua forma, através da gama e outra, sob o ponto de vista de seu suporte. Um terceira, que veremos a seguir, a faz sob a motivação da deformaçõ temporal[5].

Suponha que tomemos $X \sim Exp(\lambda)$. Sabemos que aX, $a > 0$ tem também distribuição exponencial, mas com parâmetro a/λ. O que aconteceria se tomássemos Y tal que Y^γ tenha distribuição exponencial? Antes de derivarmos sua distribuição, pensemos sobre o significado de tal propriedade. Naturalemte, por assumir valores reais positivos, podemos associar a exponencial, gama e distribuições associadas com o tempo. Y teria, nesse contexto, um sentido de deformação do tempo para adequá-lo à distribuição exponencial. Se $\gamma < 1$, o *tempo correria mais lentamente* do que uma exponencial e vice-versa.

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(X^{\frac{1}{\gamma}} \leq y) \quad = \quad \mathbb{P}(X \leq y^\gamma) \\
&= \int_0^{y^\gamma} \lambda \exp(-\lambda x) dx \\
&= 1 - exp(-\lambda y^\gamma).
\end{aligned}$$

Portanto, Y seria uma variável aleatória com densidade dada por

$$f_Y(y) = \lambda \gamma y^{\gamma-1} exp(-\lambda y^\gamma) \mathbf{1}_{\mathbb{R}_+}(y).$$

Definição 5.8.2 (Distribuição Weibull)

Seja Y uma variável aleatória com densidade dada por

$$f_Y(y) = \lambda \gamma y^{\gamma-1} exp(-\lambda y^\gamma) \mathbf{1}_{\mathbb{R}_+}(y),$$

[5]O conceito de deformação temporal é fundamental para estudos de Confiabilidade e Análise de Sobrevivência, onde se utilizam rotineiramente **testes de vida acelerados**, isto é, experimentos realizados cujo tempo natural é lento demais para a obtenção de resultados em tempo hábil. São então realizados testes laboratoriais com o **tempo acelerado**: a sofisticação da metodologia reside em recuperar desses experimentos resultados para o tempo real, sem distorção.

onde λ, γ são reais positivos.

Y é dita ter uma distribuição **Weibull** *, com parâmetros λ e γ, $Y \sim We(\lambda, \gamma)$.*

5.9 Distribuições Relacionadas à Normal

Por sua importância na Teoria Estatística, como elemento estrutural sobre o qual toda uma metodologia pode ser desenvolvida de forma assintótica[6], a *Distribuição Normal* tem vastos estudos na literatura; além disso, todas as distribuições que dela podem ser derivadas também têm destaque para as metodologias estatísticas.

Nessa seção, faremos um breve estudo exposional sobre algumas dessas distribuições. Algumas serão apenas definidas e sua forma de ligação com a Normal será enunciada mas não demonstrada[7]. Todas as distribuições têm sentido de transformações corriqueiramente feitas nas observaçoes e são chamadas de distribuições amostrais.

Definição 5.9.1 (Distribuição Qui-quadrado)

Seja X uma variável aleatória contínua em $(\Omega, \mathcal{S}, \mathbb{P})$ com a seguinte função de densidade:

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} exp - \left(-\frac{x}{2}\right) \mathbf{1}_{\mathbb{R}^+}(x).$$

X é dita ter uma distribuição qui-quadrado com n graus de liberdade, $X \sim \chi^2_n$.

Observe-se que X nada mais é do que uma variável aleatória com distribuição Gama com parâmetros $n/2$ e $1/2$.

[6] O leitor interessado é novamente convidado a consultar um bom livro de Teoria Estatística. Para nossos objetivos neste livro, basta-nos entender que existe tal metodologia e que a *Distribuição Normal* e suas relacionadas devem ser estudadas com atenção.

[7] As demonstrações formais necessitam de noções multivariadas, além dos objetivos deste livro.

Proposição 5.9.1 (Esperança e Variância da Qui-quadrado)

Seja $X \sim \chi^2_n$. *Então,*

$$\mathbb{E}(X) = n \text{ e } Var(X) = 2n.$$

Definição 5.9.2 (Distribuição t de Student)

Seja X *uma variável aleatória contínua em* $(\Omega, \mathcal{S}, \mathbb{P})$ *com a seguinte função de densidade:*

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}.$$

X *é dita ter uma distribuição* t *de Student com* n *graus de liberdade,* $X \sim t_n$.

Note que X, assim definida, tem densidade simétrica.

Definição 5.9.3 (Distribuição F de Snedecor-Fisher)

Seja X *uma variável aleatória contínua em* $(\Omega, \mathcal{S}, \mathbb{P})$ *com a seguinte função de densidade:*

$$f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{m}{n} \left(\frac{mx}{n}\right)^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \mathbf{1}_{\mathbb{R}^+}(x).$$

X *é dita ter uma distribuição* F *com* n_1 *e* n_2 *graus de liberdade,* $X \sim F(n_1, n_2)$.

Proposição 5.9.2 (Operações com Amostras)

Sejam $X_1, X_2, \ldots, X_n$ *v.a's iid* $N(0,1)$. *Então,*

$X_1^2 \sim \chi^2_1$;
$\sum_{i=1}^n X_i^2 \sim \chi^2_n$;

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(0, \frac{1}{n}\right)$;

$(n-1)S^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sim \chi^2_{n-1}$;

$\bar{X}$ *é independente de* S^2;

$\bar{X}/\sqrt{S^2} \sim t_{(n-1)}$; *e*

$\bar{X}^2/S^2 \sim F_{1,(n-1)}$

5.10 Exercícios

Exercício 5.1 *Demonstre a seguinte afirmação.*

> *Seja X uma variável aleatória $U(0,1)$. Defina uma outra v.a. $Y = (b-a)X + a$, onde $a < b$, $a, b \in \mathbb{R}$. Então, Y também tem distribuição uniforme, no intervalo (a, b).*

Exercício 5.2 *Seja $X \sim N(0,1)$. Mostre que todo momento ímpar de X (e os centrais ímpares, por conseqüência) é nulo e que todo momento par μ_k é dado pela expressão*

$$\mu_k = (k-1)(k-3)\cdots 3 \cdot 1.$$

Encontre expressões equivalentes para uma variável normal com parâmetros μ e σ^2. Ache também expressões para os momentos centrais de uma $N(\mu, \sigma^2)$.

Exercício 5.3 *Seja X uma v.a. normal-padrão. Suponha que $Y = aX + b$, onde $a \neq 0$. Prove que Y também tem distribuição normal e ache seus parâmetros. Interprete a expressão normal-padrão em vista desse fato.*

Exercício 5.4 *Utilize-se do resultado do exercício 5.3 para mostrar que, se $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ e $\sigma > 0$, a probabilidade de qualquer evento do tipo $[X \leq a]$, $[X < a]$, $a \in \mathbb{R}$, pode ser calculada utilizando a probabilidade para eventos relativos a Y , uma v.a. normal-padrão. Argumente que o mesmo pode ser dito para uma classe maior de eventos.*

Exercício 5.5 *Suponha que X tenha distribuição Normal com média 100 e variância 400. Calcule: $\mathbb{P}(X = 100)$, $\mathbb{P}(X = 400)$, $\mathbb{P}(X \leq 100)$, $\mathbb{P}(X < 100)$, $\mathbb{P}(X \leq 140)$ e $\mathbb{P}(60 \leq X \leq 140)$.*

Exercício 5.6 *Mostre que a distribuição exponencial não tem memória.*

Exercício 5.7 *Seja X uma v.a. distribuída como uma $Gama(t, \lambda)$. Estude detalhadamente o comportamento (gráficos, esperança, variância, simetria etc) de X de acordo com os valores de $t > 0$ e $\lambda > 0$.*

Exercício 5.8 *Prove que o tempo de espera para a n-ésima ocorrência num processo de Poisson de intensidade λ tem distribuição $Gama(n, \lambda)$, $n \in \mathbb{N}^*$ e $\lambda > 0$.*

Exercício 5.9 *Seja X uma v.a. com distribuição $Gama(1, 1/10)$.*

(a) Qual a esperança e variância de X?

(b) Calcule a probabilidade de cada um dos seguintes eventos: $[X < 0]$, $[X = 0]$, $[X \leq 0]$, $[X > 3]$, $[X \geq 3]$, $[8 \leq X \leq 11]$; *e*

(c) Estime, pela desigualdade de Tchebichev, $\mathbb{P}(X > x)$ e compare com o valor exato, para $x = 1, 10, 50$ e 100.

Exercício 5.10 *Seja X uma v.a. com distribuição $Gama(10, 1/2)$.*

(a) Qual a esperança e variância de X?

(b) Calcule a probabilidade dos seguintes eventos: $[X < 0]$, $[X = 0]$, $[X \leq 0]$, $[X > 3]$, $[X \geq 3]$, $[8 \leq X \leq 11]$; *e*

(c) Estime, pela desigualdade de Tchebichev, $\mathbb{P}(X > x)$ e compare com o valor exato, para $x = 1, 10, 50$ e 100

Exercício 5.11 *Seja f_n a densidade de uma t de Student com n graus de liberdade ($X \sim t_n$). Mostre que $\lim_{n\to\infty} f_n = f$, onde f é densidade de uma distribuição Normal e identifique seus parâmetros. Consulte um livro de Teoria Estatística e interprete esse resultado à luz do fato de a variância ser conhecida ou estimada. Quais as consequências práticas de tal convergência?*

Exercício 5.12 *Seja X uma v.a. $Beta(\alpha, \beta)$. Estude detalhadamente o comportamento (gráficos, esperança, variância, simetria etc) de X de acordo com os valores de $\alpha > 0$ e $\beta > 0$. Demonstre, em particular, que*

- *quando $\alpha = \beta > 1$, a distribuição é unimodal, sendo este ponto sempre $1/2$;*
- *quando $1 < \alpha < \beta$, a distribuição è unimodal, sendo sua moda estritamente menor do que $1/2$;*
- *quando $1 < \beta < \alpha$, a distribuição è unimodal, sendo sua moda estritamente maior do que $1/2$; e*
- *quando $\min \alpha, \beta < 1$, o comportamento modal não representa um verdadeiro valor-padrão da respectiva distribuição.*

Exercício 5.13 *Seja $X \sim C(m, a)$. Demonstre que m e $2a$ são respectivamente a mediana e o desvio interquartílico de X.*

Exercício 5.14 *Ache os quartis (primeiro,segundo e terceiro) das seguintes distribuições: $U(a, b)$, $N(\mu, \sigma^2)$, $EE(\lambda)$, t_n (com $n = 5, 10, 20, 40, 60$ e 120 graus de liberdade) e Cauchy-padrão. Compare os quartis e discorra qualitativamente sobre as diferenças entre as distribuições.*

Em suas comparações (gráficas e analíticas, dê atenção especial às caudas de cada uma das distribuições, isto é, às respectivas probabilidades de eventos do tipo: $\mathbb{P}(|X| > \epsilon)$. Como reflexão inicial, pense na desigualdade de Tchebichev. Depois, procure um pacote (ou, caso seja possível, programe) para obter os quantis mais afastados da mediana da distribuição de cada uma das distribuições citadas. Como utilizá-los para comparar as respectivas caudas? Qual sua conclusão? Mostre que a Cauchy não tem esperança e interprete esse fato pelos estudos acima.

Capítulo 6

Teoremas-Limite I

6.1 Introdução

No começo do século XVIII, Jacques (Jacob) Bernoulli (1654-1705) e Abraham de Moivre (1667-1754) escreveram dois trabalhos fundamentais na então incipiente Teoria das Probabilidades. Bernoulli fazia parte de uma família de matemáticos suiços, entre os quais se destacaram ele e seu irmão Jean (1667-1748). Jacques escreveu um tratado chamado *Ars Conjectandi*, que só seria publicado em 1713, após sua morte. Nessa obra, ele demonstrou que, numa série de experimentos repetidos e independentes, a freqüência relativa de um evento converge a sua verdadeira probabilidade.[1] Esse resultado, posteriormente denominado Teorema de Bernoulli, admitiu diversas generalizações, denominadas *Leis dos Grandes Números*. Dadas as poucas ferramentas teóricas disponíveis na época, Bernoulli utilizou uma demonstração bastante complicada. Na segunda metade do século XIX, o matemático russo P. L. Tchebichev introduziu uma desigualdade - hoje conhecida como Desigualdade de Tchebichev- que lhe permitiu provar e generalizar o resultado de Bernoulli de uma forma muito simples. Assim, mesmo sabendo que o Teorema de Bernoulli precede a Desigualdade de Tchebichev em aproximadamente um século e meio, o texto desse capítulo se constroi a partir da Desigualdade de Tchebichev, o que o torna mais fluido e permite demonstrações curtas e elegantes[2].

[1] De um certo modo que será especificado ao longo deste capítulo.

[2] O leitor não deve pensar que essa construção dos Teoremas-limite seja de forma alguma peculiar. A construção de teorias não é necessariamente linear no tempo. Faz-se muitas vezes necessária uma reavaliação do ferramental anterior em vista de novas técnicas que não o substituirão mas preceder-lhe-ão na construção. No específico caso do Teorema de

Abraham de Moivre foi um huguenote francês que, devido a problemas religiosos foi morar na Inglaterra. Apesar de seu talento e contribuições importantíssimas, jamais conseguiu uma cátedra de professor universitário. Na segunda edição do seu tratado *Doctrine of Chances* (1738), ele utilizou pela primeira vez na história a posteriormente denominada curva normal ou Gaussiana[3], com o objetivo específico de facilitar os cálculos da somas de termos da Distribuição Binomial, no que seria a primeira versão do Teorema Central do Limite . O estudo dos resultados de Bernoulli e de de Moivre e suas respectivas extensões, a Lei dos Grandes Números e o Teorema Central do Limite, constituem o objetivo principal deste capítulo.

6.2 Desigualdade de Tchebichev

O Teorema 6.2.1 e seu corolário 6.2.1 são versões da Desigualdade de Tchebichev, já vista no capitulo 3 (Teorema 3.5.1). Para maior conforto do leitor, enunciamos novamente esse importante resultado.

Teorema 6.2.1 (Desiguladade de Tchebichev)

Sejam X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e $f : \mathbb{R} \to \mathbb{R}$ uma função não-negativa. Dado um conjunto A tal que $\emptyset \neq A \subset \mathbb{R}$, seja $i_A = \inf\{f(x) : x \in A\}$. Nessas condições:

$$\mathbb{E}(f(X)) \geq i_A.\mathbb{P}(X \in A).$$

Demonstração:

$$\begin{aligned}
\mathbb{E}(f(X)) &= \mathbb{E}\left(f(X).\mathbf{1}_{[X \in A]} + f(X).\mathbf{1}_{[X \notin A]}\right) \\
&= \mathbb{E}\left(f(X).\mathbf{1}_{[X \in A]}\right) + \mathbb{E}\left(f(X).\mathbf{1}_{[X \notin A]}\right).
\end{aligned}$$

Bernoulli, seu estudo é motivado por razões históricas e didáticas e essas demandam que a Desigualdade de Tchebichev seja utilizada.

[3]Gauss nasceu em 1777!

Sendo f não-negativa,

$$\mathbb{E}\left(f(X).\mathbf{1}_{[X\notin A]}\right) \geq 0$$

e, portanto,

$$\mathbb{E}(f(X)) \geq \mathbb{E}\left(f(X).\mathbf{1}_{[X\in A]}\right) \geq i_A.\mathbb{E}\left(\mathbf{1}_{[X\in A]}\right) = i_A.\mathbb{P}\left(X \in A\right). \blacksquare$$

Algumas das muitas aplicações do Teorema 6.2.1 se encontram no corolário 6.2.1. O Teorema 6.2.1 será também utilizado diretamente na demonstração do Teorema de Bernoulli.

Corolário 6.2.1 *Seja X uma variável aleatória definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Então, para quaisquer $\varepsilon > 0$ e $j \in \mathbb{N}$, têm-se*

a)

$$\mathbb{P}(|X| > \varepsilon) \leq \frac{\mathbb{E}(|X|^j)}{\varepsilon^j}.$$

Em particular, para $j = 1$ e g a função identidade, obtem-se a Desigualdade de Markov (3.9)

$$\mathbb{E}((|X|) \geq \varepsilon\mathbb{P}(|X| > \varepsilon);$$

b) se $\mathbb{E}(|X|) < +\infty$,

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^j)}{\varepsilon^j}.$$

Em particular, para $j = 2$, se $\mathbb{E}\left(X^2\right) < +\infty$, tem-se a Desigualdade de Tchebichev (ou Bienaymé-Tchebichev)

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2};$$

c) se $\sigma_j := \mathbb{E}(|X - \mathbb{E}(X)|^j) < +\infty$*,*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > \varepsilon.(\sigma_j)^{1/j}\right) \leq \frac{1}{\varepsilon^j}.$$

Em particular, para $j = 2$*, se* $\sigma^2 := Var(X) < +\infty$*, tem se*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon.\sigma) \leq \frac{1}{\varepsilon^2}.$$

Demonstração:
a) Aplique o Teorema 6.2.1 a $f(x) = |x|^j$ e $A = \{x \in \mathbb{R} : |x| > \varepsilon\}$;

b) Aplique o Teorema 6.2.1 a $f(x) = |x - \mathbb{E}(X)|^j$ e $A = \{x \in \mathbb{R} : |x - \mathbb{E}(X)| > \varepsilon\}$; e

c) Use $\varepsilon = \varepsilon.(\sigma_j)^{1/j}$ em b). ■

Observação 6.2.1

i) Embora o item a) do Corolário 6.2.1 seja válido em geral, a fórmula só tem utilidade quando $\mathbb{E}(|X|^j) < +\infty$ [4]*. Analogamente, a parte b) só é útil se o* j*-ésimo momento central absoluto,* σ_j*, for finito*

ii) Os itens a) e b) do Corolário 6.2.1 nos permitem estudar a velocidade do decrescimento das respectivas probabilidades dos conjuntos $\{|X| > \varepsilon\}$ *e* $\{|X - \mathbb{E}(X)| > \varepsilon\}$*, quando* ε *cresce (as* **caudas** *da distribuição de* X*), a partir da existência dos seus momentos;*

iii) Se $\sigma_j = 0$*, os itens b) e c) do Corolário 6.2.1 são trivialmente verdadeiros e inúteis*[5]*; se* σ_j *for finito, os resultados*

[4] Na realidade, sua utilidade só existe se $\mathbb{E}(|X|^j) < \varepsilon^j$.

[5] Pense na probabilidade de um desvio com relação à esperança, caso a variável tenha momentos centrais absolutos nulos!

são equivalentes; se $\sigma_j = +\infty$, o enunciado do corolário c) não faz sentido;

iv) Se $\varepsilon^j \leq \sigma_j$, o item b) do Corolário 6.2.1 não fornece qualquer informação de interesse; o mesmo ocorre no item c), se $\varepsilon \leq 1$.

6.2.1 Teorema de Weierstrass - Uma Demonstração de Bernstein

Nesta subseção, apresentaremos uma demonstração de Bernstein para um conhecido resultado de Weierstrass (1815-1897) , em que se mostra que toda função definida num intervalo fechado pode ser aproximada uniformemente por polinômios. Trabalharemos aqui com o intervalo $[0,1]$, mas o resultado pode ser estendido facilmente a qualquer intervalo $[a,b]$, com $a < b$.

Devemos alertar o leitor que essa subseção é de interesse indireto, isto é, sua motivação é uma utilização extremamente inteligente da desigualdade de Tchebichev para demonstrar um resultado aparentemente disconexo do objetivo para o qual a desigualdade foi criada. Essa utilização de resultados *probabilísticos* em problemas onde a aleatoriedade não parece ter papel não é exclusiva desse resultado. O leitor interessado deve dar uma lida em [20] , por exemplo, para ver outras utilizações desse tipo. Finalmente, pela natureza dessa subseção, o leitor pode suprimí-la de uma primeira leitura.

Seja $\mathcal{C}([0,1])$ o conjunto das funções contínuas sobre $[0,1]$, $f : [0,1] \to R$. O n-ésimo polinômio de Bernstein correspondente à função f é definido por:

$$B_n(x) = \sum_{i=0}^{n} \binom{n}{i} f(i/n) x^i (1-x)^{n-i}.$$

Teorema 6.2.2 (Teorema de Weierstrass)

Sejam $\epsilon > 0$ e $f \in \mathcal{C}([0,1])$. Então, existe um número natural, $n_0 = n_0(\epsilon, f)$, tal que

$$n \geq n_0 \Rightarrow |B_n(x) - f(x)| \leq \epsilon \quad \forall x \in [0,1].$$

Demonstração:

Sejam $\epsilon > 0$ e $f \in \mathcal{C}([0,1])$ fixos por toda a demonstração. Dado que toda função contínua em um intervalo fechado é também uniformemente contínua, existe um $\delta > 0$, $(\delta = \delta(\epsilon, f))$ tal que:

$$x \in [0,1],\ \ y \in [0,1] \ \ e \ \ |x-y| < \delta \quad \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{2}.$$

Consideremos δ satisfazendo a condição acima fixado a partir deste momento. Por ser contínua em $[0,1]$, f é limitada, ou seja, existe um número natural c tal que:

$$|f(x)| \leq c \quad \forall x \in [0,1]$$

e, portanto,

$$|f(x) - f(y)| \leq 2c \quad \forall x \in [0,1] \ \ \forall y \in [0,1].$$

Por outro lado, dado que

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

tem-se

$$f(x) = \sum_{i=0}^{n} f(x) \binom{n}{i} x^i (1-x)^{n-i}.$$

Como conseqüência,

$$|f(x) - B_n(x)| \leq \sum_{i=0}^{n} |f(x) - f(i/n)| \binom{n}{i} x^i (1-x)^{n-i};$$

e

$$\begin{aligned} |f(x) - B_n(x)| \quad \leq \quad & \sum_{|x-i/n|<\delta} |f(x) - f(i/n)| \binom{n}{i} x^i (1-x)^{n-i} + \\ & + \sum_{|x-i/n|\geq\delta} |f(x) - f(i/n)| \binom{n}{i} x^i (1-x)^{n-i}. \end{aligned}$$

Denotemos por S_1 e S_2, respectivamente, o primeiro e o segundo somandos da expressão acima. Seja n_1 um número natural tal que $(n_1)^{-1} < \delta$ ou, equivalentemente, tal que $n_1 > \delta^{-1}$. Dada a continuidade uniforme de f e as relações entre ϵ, δ e n_1, tem-se:

$$\forall n \geq n_1 \quad S_1 \leq \frac{\epsilon}{2} \sum_{|x-i/n|<\delta} \binom{n}{i} x^i (1-x)^{n-i} \leq \frac{\epsilon}{2}.$$

Por outro lado, pela escolha da constante c:

$$S_2 \leq 2c \sum_{|x-i/n|\geq\delta} \binom{n}{i} x^i (1-x)^{n-i}.$$

Assim, dado que

$$\sum_{|x-i/n|\geq\delta} \binom{n}{i} x^i (1-x)^{n-i}$$

nada mais é do que a probabilidade do conjunto

$$\{i : |x - i/n| \geq \delta\} = \{i : |i - nx| \geq \delta \times n\},$$

sob a distribuição $Bin(n,p)$, a aplicação da desigualdade de Tchebichev resulta em:

$$S_2 \leq 2c\frac{Var(i)}{\delta^2 n^2} = \frac{2cnx(1-x)}{\delta^2 \times n^2} \leq \frac{2c}{n\delta^2}.$$

Seja $n_2 \geq \frac{4c}{\epsilon\delta^2}$:

$$\forall n \geq n_2 \quad \frac{2c}{\delta^2 \times n^2} < \frac{\epsilon}{2}.$$

Finalmente, se $n_0 = \max(n_1, n_2)$, é fácil verificar que

$$\forall n \geq n_0 \quad |B_n(x) - f(x)| \leq \epsilon \ \ \forall x \in [0,1]. \quad \blacksquare$$

Corolário 6.2.2 *Sejam a e b números reais tais que $a < b$ e $f \in \mathcal{C}([a,b])$. Nessas condições, existe uma seqüência de polinômios P_n definidos no intervalo $[a,b]$, tal que P_n converge uniformemente para f.*

Os detalhes da demonstração do teorema e a verificação do corolário ficam por conta do leitor (ver exercício 6.1).

6.3 Lei Fraca dos Grandes Números

Definição 6.3.1 (Lei Fraca dos Grandes Números)

Seja $\{X_j\}_{j\in\mathbb{N}}$ uma seqüência de variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tais que $\mathbb{E}(|X_j|) < +\infty$, para todo $j \in \mathbb{N}$. Defina $\{S_n\}_{n\in\mathbb{N}}$ como a seqüência de somas parciais, isto é, $S_n = \sum_{j=1}^{n} X_j$. Nessas condições, diremos que a seqüência $\{X_j\}_{j\in\mathbb{N}}$ satisfaz à Lei Fraca dos Grandes Números *se, para qualquer $\varepsilon > 0$,*

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| > \varepsilon\right) = 0.$$

As diversas versões existentes de *Leis Fracas dos Grandes Números* que veremos em seguida (e todas as demais que omitiremos) dizem respeito às condições suficientes para que a seqüência de variáveis aleatórias $\{X_j\}_{j\in\mathbb{N}}$ satisfaça à condição:

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| > \varepsilon\right) = 0.$$

Teorema 6.3.1 (Teorema de Bernoulli)

Consideremos uma seqüência de ensaios de Bernoulli independentes (veja Capítulo 4). Sejam p a probabilidade de sucesso em cada ensaio e S_n o número observado de sucessos nos n primeiros *ensaios. Então, para todo $\varepsilon > 0$, vale que*

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = 0.$$

Esse resultado não nos deve surpreender de todo pois nada mais é do que uma formalização matemática do que fora já visto no exemplo 4.3.1. Recordemos que, nesse exemplo, discutimos a *honestidade* de moedas e, através de números simulados de lançamentos de uma moeda, ilustramos graficamente que sua freqüência relativa de caras se aproximava do verdadeiro valor p à medida em que n crescia: isso é exatamente o que está formalizado no Teorema de Bernoulli.

Demonstração:

Dado que S_n tem distribuição $Bin(n,p)$, tem-se que $\mathbb{E}(S_n) = np$ e $Var(S_n) = np(1-p)$. Então, $\mathbb{E}(S_n/n) = p$ e $Var(Sn/n) = p(1-p)/n$; aplicando-se a Desigualdade de Tchebichev (Corolário 6.2.1 item b)) à variável S_n/n obtém-se

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Portanto, basta-nos tomar $n \to +\infty$. ∎

Observação 6.3.1 *A tese do Teorema de Bernoulli pode ser escrita na seguinte forma equivalente.*

Para qualquer $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) = 1.$$

Tendo em mente essa última versão, devemos salientar que o Teorema de Bernoulli não implica na convergência das freqüências relativas à probabilidade p, no sentido ordinário de análise. Com efeito, a afirmação $\lim_{n\to\infty} S_n/n = p$ considerada no sentido usual significaria que, para qualquer $\varepsilon > 0$, existiria um índice $n_0 \in \mathbb{N}$

$(n_0 = n_0(\varepsilon))$, tal que, para todo $n \geq n_0$, teríamos $|Sn/n - p| \leq \varepsilon$. Estaríamos, portanto, afirmando que, para qualquer $\varepsilon > 0$, existiria um índice $n_0 \in \mathbb{N}$ $(n_0 = n_0(\varepsilon))$, tal que, para todo $n \geq n_0$, o evento $\{|Sn/n - p| \leq \varepsilon\}$ seria o evento certo[6]*, numa linguagem menos formal: S_n/n estaria* arbitrariamente perto *de p se n fosse* suficientemente grande.

A tese do teorema de Bernoulli, no entanto, é mais fraca e tem o seguinte significado

Para qualquer $\varepsilon > 0$ **e qualquer** *$\delta > 0$, existe um índice $n_0 \in \mathbb{N}$ $(n_0 = n_0(\varepsilon, \delta))$, tal que, para todo $n \geq n_0$, temos $\mathbb{P}(|Sn/n - p| > \varepsilon) < \delta$ ou, equivalentemente, para qualquer $\varepsilon > 0$* **e qualquer** *$\delta > 0$, existe um índice $n_0 \in \mathbb{N}$ $(n_0 = n_0(\varepsilon, \delta))$, tal que, para todo $n \geq n_0$, temos $\mathbb{P}(|Sn/n - p| \leq \varepsilon) > 1 - \delta$.*

Informalmente, o teorema de Bernoulli afirma que a probabilidade de S_n/n estar arbitrariamente perto *de p estará* tão próxima de 1 quanto se desejar*, se n for* suficientemente grande.

Nesta altura, é interessante dizer que, em 1909, Borel provou a **Lei Forte dos Grandes Números***, nos seguintes termos:*

Consideremos uma seqüência de ensaios de Bernoulli independentes; sejam p a probabilidade de sucesso em cada ensaio e S_n o número observado de sucessos nos n primeiros ensaios. Então:

$$\mathbb{P}(\lim_{n \to +\infty} S_n/n = p) = 1.$$

Observe que esse resultado é o mais próximo possível à convergência de Sn/n a p no sentido ordinário da análise, em que $\{\lim_{n \to \infty} S_n/n = p\}$ seria o evento certo (visto que este resultado não existe).

[6]Aqui, devemos recordar que, quando nos referimos ao *evento certo*, não nos referimos apenas a um evento que tem probabilidade 1: a notação *evento certo* é muito mais do que isso: ela significa TODO O ESPAÇO AMOSTRAL.

Dado o exposto na observação anterior, é claro que o Teorema de Bernoulli não demonstra a convergência da freqüência relativa à probabilidade. Segundo Todhunter, em [35], Bernoulli jamais pensou em tal coisa, tendo inclusive afirmado que a importância do seu resultado era permitir a construção do que hoje é denominado um **intervalo de confiança** para o valor p desconhecido (veja o exemplo 6.3.1).

O tipo de convergência de que trata o Teorema de Bernoulli (e em geral as chamadas *Leis Fracas dos Grandes Números*), foi chamado de *Convergência em Probabilidade* pelo matemático italiano Cantelli, no começo do século XX, sendo até hoje assim denominado.

Quando a probabilidade é pensada como caso particular de medida, esse conceito equivale ao de *Convergência em Medida.* Não nos ocuparemos de Teoria da Medida neste texto, a não ser no caso muito particular da Teoria das Probabilidades e deixamos esta observação apenas para o esclarecimento do leitor.

Não dispondo da Desigualdade de Tchebichev, Bernoulli se defrontou com o cálculo de $\mathbb{P}(|S_n/n - p| > \varepsilon)$ ou, equivalentemente,

$$\mathbb{P}(|S_n - np| > n\varepsilon).$$

onde S_n tem distribuição binomial, cuja complexidade de manipulação cresce com n (razão aliás por que se busca um teorema limite).

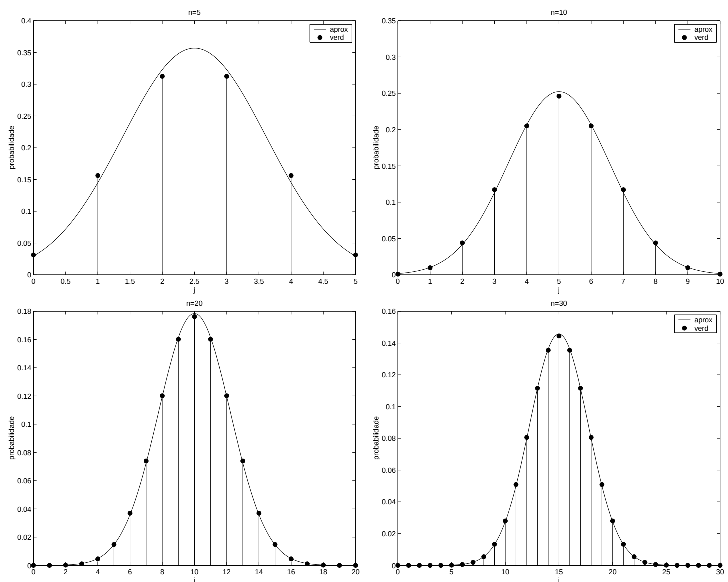
Curiosamente, na tentativa de melhorar a demonstração do Teorema de Bernoulli, de Moivre teve a idéia de substituir as probabilidades binomiais por uma expressão assintótica. Mais precisamente, ele demonstrou que, se S_n tem distribuição $Bin(n, 1/2)$, então:

$$\mathbb{P}(Sn = j) \approx (2\pi n(1/2)^2)^{-1/2} \exp\left(-\frac{(j - n/2)^2}{2n(1/2)^2}\right),$$

se n for *suficientemente grande*, introduzindo desta forma a famosa *curva Gaussiana* e dando também a primeira versão do Teorema Central do Limite : a convergência da distribuição Binomial $(n, 1/2)$ à normal.

Nos gráficos da figura 6.1, vemos o comportamento de tal aproximação. Note quão bom é seu desempenho, mesmo para aparentemente valores pequenos de n. No entanto, na tabela 6.1, vemos que o desempenho absoluto (bem retratado nos gráficos) melhora rapidamente em n mas exatamente o contrário ocorre com o erro relativo máximo, que **cresce** com n. Devemos salientar, no entanto, que essa piora acontece apenas nas caudas, não querendo dizer que a aproximação esteja piorando.

Figura 6.1: Aproximação da Binomial pela Normal, $p = 1/2$



Teorema 6.3.2 (Lei Fraca dos Grandes Números para v.a.'s iid.)

Seja $\{X_j\}_{j\in\mathbb{N}}$ *uma seqüência de variáveis aleatórias iid definidas no espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$, *tais que* $Var(X_1) < +\infty$. *Defina-se a seqüência de somas parciais* $\{S_n\}_{n\in\mathbb{N}}$, *isto é,* $S_n =$

Tabela 6.1: Erro da Aproximação da Binomial pela Normal, $p = 1/2$

n	Erro Abs. Máximo	Erro Rel. Máximo
5	$0,011$	$0,0715$
10	$0,0062$	$0,7409$
20	$0,0022$	$7,4934$
30	$0,0012$	$46,85$
50	$0,000056$	$1,7634 \times 10^3$
100	$0,000020$	$1,95 \times 10^7$
200	$0,0000070$	$3,37 \times 10^{15}$

$\sum_{j=1}^n X_j$. *Então,* $\{X_n\}_{n\in\mathbb{N}}$ *satisfaz a* **Lei Fraca dos Grandes Números** *ou, equivalentemente, para qualquer* $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}(|Sn/n - \mathbb{E}(X_1)| > \varepsilon) = 0.$$

Demonstração:

Aplicaremos a Desigualdade de Tchebichev (Corolário 6.2.1 item b) à variável aleatória S_n/n. Para tanto, precisamos calcular suas esperança e variância.

$$\mathbb{E}(S_n/n) = n^{-1}\mathbb{E}(Sn) = n^{-1}\mathbb{E}\left(\sum_{j=1}^n X_j\right) = n^{-1}\sum_{j=1}^n \mathbb{E}(X_j) = \mathbb{E}(X_1).$$

Sendo $X_1, X2, \ldots, Xn, ..$ iid, temos que

$$Var\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n Var(X_j)$$

e, portanto,

$$Var(S_n/n) = n^{-2}Var(S_n) = n^{-2}Var\left(\sum_{j=1}^{n} X_j\right) = n^{-2}\sum_{j=1}^{n} Var(X_j) = n^{-1}Var(X_1).$$

Então, pela Desigualdade de Tchebichev (pelo item b do Corolário 6.2.1):

$$\mathbb{P}(|S_n/n - \mathbb{E}(X_1)| > \varepsilon) \leq \frac{Var(X_1)}{n\varepsilon^2}.$$

Sendo a variância de X_1 finita, temos o resultado, quando $n \to +\infty$. ■

Observação 6.3.2

a) Note que no Teorema 6.3.1, S_n representava o número de sucessos nos n primeiros experimentos de Bernoulli independentes. Portanto, o teorema 6.3.2 é uma generalização do Teorema de Bernoulli, já que S_n, definida no Teorema 6.3.1, pode ser escrita $S_n = \sum_{j=1}^{n} X_j$, onde $X_1, X_2, \ldots, X_n$ são iid Bernoulli (p) e, no teorema 6.3.2, nenhuma específica distribuição é assumida para cada X_j, existindo apenas a restrição de que as variâncias devam ser finitas e as variáveis independentes e identicamente distribuídas;

b) o resultado

$$\lim_{n \to +\infty} \mathbb{P}(|Sn/n - \mathbb{E}(X_1)| > \varepsilon) = 0$$

implica em que os valores da variável aleatória S_n/n se concentram em torno do valor $\mathbb{E}(X_1)$, à medida em que $n \to +\infty$, sendo esse o motivo por que o valor $\mathbb{E}(X_1)$ pode ser legitimamente chamado de valor esperado *ou* esperança *(ver tambem o Capítulo 3 sobre a origem da palavra);*

c) o resultado intermediário,

$$\mathbb{P}(|S_n/n - \mathbb{E}(X_1)| > \varepsilon) \leq \frac{Var(X_1)}{n\varepsilon^2},$$

é de muita utilidade para a estimação de $\mathbb{E}(X_1)$, como pode ser visto no exemplo 6.3.1.

Exemplo 6.3.1 (Pesquisa Eleitoral)

Suponha que desejemos estimar a proporção do eleitorado que pretende votar em um certo candidato. Qual deve ser o tamanho amostral para garantir um certo erro entre a proporção populacional, p, e a proporção amostral, S_n/n?

Antes de resolvermos esse problema, devemos refletir sobre a definição de **erro**. *Usualmente, quando se fala em erro, está-se diante de um número real que exprime a (in)capacidade de uma certa quantidade ao representar uma outra. No caso em estudo, porém, devemos pensar que, sendo uma das quantidades baseada na amostra e a verdadeira, populacional, essa simples interpretação não nos seria possível - aqui, a interpretação é análoga à de convergência de S_n/n para p.*

Como temos, associada a uma medida amostral (que é uma variável aleatória) uma incerteza (expressa por um modelo probabilístico)[7]*, restrições de distância entre as proporções amostral e populacional só podem ser avaliadas em conjuntos contidos em Ω e nunca no próprio evento certo. Portanto, quando se fala que desejamos encontrar um tamanho amostral suficiente para um certo erro máximo, por exemplo $0,01$, temos que fazê-lo com uma medida de certeza a ela associada, expressa numa probabilidade, como por exemplo $0,95$. Matematicamente, queremos encontrar n tal que:*

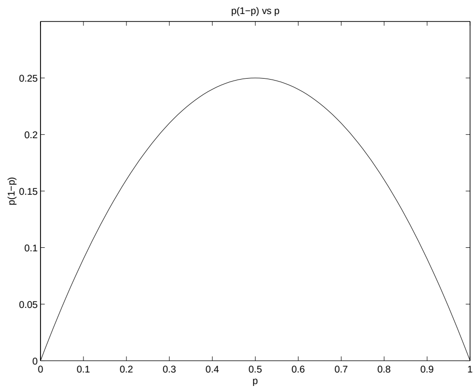
$$\mathbb{P}(|S_n/n - p| > 0,01) \leq 0,95.$$

Supondo ter S_n uma distribuição $Bin(n,p)$[8] *e X_1 distribuição*

[7] Nesse exemplo, considera-se que o voto de cada eleitor tenha uma distribuição $b(p)$

[8] Fica para o leitor a justificativa dessa hipótese.

Figura 6.2: Incerteza na Pesquisa Eleitoral



$b(p)$, a fórmula mencionada na Observação 6.3.2 c) nos diz que

$$\mathbb{P}(|S_n/n - p| > 0,01) \leq \frac{p(1-p)}{n(0,01)^2}.$$

Desconhecemos p, mas sabemos que $p(1-p)$ é uma parábola convexa, ilustrada na figura 6.2. É fácil ver que seu máximo ocorre em $1/2$, isto é, $p(1-p) \leq 1/4$. Portanto, trabalhando-se com a pior hipótese ($p = 1/2$), temos:

$$\mathbb{P}(|S_n/n - p| > 0,01) \leq \frac{1}{4n(0,01)^2}.$$

Finalmente, igualando nosso valor teórico $1/(4n(0.01)^2)$ ao valor máximo desejado de incerteza, $0,05 = 1 - 0,95$, verificamos que

$$n \geq (0,05).4(0,01)^2)^{-1}$$

nos garante que $|S_n/n - p| > 0,01$ tem probabilidade no máximo igual a 0,05 ou, equivalentemente, $n \geq (0,05.4(0,01)^2)^{-1}$ é uma *condição suficiente para que* $|S_n/n - p| \leq 0,01$ *tenha uma probabilidade no mínimo igual a 0,95;*

Podemos provar que, em geral, quando se deseja estimar uma proporção p, de tal forma que o erro na estimação seja menor do que um $\varepsilon > 0$ *com probabilidade pelo menos igual a* $1 - \alpha$ *(*ε *e* α *prefixados), é suficiente que o tamanho da amostra* n *seja maior do que ou igual a* $(4\alpha\varepsilon^2)^{-1}$.

Note que, no caso em questão, $\varepsilon = 0,01$ *e* $\alpha = 0,05$, n *deveria ser pelo menos* 50000, *um número absurdo para uma amostra. Felizmente, as técnicas de amostragem são desenvolvidas de forma bem mais sofisticada do que nesse exemplo, diminuindo-se assim o tamanho amostral necessário a valores viáveis. Para maiores detalhes, sugerimos ao leitor interessado a consulta aos excelentes [17] e [18].*

Teorema 6.3.3 (Lei Fraca dos Grandes Números de Poisson)

Seja $\{X_n\}_{n\in\mathbb{N}}$ *uma seqüência de variáveis aleatórias independentes, onde* X_j *tem distribuição* $b(p_j)$, *para cada* $j \in \mathbb{N}$. *Defina-se ainda a soma parcial* $S_n = \sum_{j=1}^n X_j$, *para cada* $n \in \mathbb{N}$ *e* $\lambda_n = n^{-1}\sum_{j=1}^n p_j$, *para todo* $n \in \mathbb{N}$. *Nessas condições, a seqüência* $\{X_n\}_{n\in\mathbb{N}}$ *satisfaz a* **Lei Fraca dos Grandes Números**, *ou seja, para qualquer* $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}(|Sn/n - \lambda_n| > \varepsilon) = 0.$$

Demonstração:

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \mathbb{E}(X_j) = \sum_{j=1}^n p_j$$

e

$$Var(S_n) = Var\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} Var(X_j) = \sum_{j=1}^{n} p_j(1-p_j),$$

dado que as variáveis aleatórias são independentes.

Portanto,

$$\mathbb{E}(S_n/n) = \lambda_n \text{ e } Var(S_n/n) = \frac{1}{n^2}\sum_{j=1}^{n} p_j(1-p_j)$$

Como temos $\mathbb{E}(S_n/n)$ e sabemos que $p(1-p) \leq 1/4$, basta-nos aplicar a Desigualdade de Tchebichev (Corolário 6.2.1 item b) à variável aleatória S_n/n. ■

Observação 6.3.3 *Tanto o teorema 6.3.2 como o teorema 6.3.3 são generalizações do Teorema de Bernoulli (Teorema 6.3.1). No entanto, o teorema 6.3.2 o faz com a flexibilização da distribuição de cada v.a. enquanto o teorema 6.3.3 relaxa a hipótese de que as variáveis sejam identicamente distribuídas. O resultado que provaremos a seguir é uma generalização dos teoremas 6.3.2 e 6.3.3 (e portanto também do Teorema 6.3.1).*

Teorema 6.3.4 (Lei Fraca dos Grandes Números de Tchebichev)

Seja $\{X_n\}_{n\in\mathbb{N}}$ uma seqüência de variáveis aleatórias. Definam-se a somas parciais, $S_n = \sum_{j=1}^{n} X_j$, para cada $n \in \mathbb{N}$. Se as variáveis forem independentes, com variâncias finitas e, além disso,

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{j=1}^{n} Var(X_j) = 0,$$

então, a seqüência $\{X_n\}_{n\in\mathbb{N}}$ satisfaz a **Lei Fraca dos Grandes Números**, *ou seja, para qualquer $\varepsilon > 0$,*

$$\lim_{n\to+\infty} \mathbb{P}(|(Sn - \mathbb{E}(S_n))/n| > \varepsilon) = 0.$$

Demonstração:

Aplique a Desigualdade de Tchebichev clássica (Corolário 6.2.1 item b) à variável S_n/n. ■

Observação 6.3.4 *É possível provar-se que, nos teoremas 6.3.2, 6.3.3 e 6.3.4, a hipótese de independência pode ser substituída pela mais fraca $Cov(X_i, X_j) = 0$ para quaisquer $i, j \in \mathbb{N}$, $i \neq j$.*

6.4 Teorema Central do Limite

6.4.1 O Teorema de de Moivre

Teorema 6.4.1 (Teorema de de Moivre)

Consideremos uma seqüência de ensaios de Bernoulli independentes (conforme a definição 4.2.1). Sejam p a probabilidade de sucesso em cada ensaio e S_n o número observado de sucessos nos n primeiros ensaios. Então, para n suficientemente grande e para qualquer $x \in \{0, \ldots, n\}$, vale que

$$\mathbb{P}(S_n = x) \asymp \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left[-\frac{(x-np)^2}{2np(1-p)}\right].$$

Demonstração:

Por conveniência, defina-se $q = 1 - p$. Seja $P_n(x)$ a probabilidade de $[S_n = x]$:

$$P_n(x) = \mathbb{P}(S_n = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}.$$

A fórmula de Stirling afirma (veja Observação 6.4.1 iii) a seguir) que, se $m \in \mathbb{N}$, então

$$m! \asymp m^m e^{-m} \sqrt{2\pi m}.$$

Portanto,

$$\begin{aligned} P_n(x) &\asymp \frac{n^n e^{-n}\sqrt{2\pi n}p^x q^{n-x}}{x^x e^{-x}\sqrt{2\pi x}(n-x)^{n-x}e^{-n+x}\sqrt{2\pi(n-x)}} \\ &= \frac{\left(\frac{np}{x}\right)^{x+1/2}\left(\frac{nq}{n-x}\right)^{n-x+1/2}}{\sqrt{2\pi npq}}. \end{aligned} \tag{6.1}$$

Sejam

$$w = \left(\frac{x}{np}\right)^{x+1/2}\left(\frac{n-x}{nq}\right)^{n-x+1/2}$$

e

$$t = \frac{x-np}{\sqrt{npq}}.$$

Conseqüentemente

$$x = np + t\sqrt{npq}$$

e

$$\begin{aligned} \ln w &= (x+1/2)\ln\left(\frac{x}{np}\right) + (n-x+1/2)\ln\left(\frac{n-x}{nq}\right) \\ &= (np + t\sqrt{npq} + 1/2)\ln\left(1 + \frac{tq}{\sqrt{npq}}\right) + \\ &\quad +(nq - t\sqrt{npq} + 1/2)\ln\left(1 - \frac{tp}{\sqrt{npq}}\right). \end{aligned} \tag{6.2}$$

Consideremos a função $\ln(1+u)$ e sua representação pelo polinômio de Mc Laurin de grau 2, ou seja,

$$\ln(1+u) \asymp u - \frac{u^2}{2}[9]. \tag{6.3}$$

Finalmente, se n for *suficientemente grande*,

[9] Veja as condições e as conseqüências dessa substituição na Observação 6.4.1 iii

$$\ln\left(1+\frac{tq}{\sqrt{npq}}\right)=\frac{tq}{\sqrt{npq}}-\frac{t^2q^2}{2npq} \tag{6.4}$$

e

$$\ln\left(1-\frac{tp}{\sqrt{npq}}\right)=-\frac{tp}{\sqrt{npq}}-\frac{t^2p^2}{2npq}. \tag{6.5}$$

Substituindo-se (6.2) e (6.5) em (6.2), temos:

$$\ln w \asymp \frac{t^2}{2}+\left(\frac{t(q-p)+t^3(p^2-q^2)}{2\sqrt{npq}}\right)-\frac{t^2(q^2+p^2)}{4npq} \tag{6.6}$$

e, finalmente, utilizando-se (6.6) em (6.1):

$$P_n(x)\asymp\frac{w^{-1}}{\sqrt{2\pi npq}}=\frac{\exp(-\ln w)}{\sqrt{2\pi npq}}=\asymp\frac{1}{\sqrt{2\pi npq}}\exp(-t^2/2),$$

que é o resultado desejado. ■

Observação 6.4.1

i) A diferença entre $m!$ e $m^m e^{-m}\sqrt{2\pi m}$ não é limitada, motivo pelo qual a fórmula de Stirling não deve ser usada para aproximar o valor de um fatorial de forma isolada;

ii) Em termos estritos, o sentido da aproximação *da Fórmula de Stirling é o seguinte:*

$$\lim_{n\to\infty}\frac{m!}{m^m e^{-m}\sqrt{2\pi m}}=1.$$

Observe que, na demonstração do Teorema de de Moivre, a Fórmula *de Stirling foi utilizada para calcular o limite de quocientes de fatoriais*[10]*; portanto não há contradição com a observação i);*

iii) Na demonstração do Teorema de de Moivre, a função $f(u) = \log(1+u)$ *foi substituída pelo correspondente polinômio de Mc Laurin de grau 2, ou seja,*

$$f(u) = \log(1+u) \asymp u - \frac{u^2}{2}, \tag{6.7}$$

prescindindo-se, portanto, do termo complementar T_c*; para que esse argumento funcione, suporemos* $t = (x-np)/(npq)^{1/2}$ *limitado, ou seja,* $|t| \leq M$*, onde* $M > 0$ *é arbitrário. Assim, dados* M *e* $0 < \alpha < 1$*, existe* $n_0 = n_0(\alpha, M)$ *tal que, para qualquer* $n \geq n_0$ *e qualquer* $|t| \leq M$*, tem se*

$$\left|\frac{tp}{\sqrt{npq}}\right| \leq \alpha$$

e

$$\left|\frac{tq}{\sqrt{npq}}\right| \leq \alpha,$$

dado que o raio de convergência da série de $f(u) = \log(1+u)$ *é* 1 *e, se* $|t| \leq M$*, é lícito aproximar* $f((tp)/\sqrt{npq})$ *e* $f((tq)/\sqrt{npq})$*, mediante a substituição de* u *por* $(tp)/\sqrt{npq}$ *e* $(tq)/\sqrt{npq}$ *em (6.7), como foi feito na demonstração do teorema.*

Observemos ainda que

$$T_c = \frac{u^3}{3!} f^{(3)}(1+\theta u) = \frac{u^3}{3(1+\theta u)^3} \leq \frac{M^3}{3(npq)^{3/2}(1-u)},$$

[10]Veja também a versão da fórmula de Stirling no apêndice.

onde $0 < \theta < 1$*:*

$$0 < \theta < 1 \ , \ |u| \leq \alpha \Rightarrow |1 + \theta u| \geq 1 - \alpha > 0.$$

Assim, dado $\varepsilon > 0$*, deve existir um* $n_1 \geq n_0$ *tal que* $|T_c| < \varepsilon$*.*

Finalmente, devemos salientar que a escolha de M *precede a escolha de* n_0 *(e portanto de* n_1*) e isto significa que as desigualdades (e portanto a aproximação) não serão válidas fora do intervalo* $|t| \leq M$*. Observe os problemas com a aproximação nas caudas da binomial pela distribuição normal, na página 253; e*

iv) Para qualquer x *fixo,* $\lim_{n\to\infty} |x-np| = +\infty$*. Logo,* $\lim_{n\to\infty} P_n(x) = 0$ *(pela Desigualdade de Tchebichev); entretanto, se* t *for fixo,* x *varia em função de* n *e* $x - np = t\sqrt{npq}$*, ou seja, a diferença entre* x *e* np *é sempre igual a* t *desvios-padrão da distribuição Binomial* (n, p) *- precisamente nessas condições, foi provado que, se* $n \geq n_1$ *e* $|t| \leq M$*,*

$$P_n(x) \asymp \frac{1}{\sqrt{2\pi npq}} \exp(-t^2/2) = \frac{1}{\sqrt{2\pi npq}} \exp - \left(\frac{(x - np)^2}{2npq} \right).$$

6.4.2 A Fórmula de Laplace

A partir do Teorema de de Moivre, a fim de calcular o valor numérico de $\sum_{x\in A} P_n(x)$, para A qualquer, bastaria apenas dispor de uma tabela da densidade normal e substituir cada somando $P_n(x)$ por sua expressão aproximada.

Entretanto, Laplace encontrou em 1801 o seguinte procedimento bem mais simples para calcular $\sum_{x\in A} P_n(x)$, para qualquer subconjunto A dos inteiros não-negativos.

Com efeito, em vez de considerar uma representação gráfica em que a probabilidade de x fosse representada por um segmento paralelo ao eixo das abcissas localizado na ordenada x de altura igual a $P_n(x)$, Laplace fez corresponder áreas às probabilidades de x, associando ao inteiro x um retângulo de altura igual a $P_n(x)$ e base unitária - o segmento $(x - 1/2, x + 1/2)$.

Quando se faz a mudança de variáveis $t = (x - np)/(npq)^{1/2}$, as abcissas são dividas por $(npq)^{1/2}$ e, portanto, para que as áreas continuem a representar probabilidades, as ordenadas devem ser multiplicadas pelo mesmo número. Agora, levando mais uma vez em consideração a fórmula

$$P_n(x) \asymp \frac{1}{\sqrt{2\pi npq}} \exp(-t^2/2) = \frac{1}{\sqrt{2\pi npq}} \exp - \left(\frac{(x-np)^2}{2npq} \right),$$

e considerando-se a e b inteiros, de forma que $0 \leq a \leq b \leq n$, temos

$$P_n(a \leq S_n \leq b) \asymp \frac{1}{\sqrt{2\pi}} \int_{(a-np)/\sqrt{npq}}^{(b-np)/\sqrt{npq}} e^{-t^2/2}\, dt. \tag{6.8}$$

Observe que $f(t) = (2\pi)^{-1/2} e^{-t^2/2}$ é a função de densidade da distribuição $N(0,1)$.

6.4.3 A Correção de Bernstein

A fórmula de Laplace (6.8) tem um novo problema: no seu membro esquerdo, aparece a variável aleatória discreta S_n e, no direito, uma expressão equivalente à probabilidade do intervalo fechado $[(a-np)/\sqrt{npq}, (b-np)/\sqrt{npq}]$, sob a distribuição (contínua) $N(0,1)$.

Com efeito, se Z é qualquer variável aleatória contínua, $\mathbb{P}(a \leq Z \leq b) = \mathbb{P}(a < Z < b)$ mas, por outro lado, a e b estão na imagem de $S_n = \{0, 1, 2, \ldots, n\}$ e, portanto, a igualdade acima não vale para S_n.

Serge Bernstein propôs a substituição de $[(a-np)/\sqrt{npq}, (b-np)/\sqrt{npq}]$ pelo intervalo $[(a-np)/\sqrt{npq} - 1/2, (b-np)/\sqrt{npq} + 1/2]$, ou seja, a nova fórmula seria:

$$P_n(a \leq S_n \leq b) \cong \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{npq}} - \frac{1}{2}}^{\frac{b-np}{\sqrt{npq}} + \frac{1}{2}} e^{-t^2/2}\, dt.$$

Esta fórmula dá aproximações boas inclusive para n pequeno. O trabalho de Bernstein recebeu revisões de W. Feller em [11]. Outras revisões podem ser feitas mas os efeitos se tornam cada vez menos expressivos, principalmente pelo já bom desempenho da correção de Bernstein até para n pequeno.

6.4.4 Velocidade de Convergência

No seu curso para a École Polytechnique francesa, Jacques Neveu, faz uma observação nos seguintes termos.

Observação 6.4.2

Voltando ao modelo de uma seqüência de ensaios de Bernoulli independentes (veja definição 4.2.1), sendo p a probabilidade de sucesso em cada ensaio e S_n o número observado de sucessos nos n primeiros ensaios, vemos que

$$\mathbb{E}\left((S_n - np)^2\right) = Var(S_n) = np(1-p).$$

Isto indica que $(S_n - np)^2$ é em média *de ordem n e, portanto, $S_n - np$ deveria ser de ordem $\sqrt{n}$. Essa intuição é confirmada pelos teorema de de Moivre e de Laplace pois esses mostram que a distribuição do quociente $(S_n - np)/\sqrt{n}$ possui um limite quando $n \to +\infty$. Dito de uma outra forma, como*

$$\frac{S_n - np}{\sqrt{n}} = \sqrt{n}\left(\frac{S_n}{n} - p\right),$$

a freqüência S_n/n converge *à probabilidade p, com* velocidade *$1/\sqrt{n}$ e a diferença $(S_n/n) - p$, após divisão por essa velocidade, possui uma distribuição-limite.*

6.5 Exercícios

Exercício 6.1 (Teorema de Weierstrass)

(a) Complete a demonstração do Teorema de Weierstrass 6.2.2; e
(b) Verifique o corolário ao Teorema de Weierstrass.

Dicas: *Seja $f \in C([a,b])$. Considere $\varphi : [0,1] \to [a,b]$ e $g \in C([0,1])$ definidas abaixo*

$$\varphi(x) \quad = \quad a + x(b-a)$$

$$g = f \circ \varphi.$$

Argumente que P_n*, o polinômio de grau n definido no intervalo* $[a, b]$*, pode ser escrito como*

$$P_n = B_n \circ \varphi^{-1},$$

onde B_n *é o n-ésimo polinômio de Bernstein correspondente a* g*.*

Finalmente, observe que

$$|P_n(t) - f(t)| = |(B_n \circ \varphi^{-1})(t) - g(\varphi^{-1}(t)| \ \ \forall t \in [a, b]$$

e que $\varphi^{-1} : [a, b] \to [0, 1]$ *é bijetora e, em conseqüência,*

$$\sup_{t \in [a,b]} |P_n(t) - f(t)| = \sup_{x \in [0,1]} |B_n(x) - g(x)|.$$

Exercício 6.2 *Analise o Teorema Central do Limite, em especial sob os seguintes aspectos:*

(a) Formação da Teoria Estatística em suas bases probabilísticas como um ramo independente da Matemática, em vez de apenas um conjunto de suas aplicações;

(b) Facilidade no cálculo de probabilidades para variáveis aleatórias em geral (pense se você conhece alguma v.a. que não possa ser aproximada pela normal); e

(c) diferenças operacionais no TCL para v.a.'s discretas e contínuas.

Exercício 6.3 *Seja* X *uma v.a.* $Bin(n, p)$*. Enuncie as respectivas aproximações pelas distribuições Poisson e Normal. Compare em detalhes as condições dos dois resultados e argumente por que ambas funcionam e isto não caracteriza uma contradição.*

Exercício 6.4 *Utilizando-se da correção para a aproximação de uma Binomial por uma Normal, ache a aproximação da soma de n v.a.'s i.i.d's $Po(\lambda)$.*

Exercício 6.5 *Suponha que X tenha distribuição Binomial com parâmetros dados abaixo. Calcule as aproximações por TCL com e sem correção e pela fórmula exata de $\mathbb{P}(X = k)$, $k = 0, 1, 2, \ldots, n$. Faça comparações (por gráficos) de quão boas são as aproximações e quão necessária é a correção de acordo com os valores de n e p. Utilize-se de $n = 5, 8, 10, 15, 20, 25, 50, 100, 200$ e $p = 0,05s$, $s = 1, 2, \ldots, 10$.*

Nota: *Para o cálculo das probabilidades exatas e aproximações, caso desejado (o ideal seria programar com a fórmula da Binomial), utilize-se de um pacote. Não deixe, no entanto, de entender o que está sendo feito pelo utilitário. Quem quiser programar, lembre-se da fórmula de Stirling:*

$$n! \asymp \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Exercício 6.6 *Faça um estudo detalhado da convergência das distribuições discretas e contínuas apresentadas no texto, segundo os teoremas-limite apresentados neste capítulo. Primeiramente, verifique qual o resultado teórico se adequa à distribuição em questão (se existe algum). Caso seja verdadeira a convergência, estude graficamente o comportamento dessas convergências. Concentre seus estudos em dois pontos principais: assimetria e variância da distribuição original e sua influência no n suficiente para uma boa aproximação. Varie os valores dos parâmetros de cada distribuição e estude as conseqüências.*

Apêndice A

Teoria Ingênua de Conjuntos

A definição exata (axiomática) de conjunto envolve grandes dificuldade conceituais. No entanto, a idéia de um conjunto visto como uma coleção ou agrupamento de objetos (chamados elementos) é bastante clara e prática, podendo ser compreendida e aplicada num nível intuitivo. Portanto, para nós, um conjunto será entendido como **a coleção de seus elementos**.

Ao leitor interessado, sugere-se a benéfica leitura dos textos de Halmos, entre eles destacando-se [16], cujo título é respeitosamente traduzido e plagiado neste apêndice. Outra referência importante é [8].

A.1 Notação

Um conjunto será sempre denotado usando-se *chaves*, isto é, { e }. Podemos descrevê-lo enumerando-se seus elementos: por exemplo, se A é um conjunto finito cujos elementos são $e_1, e_2, \ldots, e_n$, escreve-se

$$A = \{e_1, e_2, \ldots, e_n\}.$$

Se o conjunto é infinito, poderemos, em muitas situações, escrever

$$\{e_1, e_2, e_3, \ldots\}$$

como o conjunto cujos elementos são exatamente $e_1, e_2, e_3, \ldots$. Por exemplo,

$$\mathbb{N} = \{0, 1, 2, \ldots\}$$

é o conjunto de inteiros não-negativos,

$$N^* = \{1, 2, \ldots\}$$

é o conjunto de inteiros positivos e

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

é o conjunto de inteiros.

No entanto, há situações em que uma descrição de um conjunto pela enumeração de todos os seus elementos pode ser impossível ou indesejável. Uma forma alternativa existe quando um conjunto A pode ser definido por uma propriedade $\mathcal{P}$, isto é, quando A é o conjunto dos pontos e tais que $\mathcal{P}(e)$ é verdadeira. Nesse caso, escreve-se

$$A = \{e \ : \ \mathcal{P}(e)\}.$$

Por exemplo,

$$\mathbb{Q} = \{p/q \ : \ p \in \mathbb{Z}, q \in \mathbb{N}^*\}$$

é o conjunto de números racionais e

$$\mathbb{R} = \{x \ : \ x \text{ é um número real}\}$$

é o conjunto de números reais.

Notemos que tanto $\mathbb{Q}$ quanto $\mathbb{R}$ têm uma quantidade de elementos que torna inviável qualquer enumeração explícita completa enquanto a existência de uma propriedade como a dos racionais nos indica uma maneira elegante de descrevê-los.

Para representar conjuntos, usaremos letras maiúsculas como, por exemplo, A, B, C ou Ω, Θ, Λ. Os elementos serão sempre representados por letras minúsculas como, por exemplo, a, b, c , ω, θ e λ.

Para denotar que um elemento e **pertence a** um conjunto A, escrevemos

$$e \in A.$$

Chama-se $\emptyset$ o **conjunto vazio**, isto é o conjunto sem elementos.

A.2 Subconjuntos de um Conjunto

A.2.1 Subconjuntos

É natural a pergunta de, quando da existência de dois ou mais conjuntos, como se pode compará-los. A maneira de que nos utilizamos é baseada na noção de subconjuntos.

> **Definição A.2.1** *B é um* **subconjuto** *de A se qualquer elemento $e \in B$ também pertence a A ($e \in A$). Escreveremos então $B \subset A$.*

Cuidado: Não confunda $\in$ e $\subset$: o primeiro relaciona um elemento com um conjunto e o segundo dois conjuntos! Isto quer dizer que $\subset$ relaciona entidades de mesma natureza (conjuntos) enquanto $\in$ relaciona entidades de natureza hierarquicamente diferentes, no caso elementos e conjuntos.

Note que a definição de subconjunto nada diz sobre a volta, isto é, nenhuma restrição sobre os elementos de A é tomada. Quando o fazemos, temos a seguinte propriedade.

> **Propriedade A.2.1** *Sejam A e B dois conjuntos. Se $A \subset B$* **e** *$B \subset A$, então A é* **igual a** *B ($A = B$) e reciprocamente.*

Portanto, lembrando-nos de que existem conjuntos cuja totalidade de elementos não nos é possível enumerar, vemos que a propriedade A.2.1 é fundamental para verificar se dois conjuntos são iguais.

Nota: Caso as condições da propriedade A.2.1 não sejam satisfeitas, diremos que A não é igual a B, com notação $A \neq B$.

A.2.2 Subconjunto Complementar

Suponha, na seqüência, ser A um conjunto não-vazio.

> **Definição A.2.2** *Seja B um subconjunto de A. Definimos o* **complementar** *de B com relação a A, ou a* **diferença** *entre*

A e B, com notação $A \setminus B$ ou $A - B$ ou ainda B^{c_A}, o subconjunto de A definido por

$$A - B = \{x : x \in A \text{ e } x \notin B\}.$$

Uma rápida observação que se faz quanto à diferença entre A e B é a de que, se $A \subset B$, então $A - B$ é vazio. Intuitivamente, $A - B$ se compõe de todos os elementos exclusivos de A.

A.3 Operações sobre Conjuntos

Definição A.3.1 *Sejam A e B dois conjuntos. Definem-se:*

União: $A \cup B = \{x \ : \ x \in A \text{ ou } x \in B\}$; *e*

Intersecção: $A \cap B = \{x \ : \ x \in A \text{ e } x \in B\}$.

A união de dois conjuntos nada mais é do que a coleção de todos os elementos que pertençam a pelo menos um deles, enquanto a intersecção se compõe de elementos comuns aos dois conjuntos.

Se $A \cap B = \emptyset$, os conjuntos A e B são chamados de **disjuntos**.

A idéia de elementos comuns a dois conjuntos ($A \cap B$), elementos exclusivos de A $(A - B)$ ou de B $(B - A)$ nos leva naturalmente a questionar sobre os elementos exclusivos de um deles, cujo conjunto é dado por

Definição A.3.2 *Sejam A e B dois conjuntos. Define-se:*

Diferença simétrica: $A \triangle B = \{x \ : \ (x \in A \text{ e } x \notin B) \text{ ou } (x \notin A \text{ e } x \in B)\}$.

Algumas propriedades intuitivas (mas que precisam ser devidamente demonstradas) referentes às relações entre as operações de conjuntos elementares são dadas abaixo.

Propriedade A.3.1 *Sejam A e B dois conjuntos. Então,*

(i)	$A \cup A = A \cap A = A$;
(ii)	$A \cup B = B \cup A$; $A \cap B = B \cap A$; $A \triangle B = B \triangle A$;
(iii)	$A \cap B \subset A \subset A \cup B$;
(iv)	$A \subset B$ *se e somente se* $A \cup B = B$;
(v)	$A \subset B$ *se e somente se* $A \cap B = A$;
(vi)	$A \triangle B = (A \cup B) - (A \cap B)$;
(vii)	$A \triangle B = \emptyset$ *se e somente se* $A = B$.

Propriedade A.3.2 (Distributividade)

Sejam A, B e C conjuntos. Então

(i)	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
(ii)	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

A.4 Coleção de Conjuntos

A.4.1 Conjuntos-produto

Definição A.4.1 *Sejam X e Y dois conjuntos. Define-se o conjunto* **produto (Cartesiano)** *de X com Y, $X \times Y$, como*

$$X \times Y = \{(x, y) \ : \ x \in X, \ y \in Y\}.$$

A.4.2 Conjunto das Aplicações

Definição A.4.2 *Sejam E e F dois conjuntos. Define-se F^E, o* **conjunto das aplicações** *de E sobre F, como*

$$F^E = \{f \ / \ f : E \to F\}.$$

A.4.3 Conjunto das Partes

Definição A.4.3 *Seja A um conjunto. Define-se, $\mathcal{P}(A)$, o* **conjunto das partes** *ou* **conjunto dos subconjuntos** *de A como*

$$\mathcal{P}(A) = \{B \ : \ B \subset A\}.$$

Nota: É importantíssimo que se entenda que os elementos de $\mathcal{P}(A)$ são conjuntos. Esse fato em nada fere nossa visão ingênua dos conjuntos pois nenhuma restrição conceitual fora feita sobre a natureza dos elementos quando da introdução de conjuntos. Mas devemos estar alertas para níveis heirárquicos e utilização das operações e simbologias adequadas em cada situação. Por exemplo, $A \in \mathcal{P}(A)$, pois A está um nível hierárquico abaixo de $\mathcal{P}(A)$, isto é, A é um elemento de $\mathcal{P}(A)$ e, portanto é **absolutamente errado dizer** $A \subset \mathcal{P}(A)$.

Propriedade A.4.1 *Se $B \in \mathcal{P}(A)$ e $C \in \mathcal{P}(A)$, então*

$$B \cup C \in \mathcal{P}(A); \ B \cap C \in \mathcal{P}(A); \ B \triangle C \in \mathcal{P}(A).$$

A.4.4 Limites de Conjuntos

Uma noção muito importante é do comportamento de uma seqüência infinita de conjuntos.

Definição A.4.4 *Seja $\{A_i \ : \ i \in I\}$ uma coleção (talves infinita) de conjuntos. Definimos $\cup_{i \in I} A_i$, sua* **união**, *e $\cap_{i \in I} A_i$, sua* **intersecção**, *por, respectivamente*

$$\begin{aligned} \cup_{i \in I} A_i &= \{x \ : \ \exists j \in I \ \textit{tal que} \ x \in A_j\}; \\ \cap_{i \in I} A_i &= \{x \ : \ \forall j \in I \ x \in A_j\}. \end{aligned}$$

Na definição anterior, utilizamo-nos da notação $j \in I$, para indicar que um determinado conjunto pertence à coleção em questão. O conjunto I nada mais é do que uma generalização do nossos antigos conjuntos $\{1, 2, \ldots, n\}$ e $\mathbb{N}$ e será chamado de **conjunto de índices**.

Propriedade A.4.2 *Sejam I e J dois conjuntos de índices e seja $\{A_{i,j} \ : \ i \in I, \ j \in J\}$ uma coleção de conjuntos.*

$$\cup_{j\in J} \cap_{i\in I} A_{i,j} = \cap_{i\in I} \cup_{j\in J} A_{i,j}$$

Seja $A_1, A_2, A_3, \ldots$ uma seqüência de conjuntos. Para qualquer $n \geq 1$, definimos

$$C_n = \cup_{i=n}^{\infty} A_i$$

e

$$D_n = \cap_{i=n}^{\infty} A_i.$$

$\{C_n, n \geq 1\}$ e $\{D_n, n \geq 1\}$ são duas seqüências com propriedades muito particulares.

Primeiramente, temos $D_n \subset C_n$, para qualquer n. Observemos também que, para qualquer $n \geq 1$, $C_{n+1} \subset C_n$ e $D_n \subset D_{n+1}$, isto é, $\{C_n\}_{n\geq 1}$ é uma seqüência **decrescente** de conjuntos e $\{D_n\}_{n\geq 1}$ é uma seqüência **crescente** de conjuntos.

Portanto, $\cap_{n=1}^{m} C_n = C_m$, $\cup_{n=1}^{m} D_n = D_m$ e definimos os seguintes conjuntos

$$\lim_{n\uparrow\infty} C_n = \cap_{n=1}^{\infty} C_n$$

e

$$\lim_{n\uparrow\infty} D_n = \cup_{n=1}^{\infty} D_n.$$

Note que ambos os conjuntos $\lim_{n\uparrow\infty} C_n$ e $\lim_{n\uparrow\infty} D_n$ são, nesta situação, bem, definidos, pois as respectivas expressões à direita das igualdades são bem definidas.

Dessas observações, tiramos a definição a seguir.

Definição A.4.5 *Seja $\{A_k, k \geq 1\}$ uma coleção (infinita) de conjuntos. Define-se o* **limite inferior** *de $\{A_k\}$ por*

$$\liminf A_k = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k.$$

Define-se o **limite superior** *de* $\{A_k\}$ *por*

$$\limsup A_k = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k.$$

Propriedade A.4.3

(i) $\liminf A_k \subset \limsup A_k$*;*

(ii) $\liminf A_k = \{x \,:\, x \in A_i$ *a menos de um número finito de índices* $i\}$*;*

(iii) $\limsup A_k = \{x \,:\, x \in A_i$ *para uma infinidade de índices* $i\}$*;*
e

(iv) $(\limsup A_k)^c = \liminf A_k^c$.

Notemos que o limite inferior trabalha com conjuntos sempre contidos nos conjuntos utilizados pelo limite superior, intuitivamente levando-nos à primeira propriedade acima.

Definição A.4.6 *Seja* $\{A_k, k \geq 1\}$ *uma coleção (infinita) de conjuntos, de forma que* $\liminf A_k = \limsup A_k$*. Então, define-se o* **limite** *de* A_k *como*

$$\lim A_k = \liminf A_k = \limsup A_k.$$

Cuidado: O limite de uma seqüência de conjuntos nem sempre existe. Nesse caso, temos $\liminf A_k \neq \limsup A_k$.

A.5 Função Indicadora de um Conjunto

Seja A um conjunto qualquer. Define-se sua *função indicadora*, $\mathbf{1}_A$, por

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{c.c. } . \end{cases}$$

Apesar de sua aparente simplicidade, a função indicadora, assim definida, tem utilidade ampla e, entre suas interessantes propriedades, temos

Propriedade A.5.1

(i) $\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B$;

(ii) $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$;

(iii) $\mathbf{1}_{A\triangle B} = |\mathbf{1}_A - \mathbf{1}_B| = \mathbf{1}_A + \mathbf{1}_B - 2 \times \mathbf{1}_A \mathbf{1}_B$; *e*

(iv) $\mathbf{1}_{A-B} = \mathbf{1}_A - \mathbf{1}_B$.

A.5.1 Fórmula de Inclusão-Exclusão

Seja $\{A_i, i = 1, 2, \ldots, n\}$ uma coleção de conjuntos. Uma fórmula muito importante é a de **inclusão-exclusão**, descrita abaixo.

$$
\begin{aligned}
\mathbf{1}_{\cup_{i=1}^n A_i} &= \sum_{i=1}^n \mathbf{1}_{A_i} - \sum_{1\le i_1<i_2\le n} \mathbf{1}_{A_{i_1}\cap A_{i_2}} + \cdots \\
&\quad +(-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le n} \mathbf{1}_{A_{i_1}\cap\cdots\cap A_{i_k}} + \cdots \\
&\quad +(-1)^{n+1}\mathbf{1}_{A_1\cap\cdots A_n}.
\end{aligned}
$$

A.6 O conjunto $\mathbb{N}$

A.6.1 Axiomas de Peano

Tradicionalmente, o conjunto dos números naturais é introduzido pelos chamados axiomas de Peano.

Axiomas de Peano

Existe um conjunto, denominado $\mathbb{N}$, tal que, a cada $n \in \mathbb{N}$, podemos fazer corresponder um elemento denominado $n + 1$, também em $\mathbb{N}$, chamado de sucessor de n, tendo as seguintes propriedades:

(i) $n + 1 = m + 1$ implica em $n = m$;

(ii) existe um elemento em $\mathbb{N}$, denominado 0, que não é o sucessor de qualquer elemento de $\mathbb{N}$;

(iii) se $\mathcal{N} \subset \mathbb{N}$ é tal que $[0 \in \mathcal{N}]$ e $[(n \in \mathcal{N})$ implica em $(n+1 \in \mathcal{N})]$, então $\mathcal{N} = \mathbb{N}$.

A.6.2 Raciocínio por Indução

Entre as conseqüências dos Axiomas de Peano, temos uma que é fundamental para a Teoria de Probabilidade, que é o raciocínio por indução.

Teorema A.6.1 (Teorema de Indução)

Seja $P(n)$ uma propriedade (enunciado) que só depende da variável n. Suponha que

$$\left.\begin{array}{l} (i)\ P(0) \text{ é verdadeira; e} \\ (ii)\ P(n) \text{ verdadeira implica } P(n+1) \text{ verdadeira.} \end{array}\right\} \qquad (1.1)$$

Então, $P(n)$ é verdadeira para cada inteiro n.

Demonstração:

Seja $\mathcal{N} = \{n \in \mathbb{N} \ : \ P(n)$ é verdadeira$\}$. Temos $0 \in \mathcal{N}$, usando-se a hipótese (i) de (1.1). Além disso, sabemos que $n \in \mathcal{N}$ implica em $n+1 \in \mathcal{N}$, usando (1.1) (ii).

Isto quer dizer que $\mathcal{N} = \mathbb{N}$, pelo terceiro axioma de Peano. ■

Na prática, se queremos mostrar que uma propriedade $P(n)$, indexada por n, é verdadeira, para qualquer n, a partir de um certo n_0, basta-nos seguir as etapas descritas baixo.

Primeira etapa: *Inicialização*

Temos que mostrar que $P(n_0)$ é verdadeira.

Segunda etapa: *Hipótese de Indução*

Vamos supor $P(n)$ verdadeira **até** o índice $k \geq n_0$. (Isto é, $P(n_0)$ é verdadeira, $P(n_0+1)$ é verdadeira, ... , $P(k)$ é verdadeira). Temos que mostrar então que $P(k+1)$ é tambem verdadeira.

Terceira etapa: *Conclusão*

A primeira etapa mostrou que $P(n_0)$ era verdadeira. A segunda mostra então que também o é $P(n_0+1)$. Empregando a segunda etapa em seqüência, temos demonstrado o fato de que $P(n_0+2)$ é verdadeira ... . Conclui-se que $P(n)$ é verdadeira para qualquer $n \geq n_0$.

Exemplo A.6.1 (Identidade)

Queremos mostrar a fórmula bem conhecida

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}), \quad (1.2)$$

para qualquer $n \geq 2$.

Chamemos $P(n)$ *a fórmula 1.2.*

1) $P(2)$ *é verdadeira pois*

$$a^2 - b^2 = (a-b)(a+b).$$

2) Suponha-se que $P(k)$ *seja verdadeira, até* $k \geq 2$.

$$\begin{aligned} a^{k+1} - b^{k+1} &= a(a^k - b^k + b^k) - bb^k \\ &= a(a^k - b^k) + b^k(a-b) \end{aligned}$$

e, usando-se a hipótese de indução de ordem k, *temos*

$$\begin{aligned} a^{k+1} - b^{k+1} &= a(a-b)\sum_{j=0}^{k-1} a^{k-1-j}b^j + b^k(a-b) \\ &= (a-b)\left(\sum_{j=0}^{k-1} a^{k-j}b^j + b^k\right) \\ &= (a-b)\left(\sum_{j=0}^{k} a^{k-j}b^j\right), \end{aligned}$$

que é exatamente $P(k+1)$.

3) No item 1, provamos que $P(2)$ *é verdadeira e, pelo item 2, temos então* $P(3)$ *etc.* ■

Observação: Acabamos de demonstrar assim que, se $a \neq b$, *então*

$$\sum_{j=0}^{n} a^{n-j}b^j = \frac{a^{n+1} - b^{n+1}}{a-b}.$$

A.7 Cardinalidade

Definição A.7.1

(i) Dois conjuntos A *e* B *têm mesma* potência *se existe uma aplicação bijetora de* A *sobre* B*;*

(ii) Um conjunto A *é chamado de* finito *se existe um (único) número natural* n $(n \in \mathbb{N})$ *tal que* A *tem mesma potência que* $\{1, 2, \ldots, n\}$*. Caso contrário,* A *é dito* infinito*;*

(iii) Se A *é um conjunto finito, chama se* cardinalidade *de* A *o número natural* n *que define sua potência e denota-se* $|A| = n$*; e*

(iv) Se A é um conjunto infinito com mesma potência que $\mathbb{N}$, dizemos que A é enúmeravel e $|A| = |\mathbb{N}| = \aleph_0$. Se A é infinito mas não tem potência igual a de $\mathbb{N}$, A é dito não-enumerável.

Teorema A.7.1

(i) Seja A um conjunto finito. Então, qualquer que seja $B \in \mathcal{P}(A)$, B também é um conjunto finito e, se $B \neq A$, temos:

$$|B| < |A|.$$

(ii) Sejam A e $B \neq \emptyset$ dois conjuntos finitos. Então, $A \cup B$, $A \times B$, B^A e $\mathcal{P}(A)$ são também finitos e temos

$$\begin{aligned}
&|A \cup B| = |A| + |B| - |A \cap B|; \\
&|A \times B| = |A|.|B|; \\
&|B^A| = |B|^{|A|}; \text{ e} \\
&\mathcal{P}(A)| = 2^{|A|}.
\end{aligned}$$

Teorema A.7.2 *Todo conjunto infinito tem a mesma potência que um de seu subconjuntos próprios. Reciprocamente, se um conjunto tem potência igual a de um de seus subconjuntos próprios, então, esse conjunto é infinito.*

Teorema A.7.3 *Seja A um conjunto infinito. A reunião de uma coleção enumerável de subconjuntos enumeráveis de A é um subconjunto de A também enumerável.*

Apêndice B

Ferramentas de Análise

B.1 Propriedades dos Números Complexos

B.1.1 Escritura

Qualquer que seja $z \in \mathbb{C}$, existem $x, y \in \mathbb{R}$ tais que

$$z = x + iy, \tag{2.1}$$

onde i é o número (complexo) tal que $i^2 = -1$.

Define-se o **conjugado** de $z(= x + iy)$ por

$$\bar{z} = x - iy. \tag{2.2}$$

Chama-se a **parte real** de z

$$\mathcal{R}e(z) = \frac{z + \bar{z}}{2}$$

e sua **parte imaginária** é definida por

$$\mathcal{I}m(z) = \frac{z - \bar{z}}{2i}.$$

Chama-se o **módulo** de $z \in \mathbb{C}$ o número em $\mathbb{R}^+$ definido por

$$|z| = (x^2 + y^2)^{1/2}. \tag{2.3}$$

Chama-se o **argumento** de $z \in \mathbb{C}$ o número em $[0, 2\pi[$ tal que

$$Arg(z) = \arctan \frac{y}{x}. \tag{2.4}$$

Reciprocamente, dados $(\rho, \theta) \in \mathbb{R}^+ \times [0, 2\pi[$, existe um único número complexo z, tal que

$$z = \rho e^{i\theta}. \tag{2.5}$$

Tem-se famosa **Fórmula de Euler**

$$e^{i\theta} = \cos(\theta) + isen(\theta). \tag{2.6}$$

Portanto se $z = \rho e^{i\theta}$

$$\begin{aligned} &z = \rho(\cos(\theta) + isen(\theta)) \\ &\mathcal{R}e(z) = \rho\cos(\theta) \\ &\mathcal{I}m(z) = \rho sen(\theta) \end{aligned}$$

B.1.2 Funções Trigonométricas e Hiperbólicas

Qualquer que seja $z \in \mathbb{C}$,

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2} = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}; \tag{2.7}$$

$$sen(z) = \frac{e^{iz} - e^{-iz}}{2i} = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}; \tag{2.8}$$

$$\cosh(z) = \frac{e^{z} + e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}; \text{ e} \tag{2.9}$$

$$senh(z) = \frac{e^{z} - e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}. \tag{2.10}$$

Têm-se, também,

$$
\begin{aligned}
sen(z) &= \frac{1}{i} senh(iz) \\
\cos(z) &= \cosh(iz).
\end{aligned}
\tag{2.11}
$$

B.1.3 As Fórmulas de de Moivre

$$(\cos(z) + isen(z))^n = e^{inz} = \cos(nz) + isen(nz) \tag{2.12}$$

$$(\cosh(z) + isenh(z))^n = e^{nz} = \cosh(nz) + isenh(nz) \tag{2.13}$$

B.2 Séries

Teorema B.2.1 (Séries de Riemann)

A série

$$\sum_n \frac{1}{n^\alpha}$$

é convergente, se $\alpha > 1$, e divergente, se $\alpha \leq 1$.

Teorema B.2.2 (Séries de Bertrand)

A série

$$\sum_{n \geq 2} \frac{1}{n(\log n)^\alpha}$$

é convergente, se $\alpha > 1$, e divergente, se $\alpha \leq 1$.

Teorema B.2.3 (Troca de Ordem dos Termos de uma Série)

Seja $\{a_n\}$ uma série de números complexos, tais que $\sum_n |a_n|$ seja convergente. Então, qualquer rearranjo dos termos de $\sum_n a_n$ converge e tem limite igual ao da série original.

Teorema B.2.4 (Troca de Ordem de Soma)

Seja $\{a_{i,j}\}$ uma seqüência duplamente indexada, por $i = 1, 2, 3\ldots$ e $j = 1, 2, 3, \ldots$. Suponha-se que, para qualquer i

$$b_i = \sum_{j=1}^{\infty} |a_{i,j}|$$

e $\sum_i b_i$ convirja. Então,

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{i,j}.$$

Teorema B.2.5 (Teorema de Integração)

Sejam f_n integráveis em I, intervalo compacto de $\mathbb{R}$, e

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

tais que

$$\sum_{n=1}^{m} f_n(x) \overset{unif}{\Rightarrow} f(x)$$

em I.

Então,

$$\int_I f(x)\ dx = \sum_{n=1}^{\infty} \int_I f_n(x)\ dx.$$

Teorema B.2.6 (Teorema de Derivação)

Sejam $\{f_n\}$ uma seqüência de funções deriváveis num intervalo de $\mathbb{R}$, I, e

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

convergente em I. Suponha-se também que $\sum_n f_n'$ seja uniformemente convergente em I. Então, a função f é derivável em I, de forma que

$$f'(x) = \sum_{n=1}^{\infty} f_n'(x).$$

B.2.1 Fórmula de Stirling

Esta fórmula dá uma aproximação de $n!$

$$n! = n^n e^{-n}\sqrt{2\pi n}(1+\varepsilon_n), \tag{2.14}$$

onde $\varepsilon_n/n! \to 0$ quando $n \to \infty$.

Teorema B.2.7 (Fórmula de Stirling para a Função Γ)

$$\lim_{|x|\to\infty} \frac{\Gamma(x+1)}{(x/e)^x\sqrt{2\pi x}} = 1$$

B.3 Séries de Potência

Teorema B.3.1 (Teorema de Abel)

Seja $\sum_{n=0}^{\infty} c_n$ uma série convergente. Para $-1 < x < 1$, seja

$$f(x) = \sum_{n=0}^{\infty} c_n x^n.$$

Então,

$$\lim_{x\to 1} f(x) = \sum_{n=0}^{\infty} c_n.$$