A demonstração dos outros resultados pode ser decomposta em dois casos, de acordo com a variável ser discreta ou (absolutamente) contínua.

Caso Discreto:

(a) Sendo a variável aleatória positiva, sua esperança nada mais é do que uma soma de quantidades positivas, sendo portanto positiva;

(c) Sem perda de generalidade, vamos supor que $X \in \{x_1, x_2, \ldots\}$, com função de massa p_X . A v.a. aX tem imagem $\{ax_1, ax_2, \ldots\}$ e sua função de massa, p_{aX}, atende a

$$p_{aX}(ax_i) = p_X(x_i),$$

para qualquer $i \geq 1$. Portanto,

$$\mathbb{E}(aX) = \sum_{i=1}^{\infty} ax_i p_{aX}(ax_i) = a \sum_{i=1}^{\infty} x_i p_X(x_i) = a\mathbb{E}(X).$$

(d) Seja $Y = X + b$. Sendo, sem perda de generalidade, Imagem$(X) = \{x_1, x_2, \ldots\}$, temos

$$\text{Imagem}(Y) = \{x_1 + b, x_2 + b, \ldots\}.$$

Além disso, $p_Y(k) = p_X(k - b)$. Portanto,

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{k \in ImY} kp_Y(k) = \sum_{m \in ImX} (m + b)p_X(m) \\
&= \left(\sum_{m \in ImX} mp_X(m) \right) + b \\
&= \mathbb{E}(X) + b.
\end{aligned}$$

Caso (absolutamente) Contínuo:

(a) Sendo a variável aleatória positiva, sua esperança nada mais do que uma integral do produto de duas funções positivas, sendo portanto positiva;

(c) Se $a = 0$, o resultado é trivial. Se $a > 0$, seja f_X a densidade de X. A v.a. aX tem por função de distribuição acumulada

$$\mathbb{P}(aX \leq x) = \mathbb{P}(X \leq x/a) = \int_{-\infty}^{x/a} f_X(t)\ dt.$$

Usando-se a mudança de variável: $u = at$, temos

$$\mathbb{P}(aX \leq x) = \int_{-\infty}^{x} \frac{1}{a} f_X(u/a)\ du.$$

Conseqüentemente, aX é uma variável aleatória contínua, com densidade $\frac{1}{a} f_X(u/a)$. Segue que

$$\mathbb{E}(aX) = \int_{-\infty}^{+\infty} \frac{u}{a} f_X(u/a)\ du.$$

Usando-se $t = u/a$, tem-se

$$\mathbb{E}(aX) = \int_{-\infty}^{+\infty} at f_X(t)\ dt = a\mathbb{E}(X).$$

O caso $a < 0$ é semelhante e será tratado no Exercício 3.40.

(d) Seja $Y = X + b$. A distribuição acumulada de Y é dada por

$$\begin{aligned} F_Y(x) &= F_{X+b}(x) = \mathbb{P}(X + b \leq x) = \mathbb{P}(X \leq x - b) \\ &= F_{X+b}(x) = \int_{-\infty}^{x-b} f_X(t)\ dt \\ &= \int_{-\infty}^{x} f_X(u - b)\ du. \end{aligned}$$

Portanto, $Y = X + b$ tem densidade $f_X(. - b)$ e

$$\begin{aligned}
\mathbb{E}(X+b) &= \int_{-\infty}^{+\infty} t f_X(t-b)\, dt = \int_{-\infty}^{+\infty} (u+b) f_X(u)\, du \\
&= \int_{-\infty}^{+\infty} u f_X(u)\, du + b \\
&= \mathbb{E}(X) + b.
\end{aligned}$$

■

3.5.3 Relação entre Esperança e Probabilidade

Considere-se um evento $A \in \mathcal{F}$. Defina-se a v.a. indicadora do evento A, denotada por $\mathbf{1}_A$, como

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

O leitor pode verificar que $\mathbf{1}_A$ é uma v.a. em $(\Omega, \mathcal{F}, \mathbb{P})$ (veja o exercício 3.41). Além disso, é uma v.a. discreta, com dois resultados possíveis: 0 ou 1. Sua função de massa p é descrita por:

$$\begin{aligned}
p(1) &= \mathbb{P}(A) \\
p(0) &= \mathbb{P}(A^c) = 1 - \mathbb{P}(A)
\end{aligned}$$

Um resultado muito simples mas extremamente interessante que relaciona de forma direta os conceitos de probabilidade e esperança é dado na proposição a seguir.

Proposição 3.5.2 (Esperança da Função Indicadora)
Para qualquer evento $A \in \mathcal{F}$, tem-se

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A).$$

Demonstração: Por sua própria definição, $\mathbb{E}(\mathbf{1}_A)$ pode ser escrita como

$$\mathbb{E}(\mathbf{1}_A) = 0 \times (1 - \mathbb{P}(A)) + 1 \times \mathbb{P}(A).$$

■

3.5.4 Desigualdade de Markov

Uma aplicação imediata da proposição 3.5.2 gera um dos resultados fundamentais da Teoria de Probabilidade: a **Desigualdade de Markov**.

Teorema 3.5.1 (Desigualdade de Markov)

Seja X uma variável aleatória positiva e $a > 0$ real. Então,

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}$$

Demonstração:

Caso $\mathbb{E}(X)$ não seja finita, o resultado é trivial, com desigualdade estrita. Considere o evento $A = \{X > a\}$. Note que $\Omega = A \cup A^c$,

$$[X = x] = [X = x, \omega \in A] \bigcup [X = x, \omega \in A^c]$$

e, para cada $\omega \in \Omega$,

$$\mathbf{1}_A(\omega) + \mathbf{1}_{A^c}(\omega) = 1,$$

pois $\omega \in A$ ou $\omega \in A^c$ exclusivamente. Portanto,

$$\begin{aligned} X &= X\mathbf{1}_A + X\mathbf{1}_{A^c} \\ &= X\mathbf{1}_{[X>a]} + X\mathbf{1}_{A^c} \\ &\geq a\mathbf{1}_A, \end{aligned} \tag{3.4}$$

onde usamos que $X \geq 0$. Segue das propriedades de linearidade e positividade da esperança e da proposição 3.5.2 que

$$\mathbb{E}(X) \geq a\mathbb{E}(\mathbf{1}_A) \tag{3.5}$$

$$= a\mathbb{P}(X > a). \tag{3.6}$$

■

Exemplo 3.5.3 *Suponha que X possa assumir apenas os valores 0 e 1, com respectivas probabilidades $1-p$ e p, onde $0 \leq p \leq 1$. Sabemos, portanto que:*

$$\mathbb{P}(X > a) = \begin{cases} p & 0 < a < 1 \\ 0 & 1 \leq a \end{cases} \tag{3.7}$$

Mas, por outro lado, sabemos que

$$\mathbb{E}(X) = 0 \times (1-p) + 1 \times p \ = \ p,$$

e, pela desigualdade de Markov 3.5.1, temos o seguinte limite para $\mathbb{P}(X > a)$, $a > 0$:

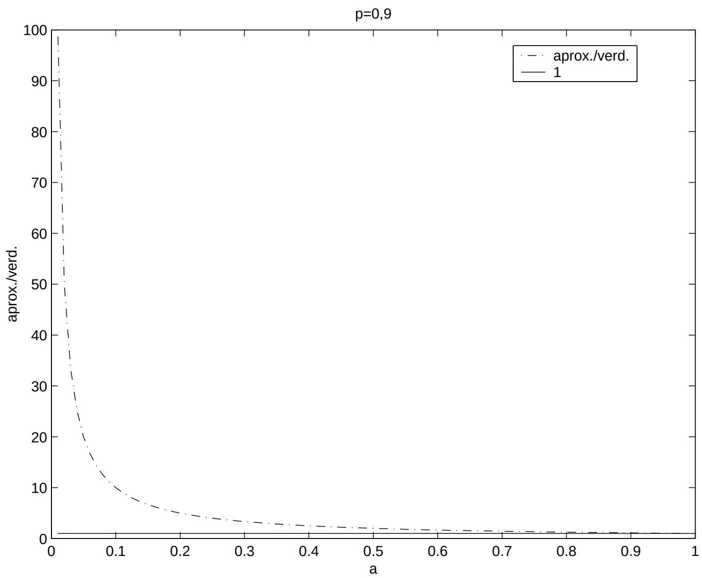
$$\mathbb{P}(X > a) < \frac{\mathbb{E}(X)}{a} \ = \ \frac{p}{a}. \tag{3.8}$$

Note que os valores dados pela desigualdade são tão maiores quanto maior for o valor da probabilidade de se obter $X = 1$ e inversamente proporcionais ao início da cauda, a. O verdadeiro valor caudal, no entanto, tem comportamento dicotômico, isto é assume apenas dois valores: p ou 0. A importância do início da cauda, a, é apenas para compará-lo com 1, cota essencial superior (veja definição 3.2.4) de X. É muito importante levar em conta que a desigualdade de Markov fornece valores universais, no sentido de que independem das distribuições das variáveis aleatórias (dependem apenas da esperança), razão por que são úteis. Porém, exatamente por isso, seu comportamento será heterogêneo, provendo valores extremamente finos por vezes e grosseiros por outras.

Na figura 3.5.3, ilustramos os comentários acima para alguns casos de p. Note que o valor fornecido por (3.8) chega a ser 100 vezes maior do que o verdadeiro valor, dado por (3.7), para $p = 0,9$, com valores aproximados de $\mathbb{P}(X > a)$ iguais a 90, ou seja, completamente inúteis. As comparações são realizadas apenas para

$0 < a < 1$, pois $\mathbb{P}(X > a) = 0$, para $a \geq 1$. Para esses valores de a, a aproximação fornecida pela desigualdade se torna extremamente ruim. O comportamento da razão valor estimado / valor verdadeiro é homogêneo para $p \in (0, 1)$.

Figura 3.4: Desigualdade de Markov para a Bernoulli



Exemplo 3.5.4 *Suponha que X tenha densidade dada por*

$$f(x) = \exp(-x)\mathbf{1}_{\mathbb{R}^+}(x).$$

Claramente $\mathbb{P}(X > 0) = 1$ (por que?). Sua esperança é dada por:

$$\mathbb{E}(X) = \int_0^{+\infty} x \exp(-x) dx = 1.$$

Observe-se que, se g fosse uma função não-decrescente, positiva e se $E(g(X))$ existisse, rescrevendo-se (3.4) como

$$g(X) = g(X)\mathbf{1}_A + g(X)\mathbf{1}_{A^c} \geq g(a)\mathbf{1}_A,$$

teríamos

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(g(X))}{g(a)}. \tag{3.9}$$

Portanto, é interessante podermos calcular valores esperados não somente de variáveis aleatórias mas de funções adequadas das mesmas. Para isso, precisamos de dois resultados muito gerais e importantes.

Suponha que X seja uma variável discreta com função de massa p: denotamos Imagem(X) sua imagem, isto é, Imagem$(X) = \{x \in \mathbb{R} | p(x) > 0\}$. Analogamente, caso X seja uma variável aleatória contínua com densidade f, Imagem$(X) = \{x \in \mathbb{R} | f(x) > 0\}$. Feito isto, podemos proceder com uma brevíssima exploração de *funções* de variáveis aleatórias e a consistência de tais procedimentos. Ao definirmos uma variável aleatória, supomos ter esssa uma determinada característica de imagem inversa na σ-álgebra. Essa necessidade de ligação traz consigo duas reflexões: quão difícil é a existência de tais ligações e podemos *definir* novas variáveis aleatórias a partir de antigas, sem precisar retornar à σ-álgebra?

A primeira questão é bastante técnica e foi respondida, de forma positiva, sem maiores justificativas no começo deste capítulo. A segunda será aqui brevemente comentada, com resposta também positiva, para um conjunto de funções tão grande que o leitor não se deve mais ocupar com tal assunto, ao nível deste curso.

Lema 3.5.1 (Transformação de Variáveis Aleatórias)

Sejam: X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$; e g uma função de Imagem(X) em $J \subset \mathbb{R}$, tal que, qualquer que seja I intervalo de $\mathbb{R}$, o conjunto $g^{-1}(I)$ também é um intervalo de $\mathbb{R}$. Então, $g(X)$ é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$.

Demonstração: Seja I um intervalo de $\mathbb{R}$. Pelas propriedades de inversas de funçoes compostas, temos

$$(g(X))^{-1}(I) = X^{-1}(g^{-1}(I)).$$

Como, por hipótese, $g^{-1}(I)$ é um intervalo de $\mathbb{R}$ e

$$X^{-1}(J) \in \mathcal{F},$$

para qualquer intervalo real J, pois X é uma v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$, concluímos que $X^{-1}(g^{-1}(I)) \in \mathcal{F}$ e portanto $g(X)$ é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$.[10] ■

Neste ponto, não nos interessa estudar em detalhes a distribuição da nova variável aleatória $g(X)$. Para nossos objetivos, basta-nos caracterizar sua esperança, como determinada na proposição a seguir.

Proposição 3.5.3 (Esperança da Transformação de V.a.'s)

Seja g uma função mensurável e X uma v.a.

(a) discreta, com funçao de massa p_X e imagem $Imagem(X)$. Então,

$$\mathbb{E}(g(X)) = \sum_{k \in Imagem(X)} g(k) p_X(k); \text{ e}$$

(b) contínua, com densidade f_X. Então,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t)\, dt.$$

No entanto, mesmo sabendo calcular $E(g(X))$, para qualquer g mensurável, o leitor deve estar questionando a utilidade prática de resultado como (3.9), pois de que maneira poderíamos decidir qual a função g adequada para uma determinada situação? Na subseção a seguir, definimos uma série de funções g simples que, no entanto, são bastante úteis para essa finalidade.

[10] O resultado anterior se estende sem dificuldades para toda função g *mensurável.* Uma função g é dita mensurável se, para qualquer intervalo real I, $g^{-1}(I)$ for um boreliano de $\mathbb{R}$ (borelianos são elementos da σ-álgebra de Borel, gerada pelos intervalos reais). Vale ressaltar que, em particular, toda função contínua é mensurável.

3.5.5 Momentos de Ordem Superior e Centrais

Uma aplicação direta da Desigualdade de Markov (em sua versão dada por (3.9)), com função g do tipo: $x \mapsto x^2$, $x \mapsto x^3, \ldots$, $x \mapsto x^k$ é a seguinte:

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X^k)}{a^k}, \tag{3.10}$$

para todo a positivo.

O conhecimento dos valores $\mathbb{E}(X^k)$ pode então revelar-se muito interessante quando queremos *controlar* a probabilidade de um evento *raro*, do tipo $\mathbb{P}(X > a)$[11] [12].

Definição 3.5.4 (Momentos de Ordem Superior)

Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$. O valor $\mathbb{E}(X^k)$ (quando existe) é chamado de **Momento de ordem** k *de X e denotado por μ_k.*

Apesar de os eventos raros em si já justificarem a importância dos momentos, pensemos no seguinte: se tivéssemos acesso a $\mathbb{E}(X)$ e $\mathbb{E}(Y)$, teríamos toda a informação necessária para comparar as v.a.'s X e Y? O exemplo a seguir dar-nos-á auxílio na resposta a essa importante pergunta[13].

---

[11] Intuitivamente, um evento raro deve ser caracterizado por uma baixa probabilidade de ocorrência. Mas, do ponto de vista prático, um evento raro também deve ser especial do ponto de vista dos valores que a v.a. estará assumindo. Portanto, para v.a's positivas, um evento do tipo $[a < X < b]$ não seria raro pois, mesmo que sua probabilidade de ocorrência fosse pequena, nada haveria de especial nos valores de X a ele associados.

[12] Note que o resultado (3.10) é válido mesmo que o termo à direita não seja finito.No entanto, quando esse não for finito, o resultado é trivial e inútil, pois já sabíamos ser a probabilidade em questão limitada por 1.

[13] Essa pergunta, aparentemente banal, tem, na realidade, importância enorme em aplicações. Quando obtemos dados sobre um determinado fenômeno, fazê-mo-lo de forma finita, isto é, adquirimos, de forma cara, um conjunto relativamente pequeno de observações. Desse diminuto conjunto, tentamos recuperar informações complexas, como a distribuição de probabilidade verdadeira. É interessante, portanto, fazê-lo de forma parcimoniosa, isto é, substituindo essa complexa entidade por um número finito de características, chamadas de parâmetros.

Exemplo 3.5.5 (É a esperança tudo?)

Tabela 3.1: Caracterização de Variáveis Aleatórias pela Esperança

x	p_x	p_y	p_w
-2	0	0	0,5
-1	0	0,25	0
0	1	0,5	0
1	0	0,25	0
2	0	0	0,5

Suponha que X, Y e W tenham respectivas funções de massa dadas na tabela 3.1, por p_X, p_Y e p_W. Claramente, as distribuições são completamente diferentes. Mas quais são os valores de $\mathbb{E}(X)$, $\mathbb{E}(Y)$ e $\mathbb{E}(W)$?

$$\begin{aligned}
\mathbb{E}(X) &= 0 \times (-2) + 0 \times (-1) + 1 \times (0) + 0 \times (1) + 0 \times (2) \;=\; 0 \\
\mathbb{E}(Y) &= 0 \times (-2) + 0,25 \times (-1) + 0,5 \times (0) + 0,25 \times (1) + 0 \times (2) \;=\; 0 \\
\mathbb{E}(W) &= 0,5 \times (-2) + 0 \times (-1) + 0 \times (0) + 0 \times (1) + 0,5 \times (2) \;=\; 0
\end{aligned}$$

Portanto, apesar de as três distribuições serem completamente diferentes, as esperanças são iguais. No entanto,se calcularmos $\mathbb{E}\left(X^2\right)$, $\mathbb{E}\left(Y^2\right)$ e $\mathbb{E}\left(W^2\right)$, encontraremos valores completamente diferentes. O leitor atento deve estar questionando se fenômeno análogo aconteceria com o segundo momento[14]*. A resposta é positiva. Portanto, podemos, através de exemplos elaborados encontrar distribuições de probabilidade diferentes para as quais os k primeiros momentos são idênticos. Tente!*

[14]Isto é, se podemos encontrar distribuições de probabilidade diferentes com momentos de ordem até 2 respectivamente iguais.

Uma outra propriedade importante dos momentos é dada na proposição 3.5.4, em que é demonstrado o fato de que, sempre que existir o n-ésimo momento de uma v.a., X, existirão também todos os momentos de ordem $k = 1, 2, \ldots, n-1$. Antes, definamos

Proposição 3.5.4 (Existência de Momentos)

Sejam X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$ e n um número natural tais que existe $E(X^{n+1})$. Então existe $E(X^n)$.

Demonstração:

Devemos provar que

$$E(|X^n|) < \infty.$$

Dado que

$$|X^n| = |X^n|.1_{\{|X| \leq 1\}} + |X^n|.1_{\{|X|>1\}} \leq 1 + |X|^{n+1},$$

chega-se a

$$E(|X^n|) = E(|X^n|.1_{\{|X|\leq 1\}}) + E(|X^n|.1_{\{|X|>1\}}) \leq 1 + E(|X|^{n+1}).$$

Finalmente, dado que, por hipótese, $E(|X|^{n+1}) < \infty$, tem-se

$$E(|X^n|) < \infty. \qquad \blacksquare$$

Uma pequena observação é a de que, existindo $\mathbb{E}(X^k)$, para alguma $k \in \mathbb{N}$, também existirá (por quê?) a seguinte esperança:

$$\mathbb{E}\left(X(X-1)\cdots(X-k+1)\right).$$

Esses momentos, definidos a seguir, aparecem naturalmente em algumas situações e serão de interesse, neste livro, na seção 4.8, como ponto de ligação entre as funções geratrizes de probabilidades e de momentos.

Definição 3.5.5 (Momentos Fatoriais)
Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\mathbb{E}(X^k)$ exista.
O k-ésimo momento fatorial de X, $\mathbb{E}\left(X_{(k)}\right)$ é dado por

$$\mathbb{E}\left(X_{(k)}\right) = \mathbb{E}(X(X-1)\cdots(X-k+1)).$$

Na hora de calcular probabilidades de eventos raros, aparece naturalmente a seguinte pergunta: se X tem por média $\mu := E(X)$ como exprimir a probabilidade de a variável se desviar por mais do que uma certa quantidade de μ? Matematicamente, dado $\delta > 0$, como estimar $\mathbb{P}(|X-\mu| > \delta)$?

Pensemos que esse problema nos leva à idéia de que, assumida a esperança como um valor característico de X, gostaríamos de quantificar incertezas (através de, entre outras coisas, suas probabilidades) de eventos do tipo $[X - \mu > 3]$ ou $[X - \mu < -3]$ ou mesmo do simétrico $|X - \mu| > 3$.

Definamos, então, primeiramente, o conceito de momento central: esse é simplemente o valor esperado (de alguma ordem k escolhida) de uma nova variável aleatória, que se utiliza dos desvios com relação à esperança μ, isto é,

Definição 3.5.6 (Momento Central de Ordem k)

Seja X uma variável aleatória em $(\Omega, \mathcal{S}, \mathbf{P})$. O k-ésimo momento central de X (quando existe), α_k, é:

$$\alpha_k = E\left((X-\mu)^k\right).$$

Caso queiramos lidar diretamente com eventos do tipo $[|X-\mu| > \delta]$, sendo $|X-\mu|$ uma v.a. positiva podemos usar a Desigualdade de Markov (3.9) com $g : x \mapsto x^k$, para obter:

$$\mathbb{P}(|X-\mu| > \delta) \leq \frac{\mathbb{E}(|X-\mu|^k)}{\delta^k}.$$

Portanto, justifica-se a definição de um conceito próximo ao da definição 3.5.6: o momento central absoluto de ordem k.

Definição 3.5.7 (Momento Central Absoluto de Ordem k)

Chama-se de momento central de ordem k o número (quando existe) $\sigma_k = \mathbb{E}(|X-\mu|^k)$. O momento central de ordem 2 é chamado de variância e denotado por σ^2.

Uma observação bem simples que pode ser feita é a de que $\sigma_k = \alpha_k$, para todo k par, em particular, para o caso da variância. Havíamos visto no exemplo 3.5.5 diferentes distribuições que tinham esperanças iguais mas segundos momentos diferentes. Note que nesse exemplo, como $\mu_X = \mu_Y = \mu_W = 0$, tínhamos:

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \mathbb{E}\left(|X-\mu_X|^2\right) \\
\mathbb{E}\left(Y^2\right) &= \mathbb{E}\left(|Y-\mu_Y|^2\right) \\
\mathbb{E}\left(W^2\right) &= \mathbb{E}\left(|W-\mu_W|^2\right),
\end{aligned}$$

ou seja, os momentos centrais absolutos (no caso, a variância) foram capazes de diferenciar as distribuições quando as esperanças não o foram[15]. Vemos que, das distribuições definidas na tabela 3.1, a de X tem a menor dispersão de probabilidades, isto é, a maior concentração da massa de probabilidade total (que é 1) em torno de seu valor esperado. Y se comporta de forma intermediária nesse aspecto, enquanto W apresenta maiores probabilidades de afastamento do valor esperado. Note que $Var(X) = 0$, $Var(Y) = 0,5$ e $Var(W) = 4$. Portanto, tomando-se a definição de momentos centrais 3.5.5, temos que variâncias maiores indicam dispersões maiores em torno do valor esperado. Por isso, diz-se que a variância é uma **medida de dispersão**. . No caso do valor esperado, vemos que ele consegue localizar o *centro* da distribuição de probabilidade, sendo portanto chamado de uma **medida de locação**.

[15] Não se deve terminar a leitura desse parágrafo com a falsa impressão de que a variância tem maior capacidade de caracterização do que a esperança. A lição que se deve tirar desse exemplo é de que O MAIOR NÚMERO POSSÍVEL de momentos, centrais ou não, deve ser analisado.

Exemplo 3.5.6 (Mínimos Quadrados)

Uma outra interpretação muito interessante do valor esperado é dada pela seguinte questão: se tivermos uma distribuição de probabilidade discreta fidedigna $p_1, p_2, \ldots, p_n$*, para* $\{x_1, x_2, \ldots, x_n\}$*, qual o valor* c *que minimiza a seguinte quantidade:*

$$S(c) = \sum_{i=1}^{n} p_i \left(x_i - c\right)^2 \ ? \tag{3.11}$$

Antes de lidarmos com a tal questão, pensemos em qual seria o interesse de respondê-la. Suponha que tenhamos uma amostra com três valores possíveis: 1, 2 *e* 3*, com respectivas freqüências relativas* $f_1 = 0,1$*,* $f_2 = 0,8$ *e* $f_3 = 0,1$*. Claramente,* 2 *seria o valor mais representativo da amostra. Se quiséssemos medir o erro cometido ao escolher* 2 *em detrimento de outros valores, como por exemplo,* $m(< 2)$*, utilizar-nos-íamos de sua diferença:* $2 - m$*, ponderada por sua freqüência relativa,* $0,8$*. Portanto, um erro de* $c/8$ *unidades com relação a dois seria, por nossa ponderação, tão desastrosa quanto um erro de* c *unidades com relação a* 3*. Esse raciocício falha, no entanto, por os erros positivos e negativos poderem se cancelar. portanto, medidas de erros devem ser sempre positivas. Uma da maneiras de o fazer é pela utilização dos quadrados dos desvios, como em (3.11). Uma última nota sobre essa metodologia é a de que os quadrados mínimos tendem a proteger desvios maiores com média ponderação em detrimento de valores médios com alta ponderação. Finalmente, achemos* c *que miniza (3.11).*

Seja $\mu = \sum_{i=1}^{n} p_i x_i$*. Podemos escrever (3.11) como:*

$$\begin{aligned} S(c) &= \sum_{i=1}^{n} p_i \left(x_i - c\right)^2 \\ &= \sum_{i=1}^{n} p_i \left(x_i - \mu + \mu - c\right)^2 \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} p_i \left\{ (x_i-\mu)^2 + (\mu-c)^2 + 2\,(x_i-\mu)\,(\mu-c) \right\} \\
&= \sum_{i=1}^{n} p_i \,(x_i-c)^2 + (\mu-c)^2 + 2\,(\mu-c) \overbrace{\sum_{i=1}^{n} p_i\,(x_i-\mu)}^{=0} \\
&= \sum_{i=1}^{n} p_i \,(x_i-c)^2 + (\mu-c)^2 .
\end{aligned}
$$

Portanto, $S(c) = S(\mu) + (\mu-c)^2$, ou seja, μ minimiza $S(c)$ para todo c real.

3.6 Função Geratriz dos Momentos

A definição a seguir nos auxilia, entre outras coisas, no cálculo dos momentos, numa forma compacta, além de prover uma desigualdade muito interessante, a **Desigualdade de Bernstein**, vista no final desta seção.

Definição 3.6.1 (Função Geratriz de Momentos)

Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$. Define-se sua função geratriz dos momentos por

$$M_X(t) = \mathbb{E}\left(e^{tX}\right),$$

para os valores de t de forma que essa esperança exista.

Observe-se que $M_X(0) = 1$, pela propriedade (b) da proposição 3.5.1.

Se X fosse uma v.a. discreta com um número finito de valores possíveis, $\{0, 1, 2, \ldots, n\}$, e função de massa p_X, então:

$$M_X(t) = \sum_{k=0}^{n} e^{tk} p_X(k).$$

Se derivamos em t essa expressão, temos

$$M'_X(t) = \sum_{k=1}^{n} k e^{tk} p_X(k)$$

e, tomando-se $t = 0$,

$$M'_X(t)|_{t=0} = \sum_{k=1}^{n} k p_X(k) = \mathbb{E}(X).$$

Usando raciocínio análogo,

$$M_X^{(m)}(t) = \sum_{k=1}^{n} k^m e^{tk} p_X(k)$$

e

$$M_X^{(m)}(t)|_{t=0} = \sum_{k=1}^{n} k^m p_X(k) = \mathbb{E}(X^m).$$

Notemos que o que se demonstrou para a esperança de uma v.a. discreta em nada utilizou propriedades específicas de sua função de massa[16]. Foi apenas necessária a manipulação de uma fgm. Portanto, podemos estender essa propriedade para qualquer variavel X cuja $M_X(t)$ exista e esteja definida sobre $[0, s)$, onde $s > 0$.

Teorema 3.6.1 (Cálculo de Momentos pela fgm)

Seja X uma v.a. cuja fgm, $M_X(t)$, exista e esteja definida sobre $[0, s)$, para algum $s > 0$. Então

$$M_X^{(m)}(t)|_{t=0} = \mathbb{E}(X^m).$$

O teorema 3.6.1 justifica a denominação de função geratriz de momentos: a partir dela, podemos achar todos os momentos da distribuição de X (desde que existam).

O problema inverso de, a partir dos momentos, reconstituirmos sua fgm é bem mais complexo e em geral falso.

[16]A derivação dentro da soma, entre outras coisas, é possível sob certas condições, em geral atendidas pelas integrais com que usualmente lidamos

3.6.0.1 A Desigualdade de Bernstein

Seja X uma v.a. positiva. Usando-se a desigualdade de Markov, (3.9) com $g : x \mapsto \exp(tx)$, temos:

$$\mathbb{P}(X > x) \leq \frac{\mathbb{E}(\exp(tX))}{\exp(tx)} = \exp(-tx) M_X(t),$$

para qualquer t tal que $M_X(t)$ existe.

Portanto, tomando-se o menor valor possível para essas cotas, temos o resultado:

Teorema 3.6.2 (Desigualdade de Bernstein)

Seja X uma v.a. positiva com função geratriz dos momentos $M_X(t)$, definida sobre $[0, a)$, ($a \in (0, +\infty]$). Então,

$$\mathbb{P}(X > x) \leq \inf_{t \in [0,a)} \{\exp(-tx) M_X(t)\}.$$

Exemplo 3.6.1 *Seja X uma variável aleatória contínua, com densidade dada por:*

$$f(x) = \exp(-x) \mathbf{1}_{\mathbb{R}^+}(x).$$

X é dita uma exponencial de parâmetro 1[17]*.Então,*

$$\begin{aligned} P(X > x) &= \int_x^{+\infty} u f(u) du \\ &= \int_x^{+\infty} u \exp(-u) du \\ &= \exp(-x). \end{aligned}$$

No entanto,

$$M_X(t) = E(\exp(tX)) = \int_0^{+\infty} \exp(tu) f(u) du$$

[17] Veja a seção 5.4 para detalhes.

$$
\begin{aligned}
&= \int_0^{+\infty} \exp(tu)\exp(-u)du \;=\; \exp((t-1)u) \\
&= \frac{1}{1-t}, \qquad (3.12)
\end{aligned}
$$

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

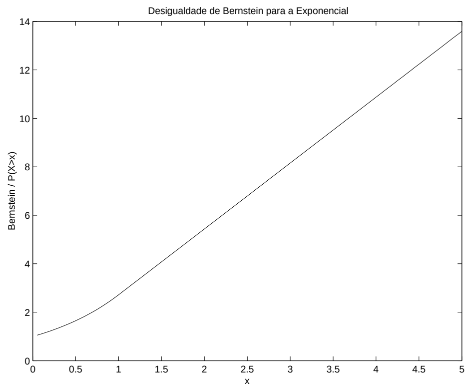
A Desigualdade de Bernstein nos diz, portanto, que:

$$
\begin{aligned}
P(X > x) &\leq \inf_{t\in[0,1)} \{\exp(-tx) M_X(t)\} \\
&= \inf_{t\in[0,1)} \frac{\exp(-tx)}{1-t}. \qquad (3.13)
\end{aligned}
$$

Como se comparam o verdadeiro valor de $P(X > x)$*, dado por (3.12), com o* chute*, dado por (3.13)?*

A figura 3.5 nos mostra o comportamento da desigualdade relativamente ao verdadeiro valor, ou seja, sua eficiência (dada pela razão entre o valor obtido pela desigualdade e o verdadeiro). Note que a desigualdade é tão mais eficiente quanto menor for o valor de x *e os valores por ela obtidos superestimam a cauda em até quatorze vezes, para* $x = 5$*. Portanto, a desigualdade de Bernstein, nesse caso, se mostra limitada, principalmente nos casos mais interessantes, os valores das caudas. O leitor não deve desanimar no uso de tais desigualdades. No entanto, deve fazê-lo criteriosamente: sempre que for possível, devemos utilizar resultados elaborados para a distribuição em estudo pois resultados gerais como a desigualdade de Bernstein, por mais espetaculares que pareçam, serão sempre piores do que os que podem ser obtidos especificamente para uma distribuição.*

Figura 3.5: Desigualdade de Bernstein para a Distribuição Exponencial



3.7 Outras Medidas de Locação e Dispersão

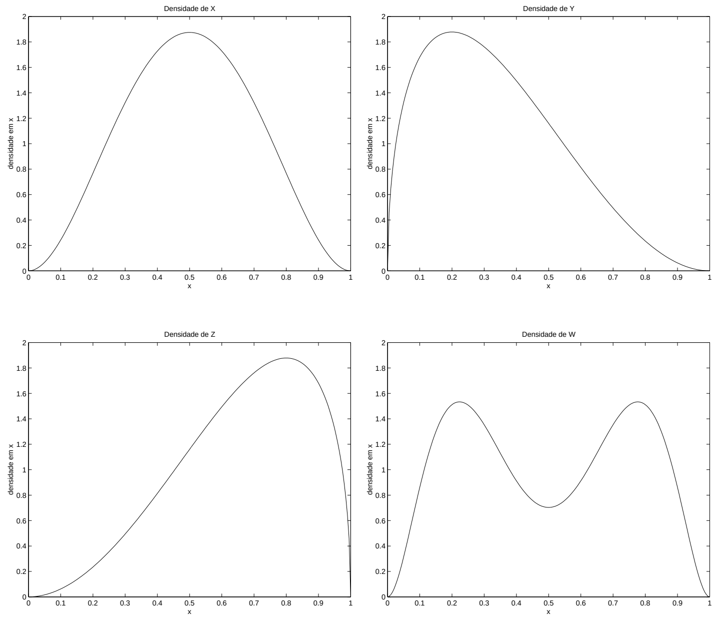
No exemplo 3.5.6, chegamos à conclusão que a esperança minimizava uma certa medida de erro, chamada de $S(.)$, motivando-nos a considerá-la como um valor característico da variável aleatória em questão. No entanto, nunca se discutiu a conveniência e relevância (sob aspectos variados) da função S. Essas discussões se dão em nível elevado e não nos cabe fazê-las em detalhes. Exporemos no exemplo a seguir não exatamente o que elas discutem mas por quê.

Exemplo 3.7.1 (Assimetria e Medidas de Locação)

Suponha que tenhamos as seguintes variáveis aleatórias e suas respectivas densidades: X, Y, Z, W, f_X, f_Y, f_Z *e* f_W. *Na figura 3.6, temos as quatro densidades ilustradas.*

Note que todas são simétricas com relação a $0,5$. *Portanto, em qualquer um dos casos, temos esperanças iguais:* $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = \mathbb{E}(W) = 0,5$, *mas claramente* $0,5$ *tem relevância diferenciada em cada uma das densidades. No caso de* X, *o ponto* $0,5$ *é central e parece ser o mais relevante no seguinte sentido:*

Figura 3.6: Assimetria e Medidas de Locação



não existe outro ponto real c que o ultrapasse em probabilidades do tipo:

$$\mathbb{P}(0,5-\delta \leq X \leq 0,5+\delta), \tag{3.14}$$

por essas probabilidades estarem diretamente ligadas à definição de densidades, quando as variáveis aleatórias são contínuas. Claramente, esse não é o caso para Y, Z *ou* W. *Além disso, note que, no caso de* W, *não só* $0,5$ *não apresenta essa propriedade de altas probabilidades como ele é um ponto de vale, isto é,* **TODOS** *os pontos em torno de* $0,5$ *terão associadas probabilidades do tipo (3.14) maiores do que as dele.*

Portanto, esses quatro exemplos ilustram qualitativamente as si-

tuações de caracterização de uma variável aleatória por sua esperança[18]*: sua relevância muito grande (caso de X), sua relevância discreta (casos de Y e Z) e sua completa irrelevância (caso de W).*

O exemplo 3.7.1 deve ser analisado sob a ótica de que, como toda medida caracterizadora, a esperança funcionará muito bem em muitos casos, bem em outros casos e muito mal em muitos casos e sua boa utilização se faz quando o usuário tiver conhecimento de quando e como usá-la. Essas deficiências da esperança motivam a utilização de outras medidas de locação. A primeira delas, diretamente retirada do exemplo 3.7.1 se refere ao tal ponto de *máxima probabilidade*. Ele é denominado de **moda** e pode ser definido como:

Definição 3.7.1 (Moda de uma Variável Aleatória)

Seja X uma variável aleatória contínua (discreta) com função de densidade (massa) f_X (p_X). Um ponto M é uma moda de X se for um ponto de máximo local de f_X (p_X).

Note que, através da definição 3.7.1, teríamos $0,5$ como uma moda de X mas não como uma moda de Z, Y ou W. Z e Y teriam uma moda cada e W teria duas. O leitor pode estar questionando a utilidade da moda: por exemplo, caso a densidade seja idêntica em todos os pontos de um intervalo, não há modas ou quando muitos pontos são modas. Na realidade, a moda permanece útil pois diz-nos em ambos os casos que não existe um valor destacadamente relevante em qualquer dos casos.

Portanto, vimos que tanto a moda quanto a esperça podem ser definidos como valores que otimizam algum critério: para a primeira, as *probabilidades locais* e, para a segunda, a função de quadrados mínimos. Iremos, agora ver uma outra quantidade que otimiza uma função extremamente natural. No exemplo 3.5.6, foi natural restringirmos os erros a funções positivas mas muitos se devem ter perguntado por que exatamente os quadrados. Na realidade,

[18] A adaptação para o caso discreto é bem simples.

a razão por que se faz essa escolha é operacional e não filosófica[19]. Como comentado anteriormente, quando utilizados os quadrados dos desvios, geramos uma deformação com conseqüências sérias para os estimadores[20]. O caminho natural, em vista dos cancelamentos dos desvios positivos e negativos, seria o de tomar os desvios absolutos, isto é, buscar-se o valor real c que minimizasse a seguinte soma:

$$A(c) = \sum_{i=1}^{n} p_i |x_i - c|. \tag{3.15}$$

Como achar o valor c que minimiza (3.15)? Definamos primeiramente o que se chama de valor mediano de uma distribuição. Aqui, o termo mediano tem o sentido usual: aquele que se coloca no meio, nem entre os maiores nem entre os menores.

Definição 3.7.2 (Mediana de uma Variável Aleatória)

Seja X uma variável aleatória em $(\Omega, \mathcal{S}, \mathbb{P})$, com fd F_X. A mediana é qualquer valor, m, tal que:

$$m = \arg_{c \in \mathbb{R}} \{\mathbb{P}(X \leq c) \geq 1/2, \mathbb{P}(X \geq c) \geq 1/2\}.$$

Observe que a mediana é um número real que tenha pelo menos metade da probabilidade acima (incluindo-se m) e metade da ocorrência abaixo (incluindo-se m). No caso contínuo, essa noção pode ser redefinida simplesmente por: m é tal que

$$F_X(m) = 0,5.$$

No caso discreto, por haver saltos, faz-se necessária a utilização de ambas as desigualdades apresentadas na definição. Tanto no caso discreto como no contínuo, podemos ter uma ou muitas medianas. Em ambos os tipos de distribuição, teremos uma entre duas situações: uma mediana ou um intervalo

[19] Quem estiver interessado em aprofundar seus conhecimentos estatísticos, deve consultar um livro de modelos lineares ou de regressão.

[20] Essas conseqüências são extremamente corriqueiras e motivam toda uma vasta área de estudos estatísticos denominada Estatística Não-paramétrica.

real cujos elementos são todos medianas. Pense por quê (Veja o exercício 3.44).

Finalmente, vejamos que a mediana realmente minimiza (3.15).

Seja m uma mediana de X. Utilizaremos o caso discreto (para o qual (3.15) é natural) mas o resultado pode ser facilmente estendido. Queremos calcular $A(c)$ e mostrar que ele atinge seu mínimo para $c = m$. Seja c um número real e $r = c - m$. Iremos demonstrar o resultado para r positivo e o caso de $r < 0$ é análogo.

$$\begin{aligned}
A(c) &= \sum_{i=1}^{n} p_i |x_i - c| \\
&= \sum_{i=1}^{n} p_i |x_i - m + m - c| \\
&\quad \text{note que } |x_i - m| = |x_i - c| - r \text{ se } x_i \leq m \\
&\quad \qquad |x_i - m| \leq |x_i - c| + r \text{ se } x_i > m \\
&\geq \sum_{x_i \leq m} \left(p_i \left(|x_i - m| + r\right)\right) + \sum_{x_i > m} \left(p_i \left(|x_i - m| - r\right)\right) \\
&= \sum_{i=1}^{n} p_i |x_i - m| + r \left(\sum_{x_i \leq m} \left(p_i - \sum_{x_i > m} p_i \right) \right) \\
&\quad \text{como } m \text{ é uma mediana} \\
&\geq A(m) + r(0,5 - 0,5) \\
&\geq A(m),
\end{aligned}$$

com igualdade somente se c também for uma mediana de X.

Essa interessante dicotomia mediana-média como provedores de valores que minimizam somas de desvios absolutos-quadrados tem reflexos muito importantes na Teoria Estatística. Além disso, o aumento extraordinário da capacidade de cálculo propiciou o acesso a técnicas estatísticas anteriormente inviáveis, incluindo-se aqui as *sucessoras* da mediana em problemas de regressão. Aqui, limitar-nos-emos a definir medidas relacionadas à mediana que nos auxiliam a estudar distribuições teóricas e orientam-nos no estudo de dados.

Definição 3.7.3 (Quartis, Percentis e Quantis)

Seja X uma variável aleatória em $(\Omega, \mathcal{S}, \mathbb{P})$, com fd F_X. Seu α-ésimo quantil, q_α, para $0 < \alpha < 1$, é qualquer valor tal que:

$$\arg_{c\in\mathbb{R}}\{\mathbb{P}\,(X \leq c) \geq \alpha, \mathbb{P}\,(X \geq q) \geq 1-\alpha\}.$$

Em particular, utiliza-se a seguinte notação:

α	*Nome*	*Notação*
$0,25$	*Primeiro Quartil*	Q_1
$0,50$	*Segundo Quartil (Mediana)*	Q_2
$0,75$	*Terceiro Quartil*	Q_3
$0,01$	*Primeiro Percentil*	P_1
$0,02$	*Segundo Percentil*	P_2
$\ldots$	$\ldots$	
$0,01 \times j$	*j-ésimo Percentil*	P_j
$\ldots$	$\ldots$	
$0,98$	*Nonagésimo Oitavo Percentil*	P_{98}
$0,99$	*Nonagésimo Nono Percentil*	P_{99}

No capítulo 5, o leitor irá conhecer uma distribuição contínua muito importante, chamada de **Distribuição Cauchy**, que não possui esperança. A mediana é seu valor de locação natural mas o que se pode fazer quanto a sua dispersão? Obviamente, o desvio-padrão não pode ser cogitado: essa mesma distribuição, no entanto, nos apresenta uma outra medida de dispersão natural, o desvio interquartílico, que é definida a seguir. Outra medida natural de dispersão nos foi motivada pela própria definição de mediana, o desvio médio,[21] definido a seguir para as variáveis aleatórias cujas esperanças sejam bem definidas.

[21] Note que, no entanto, o desvio médio é definido utilizando-se a esperança e não a mediana mas podemos fazê-lo também com esta em vez daquela. A visão alternativa do desvio médio se dá por como se calcula o desvio e não em torno do que ele é calculado.

Definição 3.7.4 (Desvio Interquartílico e Desvio Médio)

Seja X uma variável aleatória em $(\Omega, \mathcal{S}, \mathbb{P})$, com fd F_X. O desvio interquartílico, D, é dado por $DI = Q_3 - Q_1$.

Suponha que X seja contínua (discreta), com densidade (massa) f_X (p_X) e que possua esperança $\mathbb{E}(X)$. Seu desvio médio, DM, é dado por:

$$\int_{-\infty}^{+\infty} |u - \mathbb{E}(X)| f_X(u) du,$$

para v.a.'s contínuas, e

$$\sum_{i=-\infty}^{+\infty} |x_i - \mathbb{E}(X)| p_{x_i},$$

para v.a.'s discretas.

O desvio médio é uma excelente medida para se comparar os efeitos, no desvio-padrão, do uso do quadrado dos desvios com relação à esperança em vez de se utilizarem os desvios absolutos. Uma interessante ilustração é desenvolvida no capítulo 4, no exemplo 4.3.3. O desvio interquartílico utiliza as próprias probabilidades de pontos *distantes* da distribuição de uma variável aleatória para definir sua dispersão, forma bastante robusta e, por isso, de grande interesse prático.

3.8 Exercícios

Exercício 3.1 (Pontos Isolados)

(a) Prove que, se A for um conjunto finito, então A é um conjunto de pontos isolados;

(b) Prove que o conjunto $\mathbb{Z}$ (dos números inteiros) é um conjunto de pontos isolados; e

(c) Dê um exemplo de um conjunto enumerável que não seja um conjunto de pontos isolados.

Exercício 3.2 *Demonstre rigorosamente as afirmações i), ii) e iii) da observação 3.1.1.*

Exercício 3.3 *Complete os detalhes da demonstração do teorema 3.2.1.*

Exercício 3.4 *Demonstre o seguinte:*

> *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória definida em $(\Omega, \mathcal{F}, \mathbb{P})$. A função de conjunto $\mathbb{P}_X$, definida nos borelianos de $\mathbb{R}$, por:*
>
> $$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega / X(\omega) \in B\})$$
>
> *é uma probabilidade e, portanto, $(\mathbb{R}, Borelianos, \mathbb{P}_X)$ é um espaço de probabilidade (Lembre-se que os Borelianos de $\mathbb{R}$ formam a menor σ-álgebra que contém a família dos intervalos reais).*

Exercício 3.5 *Complete os detalhes da demonstração do Teorema 3.2.3.*

Exercício 3.6 *Seja X uma variável aleatória cuja função de distribuição acumulada $F(\cdot)$ é dada por:*

$$F(t) = \begin{cases} 0 & se\ t < -2; \\ 0,1 & se\ -2 \leq t < 0; \\ 0,1 + t^2 & se\ 0 \leq t < 0,8; \\ 0,7 & se\ 0.8 \leq t < 2; \\ 0,7 + 0,25(t-2) & se\ 2 \leq t < 3,2;\ e \\ 1 & se\ t \geq 3,2. \end{cases}$$

Esboce o gráfico de F e calcule as probabilidades dos seguintes eventos: $[X \leq -6]$, $[X \leq -2]$, $[X \leq 0]$, $[X \leq 0,8]$, $[X \leq 1]$, $[X \leq 1,5]$, $[X \leq 3]$, $[X \leq 3,2]$, $[X \leq 4]$, $[X \leq 84]$, $[X \leq 86]$, $[X = -5]$, $[X = -2]$, $[X = 0]$, $[X = 0,5]$, $[X = 0,8]$, $[X = 2]$, $[X = 3]$, $[X = 55]$, $[X = 64]$, $[X < -6]$, $[X < -2]$, $[X < 0]$, $[X < 0,8]$, $[X < 1]$, $[X < 1,5]$, $[X < 3]$, $[X < 3,2]$, $[X < 4]$, $[X < 84]$, $[X < 86]$, $[0,5 < X \leq 2]$, $[0,5 \leq X \leq 2]$, $[0,5 \leq X < 2]$,

$[0,5 < X < 2]$, $[0,8 < X \leq 3,2]$, $[0,8 \leq X \leq 3,2]$, $[0,8 \leq X < 3,2]$, $[X > -3]$, $[X > -2]$, $[X > 0]$, $[X > 0,8]$, $[X > 1]$, $[X > 2]$, $[X > 3]$, $[X > 3,2]$, $[X > 8,4]$, $[X > 31]$, $[X > 46]$, $[X > 2723]$, $[X \geq -3]$, $[X \geq -2]$, $[X \geq 0]$, $[X \geq 0,8]$, $[X \geq 1]$, $[X \geq 2]$, $[X \geq 3]$, $[X \geq 3,2]$, $[X \geq 8,4]$, $[X \geq 31]$, $[X \geq 46]$, $[X \geq 2723]$.

Exercício 3.7 *Prove que:*
i) se $a < b$,

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a) + \textit{salto de } F \textit{ em } a;$$

ii) se $a < b$,

$$\mathbb{P}(a \leq X < b) = F(b) - F(a) + \textit{salto de } F \textit{ em } a - \textit{ salto de } F \textit{ em } b;$$

iii) se $a < b$,

$$\mathbb{P}(a < X < b) = F(b) - F(a) - \textit{ salto de } F \textit{ em } b; \textit{ e}$$

iv) para todo $t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(X > t) &= 1 - F(t); \textit{ e} \\ \mathbb{P}(X \leq t) &= 1 - F(t) + \textit{ salto de } F \textit{ em } t. \end{aligned}$$

Exercício 3.8
a) Seja X *uma variável aleatória inferiormente limitada e seja* $m \in \mathbb{R}$ *uma cota inferior de* X*; prove que:* $F(t) = 0$ *para todo* $t < m$*;*
b) Demonstre que a conclusão em a) permanece válida se substituirmos a hipótese $X \geq m$ *por* $\mathbb{P}(X \geq m) = 1$*;*
c) Reciprocamente, seja X *uma variável aleatória tal que* $F(t) = 0$ *para todo* $t < m$*. Prove que* $\mathbb{P}(X \geq m) = 1$*;*
d) Formule e demonstre resultados simétricos aos de a), b) e c) para variáveis aleatórias superiormente limitadas e cotas superiores.

Exercício 3.9 *Demonstre a seguinte variante do exercício 3.8.*

> *Seja X uma variável aleatória:*
>
> *a) $m \in \mathbb{R}$ é uma cota inferior essencial de X se e somente se $F(t) = 0$ para todo $t < m$;*
>
> *b) $k \in \mathbb{R}$ é uma cota superior essencial de X se e somente se $F(t) = 1$ para todo $t \geq k$; e*
>
> *c) $c \in \mathbb{R}$ é uma cota essencial de X se e somente se $F(t) = 1$ se $t \geq c$ e $F(t) = 0$ se $t < -c$.*

Exercício 3.10 *Considere variáveis aleatórias X, Y, Z, W, U e V cujas respectivas acumuladas estão dadas por:*
i) F_X dada no exercício 3.6;

ii)

$$F_Y(t) = \begin{cases} 0 & t < 1, \\ e^{-2} & 1 \leq t < 2, \\ 1 - e^{-t} & t \geq 2; \end{cases}$$

iii)

$$F_Z(t) = \begin{cases} 0 & t < 0, \\ t/5 & 0 \leq t < 5, \\ 1 & t \geq 5; \end{cases}$$

iv)

$$F_W(t) = \begin{cases} 0 & t < -5, \\ 0,3 & -5 \leq t < 2, \\ 0,8 & 2 \leq t < 8, \\ 1 & t \geq 8; \end{cases}$$

v)

$$F_U(t) = \begin{cases} 0 & t < 1, \\ 1 - 2^{-n} & n \leq t < n - 1, \quad \forall n \in \mathbb{N}; \end{cases}$$

vi)

$$F_V(t) = \frac{1}{2} + \frac{1}{\pi} arctg(t),$$

para todo $t \in \mathbb{R}$, *onde arctg é a função inversa de* $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$.

Esboce os gráficos de F_X, F_Y, F_Z, F_W *e* F_V. *Determine quais variáveis são essencialmente inferiormente limitadas e quais essencialmente superiormente limitadas. É possível reconhecer se elas são inferiormente limitadas ou superiormente limitadas a partir de suas acumuladas?*

Exercício 3.11 *Ache as funções de massa de probabilidade das variáveis aleatórias do exercício 3.10. Confirme que, em todos os casos,* $\sum_{r\in\mathbb{R}} p(r) \leq 1$, *sendo que, no caso de* W *e* U, *a soma é exatamente* 1.

Exercício 3.12 *O que significa* $\sum_{r\in\mathbb{R}} p(r) \leq 1$*? Prove, usando o teorema 3.2.3, que o conjunto* $\{\omega/\mathbb{P}(\{\omega\}) > 0\}$ *é no máximo enumerável.*

Exercício 3.13 *Seja* X *uma variável aleatória definida no espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$, *com função de massa de probabilidade* p *e função de distribuição acumulada* F. *Prove que:*

a) $\sum_{r\in\mathbb{R}} p(r) = \sum_{r\in\mathbb{R}} \mathbb{P}(X = r) = \sum_{r\in\mathbb{R}}\{\textit{salto de } F \textit{ em } r\} \leq 1$;

b) $\sum_{r\leq c} p(r) = \sum_{r\leq c} \mathbb{P}(X = r) = \sum_{r\leq c}\{\textit{ salto de } F \textit{ em } r\} \leq F(c)$; *e*

c) Se $\sum_{r\in\mathbb{R}} p(r) = \sum_{r\in\mathbb{R}} \mathbb{P}(X = r) = \sum_{r\in\mathbb{R}}\{\textit{ salto de } F \textit{ em } r\} = 1$, *então*

$$\sum_{r\leq c} p(r) = \sum_{r\leq c} \mathbb{P}(X = r) = \sum_{r\leq c}\{\textit{salto de } F \textit{ em } r\} = F(c),$$

para todo $c \in \mathbb{R}$.

Exercício 3.14 *Considere as variáveis definidas no exercício 3.10.*
i) Verifique que W *e* U *são as únicas variáveis que satisfazem o item c do exercício 3.13;*

ii) Seja $S = \{-5, 2, 8\} = \{t/\mathbb{P}(W = t) > 0\}$. *Verifique também que, para todo* $A \subset \mathbb{R}$, *tem-se que*

$$\begin{aligned} \mathbb{P}(W \in A) &= \mathbb{P}(W \in A \cap S) \\ &= \mathbf{1}_A(-5).\mathbb{P}(W = -5) + \mathbf{1}_A(2).\mathbb{P}(W = 2) + \mathbf{1}_A(8).\mathbb{P}(W = 8), \end{aligned}$$

onde $\mathbf{1}_A$ *denota a função indicadora do conjunto* A;
iii) No caso da variável aleatória U, *verifique que para todo* $A \subset \mathbb{R}$, *tem-se que:*

$$\mathbb{P}(U \in A) = \mathbb{P}(U \in A \cap \mathbb{N}) = \sum_{n \in \mathbb{N}} \mathbf{1}_A(n) 2^{-n}; \text{ e}$$

iv) No caso geral, seja X *uma variável aleatória tal que:*

$$\sum_{r \in \mathbb{R}} p(r) = \sum_{r \in \mathbb{R}} \mathbb{P}(X = r) = 1.$$

Então, para todo $A \subset \mathbb{R}$, *vale que*

$$\mathbb{P}(X \in A) = \sum_{r \in A} p(r) = \sum_{r \in A} \mathbb{P}(X = r) = \sum_{r \in \mathbb{R}} 1_A(r) \mathbb{P}(X = r).$$

Exercício 3.15

a) Prove que uma variável aleatória discreta (definição 3.3.1) satisfaz a definição 3.1.1 de uma variável aleatória;
b) Verifique que toda variável aleatória discreta X *satisfaz a condição iv) do exercício 3.14, ou seja:*

$$\sum_{r \in \mathbb{R}} p(r) = \sum_{r \in \mathbb{R}} \mathbb{P}(X = r) = 1,$$

onde p *é a função de massa de probabilidade de* X *e* $\mathbb{P}(X = r) > 0$ *se e somente se* $r \in$ *Imagem*(X); *e*
c) Seja X *discreta e Imagem*$(X) = \{\alpha_i\}_{i \in \mathbb{N}}$; *demonstre que:*

$$X = \sum_{i \in \mathbb{N}} \alpha_i \mathbf{1}_{A_i},$$

onde $A_i = X^{-1}(\alpha_i)$ *para todo* $i \in \mathbb{N}$ *e que, se* F *é a função de distribuição acumulada da variável aleatória* X*, tem-se que:*

$$F(t) = \sum_{r \leq t} p(r) = \sum_{\alpha_i \leq t} \mathbb{P}(X = \alpha_i) = \sum_{i \in \mathbb{N}} \mathbb{P}(X = \alpha_i) \mathbf{1}_{[\alpha_i, -\infty)}(t).$$

Exercício 3.16 (Variante do Exemplo 3.1.4)

Um alvo tem centro preto, C*, e uma parte externa branca,* B*, em volta. Um atirador lança um dardo na direção do alvo. Se o centro é atingido, ganham-se dezoito pontos; se a parte branca é atingida, ganham-se oito pontos. Caso o dardo não atinja parte alguma do alvo, perdem-se dois pontos. Defina formalmente a variável número de pontos e construa sua função de probabilidade, função de distribuição acumulada nos dois casos seguintes:*

(a) o atirador João com $\mathbb{P}(C) = 0,2$ *e* $\mathbb{P}(B) = 0,7$*; e*
(b) a atiradora Paula, com $\mathbb{P}(C) = 0,4$ *e* $\mathbb{P}(B) = 0,45$*.*

Compare os desempenhos dos atiradores, primeiramente através de uma análise das respectivas fd's e, depois, pelas médias e variâncias e as probabilidades dos seguintes eventos:

1. $\mathbb{P}($*"João pontuar positivamente e Paula negativamente"*$)$;
2. $\mathbb{P}($*"João pontuar negativamente e Paula positivamente"*$)$; *e*
3. $\mathbb{P}($*"João ganhar mais pontos do que Paula"*$)$.

Exercício 3.17

(a) Seja X *uma v.a. tal que:* $\mathbb{P}(X = -1) = 0,2$*;* $\mathbb{P}(X = 0) = 0,1$*; e* $\mathbb{P}(X = 6) = 0,7$*. Calcule:* $\mathbb{E}(X)$*;* $\mathbb{E}(3X + 2)$*;* $\mathbb{E}(-2X + 1)$*;* $\mathbb{E}(X^2)$*;* $\mathbb{E}(X^3)$*; e a função de distribuição acumulada de* X*. Verifique também que* $Var(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$ *e demonstre este resultado no caso geral.*

(b) Seja Y *uma v.a. tal que* $\mathbb{P}(Y = -1) = \mathbb{P}(Y = 7) = 0,3$ *e* $\mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = 0,2$*. Responda às questões análogas às do item (a) e generalize, calculando formalmente a função de probabilidade de um* $h(Y)$ *qualquer, sendo* Y *uma v.a. discreta.*

Exercício 3.18
a) Mostre que a função de distribuição acumulada da variável aleatória Y do exemplo 3.1.4 é dada por

$$F_Y(t) = \begin{cases} 0 & se\ t < -5; \\ \mathbb{P}(Y=-5) & se\ -5 \leq t < 2; \\ \mathbb{P}(Y=-5)+\mathbb{P}(Y=2) & se\ 2 \leq t < 8;\ e \\ 1 & se\ 8 \leq t; \end{cases}$$

b) Mostre que a função de distribuição acumulada da variável aleatória X do exemplo 3.1.3 é dada por:

$$F_X(t) = \begin{cases} 0 & se\ t < 0; \\ t & se\ 0 \leq t < 5;\ e \\ 1 & se\ 5 \leq t. \end{cases}$$

(c) Esboce os gráficos de F_X e F_Y e comente.

Exercício 3.19 *Seja W uma v.a. cuja fd é dada por:*

$$F_W(x) = \begin{cases} 0 & x < -3; \\ 0,2 & -3 \leq x < 4; \\ 0,9 & 4 \leq x < 8;\ e \\ 1 & 8 \leq x. \end{cases}$$

(a) Calcule, através de F diretamente, as probabilidades dos seguintes conjuntos: $[3 < W \leq 7]$, $[3 \leq W \leq 7]$, $[3 \leq W < 7]$, $[3 < W < 7]$, $[-3 < W \leq 5]$, $[-3 \leq W \leq 5]$, $[-3 \leq W < 5]$, $[-3 < W < 5]$, $[W \leq 6]$, $[W < 6]$, $[W \leq 4]$, $[W < 4]$, $[W > 3]$, $[W \geq 3]$, $[W > 4]$, $[W \geq 4]$, $[W > 10]$, $[W \geq 10]$, $[W \geq -4]$, $[W < -10]$ *e* $[W \leq 14]$.

(b) Faça os cálculos análogos aos de (a), desta vez utilizando-se da densidade de W.

(c) Compare os dois métodos dos itens (a) e (b) e, pelo que julgar mais conveniente, calcule $\mathbb{E}(W)$ e $Var(W)$. Justifique.

Exercício 3.20 *Demonstre as seguintes propriedades de* $\mathbb{E}$ *e* Var*:*

(i) $X \geq 0 \Rightarrow \mathbb{E}(X) \geq 0$;
(ii) $X = c \Rightarrow \mathbb{E}(X) = c$;
(iii) $\mathbb{E}(cX) = c\mathbb{E}(X)$;
(iv) $Var(X) \geq 0$;
(v) $X = c \Leftrightarrow Var(X) = 0$;
(vi) $Var(X + c) = Var(X)$; *e*
(vii) $Var(cX) = c^2 Var(X)$.

Exercício 3.21 *Demonstre a seguinte variante da propriedade da esperança, dada pela proposição 3.5.1 item (a).*

Seja X *uma variável aleatória não-negativa.*

(i) Então, $\mathbb{E}(X) \geq 0$; *e*

(ii) Se $\mathbb{E}(X) = 0$, *então* $\mathbb{P}(X > 0) = 0$.

Explique em como essa variante enriquece o itm (a) da proposição 3.5.1.

Dicas - *Observe que:*

$$[X > 0] = \bigcup_{n=1}^{+\infty} [X > n^{-1}]$$

e, se existise $m \in \mathbb{N}$ *tal que* $\mathbb{P}\left(X > n^{-1}\right) > 0$, *então teríamos que*

$$\mathbb{E}(X) \geq \frac{1}{m}\mathbb{P}\left(X > \frac{1}{m}\right) > 0.$$

Exercício 3.22 *Sabe-se que uma determinada moeda apresenta cara três vezes mais freqüentemente que coroa. Essa moeda é jogada três vezes. Seja* X *o número de caras resultante. Estabeleça a distribuição de probabilidade e fd de* X. *Faça esboços dos respectivos gráficos. Proceda, de forma análoga (sem utilizar informações sobre* X*) para a v.a.* Y, *número de coroas resultante no mesmo experimento. Finalmente, utilize o fato de que* Y *pode ser escrito como função de* X *(como?) e calcule novamente a distribuição de probabilidades e fd de* Y.

Exercício 3.23 *Suponha que a v.a. X tenha como possíveis valores $1, 2, \ldots$ e função de probabilidade dada por: $p_j \equiv \mathbb{P}(X = j) = 2^{-j}$, $j = 1, 2, \ldots$. Verifique que $\{p_j\}_{j=1}^{\infty}$ é uma função de probabilidade fidedigna. Ache a fd, F, de X e calcule de ambas as maneiras (a partir dos p_j's e de F):*

(a) $\mathbb{P}(X$ *ser par);*
(b) $\mathbb{P}(X \geq 5)$*; e*
(c) $\mathbb{P}(X$ *ser divisível por 3).*

Compare os dois métodos de cálculo.

Exercício 3.24 *Suponha que f e g sejam densidades de v.a.'s contínuas definidas em um intervalo $[a, b]$. Mostre que:*

(a) $f + g$ não é uma densidade fidedigna de uma v.a contínua no intervalo $[a, b]$;

(b) $\beta f + (1 - \beta)g$ define uma densidade fidedigna de uma v.a. no intervalo $[a, b]$ sempre que $0 \leq \beta \leq 1$; e

(c) $\beta f + (1 - \beta)g$ pode ou não definir uma densidade fidedigna de uma v.a. no intervalo $[a, b]$, se $\beta < 0$ ou $\beta > 1$. Crie um exemplo e um contraexemplo.

Exercício 3.25 *Sejam X e Y duas variáveis aleatórias discretas, com respectivas funções de massa p_X e p_Y e imagens $\{x_1, x_2, \ldots, x_n\}$ e $\{y_1, y_2, \ldots, y_m\}$.*

(a) Mostre que a função $p_X + p_Y$ não é uma função de massa;

(b) Mostre que, qualquer que seja $0 \leq \beta \leq 1$, $\beta p_X + (1 - \beta)p_Y$ define uma função de massa fidedigna;

(c) Qual a interpretação de $\beta p_X + (1 - \beta)p_Y$? Pense num **sorteio** *em dois estágios;*

(d) $\beta p_X + (1 - \beta)p_Y$ pode ou não definir uma função de massa fidedigna, se $\beta < 0$ ou $\beta > 1$. Crie um exemplo e um contraexemplo.

Exercício 3.26 *A percentagem de álcool em um certo composto pode ser considerada uma variável aleatória,* $Y = 100X$, *onde* X *tem a seguinte densidade:*

$$f(x) = \begin{cases} 20x^3(1-x) & 0 < x < 1; \text{ e} \\ 0 & \text{caso contrário.} \end{cases}$$

(a) Estabeleça que f *é uma densidade fidedigna, encontre a fd equivalente e esboce o seu gráfico;*

(b) Calcule $\mathbb{P}(X \leq 2/3)$*; e*

(c) Suponha que o preço de venda deste composto dependa do conteúdo de álcool. Especificamente, se o percentual de álcool estiver entre $100/3$ *e* $200/3$, *o composto se vende por* C_1 *reais/galão e, nos outros casos,* C_2 *reais/galão. Se o custo de produção for de* C_3 *reais/galão, onde* $C_1 > C_2 > C_3$, *encontre a distribuição do lucro por galão.*

Exercício 3.27 *Seja* X *uma v.a. contínua com densidade dada por:*

$$f(x) = \begin{cases} ax & 0 \leq x \leq 1 \\ a & 1 \leq x \leq 2 \\ -ax + 3a & 2 \leq x \leq 3 \\ 0 & \text{caso contrário.} \end{cases}$$

(a) Determine a constante a; e
(b) Determine a fd F *e esboce o seu gráfico.*

Exercício 3.28 *Sejam* X, Y *e* W *v.a.'s com respectivas fd's:*

$$F_X(t) = \begin{cases} 0 & t < 0 \\ \sqrt(t) & 0 \leq t \leq 1 \\ 1 & 1 < t; \end{cases}$$

$$F_Y(t) = \begin{cases} 0 & t < 0 \\ \ln(t) & 1 \leq t \leq e \\ 1 & e < t \end{cases} \quad ; \text{ e}$$

$$F_W(t) = \begin{cases} 0 & t < 0 \\ 2^t - 1 & 0 \le t \le 1 \\ 1 & 1 < t. \end{cases}$$

Verifique que cada uma das funções acima é uma fd e ache a respectiva função de densidade. Esboce os gráficos das densidades e fd's para cada caso. Compare as três distribuições.

Exercício 3.29 *Sejam X, Y e W v.a.'s com respectivas densidades:*

$$g_X(s) = \begin{cases} a(1-s) & 0 < s < 1 \\ 0 & c.c.; \end{cases}$$

$$g_Y(s) = \begin{cases} b \exp -10s & 0 \le s \\ 0 & c.c. \end{cases} \quad ; e$$

$$g_W(s) = \begin{cases} c - \frac{2s}{225} & 0 < s < 15 \\ 0 & c.c.. \end{cases}$$

Ache as constantes a, b e c que tornam cada uma das funções acima uma densidade fidedigna. Encontre as respectivas fd's. Esboce os gráficos das densidades e fd's. Compare as três distribuições.

Exercício 3.30 *Existem N tipos de cupom e, cada vez que uma extração é realizada, o cupom será do tipo i com probabilidade p_i, $i = 1, 2, \ldots, N$, independentemente das extrações anteriores,. Seja T o número necessário de extrações para que uma pessoa consiga obter pelo menos um cupom de cada tipo. Calcule $\mathbb{P}(T = n)$.*

Exercício 3.31 *Seja N uma v.a. inteira e não-negativa. Mostre que seu valor esperado pode ser escrito como:*

$$\mathbb{E}(N) = \sum_{i=1}^{\infty} \mathbb{P}(N \ge i).$$

Dica: Argumente que $\sum_{i=1}^{\infty} \mathbb{P}(N \geq i) = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \mathbb{P}(N = k)$ *e troque a ordem dos somatórios. Por que você o pode fazer?*

Exercício 3.32 *Uma família tem* n *filhos com probabilidade* αp^n, *onde* $\alpha \leq (1-p)/p$, $n \in \mathbb{N}$.

(a) Qual a proporção de famílias que não têm filhos?

(b) Se as probabilidades de cada um dos filhos ser do sexo masculino ou feminino são iguais (e independência é válida), qual a proporção de famílias consiste de k *meninos (e qualquer número de meninas), para* $k \in \mathbb{N}$*?*

Exercício 3.33 *Suponha que um dado seja lançado duas vezes. Liste os possíveis valores que cada uma das seguintes v.a's possa assumir:*

(a) o valor máximo que aparece nos dois lançamentos;
(b) o valor mínimo que aparece nos dois lançamentos;
(c) a soma dos valores dos dois lançamentos; e
(d) a diferença entre o valor do primeiro e segundo lançamentos.

Ache também as densidades de probabilidade para cada uma das v.a.'s.

Exercício 3.34 *Em um jogo chamado de* Morra, *dois jogadores mostram um ou dois dedos e, simultaneamente, gritam o número de dedos que imaginem que o adversário esteja mostrando. Se apenas um dos jogadores acertar a previsão, ele ganha a soma do número de dedos mostrados por ambos, em reais. Caso nenhum ou ambos acertem, não há prêmio. Escolha um dos jogadores e defina como* X *a quantidade de dinheiro ganha por ele em um desses jogos.*

(a) Se cada jogador age independemente do adversário e cada jogador escolhe o número de dedos a mostrar e sua previsão de forma que as quatro possíveis combinações de resultados sejam equiprováveis, quais os possíveis valores de X *e suas respectivas probabilidades?*

(b) Suponha que cada jogador aja independemente. Se cada jogador decide mostrar e prever números iguais e que cada jogador tem probabilidades iguais

de mostrar um ou dois dedos, quais os possíveis valores de X e respectivas probabilidades?

Exercício 3.35 *Cinco números distintos são aleatoriamente distribuídos a jogadores numerados de 1 a 5. Cada vez que dois jogadores comparam seus números, aquele com o maior é declarado vencedor. Supondo que todos as possíveis comparações sejam feitas, seja X o número de vitórias do jogador número 1. Ache a função de probabilidade de X.*

Exercício 3.36 *Um livro sobre jogos recomenda a seguinte estratégia para vencer na roleta. Joga-se um real no vermelho; caso saia vermelho (cuja probabilidade é $18/37$), o jogador deve pegar seu real de lucro e abandonar a mesa. Caso perca (cuja probabilidade é $19/37$), deve-se apostar um real no vermelho em cada uma das duas rodadas seguintes e abandonar a mesa em seguida. Seja X o lucro do jogador quando ele abandona a mesa.*

(a) Ache $\mathbb{P}(X > 0)$;
(b) Você considera esta estratégia vencedora? Explique!
(c) Ache $\mathbb{E}(X)$.

Exercício 3.37 *A cada noite, os meteorologistas nos dão probabilidades de que vá ou não chover no dia seguinte. De forma a julgar quão boas são suas previsões, os metereologistas receberão pontos da seguinte forma: se sua previsão for de que choverá com probabilidade p, receber-se-ão $1 - (1 - p^2)$ pontos caso chova ou $1 - p^2$ caso não chova.*

Acumular-se-ão os pontos de cada meteorologista durante um certo tempo e será declarado o(a) melhor aquele(a) com maior pontuação. Suponha que uma metereologista tenha descoberto nossa avaliação e que, portanto, queira maximizar sua pontuação esperada. Se sua real estimativa para a probabilidade de chuva for $p^\star$, qual(is) o(s) valor(es) de p ela deve usar para assegurar que tenha maximizada sua pontuação esperada? Esboce os gráficos que julgar convenientes para sua solução.

Exercício 3.38 *Cada turbina de um avião irá falhar com probabilidade $1-p$, independentemente das outras. Se a aeronave precisa da maioria de seus* *motores em funcionamento para manter-se no ar, quais os valores de p para os quais um avião de cinco turbinas será preferível a um de três. Assuma que ambos os modelos se utilizam de turbinas de mesma especificação. Esboce os gráficos das probabilidades de falha, para cada uma dos casos, como função de p.*

Exercício 3.39 *Prove que:*

Se g é uma função tal que $g(t) \geq 0$ para todo $t \in \mathbb{R}$ e $\int_{-\infty}^{\infty} g(t)dt = 1$, então $G(t)$, dada por

$$G(t) = \int_{-\infty}^{t} g(u)du,$$

é uma função não decrescente, contínua à direita, $\lim_{x\to-\infty} G(x) = 0$ e $\lim_{x\to\infty} G(x) = 1$.

Dadas as três propriedades acima, existe alguma variável aleatória X cuja função de distribuiçaõ acumulada seja G (veja teorema 3.2.2), sendo óbvio que X é absolutamente contínua e que g é uma densidade de X.

Exercício 3.40 *Seja X uma variável aleatória com densidade f_X e esperança $\mathbb{E}(X)$. Defina uma outra variável aleatória Y da seguinte forma*

$$Y = aX + b,$$

onde $a, b \in \mathbb{R}$. Mostre que

(a) Y é uma variável aleatória absolutamente contínua se e somente se $a \neq 0$. Qual sua densidade nesse caso? Caso $a = 0$, como poder-se-ia classificar Y? Explicite sua distribuição;

(b) $\mathbb{E}(Y) = a\mathbb{E}(X) + b$.

Exercício 3.41 *Considere $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidades e um evento $A \in \mathcal{F}$. Defina-se a v.a. (indicadora do evento A), denotada por $\mathbf{1}_A$,* *por*

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

Mostre que $\mathbf{1}_A$ *é uma variável aleatória em* $(\Omega, \mathcal{F}, \mathbb{P})$. *Mostre que* $\mathbf{1}_A$ *é discreta e que sua função de massa é dada por*

$$\begin{aligned} p(1) &= \mathbb{P}(A) \\ p(0) &= \mathbb{P}(A^c) = 1 - \mathbb{P}(A) \end{aligned}$$

Exercício 3.42 *Demonstre a seguinte versão contínua da proposição 3.5.4:*

> *Seja* X *uma variável aleatória absolutamente contínua, com densidade* f. *Então, para todo* n *natural,*
>
> $$\exists \mathbb{E}\left(X^{n+1}\right) \Rightarrow \exists \mathbb{E}\left(X^n\right).$$

Exercício 3.43
(a) Mostre que, se existe $\mathbb{E}(X^n)$, *então também existe* $\mathbb{E}((X-c)^n)$, $\forall c \in \mathbb{R}$.
(b) Mostre que, se existe $\mathbb{E}((X-k)^n)$, *para algum* $k \in \mathbb{R}$, *então também existe* $\mathbb{E}((X-c)^n)$, $\forall c \in \mathbb{R}$.

Dicas: *em (a), use o binômio de Newton para* $(X-c)^n$; *em (b), use que* $(X-c) = (X-k) + (k-c)$.

Exercício 3.44 (Mediana para Variáveis Contínuas e Discretas)
(a) Seja X *uma v.a. contínua com fd* F; *sejam*

$$u = \inf\{t \in \mathbb{R} / F(t) = 1/2\}$$

e

$$v = \sup\{t \in \mathbb{R} / F(t) = 1/2\}.$$

Mostre que m *é uma mediana de* X *se e somente se*

$$u \leq m \leq v. \tag{3.16}$$

(b) Seja X uma v.a. discreta com fd F tal que existe u que satisfaz

$$F(u) = \frac{1}{2}.$$

Neste caso,

$$u = \sup\{t \in \mathbb{R}/F(t) < 1/2\} \ = \ \inf\{t \in \mathbb{R}/F(t) = 1/2\}$$

e definimos

$$v = \sup\{t \in \mathbb{R}/F(t) = 1/2\} \ = \ \inf\{t \in \mathbb{R}/F(t) > 1/2\}.$$

Mostre que m é uma mediana de X se e somente se

$$u < m < v.$$

(c) Seja X uma v.a. discreta com fd F tal que $1/2$ não pertence à imagem de F. Mostre que

$$\sup\{t \in \mathbb{R}/F(t) < 1/2\} \ = \ \inf\{t \in \mathbb{R}/F(t) > 1/2\} \tag{3.17}$$

e que o número real, por (3.17) definido, é a única mediana de X.

Capítulo 4

Principais Modelos Discretos

4.1 Introdução

Em muitos ramos de matemática, quando se falando em modelos discretos, há uma tendência em pensá-los de forma dissociada. Isto é razoável pela natureza singular de cada problema. No caso de modelos de probabilidade discretos, há, no entanto, uma forma de reuní-los, em sua maioria, sob uma única perspectiva. Cada modelo deste capítulo pode ser definido em função de resultados de uma seqüência do experimento aleatório mais simples imaginável: *o Experimento de Bernoulli.*

Suponha que um experimento tenha apenas dois possíveis resultados: $\mathcal{S}$ e $\mathcal{F}$, sucesso e fracasso respectivamente. À primeira vista, muitos casos de interesse seriam complexos demais para tal experimento (ou seqüência destes). Um observador incauto até poderia conjecturar que tal experimento seria irrelevante para todos os *casos interessantes.*

Na realidade, o que veremos ao longo deste capítulo é exatamente o contrário. As principais variáveis aleatórias discretas têm definição e interpretação baseadas completamente em seqüências de experimentos de Bernoulli. O *tamanho* de uma tal seqüência e o algebrismo necessários variarão com a complexidade do problema em questão mas o ponto central, a interpretação como simples funções de resultados dos experimentos de Bernoulli, permanece. Isto é verdadeiro desde os problemas mais triviais aos mais complexos. O espírito do Experimento de Bernoulli tem lugar desde aplicações estatísticas simples até pesquisas de ponta.

4.2 O Experimento de Bernoulli

Definição 4.2.1 (Experimento de Bernoulli)

Seja $\mathbf{E}$ *um experimento nos moldes da definição 1.1.2. Suponha que* $\mathbf{E}$ *tenha um conjunto de resultados possíveis de cardinalidade dois. Um dos resultados será chamado de sucesso,* $\mathcal{S}$*, e o outro de fracasso,* $\mathcal{F}$*. Um sucesso ocorrerá com probabilidade* $0 \leq p \leq 1$ *e, conseqüentemente, um fracasso ocorrerá com probabilidade* $1 - p$*.*

Portanto, um *Experimento de Bernoulli* tem suas propriedades estatísticas completamente caracterizadas pela probabilidade de sucesso, p. Note que esse experimento pode ser tanto de natureza simples como de natureza abstrata, de acordo com o que se entende por sucesso ou fracasso. Mesmo assim, sua natureza aleatória, em qualquer grau de abstração, se manifesta única e exclusivamente através da probabilidade p. Apesar de estudados como casos particulares, experimentos de Bernoulli triviais, isto é, em que $p = 0$ ou $p = 1$, são, na realidade, experimentos determinísticos, em completo acordo com a definição 1.1.1 .

Como nosso maior interesse, em geral, reside no estudo das propriedades estatísticas de um experimento, faz-se interessante a caracterização dos sucessos e fracassos dissociada da complexidade de sua natureza. Para isso, associa-se a um *Experimento de Bernoulli*, uma variável aleatória (discreta), cuja distribuição de probabilidade é definida a seguir.

Definição 4.2.2 (Distribuição Bernoulli)

Seja $\mathbf{E}$ *um experimento de Bernoulli, nos moldes da definição 4.2.1, com probabilidade de sucesso* p*. Diz-se que* X *tem distribuição Bernoulli com parâmetro* p*,* $0 \leq p \leq 1$ *(*$X \sim b(p)$*), se* X *associa valores ao resultado de* $\mathbf{E}$ *da seguinte forma*

$$X = \begin{cases} 0 & \text{se } \mathbf{E} \text{ resultou em fracasso} \\ 1 & \text{se } \mathbf{E} \text{ resultou em sucesso.} \end{cases}$$

Observe-se que, associada a uma variável aleatória, X (pela definição 4.2.2), há uma outra variável aleatória $1 - X$, que também tem distribuição Bernoulli, mas com parâmetro $1-p$. Isto nada mais é do que conseqüência da utilização arbitrária dos conceitos de **sucesso** e **fracasso** na definição 4.2.1.

As principais propriedades de uma variável aleatória com distribuição Bernoulli serão casos particulares das respectivas propriedades de distribuição binomial, definida por 4.3.1 com $n = 1$ e, portanto, não serão apresentadas em separado.

4.3 Distribuição Binomial

Seja Υ uma seqüência de experimentos de Bernoulli independentes, no sentido da definição 4.2.1. Realizados n desses experimentos, seria de interesse saber quantos sucessos ocorreram.

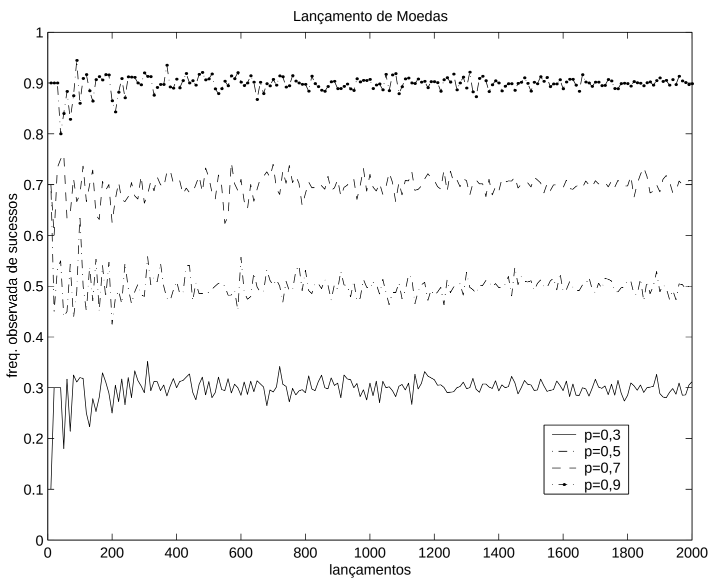
Exemplo 4.3.1 (Quão honesta é uma moeda)

As seguintes simulações (figura 4.1) mostram como se comportariam seqüências de lançamentos de moedas, de acordo com o tamanho da seqüência e o valor do parâmetro p. Notemos que, quando as seqüências têm tamanhos pequenos, há uma oscilação razoável em torno do verdadeiro valor de p. Esse comportamento das freqüências observadas é umas das motivações da definição freqüentista de probabilidade e conseqüência direta da relação entre esperança e probabilidade, vista no capítulo 3, pág. 90.

Definição 4.3.1 (Distribuição Binomial)

Suponha que um experimento de Bernoulli seja repetido, sob as mesmas condições e de forma independente, n vezes. Seja X o número de sucessos obtidos. X é dita ter distribuição binomial de parâmetros n e p, onde p é a probabilidade de sucesso no primeiro experimento de Bernoulli.

Figura 4.1: Honestidade de Moedas



Nota: Escreve-se $X \sim Bin(n,p)$.

Para sabermos calcular as probabilidades de uma v.a. binomial, poderíamos utilizar o fato de que tal v.a. nada mais é do que a soma de n v.a.'s de Bernoulli independentes. No entanto, cálculos desta natureza requerem manuseio de distribuições multidimensionais, claramente fora de propósito para um curso de primeiro semestre em probabilidade. As técnicas combinatórias que iremos utilizar na dedução das probabilidades da binomial tem caráter extremamente construtivo. Procedimentos análogos nos levarão à solução de diversos problemas relevantes relacionados aos experimentos de Bernoulli, ao longo deste capítulo.

O problema de calcular a probabilidade de termos k sucessos em n tentativas pode ser dividido em dois: qual a probabilidade de que uma particular configuração com k sucessos ocorra e quantas destas configurações existem.

Primeiramente, por os experimentos serem realizados em condições idênticas e de forma independente, as probabilidades de que uma particular configuração ocorra não depende da ordem em que sucessos e fracassos aparecem mas apenas de quantos sucessos contém, isto é,

$$\mathbb{P}((i_1, i_2, \dots, i_n)) = \mathbb{P}((j_1, j_2, \dots, j_n)),$$

onde

$$(i_1, i_2, \dots, i_n), (j_1, j_2, \dots, j_n) \in (\mathcal{S}, \mathcal{F})^n$$

e

$$\sum_{k=1}^{n} \mathbf{1}(i_k = \mathcal{S}) = \sum_{k=1}^{n} \mathbf{1}(j_k = \mathcal{S}).$$

Além disto, sabemos que a probabilidade de uma particular configuração com k sucessos será, por independência dos eventos, dada por $p^k(1-p)^{n-k}$, para $k = 0, 1, 2, \dots, n$. Finalmente, basta contar o número da tais configurações: pela irrelevância da ordem de ocorrência dos sucessos no processo de contagem, vemo-nos diante de $\binom{n}{k}$ diferentes combinações. Portanto:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \mathbf{1}_{\{0,1,2,\dots n\}}(k). \tag{4.1}$$

Cada umas das probabilidades acima definidas é não-negativa. Portanto, para que (4.1) defina uma distribuição de probabilidade fidedigna, basta que:

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = 1.$$

Observe-se que cada probabilidade em (4.1) nada mais é que o coeficiente de ordem k do Binômio de Newton $(p + (1-p))^n$, ou seja,

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = (p + (1-p))^n = 1.$$

Proposição 4.3.1 (Esperança e Variância da Binomial)
Seja X uma v.a. com distribuição binomial de parâmetros n e p. Então,

$$\begin{aligned} \mathbb{E}(X) &= np & (4.2)\\ Var(X) &= np(1-p) & (4.3) \end{aligned}$$

Demonstração:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= \sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)!(n-1-(k-1))!}p^k(1-p)^{n-k}\\ &= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{(n-1)-(k-1)} = np, \end{aligned}$$

sendo a penúltima soma o Binômio de Newton $(p+(1-p))^{n-1}$, i.e., 1.

Sabemos que $Var(X)$ pode ser escrita como $\mathbb{E}(X^2)-(\mathbb{E}(X))^2$ mas, pelas características dos coeficientes binomiais, é mais fácil lidar com $\mathbb{E}\,(X(X-1))$:

$$\begin{aligned} \mathbb{E}\,(X(X-1)) &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k}\\ &= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= \sum_{k=2}^{n} n(n-1)\frac{(n-2)!}{(k-2)!(n-2-(k-2))!}p^k(1-p)^{n-k}\\ &= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{(n-2)-(k-2)} = n(n-1)p^2, \end{aligned}$$

sendo a penúltima soma o Binômio de Newton $(p + (1-p))^{n-2}$, i.e., 1.

Finalmente,

$$\begin{aligned}
Var(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= np - np^2 \quad = \quad np(1-p). \blacksquare
\end{aligned}$$

Como é comum com as distribuições de variáveis aleatórias discretas, a mediana não é computacionalmente simples no caso de uma v.a. binomial. Na realidade, aquela tem que ser calculada por um processo numérico. Primeiramente, vamos recordar a definição de mediana, 3.7.2:

$$m = \arg_{c\in\mathbb{R}}\{\mathbb{P}(X \leq c) \geq 1/2, \mathbb{P}(X \geq c) \geq 1/2\}.$$

No caso da binomial, estas duas probabilidades são:

$$\begin{aligned}
\mathbb{P}(X \leq m) &= \sum_{j=0}^{m} \binom{n}{k} p^k (1-p)^{n-k} \\
\mathbb{P}(X \geq m) &= \sum_{j=m}^{n} \binom{n}{k} p^k (1-p)^{n-k},
\end{aligned}$$

indicando que a mediana m será aquele ponto

A princípio, o cálculo da moda teria o mesmo problema. Como veremos a seguir, no entanto, há uma solução simples e elegante:

Proposição 4.3.2 (Moda da Binomial)

Seja X uma v.a. $Bin(n,p)$. Se $(n+1)p \in \mathbb{N}$, a moda de X é dupla e se localiza em $(n+1)p - 1$ e $(n+1)p$. Caso contrário, a moda é única, localizando-se em $\lfloor (n+1)p \rfloor$. Mais ainda, a função de probabilidade da $Bin(n,p)$ é monotamente crescente até sua moda e monotamente decrescente a partir daí.

Demonstração:

É suficiente estudar em quais pontos a razão $\mathbb{P}(X = k)/\mathbb{P}(X = k-1)$ é maior do que, menor do que ou igual a 1 para os dois casos: $(n+1)p$ inteiro ou não. Mas, para $k = 1, 2, \ldots, n$:

$$\begin{aligned}
\frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} &= \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1}(1-p)^{n-(k-1)}} \\
&= \frac{\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}}{\frac{n!}{(n-(k-1))!(k-1)!} p^{k-1}(1-p)^{n-(k-1)}} \\
&= \frac{(n-k+1)p}{k(1-p)},
\end{aligned}$$

isto é,

$$\begin{aligned}
\mathbb{P}(X = k) \geq \mathbb{P}(X = k-1) &\Leftrightarrow (n-k+1)p \geq k(1-p) \\
&\Leftrightarrow k \leq (n+1)p. \qquad (4.4)
\end{aligned}$$

Note que, de forma análoga, podemos concluir que:

$$\begin{aligned}
\mathbb{P}(X = k) > \mathbb{P}(X = k-1) &\Leftrightarrow (n-k+1)p > k(1-p) \\
&\Leftrightarrow k < (n+1)p. \qquad (4.5)
\end{aligned}$$

Portanto, há um inteiro $(n+1)p \leq m_1 < (n+1)p + 1$ cuja probabilidade é maior do que todas as antecedentes e não é menor do que as posteriores. Mais ainda, quando $m_2 = (n+1)p$ for inteiro, $\mathbb{P}(X = m_2) = \mathbb{P}(X = m_2 - 1)$, sendo o valor máximo atingido por $\mathbb{P}(X = k)$, $k = 1, 2, \ldots, n$. Além disso, por (4.4) e (4.5), temos que $\mathbb{P}(X = k)$ é estritamente crescente até m_1 $(m_2 - 1)$ e estritamente decrescente a partir de m_1 (m_2) se m_2 não for inteiro $(m_2 \in \mathbb{N})$.
■

O exemplo a seguir nos mostra alguns valores de n e p e quais as modas das respectivas distribuições binomiais. Além disso, temos também uma ilustração do grau de assimetria da distribuição binomial de acordo com os valores de seus parâmetros.

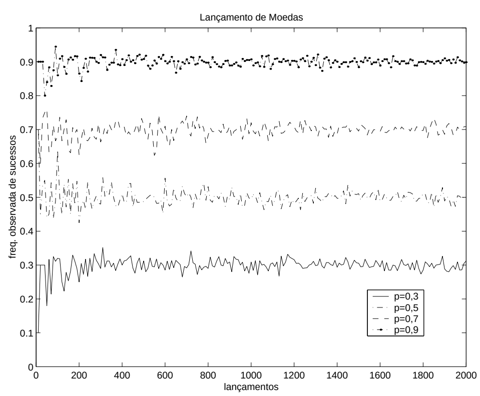
Exemplo 4.3.2 (Modas e Assimetria da Binomial)

Notemos, na tabela 4.1 que apenas para o valor de $n = 99$*, isto é, quando* $(n+1)p$ *são números inteiros,* $p = 0,3; 0,5; 0,7; 0,9$*, temos moda dupla. Na figura 4.2, vemos como se comporta a distribuição binomial do ponto de vista de suas assimetrias.*

Tabela 4.1: Modas da Distribuição Binomial

$n \setminus p$	$0,3$	$0,5$	$0,7$	$0,9$
10	3	5	7	9
20	6	10	14	18
50	15	25	35	45
99	29 30	49 50	69 70	89 90

Figura 4.2: Modas da Distribuição Binomial



O exemplo abaixo, cuja motivação é devida a uma nota de N.L. Johnson, [21], mostra uma comparação interessante entre os conceitos de desvio médio e desvio padrão aplicados à distribuição binomial.

Exemplo 4.3.3 (Uma Nota no Desvio Médio da Distribuição Binomial)

Primeiramente, citemos o resultado devido a Gruder, [15]:

$$\begin{aligned}
\sum_{r=m}^{n} \binom{n}{r} p^r q^{n-r}(r-np) &= \sum_{r=m}^{n} \binom{n}{r} p^r q^{n-r}(rq-(n-r)p) \\
&= \sum_{r=m}^{n} \left[n\binom{n-1}{r-1} p^r q^{n-r+1} - n\binom{n-1}{r} p^{r+1} q^{n-r} \right] \\
&= n\binom{n-1}{m-1} p^m q^{n-m+1} \\
&= m\binom{n}{m} p^m q^{n-m+1} \qquad (4.6) \\
& \qquad (4.7)
\end{aligned}$$

A identidade (4.6) foi utilizada por N.L. Johnson no cálculo do desvio médio de uma v.a. binomial. O desvio médio é dado, no caso da binomial, por (veja definição 3.7.4):

$$\begin{aligned}
\sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} |r-np| &= -\sum_{r=0}^{\lceil np-1 \rceil} \binom{n}{r} p^r q^{n-r} (r-np) + \\
&\quad + \sum_{\lceil np \rceil}^{n} \binom{n}{r} p^r q^{n-r} (r-np) . \quad (4.8)
\end{aligned}$$

Outra identidade extremamente útil é dada pela aplicação do fato de a soma dos desvios em relação à esperança ser nula, em *particular*

$$\sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} (r - np) = 0,$$

que implica que (4.8) pode ser reescrita como

$$\sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} |r - np| = 2 \sum_{r=\lceil np \rceil}^{n} \binom{n}{r} p^r q^{n-r} (r - np) . \quad (4.9)$$

Portanto, utilizando-se (4.6) em (4.9), temos que o desvio médio de uma binomial é simplesmente:

$$2\lceil np \rceil \binom{n}{\lceil np \rceil} p^{\lceil np \rceil} q^{n-\lceil np \rceil+1}.$$

Faz-se interessante utilizar $\xi = \lceil np \rceil - np$*. Com isto, a razão entre os desvios médio e padrão de uma binomial é dada por:*

$$\begin{aligned} R &= 2\,(np+\xi) \binom{n}{np+\xi} p^{np+\xi} q^{nq-\xi+1} / \sqrt{npq} \\ &= \frac{2\,(np+\xi)}{\sqrt{n}} \binom{n}{np+\xi} p^{np+\xi-\frac{1}{2}} q^{nq-\xi+\frac{1}{2}}. \end{aligned} \quad (4.10)$$

Aplica-se, então, a fórmula de Stirling (como no teorema 1.3.2), na forma

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} exp(-n) \left(1 + \frac{1}{12n}\right) {}^{1},$$

obtendo-se

[1] Note que aqui, a fórmula de Stirling é diferente da definida por (2.14), na página 287. No entanto, a diferença entre as duas é apenas num termo de grandeza menor, isto é, a presente fórmula de Stirling é mais precisa do que a utilizada normalmente

$$\begin{aligned} R &\approx \sqrt{\frac{2}{\pi}}\left(1+\frac{\xi}{np}\right)^{\frac{1}{2}-\xi}\left(1-\frac{\xi}{nq}\right)^{\xi-\frac{1}{2}}\left(1+\frac{\xi}{np}\right)^{-np}\times \\ &\left(1-\frac{\xi}{nq}\right)^{-nq}\left(1+\frac{1}{12n}\right)\left(1+\frac{1}{12(np+\xi)}\right)^{-1}\left(1+\frac{1}{12(nq-\xi)}\right)^{-1}. \end{aligned}$$

Para np e nq suficientemente grandes, temos as seguintes aproximações:

$$\begin{aligned} \left(1+\tfrac{\xi}{np}\right)^{\frac{1}{2}-\xi} &\approx 1+\tfrac{\xi\left(\frac{1}{2}-\xi\right)}{np} \\ \left(1-\tfrac{\xi}{nq}\right)^{\xi-\frac{1}{2}} &\approx 1+\tfrac{\xi\left(\frac{1}{2}-\xi\right)}{nq} \\ \left(1+\tfrac{\xi}{np}\right)^{-np} &\approx exp-\xi\left(1+\tfrac{\xi^2}{2np}\right) \\ \left(1-\tfrac{\xi}{nq}\right)^{-nq} &\approx exp\xi\left(1+\tfrac{\xi^2}{2nq}\right) \end{aligned}$$

Podemos, portanto, finalmente reescrever a razão aproximada entre os desvios médio e padrão de uma v.a. binomial como:

$$\begin{aligned} R &\approx \sqrt{2}\pi\left(1+\frac{\xi^2}{2npq}+\frac{\xi\left(\frac{1}{2}-\xi\right)}{npq}+\frac{1}{12n}-\frac{n}{12\left(np+\xi\right)\left(nq-\xi\right)}\right) \\ &\approx \sqrt{2}\pi\left(1+\frac{\xi\left(1-\xi\right)}{2npq}-\frac{1-pq}{12npq}\right). \qquad (4.11) \end{aligned}$$

A seguinte tabela é provida em N.L Johnson que nos mostra a qualidade dessa aproximação para o cálculo de R. Note que, para $n \geq 50$, a aproximação coincide com o valor exato em quatro casas decimais.

De modo geral, os seis gráficos da figura 4.3 nos mostram como se comportam a razão entre os desvios médio e padrão, dada por (4.10), e a aproximação proposta para seu cálculo, dada por (4.11). No primeiro gráfico, acima à esquerda, temos o gráfico da verdadeira razão entre os desvios médio e padrão: note quão

Tabela 4.2: Relação entre Desvios Médio e Padrão para a Binomial

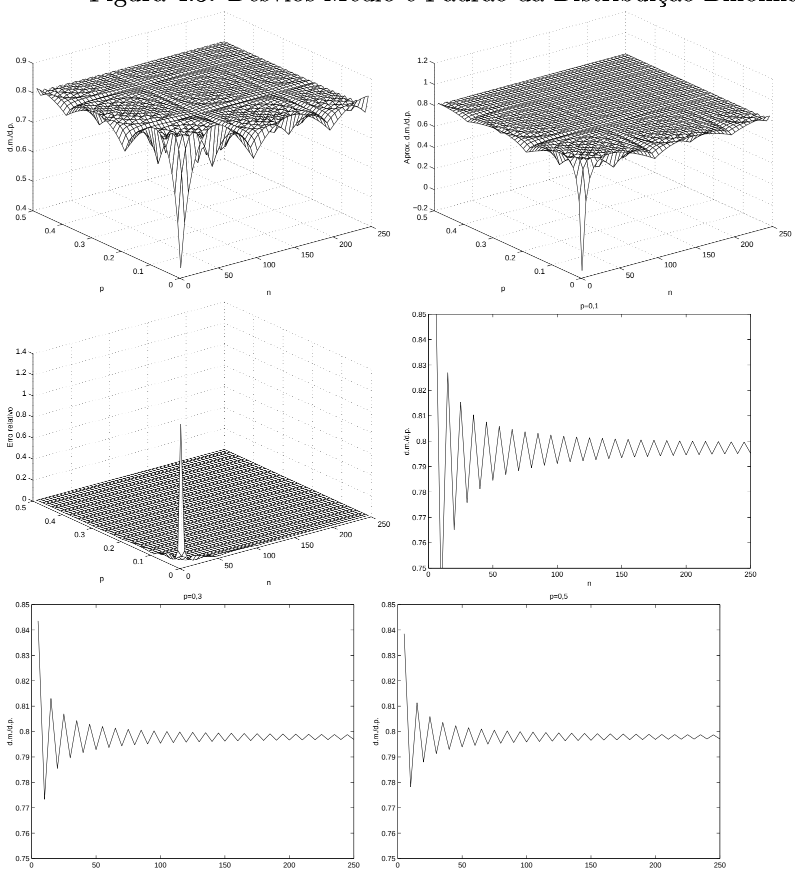
p / n	0,1 (A) (B)	0,2 (A) (B)	0,3 (A) (B)	0,4 (A) (B)	0,5 (A) (B)
10	0,7351	0,7640	0,7733	0,7771	0,7782
	0,7313	0,7630	0,7729	0,7768	0,7779
20	0,7652	0,7807	0,7855	0,7874	0,7879
	0,7642	0,7804	0,7854	0,7874	0,7879
50	0,7846	0,7909	0,7929	0,7937	0,7939
	0,7846	0,7909	0,7929	0,7937	0,7939
100	0,7912	0,7949	0,7954	0,7958	0,7959
	0,7912	0,7949	0,7954	0,7958	0,7959

(A) Valor exato dado pela equação 4.10

(B) Valor aproximado dado pela equação 4.11

rápida, em n e p (com o sentido de p se aproximar de $0,5$), é a convergência dessa razão para a vizinhança de $0,795$. No segundo gráfico, acima à direita, temos a mesma ilustração para a proximação: note que o comportamento é parecido com o do primeiro mas, aqui, temos uma oscilação um pouco menor. O terceiro gráfico, em que se mostra o erro relativo da aproximação, retrata quão boa é a aproximação, funcionando muito mesmo para pequenos valore de n e p (as diferenças só são razoáveis se n for bem menor do que 50 **e** *p menor do que $0,1$). Finalmente, os três últimos gráficos ilustram a verdadeira razão entre os desvios para $p = 0,1$; $0,3$ e $0,5$. Duas características são interessantes: todos os gráficos mostram mudanças bruscas de direção, exatamente pela função teto na fórmula; e a estabilização da razão entre os desvios é muito mais rápida para $p = 0,5$.*

Figura 4.3: Desvios Médio e Padrão da Distribuição Binomial



4.4 Distribuição Geométrica

Definição 4.4.1 (Distribuição Geométrica)

Suponha que uma seqüência de experimentos independentes de Bernoulli seja realizada. Seja X o número de realizações até a ocorrência do primeiro sucesso. A distribuição da variável aleatória X é chamada de Geométrica de parâmetro p ($X \sim G(p)$).

Algumas observações devem ser feitas antes de obtermos a função de probabilidade de X. Primeiramente, note que X pode assumir qualquer valor inteiro positivo. É natural que se pense que o número *médio* de realizações seja tanto maior quanto menor for a probabilidade de sucesso, p. Defina A_j como o evento em que os j primeiros experimentos resultam em fracasso. O evento B_k, o primeiro sucesso ocorre exatamente na k-ésima realização, pode ser escrito como:

$$B_k = A_{k-1} \bigcap \{\text{Sucesso na } k\text{-ésima realização}\}, \tag{4.12}$$

$k = 1, 2, \ldots$, sendo $A_0 = \Omega$.

Note-se que, dada a independência dos experimentos,

$$\mathbb{P}(B_k) = (1-p)^{k-1} p \mathbf{1}_{\{1,2,\ldots\}}(k). \tag{4.13}$$

Para demonstrar que (4.13) define uma função de probabilidade fidedigna, temos que provar que as propriedades de uma função de massa são atendidas. Obviamente, todas as probabilidades são não-negativas. Quanto a as probabilidades somarem 1, note que $\{\mathbb{P}(X = k)\}_{k=1}$ é uma progressão geométrica de razão $1 - p$, logo:

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}(X = k) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\ &= \frac{(1-p)^{1-1} p}{1 - (1-p)} = 1. \end{aligned}$$

Note que o estudo de tal variável aleatória só é de interesse quando $0 < p < 1$. No casos triviais de $p = 1$ ou $p = 0$, haverá sucesso na primeira realização ou fracassos em todas as realizações, respectivamente, com probabilidade 1. A proposição 4.4.1 nos mostra o comportamento da esperança e variância de uma v.a. geométrica nos casos não-triviais.

Proposição 4.4.1 (Esperança e Variância da Geométrica)

Seja $X \sim G(p)$, $0 < p < 1$. *Então,*

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{p} \;\; ; e \\ Var(X) &= \frac{1-p}{p^2}. \end{aligned}$$

Demonstração:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=1}^{\infty} n(1-p)^{n-1}p &= \sum_{n=0}^{\infty} p\left(\frac{d}{dq}q^n\right) &= p\frac{d}{dq}\sum_{n=0}^{\infty} q^n \\ &= p\frac{d}{dq}\frac{q^0}{1-q} &= p\frac{(-1)}{(1-q)^2} &= \frac{1}{p}. \end{aligned}$$

Como usual, utilizar-nos-emos da $\mathbb{E}(X^2)$ no cálculo de $Var(X)$:

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{n=1}^{\infty} n^2(1-p)^{n-1}p &= \sum_{n=0}^{\infty} p\left(\frac{d}{dq}nq^n\right) &= p\frac{d}{dq}\sum_{n=0}^{\infty} nq^n \\ &= p\frac{d}{dq}\left(\frac{q}{p}\sum_{n=1}^{\infty} nq^{n-1}p\right) &= p\frac{d}{dq}\left(\frac{q}{p}E(X)\right) \\ &= p\frac{d}{dq}\left(\frac{q}{p^2}\right) &= p\frac{d}{dq}\left(2q(1-q)^{-3}+(1-q)^{-2}\right) \\ &= \frac{1}{p}\left(\frac{2}{p}-1\right) \end{aligned}$$

Portanto,

$$Var(X) \;=\; \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 \;=\; \frac{1}{p}\left(\frac{2}{p} - 1\right) - \frac{1}{p^2} \;=\; \frac{1-p}{p^2}. \blacksquare$$

Observe-se que tanto maiores serão a esperança e a variância do tempo para primeira realização quanto menor for a probabilidade de ocorrência do evento em questão. É natural que X assuma valores estocásticos maiores visto que a probabilidade de ocorrência do evento de interesse é pequena. A variância maior pode ter uma interpretação equivocada, no sentido de que as probabilidades *para valores distantes* da média são maiores quando p é menor. Na realidade, o que acontece é que a seqüência $(\{1-p)^k p\}_{k=1}^{\infty}$ decai muito lentamente quando p é muito pequeno e a existência de altos valores com *grandes* probabilidades é portanto a razão de maiores valores para a variância.

Nota: Existe na literatura uma versão alternativa da distribuição geométrica em que, em vez de contarem-se as tentativas até que surja o primeiro sucesso, contam-se os fracassos até o primeiro sucesso. Sendo Y distribuída como tal, é fácil ver que Y pode ser escrita como $Y = X - 1$, onde $X \sim G(p)$. A escolha de qual definição utilizar não é arbitrária. Neste livro, quando se falar em $G(p)$, refere-se à definida em 4.4.1. No entanto, existem várias razões por que escolher a versão alternativa. O nível de discussão é elevado e o leitor é convidado a consultar [10], [12] ou [13].

4.5 Distribuição Binomial Negativa

Uma questão que naturalmente se segue à respondida pela distribuição geométrica é a do comportamento da variável aleatória *Tempo de espera pelo r-ésimo sucesso*, para $r \in \mathbb{N}$.

Definição 4.5.1 (Distribuição Binomial Negativa)

Suponha que uma seqüência de experimentos independentes de Bernoulli seja realizada. Seja X o número de realizações até a ocorrência do r-ésimo sucesso ($r \in \mathbb{N}$). A distribuição da variável *aleatória X é chamada de Binomial negativa de parâmetros r e p $(X \sim BN(r,p))$.*

A primeira observação sobre uma variável aleatória cuja distribuição de probabilidades é binomial negativa com parâmetros r e p é a de que esta pode ser escrita como a soma de r v.a.'s geométricas independentes de parâmetro p. Para entender esta afirmação, note que, com X, estamos contando o número de experimentos até que o n-ésimo sucesso ocorra (veja (4.14)).

$$\underbrace{\overbrace{\mathcal{F}\mathcal{F}\cdots\mathcal{F}}^{\text{fracassos}}\mathcal{S}}_{Y_1}\quad\underbrace{\overbrace{\mathcal{F}\mathcal{F}\cdots\mathcal{F}}^{\text{fracassos}}\mathcal{S}}_{Y_2}\quad\cdots\quad\underbrace{\overbrace{\mathcal{F}\mathcal{F}\cdots\mathcal{F}}^{\text{fracassos}}\mathcal{S}}_{Y_r} \tag{4.14}$$

A tarefa de *contar* pode ser dividida em r partes: tempo até o primeiro sucesso, Y_1, tempo entre o primeiro e segundo sucessos, Y_2 etc. X pode ser, então definida simplesmente como

$$X = \sum_{i=1}^{r} Y_i.$$

Como a seqüência é formada por experimentos de Bernoulli independentes com probabilidade de sucesso igual a p, todos os tempos descritos têm a mesma distribuição. Além disso, a seqüência de tempos é independente. Finalmente, cada distribuição é geométrica de parâmetro p, como definido na seção 4.4.1.

Proposição 4.5.1 (Função de Massa da Binomial Negativa)

Seja X uma v.a. distribuída como uma $BN(r,p)$. Então,

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \mathbf{1}_{\{r,r+1,\ldots\}}(n). \tag{4.15}$$

Demonstração:

Note que o evento $A = \{$O r-ésimo sucesso ocorre exatamente na n-ésima tentativa$\}$ pode ser escrito como:

$$A = B \cap C,$$

onde $B = \{$há $(r-1)$ sucessos nas $(n-1)$ primeiras tentativas$\}$ e $C = \{$há sucesso na n-ésima tentativa$\}$. B e C são eventos independentes; mais ainda, $\mathbb{P}(B) = \mathbb{P}(X = r - 1)$, onde $X \sim Bin(n-1, p)$ (veja definição 4.3.1) e $\mathbb{P}(C) = p$. Logo, $\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(C)$ e a demonstração está completa. ■

Note que $\mathbb{P}(X = x)$ é não-negativa para todo x real. Logo, basta provar que $\sum_{n=r}^{\infty} \mathbb{P}(X = n) = 1$ para que se tenha uma função de probabilidade fidedigna. Para isto, iremos utilizar a versão alternativa da binomial negativa. Esta, analogamente à versão alternativa da geométrica, conta apenas o número de fracassos até que o r-ésimo sucesso ocorra. Desta forma Y, assim definida, pode ser escrita como $Y = X - r$, onde $X \sim BN(r, p)$.

Proposição 4.5.2 (Massa da Bin. Negativa Alternativa)

Sejam $X \sim BN(r, p)$ e $Y = X - r$. Então,

$$\mathbb{P}(Y = n) = \binom{-r}{n} p^r (p-1)^n \mathbf{1}_{\{0,1,\ldots\}}(n).$$

Demonstração:

O evento $[Y = n]$ é equivalente a $A = \{$O r-ésimo sucesso ocorre exatamente na $(n+r)$-ésima tentativa$\}$ que, por sua vez, pode ser escrito como:

$$A = B \cap C,$$

onde $B = \{$há n fracassos nas $(n+r-1)$ primeiras tentativas$\}$ e $C = \{$há sucesso na $(n+r)$-ésima tentativa$\}$. B e C são eventos independentes; além disso, $\mathbb{P}(B) = \mathbb{P}(X = n)$, onde $X \sim Bin(n+r-1, 1-p)$ (veja definição 4.3.1)) e $\mathbb{P}(C) = p$. $\mathbb{P}(B)$ pode ser calculada como:

$$\mathbb{P}(B) = \binom{n+r-1}{n} p^{r-1} (1-p)^n \mathbf{1}_{\{0,1,\ldots,n+r-1\}}(n).$$

Utilizando-nos da definição generalizada dos coeficientes binomiais,

$$\begin{aligned}
\binom{n+r-1}{n} &= \frac{(n+r-1)(n+r-2)\cdots(n+r-1-n+1)}{n!} \\
&= \frac{(n+r-1)(n+r-2)\cdots r}{n!} = \frac{r(r+1)\cdots(r+n-1)}{n!} \\
&= \frac{(-1)(-r)(-1)(-r-1)\cdots(-1)(-r-n+1)}{n!} \\
&= \binom{-r}{n}(-1)^n
\end{aligned}$$

e o fato de que $n = 0, 1, \ldots$ aplicados sobre $\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(C)$, a fórmula para $\mathbb{P}(Y = n)$ esta completa.

Finalmente, mostrar que $\sum_{n=r}^{\infty} \mathbb{P}(X = n) = 1$ é obviamente equivalente a fazê-lo para $\sum_{n=0}^{\infty} \mathbb{P}(Y = n)$, se $Y = X - r$. Mas

$$\sum_{n=0}^{\infty} \mathbb{P}(Y = n) = \sum_{n=0}^{\infty} \binom{-r}{n} p^r (p-1)^n = p^r \sum_{n=0}^{\infty} \binom{-r}{n} (p-1)^n = 1,$$

sendo a série no penúltimo termo nada mais do que o Binômio de Newton para p^{-r} (veja o exercício 1.13).

Portanto, a definição 4.5.1 fornece uma função de probabilidade fidedigna.

■

Proposição 4.5.3 (Esperança e Variância da Binomial Negativa) [2]

Seja $X \sim BN(r, p)$*:*

$$\begin{aligned}
\mathbb{E}(X) &= \frac{r}{p}; \ e \\
Var(X) &= \frac{r(1-p)}{p^2}.
\end{aligned}$$

[2] O resultado expresso em 4.5.3 não deve surpreender, visto que X pode ser escrito como a soma de r v.a's geométricas de parâmetro p independentes

Demonstração:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=r}^{\infty} n\binom{n-1}{r-1}p^r(1-p)^{n-r} \;=\; \sum_{n=r}^{\infty} r\binom{n}{r}p^r(1-p)^{n-r} \\
&= \frac{r}{p}\sum_{n=r}^{\infty}\binom{n}{r}p^{r+1}(1-p)^{n-r} \;=\; \frac{r}{p}\sum_{m=r+1}^{\infty}\binom{m-1}{s-1}p^s(1-p)^{m-s} \\
&= \frac{r}{p},
\end{aligned}$$

sendo esta última série a probabilidade total de uma distribuição $BN(m,r)$, isto é, 1.

Para calcular a variância de X, é mais fácil utilizar $E(X(X+1))$ e a identidade $VAR(X) = \mathbb{E}(X(X+1)) - \mathbb{E}(X) - (\mathbb{E}(X))^2$:

$$\begin{aligned}
\mathbb{E}(X(X+1)) &= \sum_{n=r}^{\infty} n(n+1)\binom{n-1}{r-1}p^r(1-p)^{n-r} \\
&= \sum_{n=r}^{\infty}(r+1)r\binom{n+1}{r+1}p^r(1-p)^{n-r} \\
&= \frac{(r+1)r}{p^2}\sum_{m=s}^{\infty}\binom{m-1}{s-1}p^s(1-p)^{m-s} \;=\; \frac{(r+1)r}{p^2},
\end{aligned}$$

sendo esta última série a probabilidade total de uma distribuição $BN(m,r)$, isto é, 1, e

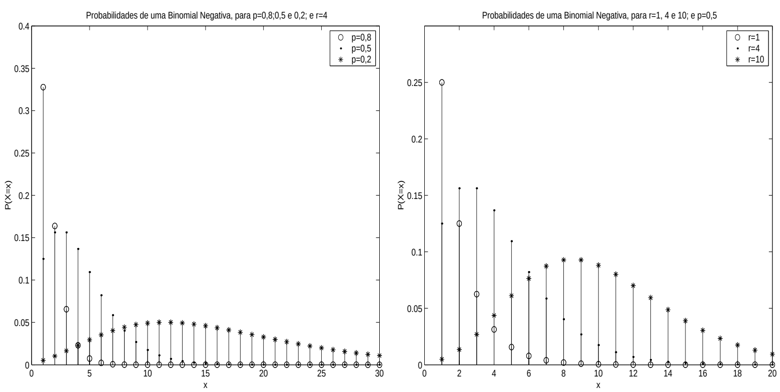
$$Var(X) = \frac{(r+1)r}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} \;=\; \frac{r(1-p)}{p^2}. \blacksquare$$

Exemplo 4.5.1 *Suponha que um dado equilibrado seja lançado até que* 1 *seja a face superior pela quarta vez. Seja* X *o número de lançamentos realizados.*

X *tem distribuição binomial negativa, com parâmetros* $r = 4$ *e* $p = 1/6$*. Portanto, o número esperado de lançamentos é* $\mathbb{E}(X) = 24 (= r/p)$ *e a variância do número de lançamentos,* $Var(X) = 120 (rp/p^2)$*.*

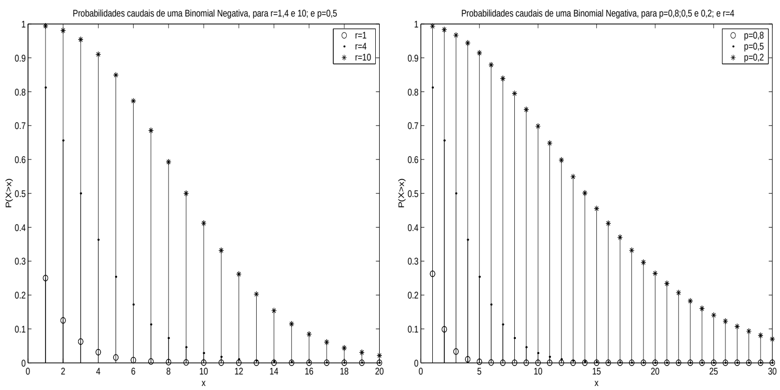
Seria também interessante saber quão provável seria a ocorrência de um número grande de lançamentos. Uma maneira eficiente de apresentação destas probabilidades seria pelos gráficos de $p_X(x)$ vs x, vistos na figura 4.4 e $1 - F(x)$ vs x, vistos na figura 4.5.

Figura 4.4: Função de Massa da Distribuição Binomial Negativa



Note, na figura 4.4, que a distribuição de probabilidades do número de lançamentos muda radicalmente com mudanças nos valores dos parâmetros: quanto maior for r ou menor for p, maior a probabilidade de um número *grande* de lançamentos. Além disso, valores de p muito próximos de 1 causam uma excessiva assimetria pois, intuitivamente, os r sucessos tendem a

Figura 4.5: Probabilidades Caudais da Distribuição Binomial Negativa



acontecer *rapidamente*. Essa noção de *grandeza* de p depende do valor de r da distribuição. Nota-se um ponto de equilíbrio no produto de r por p, abaixo do qual teríamos uma assimetria à direita e acima do qual uma assimetria à esquerda. Note-se também, na figura 4.5, que o decaimento das distribuições dependem muito do valor de r, isto é quanto maior for o número de sucessos, menor o decaimento, representado nos gráficos por um lento decréscimo dos valores das probabilidades caudais.

4.6 Distribuição Hipergeométrica

Em todas as seções deste capítulo, relacionadas ao Experimento de Bernoulli, foram supostas a independência e equivalência dos experimentos da seqüência. Isto queria dizer que as condições de experimento não sofriam qualquer tipo de *influência* dos resultados anteriores.

Uma das vantagens mais imediatas de tal situação é a de que um observador não precisa saber em que momento da seqüência começa a observá-la para obter as condições dos experimentos futuros, isto é, todo experimento da seqüência pode ser considerado o primeiro experimento de alguma seqüência *estocasticamente idêntica* de experimentos de Bernoulli.

No entanto, por mais atraentes que essas condições sejam, elas não deixam de ser restrições à modelagem de fenômenos reais, muito deles extremamente simples. Por exemplo, a extração de números em uma loteria (Sena, Mega-sena, Quina, Bingo) pode ser colocada como uma seqüência finita (número de dezenas sorteadas) de Experimentos de Bernoulli. Ao contrário dos casos anteriores, não há independência entre as extrações.

Como ilustração para esse fato, suponha que João tenha preparado um cartão com dez dezenas para o sorteio da Mega-Sena, $\mathcal{D} = \{d_1, d_2, \dots, d_{10}\}$, escolhidas entre as sessenta possíveis: $\mathcal{E} = \{01, 02, \dots, 60\}$. Para que ele seja um dos vencedores, as seis dezenas sorteadas devem ser elementos do conjunto $\mathcal{D}$. Defina como *sucesso* na i-ésima extração que a i-ésima dezena sorteada seja elemento de $\mathcal{D}$ e *fracasso* seu evento complementar, $\mathcal{E} - \mathcal{D}$. Para ilustrar a dependência entre os experimentos de Bernoulli, utilizar-nos-emos da primeira e segunda extrações: na primeira, a probabilidade de sucesso é $10/60$ ($|\mathcal{D}|/|\mathcal{E}|$). Condicionada a sucesso na primeira extração, a probabilidade de sucesso na segunda é $9/59$ mas, condicionada a fracasso, a probabilidade de sucesso na segunda é $10/59$.

É muito importante que se ressalte que esse estado de dependência não inviabiliza ou restringe a utilização de Experimentos de Bernoulli e contagens associadas. Faz-se apenas necessária a adaptação das técnicas combinatórias do caso independente para o caso dependente, sendo essa análoga à distinção entre potenciação e combinação.

Definição 4.6.1 (Seqüência de Experimentos de Bernoulli Finita)

Seja Ξ uma seqüência de N experimentos de Bernoulli com as seguintes propriedades:

(i) O número de sucessos nos N experimentos é exatamente $0 \leq m \leq N$ e, conseqüentemente, o número de fracassos é exatamente

$N - m$;

(ii) A probabilidade de sucesso no i-ésimo experimento é $(m - s)/(N-i+1)$, onde s é o número de sucessos nos $(i-1)$ primeiros experimentos, $i = 1, 2, \ldots, N$[3].

Exemplo 4.6.1 *Experimentos do tipo definido em 4.6.1 têm extensa utilização em Estatística Aplicada. Umas das áreas de maior importância na indústria moderna é a de controle estatístico de qualidade. Busca-se, por aplicação de metodologias adequadas, uma uniformização da produção. Para isto, é necessário testar a qualidade dos produtos. Há dois empecilhos a testar todos os itens: no caso de testes destrutivos, é óbvia a limitação mas, mesmo com testes não-destrutivos, o custo de mão-de-obra, tempo e materiais seria excessivo. Portanto, uma amostra é fundamental. O objetivo seria obter a menor amostra possível que garantisse níveis mínimos de confiança nos resultados.*

Definição 4.6.2 (Distribuição Hipergeométrica)

Tome uma seqüência de experimentos Ξ como na Definição 4.6.1. Seja X o número de sucessos nos $0 < n \leq N$ primeiros experimentos de Bernoulli. A distribuição de probabilidades de X é chamada de Hipergeométrica com parâmetros N, m e n e dita $H(N, m, n)$.

Proposição 4.6.1 (Massa, Esp. e Var. da Hipergeométrica)

Seja X uma v.a. com distribuição $H(N, m, n)$. Então,

$$(i) \quad \mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

[3]Todos os experimentos após o m-ésimo sucesso ou $(N - m)$-ésimo fracasso são triviais, isto é, as probabilidade de fracasso ou sucesso são respectivamente iguais a 1, em cada caso.

$$(ii)\quad \mathbb{E}(X) = \frac{nm}{N}$$

$$(iii)\quad Var(X) = \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right)$$

Muitas vezes, a probabilidade de que uma v.a. com distribuição $H(N, m, n)$ assuma um valor i é denotada por $h_i(N, m, n)$. Antes de demonstramos as propriedades da hipergeométrica descritas na proposição 4.6.1, vejamos o seguinte exemplo de sua utilização em problemas de controle de qualidade.

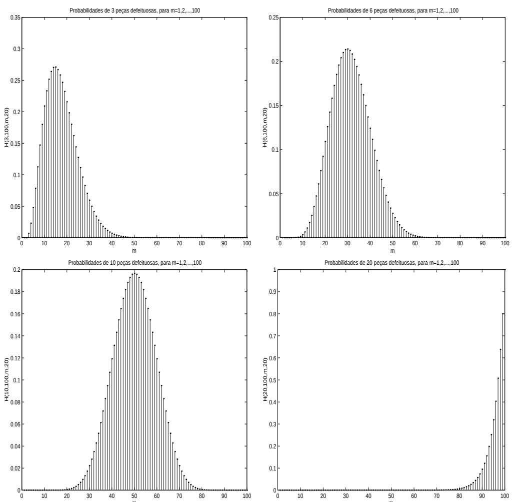
Exemplo 4.6.2 (Hipergeométrica e Controle de Qualidade)

Suponha que uma empresa **A** *produza peças para serem vendidas à empresa* **B**. *A empresa* **B** *deseja fechar negócio com* **A** *se tiver indícios que os lotes de peças que irá receber terão proporção de peças defeituosas não superior a 5%. Sabendo-se que as peças são vendidas em lotes de cem, quantas peças de um lote teriam que ser testadas e qual seria um número de peças defeituosas acima do qual* **B** *deveria rejeitar o negócio?*

O problema apresentado é típico de situações em que testes destrutivos têm que ser utilizados. Por exemplo, suponha que, por razões econômicas, decida-se que vinte peças devam ser testadas. Note que, sendo m o número de peças defeituosas no lote(desconhecido), o número de defeituosas entre as vinte testadas seguiria uma distribuição hipergeométrica com parâmetros $N = 100$, m (parâmetro de interesse, desconhecido) e $n = 20$. Trabalharemos com quatro situações: $k = 3, 6, 10$ ou 20 peças defeituosas são encontradas.

Uma maneira de se inferir sobre a verdadeira proporção, p, de defeituosas no lote, é verificar qual m maximiza $\mathbb{P}(X = k)$, isto é, qual das distribuições teóricas melhor aproxima os dados. Este conceito de proximidade é importantíssimo na Teoria Estatística e *é chamado de* **Método de Máxima Verossimilhança** . *Foram calculadas as probabilidades de* $h_k(100, m, 20)$ *(ou* $H(k, 100, m, 20)$*) para todos os possíveis valores de* m *(*$1, 2, \ldots, 100$*), para cada valor de* k *em* $\{3, 6, 10, 20\}$.

Figura 4.6: Probabilidades da Hipergeométrica



Como pode ser notado na figura 4.6, para cada valor de k*, há um valor de defeituosas,* m*, num lote de* $N = 100$ *peças, cuja distribuição associada dá valor máximo à ocorrência de* k *defeituosas em* $n = 20$ *peças testadas. Especificamente, no caso de haver três defeituosas, o valor de máximo para* $h_3(100, m, 20)$ *é* $h_3(100, 15, 20) = 0,2711$*, ou seja, quinze defeituosas entre as cem peças do lote parece ser a situação mais verossímel. Para seis defeituosas na amostra,* $m = 30$ *nos dá o valor,* $0,2141$*, máximo de* $h_6(100, m, 20)$. *Se a amostra apresentar dez defeituosas, o valor mais verossímel para* m *seria* 50*, cuja probabilidade seria* $0,1969$.

Finalmente, caso todas as peças amostradas fossem defeituosas, isto é, $k = 20$, teríamos como valor mais verossímel $m = 100$, isto é, todas as peças do lote serem defeituosas.

No exemplo 4.6.2, a questão de qual dos possíveis valores de m seria o mais verossímel foi respondida de forma numérica, isto é, calcularam-se todas as probabilidades não-triviais para o número de defeituosas na amostra e escolheu-se entre todos o de maior probabilidade como o indicador da distribuição *geradora* do resultado. Essa resposta, no entanto, pode ser dada de forma analítica. Notemos, antes de encontrar a resposta analiticamente, que, em cada um dos três primeiros gráficos da figura 4.6, isto é para $k = 3$, 6 ou 10 defeituosas numa amostra de tamanho 20, existe um valor até o qual o gráfico se comporta de forma monotonamente crescente e depois do qual o gráfico se comporta de forma monotonamente decrescente. No quarto caso, quando há $k = 20$ defeituosas, isto é, todas as peças amostradas foram defeituosas, a função é monotonamente crescente até seu último valor (que também é seu ponto de máximo), $m = 100$. Em ambos os casos, se tomarmos a fração entre probabilidades consecutivas, $h_k(N, m, n)/h_k(N, m+1, n)$, encontraremos um valor $\hat{m}$, acima do qual essa fração será sempre estritamente maior do que 1 e abaixo do qual estritamente menor do que 1[4][5], sendo $\hat{m}$ o ponto de máximo desejado.

Portanto, dada uma distribuição hipergeométrica, $X \sim H(N, m, n)$ e k ocorrências observadas, temos que, fixados N, n e k, basta-nos descobrir como se comporta a fração $\mathbb{P}(X = k|m)/\mathbb{P}(X = k|m-1)$, onde $\mathbb{P}(X = k|j)$ representa a probabilidade de k ocorrências em uma $H(N, j, n)$.

$$\frac{\mathbb{P}(X = k|m)}{\mathbb{P}(X = k|m+1)} = \frac{h_k(N, m, n)}{h_k(N, m+1, n)}$$

[4]Assume-se aqui que para $m = 1$ ($m = 100$), caso atenda à condição de ter acima (abaixo) as frações estritamente menores (maiores) do que 1, esse valor será $\hat{m}$.

[5]Deve ser utilizada a convenção de que não se calculam as razões quando um dos valores for nulo.

$$
\begin{aligned}
&= \frac{\dfrac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}}{\dfrac{\binom{m+1}{k}\binom{N-(m+1)}{n-k}}{\binom{N}{n}}} \\
&= \frac{m+1-k}{m+1}\frac{N-m}{N-m-(n-k)} \\
&= \frac{(m+1)(N-m-(n-k)-k(N-m)+(m+1)(n-k)}{(m+1)(N-m-(n-k))} \\
&= 1-\frac{k(N-m)-(m+1)(n-k)}{(m+1)(N-m-(n-k))},
\end{aligned}
$$

ou seja,

$$\mathbb{P}(X=k|m+1) \geq \mathbb{P}(X=k|m)$$

se, e somente se,

$$
\begin{aligned}
&m \geq N-(m-k) \\
&\text{e} \\
&1-\frac{k(N-m)-(m+1)(n-k)}{(m+1)(N-m-(n-k))} \leq 1,
\end{aligned}
$$

o que significa que

$$m \leq \frac{1+k(N-1)}{n}.$$

Logo:

$$arg \max_{0\leq m\leq N} \mathbb{P}(X=k|m) = \left\lfloor 1+\frac{k(N-1)}{(n)} \right\rfloor. \tag{4.16}$$

A equação (4.16) nos dá uma maneira sistemática de, dado o tamanho do lote, N, o tamanho amostral, n, e o número de defeituosas, k, encontrar o valor *mais razoável* (segundo o critério de máxima verossimilhança) para o verdadeiro número de defeituosas em todo o lote. Outras interpretações em exemplos variados podem ser feitas a partir de (4.16) mas deixá-las-emos para textos mais especializados, como, por exemplo, [2] e [26]. Resta-nos provar as propriedades da Proposição 4.6.1.

Demonstração:

(i) segue imediatamente dos seguintes fatos:

(a) Há $\binom{N}{n}$ diferentes possíveis retiradas de n elementos de uma população de tamanho N;

(b) O número de amostras favoráveis (isto é, aqueles que têm i sucessos e, conseqüentemente, $n-i$ fracassos) é dado pelo produto da escolha dos i entre os m disponíveis sucessos e dos $n-i$ entre os $N-m$ disponíveis fracassos - $\binom{m}{i}\binom{N-m}{n-i}$; e

(c) A probabilidade pode ser calculada pela razão do número de casos favoráveis pelo de casos possíveis.

(ii) Note que

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=0}^{n} i \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \\
&= \frac{nm}{N} \sum_{i=1}^{n} \frac{\binom{m-1}{i-1}\binom{N-m}{n-i}}{\binom{N-1}{n-1}} \\
&= \frac{nm}{N},
\end{aligned}$$

onde a última soma é 1, pois nada mais é do que a soma das probabilidades de todos os possíveis valores de uma v.a. hipergeométrica de parâmetros $N-1$,

$m-1$ e n. A primeira passagem faz uso da identidade:

$$j\binom{k}{j} = k\binom{k-1}{j-1},$$

com, respectivamente, $j = i$, $k = m$ para $\binom{m}{i}$ e $j = n$, $k = N$ para $\binom{N}{n}$.

Temos também que:

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \sum_{i=0}^{n} i^2 \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \\
&= \frac{nm}{N}\sum_{i=1}^{n} i \frac{\binom{m-1}{i-1}\binom{N-m}{n-i}}{\binom{N-1}{n-1}} \\
&= \frac{nm}{N}\sum_{i=0}^{n-1} (i+1) \frac{\binom{m-1}{i}\binom{N-m}{(n-1)-i}}{\binom{N-1}{n-1}} \\
&= \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1}+1\right),
\end{aligned}$$

onde as passagens são feitas seguindo raciocínios análogos aos feitos para $\mathbb{E}(X)$ e, portanto,

$$\begin{aligned}
Var(X) &= \mathbb{E}\left(X^2\right) - \left(\mathbb{E}(X)\right)^2 \\
&= \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1}+1\right) - \left(\frac{nm}{N}\right)^2
\end{aligned}$$

$$= \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right).$$

■

Um ponto muito interessante sobre a Hipergeométrica é a de que essa pode ser pensada como a *equivalente* à distribuição Binomial quando os números de sucessos e fracassos estão previamente fixados. De maneira inversa, a Binomial pode ser pensada como a Hipergeométrica quando o universo amostral tem cardinalidade infinita. Intuitivamente, como a principal diferença entre as duas situações, a dependência do caso finito, ocorre pela variação dos casos favoráveis quando experimentos são realizados, se tomarmos $N \to \infty$, aquela deveria desaparecer , desde que $m/N \to p \in [0,1]$ e $n/N \to 0$[6].

Voltemos à esperança e variância de uma v.a. hipergeométrica, X. Se tomarmos $p = m/N$, podemos escrever, através de tedioso algebrismo (veja exercício 4.1):

$$\mathbb{E}(X) = np$$

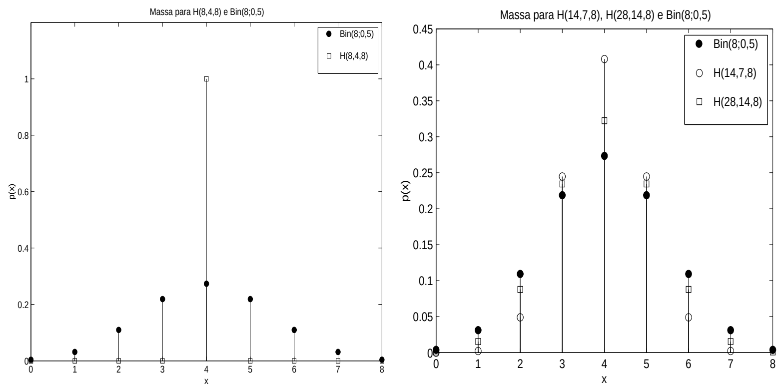
e

$$Var(X) = np(1-p)\frac{N-n}{N-1}.$$

Note que sua esperança é idêntica à de uma v.a. $Bin(n,p)$ e que sua variância difere da dessa por um fator de $(N-n)/(N-1)$ cujo limite, quando as condições assintóticas assinaladas no parágrafo acima forem atendidas, é 1.

[6] A condição de que $m/N \to p \in [0,1]$ nos diz que a proporção de defeituosas se estabiliza em p enquanto a condição de que $n/N \to 0$ nos diz que a amostra tomada deve ter tamanho, n, irrelevante, quando comparado ao tamanho populacional, o que, intuitivamente, mantém condições equivalentes em todas as retiradas, ou seja, torna o problema equivalente ao de n experimentos de Bernoulli.

Figura 4.7: A Hipergeométrica e a Binomial - I



Proposição 4.6.2 (Conv. da Hipergeométrica para a Binomial)

Seja X uma variável aleatória distribuída como uma hipergeométrica com parâmetros $N, m, n \in \mathbb{N}$:

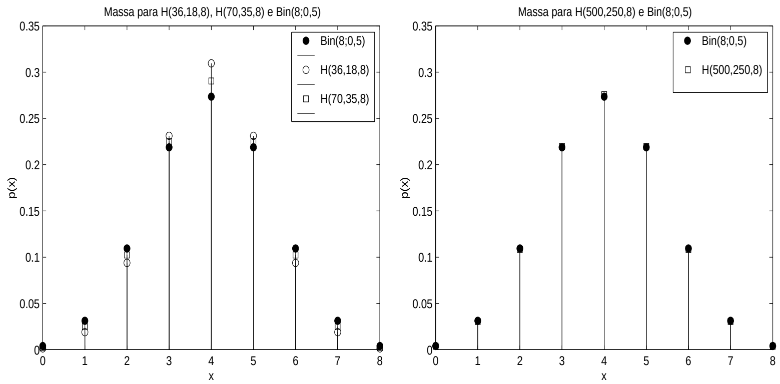
$$\lim_{N,m\to\infty} \frac{\mathbb{P}(X=k)}{\binom{n}{k}p^k(1-p)^{n-k}} = 1, \quad para\ k = 1, 2, \ldots, n, \qquad (4.17)$$

onde $p = lim_{N,m\to\infty} m/N$ e $m/N \to 0$.

Nas figuras 4.7 e 4.8, ilustramos o resultado dado pela proposição 4.6.2 através de algumas comparações entre as funções de massa de uma binomial com parâmetros $n = 8$ e $p = 0,5$ e distribuições hipergeométricas equivalentes, no sentido de que as respectivas proporções de indivíduos da população que têm a característica de interesse é sempre igual a 50%. Especificamente, retrataremos populações de tamanho 8, 14, 28, 36, 70 e 500.

Notemos que, em todos os gráficos, as distribuições hipergeométricas retratadas são simétricas em torno de um mesmo valor, 4 que é exatamente a

Figura 4.8: A Hipergeométrica e a Binomial - II



esperança de qualquer uma dessas distribuições e também a esperança da binomial em questão. Esse fato advém de a esperança de uma hipergeométrica depender somente da proporção de favoráveis na população e do tamanho amostral. Por outro lado, o que se evidencia nessa seqüência de gráficos é que os valores das funções de massa das hipergeométricas se aproximam monotonamente dos respectivos valores da função de massa da binomial, quando o tamanho populacional cresce. Isto se reflete em variâncias mais próximas da variância de uma $Bin(8; 0, 5)$ e na convergência de sua função de massa, como exposto na proposição 4.6.2. Um fato curioso é o do comportamento de $H(8, 4, 8)$: claramente, só há um resultado possível - 4 - e, portanto, estamos diante de uma variável aleatória degenerada, que assume o valor 4 com probabilidade 1.

Antes de demonstrarmos a proposição 4.6.2, tentemos entender as implicações de tal resultado. Em linhas gerais, se tivermos uma população finita e tomarmos uma amostra de tamanho n com interesse em uma particular característica C, presente em m dos N indivíduos, utilizar-nos-íamos de uma $H(N, m, n)$ para calcular as respectivas probabilidades dos números de amostrados em que tal característica estivesse presente. No entanto, se N for suficientemente grande, podemos utilizar uma $Bin(N, m/N)$ para tais cálculos. É importante notar que N *suficientemente grande* simplesmente significa que o tamanho da amostra, n, é pequeno quando comparado a N. Por exemplo, se $N = 500$, poderíamos aproximar X por uma binomial se $n = 8$ (como nas figuras 4.7 e 4.8) mas não se $n = 500$ (pense que isso seria de certa forma análogo a tomar a *primeira tentativa de aproximação* da figura, onde $n = N = 8$) .

Demonstração:

Seja $p = \lim_{N\to\infty} \frac{m}{N}$. Podemos escrever:

$$\begin{aligned}\mathbb{P}(X = i) &= \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} = \frac{m(m-1)\cdots(m-i+1)}{i!} \times \\ &\frac{(N-m)(N-m-1)\cdots(N-m-(n-i)+1}{(n-i)!} \times \\ &\frac{n!}{N(n-1)\cdots(N-n+1)} \asymp \binom{n}{i}\frac{m^i}{N^i}\frac{(N-m)^{n-i}}{N^{n-i}},\end{aligned}$$

para N e m suficientemente grandes (para detalhes, veja o exercício 4.3). Logo, para todo $i = 1, 2, \ldots, n$:

$$\lim_{N,m\to\infty} \mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}. \blacksquare$$

4.7 Distribuição Poisson

Finalizamos a seção anterior, mostrando que o cálculo de probabilidades do número de sucessos no caso de populações finitas pode ser aproximado para uma binomial desde que o tamanho da amostra seja *pequeno* quando comparado com o tamanho da população. Esse resultado é muito importante pois, nos aspectos numéricos, a distribuição hipergeométrica se mostra bem menos interessante do que a binomial. No entanto, para n extremamente grande[7], os cálculos com a binomial se tornam também inviáveis. A questão que se coloca é a da possibilidade de substituirem-se tais cálculos por outros mais simples. A resposta é positiva, sob algumas condições, e a distribuição-limite uma das mais importantes de toda a Teoria de Probabilidade.

Proposição 4.7.1 (Convergência da Binomial para a Poisson)

Seja X uma v.a. que segue uma distribuição $Bin(n,p)$, onde $p = p_n$. Suponha que $\lim_{n\to\infty} np_n = \lambda > 0$. Então, para todo $i = 1, 2, \ldots$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(X=i)}{\frac{\lambda^i \exp -\lambda}{i!}} = 1. \tag{4.18}$$

Definição 4.7.1 (Distribuição Poisson)

Seja X uma variável aleatória com a seguinte função de massa:

$$\mathbb{P}(X=i) = \frac{\lambda^i e^{-\lambda}}{i!}\mathbf{1}_{\mathbb{N}}(i).$$

X é dita ter distribuição Poisson de parâmetro λ, $X \sim Po(\lambda)$.

Demonstração:

$$\mathbb{P}(X=i) = \binom{n}{i} p^i(1-p)^{n-i} = \frac{n!}{i!(n-i)!}p^i(1-p)^{n-i}$$

usando a fórmula de Stirling

[7] Quando se fala em n grande, não há qualquer contradição com as condições para convergência de $h_k(N,m,n)$, ou seja, que n era pequeno comparado com N: em vez de pensarmos nos tamanhos relativos de n e N, aqui, falamos apenas do tamanho absoluto de n.

$$
\begin{aligned}
&\asymp \frac{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}{i!\sqrt{2\pi}(n-i)^{n-i+\frac{1}{2}}e^{-(n-i)}}p^i(1-p)^{n-i} \\
&\asymp \frac{e^{-i}}{i!}\frac{n^{n+\frac{1}{2}}}{(n-i)^{n-i+\frac{1}{2}}}p^i(1-p)^{n-i} \\
&= \frac{e^{-i}}{i!}\frac{n^{n-i+\frac{1}{2}}}{(n-i)^{n-i+\frac{1}{2}}}(np)^i(1-p)^{n-i} \\
&= \frac{e^{-i}}{i!}(np)^i\left(\frac{n-np}{n-i}\right)^{n-i}\sqrt{\frac{n}{n-i}} \\
& \quad np \to \lambda \\
&\asymp \frac{\lambda^i e^{-\lambda}}{i!},
\end{aligned}
$$

pois

$$
\left(\frac{n-np}{n-i}\right)^{n-i} = \left(\frac{n-i-(np-i)}{n-i}\right)^{n-i} = \left(1-\frac{np-i}{n-i}\right)^{n-i} \to e^{-\lambda+i},
$$

quando $n \to \infty$. ■

As conseqüências da proposição 4.7.1 são fantásticas. Suponha, por exemplo, que tenhamos uma cidade com um milhão de habitantes, em um país cuja taxa de suicídio seja de $p = 0,3/100000$ por ano, isto é, $0,3/100000 hab * ano$ e queiramos saber qual a probabilidade de que haja cinco suicídios neste ano. Em vez de lidar com a ridícula $\binom{1000000}{5}$, conseguimos fazê-lo pela aproximação:

$$
\mathbb{P}(X=i) \approx \frac{\lambda^5 exp{-\lambda}}{5!} \cong 0,1008,
$$

onde $\lambda = 1000000 \times p \quad = 3$.

Em geral, podemos utilizar essa aproximação sempre que np for moderado, sendo n grande e p pequeno. O leitor pode estar questionando quão realístico tal resultado é, principalmente pela artificialidade que pode ser atrelada ao fato de que p tem que variar inversamente ao aumento de n. Na realidade, existem inúmeras situações em que esta variação não somente é razoável como a única possível.

Exemplo 4.7.1 (Contagem de bactérias numa Coorte)

Suponha que estejamos interessados no desenvolvimento de bactérias em espaços limitados. Para estudá-lo, monta-se um experimento laboratorial em um ambiente propício e contam-se as bactérias. Qualquer que seja o processo de contagem utilizado, dividir-se-á o plano onde estão as bactérias em regiões, contando-se então as bactérias.

Note que, quanto menor for cada região, menor será a probabilidade de encontrar uma bactéria. Mais ainda, é razoável supor que esta probabilidade seja aproximadamente inversamente proporcional ao número de regiões, n. Portanto, ao analisarmos a variável **Soma do número de bactérias em cada região**, *teremos uma v.a. com distribuição $Bin(n, p)$.*

Finalmente, note que duas das hipóteses do teorema 4.7.1 já podem ser caracterizadas como razoáveis: proporcionalidade inversa de p em relação a n e distribuição binomial.

A característica mais importante, de certa forma a justificativa da proposição 4.7.1, é a de que o número de bactérias numa coorte independe de n e isso se expressa tanto por p ser inversamente proporcional a n quanto por a soma do número de bactérias convergir[8] para algum valor (ou v.a.).

A proposição 4.7.1 pode ser utilizada como justificativa para a existência de uma distribuição como a Poisson. No entanto, devemos alertar o leitor

[8] Quando se fala em convergência nesse problema, referimo-nos ao fato de que o número de somandos (mesmo que muitos deles sejam zeros) cresce com o número de parcelas e, portanto, no caso geral, não necessariamente, todas as maneiras de somar dar-nos-ão o mesmo resultado. Aqui, as condições especiais de p e n nos garantem que existe o limite.

que essa motivação nada mais é do que um caso particular da definição de uma família de processos estocásticos extremamente rica.

A idéia básica por trás da definição 4.7.2 é a de que estejamos estudando uma estrutura similar à dos experimentos de Bernoulli. Pensemos no processo de Poisson da seguinte forma: temos um evento de interesse, E, cujas características desejamos estudar.

Consideremos que não possamos controlar o processo de realização dos experimentos de Bernoulli, isto é, sabemos que algo está acontecendo mas, agora, ao contrário do que vinha sendo suposto neste capítulo, não somos os *Senhores do Tempo*, os *donos dos dados*. A primeira conseqüência é que contar experimentos ficaria muito mais *difícil*, já que não os realizamos nós mesmos.

O que se faz nesse caso é assumir que, por sua própria impossibilidade prática, em vez de contarmos quantos experimentos falharam, preocuparnos-emos apenas com os bem sucedidos, ou seja, estaremos contando quantas vezes E ocorre em um determinado período de tempo. Para ter uma idéia de tempo (que nos era, automaticamente, fornecida pela contagem dos experimentos de Bernoulli realizados), podemos utilizar o próprio tempo *real*.

O leitor atento já deve ter notado que o processo de Poisson parece destinar menos restrições à estrutura dos experimentos. O que se faz, na realidade, é trocar-se a estrutura da dicotomia **realização vs não-realização** pela estrutura de medição dos tempos de ocorrência (quando houver).

Definição 4.7.2 (Processo de Poisson)

Seja $N(t)$ o número de ocorrências de um evento E no intervalo $[0,t]$. $\{N(t)\}_{t>0}$ é dito um processo de Poisson se as seguintes condições foram satisfeitas:

(1) $P\,(E$ *ocorra em um intervalo de tempo de tamanho* $h) = \lambda h + o(h)$*;*

(2) $P\,($*duas ou mais ocorrências de* E *em um intervalo de tempo* $h) = o(h)$*; e*

(3) $\forall t \in \mathbb{R}^+$, $\forall n, j_1, j_2, \ldots, j_n \in \mathbb{Z}$, $\{I_k\}_{k=1}^{n}$ *qualquer partição de* $[0,t]$, *sendo* E_k *o evento exatamente* j_k *dos eventos ocorrem em* I_k, $k = 1$, $E_1, E_2, \ldots, E_n$ *formam uma família de eventos independentes.*

Informalmente, a definição 4.7.2 nos traz as seguintes restrições:

1. a probabilidade de ocorrência de um evento em um tempo *pequeno* é proporcional ao tamanho do intervalo;
2. não são *permitidas* ocorrências simultâneas; e
3. eventos em intervalos disjuntos são independentes.

A proposição 4.7.2 nos diz que as restrições impostas na definição 4.7.2 nos levam a que, para cada $t > 0$, $N(t)$ tenha distribuição Poisson com esperança λt.

Proposição 4.7.2 (Distribuição de um Processo de Poisson)

Seja $\{N(t)\}_{t>0}$ um processo definido por 4.7.2. Então, para cada $t > 0$, $N(t)$ tem distribuição Poisson com média λt, isto é, para cada $t > 0$,

$$P\left(N(t) = k\right) = \frac{exp\{-\lambda t\}\left(\lambda t\right)^k}{k!}\mathbf{1}_{\{0,1,2,\ldots,\}}(k).$$

Demonstração:

Seja $\{I_1, I_2, \ldots, I_n\}$ uma partição intervalar de $[0, t]$, isto é, $I_1 = [0, i_1]$, $I_2 = (i_1, i_2]$, …, $I_n = (i_{n-1}, t]$, onde $0 < i_1 < i_2 < \cdots < i_{n-1} < t$. Seja $N_n(t)$ a soma do número de ocorrências de E em cada intervalo I_1, I_2, … , I_n. Note que, qualquer que seja $\{I_1, I_2, \ldots, I_n\}$, o número de ocorrências TEM QUE SER O MESMO, isto é, $N(t)$! Mesmo assim, alguns detalhes técnicos do intercâmbio das probrabilidades, que são importantes, serão omitidos.

Definam-se os seguintes eventos:

$$\begin{aligned} A_n &= \{k \text{ intervalos contenham uma ocorrência de } E \text{ e} \\ &\qquad n-k \text{ intervalos contenham 0 ocorrências}\} \end{aligned}$$

e

$$\begin{aligned} B_n &= \{N(t) = k \text{ e pelo menos um intervalo contenha duas ou} \\ &\quad \text{mais ocorrências de } E\}. \end{aligned}$$

Seja $B^\star$ o seguinte evento:

$$B^\star = \lim_{n\to\infty} \{\text{pelo menos um intervalo contém duas ou mais ocorrências de } E\}.$$

Sejam A e B os respectivos limites de $\{A_n\}_{n\geq 1}$ e $\{B_n\}_{n\geq 1}$. Claramente, $\mathbb{P}(B) \leq P(B^\star)$ e

$$\begin{aligned} P(B^\star) &= P\left(\cup_{i=1}^n \{\, i\text{-ésimo intervalo contém duas ou mais ocorrências}\}\right) \\ &\leq \sum_{i=1}^n P\left(i\text{-ésimo intervalo contém duas ou mais ocorrências}\right) \\ &= \sum_{i=1}^n o\left(\frac{t}{n}\right) \quad = \quad no\left(\frac{t}{n}\right) \\ &= t\frac{o\left(t/n\right)}{t/n} \quad \to 0, \end{aligned}$$

quando $n \to \infty$. Por outro lado,

$$\begin{aligned} P\left(0 \text{ ocorrências em um intervalo de tamanho } h\right) &= 1 - [\lambda h + o(h) + o(h)] \\ &= 1 - \lambda h + o(h) \end{aligned}$$

e, conseqüentemente,

$$\begin{aligned} \mathbb{P}(A_n) &= \binom{n}{k}\left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right)\right]^k \times \\ &\quad \left(\left[1 - \frac{\lambda t}{n}\right] + o\left[\frac{t}{n}\right]\right)^{n-k} \end{aligned}$$

Como

$$\begin{aligned} n\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right] &= \lambda t + t\left[\frac{o(t/n)}{t/n}\right] \\ &\to \lambda t, \quad \text{quando } n \to \infty \end{aligned}$$

e, quando $a_n \to a$

$$(1-a_n)^{n-k} \to exp\{-a\},$$

$$\mathbb{P}(A) = \frac{\exp\{-\lambda t\}(\lambda t)^k}{k!}\mathbf{1}_{\{0,1,2\ldots\}}(k)$$

e, finalmente,

$$\begin{aligned} P(N(t)=k) &= \mathbb{P}(A)+\mathbb{P}(B) \\ &= \frac{\exp\{-\lambda t\}(\lambda t)^k}{k!}\mathbf{1}_{\{0,1,2\ldots\}}(k), \end{aligned}$$

pelo argumento de simetria de contagem, isto é, que o processo de contagem não interfere no número de ocorrências. ■

4.8 Função Geratriz de Probabilidades

Existem funções fundamentais para caraterizarem-se distribuições de variáveis aleatórias, além de eficientes em problemas de convergência, como no do teorema central do limite e lei de grandes números: elas são coerentemente chamadas de funções características.

No caso específico de variáveis aleatórias discretas, existe uma função caracterizadora de natureza simples, chamada de função geratriz de probabilidades. Iremos definir esta função, a partir de agora chamada de fgp, para variáveis aleatórias inteiras não-negativas. No entanto, isto não se caracteriza como uma restrição de seu uso. Podemos definir a fgp para qualquer variável aleatória discreta, desde que possua mínimo.

Definição 4.8.1 (Função Geratriz de Probabilidades)

Seja X uma variável aleatória inteira não-negativa. A função geratriz de probabilidades, associada a X, fgp de X, é dada por:

$$\phi_X(t) = \sum_{x=0}^{+\infty} \mathbb{P}(X = x)t^x, \tag{4.19}$$

para os valores de $t \in \mathbb{R}$ para os quais a série seja finita.

Exemplo 4.8.1 (Distribuição Bernoulli)

Seja $X_1 \sim b(p)$. Então, sua fgp pode ser escrita como:

$$\phi_{X_1}(t) = (1-p) + pt,$$

$\forall t \in \mathbb{R}$.

Exemplo 4.8.2 (Distribuição Uniforme Discreta)

X_2 é dita uma variável aleatória discreta em $\{1, 2, \ldots, n\}$, $X_2 \sim U_n\{1, 2, \ldots, n\}$, se

$$\mathbb{P}(X = x) = \frac{1}{n}\mathbf{1}_{\{1,2,\ldots,n\}}(x). \tag{4.20}$$

Sua fgp é dada por:

$$\begin{aligned} \phi_{X_2}(t) &= \sum_{x=1}^{n} \frac{1}{n} t^x = \frac{1}{n}\frac{t - t^{n+1}}{1-t} \\ &= \frac{t}{n}\frac{1-t^n}{1-t}, \end{aligned}$$

para $|t| < 1$.

Exemplo 4.8.3 (Distribuição Geométrica)

Seja $X_3 \sim G(p)$. *Sua fgp pode ser escrita como:*

$$\begin{aligned}
\phi_{X_3}(t) &= \sum_{x=1}^{+\infty} (1-p)^{x-1} pt^x = \frac{p}{1-p} \sum_{x=1}^{+\infty} [(1-p)\, t]^x \\
&= \frac{p}{1-p} \frac{(1-p)\, t}{1-(1-p)\, t} \\
&= \frac{pt}{1-(1-p)\, t},
\end{aligned}$$

para $|t| < (1-p)^{-1}$.

Exemplo 4.8.4 (Distribuição Binomial)

Seja $X_4 \sim Bin(n, p)$. *Sua fgp é dada por:*

$$\begin{aligned}
\phi_{X_4}(t) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} t^x \\
&= \sum_{x=0}^{n} \binom{n}{x} (pt)^x (1-p)^{n-x} \\
&= [(1-p) + pt]^n,
\end{aligned}$$

$\forall t \in \mathbb{R}$.

Exemplo 4.8.5 (Distribuição Poisson)

Seja $X_5 \sim Po(\lambda)$. *Sua fgp é dada por:*

$$\begin{aligned}
\phi_{X_5}(t) &= \sum_{x=0}^{+\infty} \frac{\lambda^x e^{-\lambda}}{x!} t^x = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{(\lambda t)^x}{x!} = e^{-\lambda} e^{\lambda t} \\
&= e^{\lambda(t-1)},
\end{aligned}$$

$\forall t \in \mathbb{R}$.

Exemplo 4.8.6 (Binomial Negativa)

Seja $X_6 \sim BN(r,p)$. *Sua fgp é dada por:*

$$\begin{aligned}
\phi_{X_6}(t) &= \sum_{x=r}^{+\infty} \binom{x-1}{r-1} p^x (1-p)^{x-r} t^x \\
&= \sum_{x=0}^{+\infty} \binom{-r}{x} p^r (1-p)^x (-1)^x t^{x+r} \quad = \quad (pt)^r \sum_{x=0}^{+\infty} \binom{-r}{x} [-t(1-p)]^x \\
&= \frac{(pt)^r}{[1-t(1-p)]^r},
\end{aligned}$$

para $|t| < (1-p)^{-1}$.

Um fato fundamental, que será utilizado na proposição 4.8.2, comum a todas as distribuições anteriores é a de que a função geratriz de probabilidade é sempre bem definida numa vizinhança de 0.

O fato de classificarmos as funções geratrizes de probabilidades como *funções caracterizadoras* é justificado pela seguinte proposição.

Proposição 4.8.1 (Função de Massa e fgp)

Sejam X *e* Y *variáveis aleatórias inteiras não-negativas, definidas no mesmo espaço de probabilidade. Então,* X *e* Y *terão a mesma distribuição de probabilidade, isto é,*

$$\mathbb{P}(X = z) = \mathbb{P}(Y = z), \quad \forall z \in \mathbb{Z}^+,$$

se, e somente se,

$$\phi_X(t) = \phi_Y(t),$$

para todo t *onde cada uma das fgp's estiver definida .*

Demonstração:

Note que:

$$\phi_X(t) = \sum_{x=0}^{+\infty} a_x t^x$$

e

$$\phi_Y(t) = \sum_{x=0}^{+\infty} b_x t^x,$$

onde $a_x = \mathbb{P}(X = x)$ e $b_x = \mathbb{P}(Y = x)$, $x = 0, 1, 2, \ldots$.

Note ainda que $\phi_X = \phi_Y$ se, e somente se, $\mathbb{P}(X = z) = a_z = b_z = \mathbb{P}(Y = z)$, $z = 0, 1, 2 \ldots$, isto é, se, e somente se, X e Y tem distribuições de probabilidades iguais.[9] ■

A proposição 4.8.1 nos mostra que, para variáveis aleatórias discretas, basta-nos conhecer as respectiva fgp's para que possamos inferir sobre equidistribuição. No entanto, mais ainda pode ser feito com as funções geratrizes de probabilidade. Como ver-se-á, a seguir, na proposição 4.8.2, existe um algoritmo relativamente simples para que recuperemos qualquer probabilidade a partir de sua fgp (sendo, esse, inclusive a justificativa para o termo *geratriz de probabilidades*).

Proposição 4.8.2 (Cálculo das probabilidades baseado na fgp)

Seja X uma variável aleatória inteira não-negativa, com fgp $\phi_X(\cdot)$, definida em uma vizinhança de 0. Então, as probabilidades de X podem ser recuperadas da seguinte forma:

$$\mathbb{P}(X = k) = \frac{1}{k!}\, \phi_X^{(k)}(t)\Big|_{t=0} \mathbf{1}_{\{0,1,2,\ldots\}}(k),$$

onde $\phi_X^{(k)}$ representa a k-ésima derivada de ϕ_X, para $k = 1, 2, \ldots$ e $\phi_X^{(0)} = \phi_X$.

[9] Há certos detalhes técnicos, quando a cardinalidade é infinita, que fogem ao nível geral do texto e serão omitidos. No entanto, deve estar clara ao leitor essa associação entre as fgp's e as distribuições, de forma completa, para o caso de cardinalidade finita.

Demonstração:

Note que ϕ_X pode ser escrita como:

$$\begin{aligned} \phi_X(t) &= \sum_{x=0}^{+\infty} \mathbb{P}(X=x)t^x \\ &= \mathbb{P}(X=0) + t\sum_{x=1}^{+\infty} \mathbb{P}(X=x)t^{x-1}, \end{aligned}$$

sendo claro que $\phi_X^{(0)}(t)\Big|_{t=0} = \mathbb{P}(X=0)$.

Podemos também escrever que:

$$\phi_X(t) = \mathbb{P}(X=0) + t\mathbb{P}(X=1) + t^2\mathbb{P}(X=2) + \ldots$$

e, portanto, podemos deduzir as seguintes expressões para as derivadas de ϕ_X:

$$\begin{aligned} \phi_X^{(1)}(t) &= \mathbb{P}(X=1) + 2t\mathbb{P}(X=2) + 3t^2\mathbb{P}(X=3) + \ldots \\ \phi_X^{(2)}(t) &= 2\mathbb{P}(X=2) + 3\times 2t\mathbb{P}(X=3) + 4\times 3t^2\mathbb{P}(X=4) + \ldots \\ &\ldots \\ \phi_X^{(k)}(t) &= k!\mathbb{P}(X=k) + (k+1)k\ldots 2t\mathbb{P}(X=k+1) + \\ &+ (k+2)(k+1)k\ldots 3t^2\mathbb{P}(X=k+2) + \ldots \\ &\ldots \end{aligned}$$

Podemos, portanto, escrever

$$\phi_X^{(k)}(t)\Big|_{t=0} = k!\ \mathbb{P}(X=k)\mathbf{1}_{\{0,1,2,\ldots\}}(k).$$

■

Como exemplo operacional da proposição 4.8.2, vejamos

Exemplo 4.8.7 *Seja X uma v.a. cuja fgp tenha a seguinte expressão:*

$$\phi_X(t) = \exp\left(\lambda\left(t-1\right)\right).$$

Note que:

$$\begin{aligned}
\phi_X^{(1)}(t) &= \lambda \exp(\lambda(t-1)) \\
\phi_X^{(2)}(t) &= \lambda^2 \exp(\lambda(t-1)) \\
&\cdots \\
\phi_X^{(k)}(t) &= \lambda^k \exp(\lambda(t-1)) \\
&\cdots
\end{aligned}$$

Conseqüentemente:

$$\mathbb{P}(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}\mathbf{1}_{\{0,1,2,\ldots\}}(k),$$

isto é, X tem distribuição Poisson com parâmetro λ, confirmando-se, assim, os cálculos realizados no exemplo 4.8.5.

Por mais interessante que seja o resultado exposto na proposição 4.8.2, é importante ressaltar seu impacto prático pois permite-nos *guardar* probabilidades de uma forma *compacta* (fgp), recuperando-as de acordo com a necessidade. O exemplo a seguir nos mostra como esta recuperação é trivial quando a fgp tem forma polinomial.

Exemplo 4.8.8 (Fs. Geratrizes de Probabilidades Polinomiais)

Suponha que X tenha fgp da forma:

$$\phi_X(t) = c\left[14t^5 + 7t^3 + 3t\right].$$

A proposição 4.8.2 consegue recupar as probabilidades de X e definir o valor de c que a torna uma distribuição de probabilidades:

$$\mathbb{P}(X=1) \quad = \quad 3c$$

$$\begin{aligned}\mathbb{P}(X=3) &= \frac{7c}{3!}\\ \mathbb{P}(X=5) &= \frac{14c}{5!}.\end{aligned}$$

Sabemos que $c \times (3+7+14) = 1$. *Temos finalmente que* $c = 1/24$ *e*

$$\begin{aligned}\mathbb{P}(X=1) &= \frac{7}{12}\\ \mathbb{P}(X=3) &= \frac{7}{24}\\ \mathbb{P}(X=5) &= \frac{1}{8}.\end{aligned}$$

Esta metodologia funciona para qualquer caso em que ϕ *possa ser escrita na forma polinomial. Isto não nos deve ser surpreendente pois, se conseguimos escrever* ϕ *como um polinômio (ou uma série de potências), os coeficientes são exatamente as probabilidades das respectivas potências.*

Uma outra utilidade das fgp's é dada pela seguinte proposição:

Proposição 4.8.3 (Cálculo dos Momentos baseado na fgp)

Seja X *uma variável aleatória inteira não-negativa, com fgp* $\phi_X(\cdot)$, *definida em uma vizinhança de* 1. *Então, os momentos fatoriais de* X *podem ser recuperados da seguinte forma:*

$$\mathbb{E}(X_{(k)}) = \phi_X^{(k)}(t)\Big|_{t=1} \mathbf{1}_{\{0,1,2,\ldots\}}(k),$$

onde $\phi_X^{(k)}$ *representa a* k*-ésima derivada de* ϕ_X, *para* $k = 1, 2, \ldots$ *e* $\phi_X^{(0)} = \phi_X$.

Demonstração:
Basta-nos notar que:

$$\begin{aligned}
\phi_X^{(1)}(t) &= \mathbb{P}(X=1) + 2t\mathbb{P}(X=2) + 3t^2\mathbb{P}(X=3) + \dots \\
\phi_X^{(2)}(t) &= 2\mathbb{P}(X=2) + 3\times 2t\mathbb{P}(X=3) + 4\times 3t^2\mathbb{P}(X=4) + \dots \\
&\dots \\
\phi_X^{(k)}(t) &= k!\mathbb{P}(X=k) + (k+1)k\dots 2t\mathbb{P}(X=k+1) + \\
&+ (k+2)(k+1)k\dots 3t^2\mathbb{P}(X=k+2) + \dots \\
&\dots
\end{aligned}$$

e que, portanto,

$$\mathbb{E}(X_{(k)}) = \phi_X^{(k)}(t)\Big|_{t=1}.$$

Os detalhes são deixados para o leitor, no exercício 4.19. ■

Podemos, baseados em todos esses exemplos e resultados expostos, tentar explicar parcialmente a relevância do estudo de funções geratrizes de probabilidades. O leitor já deve ter notado que a fgp é uma forma compacta que guarda todas as informações sobre a distribuição de uma determinada variável aleatória.

No entanto, sob o ponto de vista matemático, isto, mesmo que surpreendente e atraente, não é razão suficiente para a utilização de funções geratrizes. Por quê? Simplesmente porque tal utilização seria limitada e estéril, no sentido de que, uma vez calculada a função geratriz (a partir das probabilidades), podemos apenas calcular as probabilidades (que já eram conhecidas).

A pergunta se mantém: qual a justificativa para o estudo de funções geratrizes? Esse se justifica somente se for possível operar no âmbito das funções e se essas operações forem mais simples do que as análogas realizadas de maneira usual. O que se quer dizer com *operações realizadas no âmbito das funções geratrizes*?

Esta metodologia se justifica se conseguimos fazer transformações de variáveis, de forma que conheçamos a(s) fgp(s) original(is) e utilizemo-nas para encontrar a fgp resultante e, a partir dessa, calcular as probabilidades. No caso de variáveis discretas, transformações são facilmente interpretadas mas, em geral, as trocas de índice nos somatórios se mostram extremamente tediosas, havendo piora sensível com o aumento da dimensão.

No caso das fgp's, dois grupos de aplicações, extremamente úteis, têm soluções elegantes e relativamente fáceis.

Proposição 4.8.4 (Soma de Variáveis Aleatórias Independentes)

Sejam X e Y variáveis aleatórias inteiras não-negativas independentes. Então:

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Demonstração:

$$\begin{aligned}
\phi_{X+Y}(t) &= \sum_{z=0}^{+\infty} \mathbb{P}\left(X+Y=z\right) t^z \\
&= \sum_{z=0}^{+\infty} \mathbb{P}\left(\bigcup_{x=0}^{z} \{X=x, Y=z-x\}\right) t^z \\
&= \sum_{z=0}^{+\infty} \sum_{x=0}^{z} \mathbb{P}\left(X=x\right) P\left(Y=z-x\right) t^z \\
&\ \text{fazendo-se } y = z - x \text{ e invertendo-se a ordem} \\
&= \sum_{x=0}^{+\infty} \sum_{y=0}^{\infty} \mathbb{P}\left(X=x\right) P\left(Y=y\right) t^x t^y \\
&= \phi_X(t)\phi_Y(t).
\end{aligned}$$

■

A proposição 4.8.4 nos diz que basta conhecer as distribuições de probabilidade de duas variáveis aleatórias que ser-nos-á fácil achar a distribuição de sua soma, pela fgp.

Exemplo 4.8.9 *Suponha que X e Y sejam v.a.'s independentemente distribuídas como $b(p)$. Qual a distribuição de $X+Y$? A proposição 4.8.4 nos diz que $X+Y$ terá fgp:*

$$\begin{aligned}\phi_{X+Y}(t) &= (\phi_X(t))^2 \\ &= [(1-p)+pt]^2,\end{aligned}$$

que é exatamente, pelo exemplo 4.8.4, a fgp de uma $Bin(2,p)$, ou seja, como já sabíamos, a soma de variáveis independentes Bernoulli com parâmetros iguais tem distribuição binomial de mesmo parâmetro.

Como o leitor já deve estar imaginando, a proposição 4.8.4 pode ser facilmente estendida para n variáveis independentes, por indução

Proposição 4.8.5 (Fgp de Soma de V.a.'s Independentes)

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes, com respectivas fgp's $\phi_{X_1}, \phi_{X_2}, \ldots, \phi_{X_n}$. Defina $Y = X_1 + X_2 + \ldots + X_n$. Então:

$$\phi_Y(t) = \prod_{k=1}^{n} \phi_{X_k}(t).$$

Exemplo 4.8.10 (Soma de Bernoulli's)

Seja $X_1, X_2, \ldots, X_n$ um conjunto de variáveis aleatórias independentes identicamente distribuídas $b(p)$. Então, $Y = X_1 + X_2 + \ldots + X_n$ tem fgp

$$\begin{aligned}\phi_Y(t) &= (\phi_{X_1}(t))^n \\ &= [(1-p)+pt]^n,\end{aligned}$$

ou seja, Y tem, como já conhecido, distribuição $Bin(n,p)$.

Exemplo 4.8.11 (Soma de Geométricas)

Seja $X_1, X_2, \ldots, X_r$ um conjunto de variáveis aleatórias independentes identicamente distribuídas $G(p)$. Então, $Y = X_1 + X_2 + \ldots + X_r$ tem fgp

$$\begin{aligned} \phi_Y(t) &= (\phi_{X_1}(t))^r \\ &= \left(\frac{pt}{1-(1-p)\,t}\right)^r, \end{aligned}$$

ou seja, Y tem, como já conhecido, distribuição $BN(r,p)$.

A proposição seguinte nos mostra um resultado ainda mais interessante do que o obtido na proposição 4.8.5 (este é na realidade um caso particular daquele). Em vez de supormos somar n variáveis independentes, estamos somando N v.a's independentes, sendo N também uma variável aleatória.

Proposição 4.8.6 (Soma de N Variáveis Independentes)

Sejam N, $X_1, X_2, \ldots, X_n, \ldots$ variáveis aleatórias inteiras não-negativas. Suponha que N seja independente de $X_1, X_2, \ldots$ e que $X_1, X_2, \ldots$ seja uma seqüência de v.a.'s independentes e identicamente distribuídas, condicionalmente a N[10]. Defina a seguinte variável aleatória:

$$S_N = X_1 + X_2 + \ldots + X_N.$$

A fgp de S_N pode ser escrita como:

$$\phi_{S_N}(t) = \phi_N\left(\phi_{X_1}(t)\right), \tag{4.21}$$

onde ϕ_N e ϕ_{X_1} são as respectivas fgp's de N e X_1.

Antes de mais nada, vejamos qual a relevância de definir-se uma v.a. como S_N, na proposição 4.8.6. O leitor pode estar se perguntando qual a razão para nos preocuparmos com N aleatório. Vamos agora dar dois exemplos em que é clara a relevância de tal hipótese.

[10]A única utilização de N é na escolha do número de somandos

Exemplo 4.8.12 *Suponha que um fonte emissora A de partículas radiativas as libere de acordo com um Processo de Poisson de intensidade λ. Cada partícula emitida em A será absorvida em B com probabilidade p, de forma independente das demais. Qual a distribuição de probabilidade da absorção de partículas em A no intervalo $[0, T]$?*

Denominaremos o processo de Poisson por $\{N_t\}_{t\geq 0}$. Sabemos que $N_T \sim P(\lambda T)$. Seja Y o número de partículas absorvidas em A, no período $[0, T]$:

$$Y = X_1 + X_2 + \ldots + X_{N_T}.$$

Sabemos também que $X_1, X_2, \ldots, X_{N_T}$ são v.a's iid's $b(p)$ condicionalmente a N_T. Portanto,

$$\begin{aligned} \phi_{N_T}(s) &= \exp(\lambda T(s-1)) \\ \phi_{X_1}(s) &= (1-p) + ps. \end{aligned}$$

A fgp de Y pode ser escrita como:

$$\begin{aligned} \phi_Y(s) &= \phi_{N_T}(\phi_{X_1}(s)) \\ &= \exp(\lambda T(\phi_{X_1}(s) - 1)) \\ &= \exp(\lambda T((1-p) + ps - 1)) \\ &= \exp(\lambda T(p(s-1))) \\ &= \exp(\lambda T p(s-1)), \end{aligned}$$

ou seja, Y tem distribuição Poisson com parâmetro $\lambda T p$. Isso é um resultado muito interessante. O leitor deve atentar para o detalhe que se está realizando um processo de duas etapas: emissão e absorção, em seqüência. O parâmetro da Poisson mede *a tendência média (ou o nível médio) da contagem. É natural, portanto, que o nível médio de absorção seja o nível médio de emissão corrigido pela probabilidade de absorção, visto que supõe-se que os dois processos sejam independentes.*

Exemplo 4.8.13 *Suponha que cinqüenta pessoas formem um grupo de jogo que funciona da seguinte forma. A cada mês, joga-se uma moeda e, se der cara, cada participante joga uma outra moeda e é considerado um vencedor aquele que tirar cara novamente. Qual a distribuição do número de vencedores neste mês?*

Demonstração:

Antes de mais nada, escreveremos expressões para $\mathbb{P}(S_N = x)$ adequadas à obtenção de (4.21).

$$\begin{aligned}
\mathbb{P}(S_N = x) &= \sum_{n=0}^{+\infty} \mathbb{P}\left(S_N = x, N = n\right) = \sum_{n=0}^{+\infty} \mathbb{P}\left(S_n = x, N = n\right) \\
&= \sum_{n=0}^{+\infty} \mathbb{P}\left(N = n\right) \mathbb{P}\left(S_n = x\right)
\end{aligned}$$

Segue, portanto, que:

$$\begin{aligned}
\phi_{S_N}(t) &= \sum_{x=0}^{+\infty} t^x \mathbb{P}\left(S_N = x\right) = \sum_{x=0}^{+\infty} t^x \sum_{n=0}^{+\infty} \mathbb{P}\left(N = n\right) \mathbb{P}\left(S_n = x\right) \\
&= \sum_{n=0}^{+\infty} \mathbb{P}\left(N = n\right) \sum_{x=0}^{+\infty} t^x \mathbb{P}\left(S_n = x\right) = \sum_{n=0}^{+\infty} \mathbb{P}\left(N = n\right) \phi_{S_n}(t) \\
&= \sum_{n=0}^{+\infty} \mathbb{P}\left(N = n\right) \left(\phi_{X_1}(t)\right)^n \\
&= \phi_N\left(\phi_{X_1}(t)\right). \qquad \blacksquare
\end{aligned}$$

Existem formas análogas às apresentadas nesta seção para conjuntos de índices discretos mas em que cada variável por aquela indexada tem distribuição contínua. Esta distribuição poderá ter distribuição contínua ou mista mas, em espírito, tem propriedades semelhantes às das v.a's aqui apresentadas.

4.9 Exercícios

Exercício 4.1 *Seja $X \sim H(N, m, n)$. Mostre que:*

$$\mathbb{E}(X) = np$$

e

$$Var(X) = np(1-p)\frac{N-n}{N-1},$$

onde p é a proporção de defeituosas m/N.

Exercício 4.2 *Defina, com suas próprias palavras, um experimento de Bernoulli (e, também, uma seqüência de experimentos de Bernoulli). Explique a importância de cada uma das hipóteses tomadas e quais as questões relevantes que podem ser respondidas através desse experimento.*

Exercício 4.3 *Complete os detalhes da demonstração da proposição 4.6.2.*

Exercício 4.4 *Seja Π um experimento formado por uma seqüência de experimentos independentes de Bernoulli. Suponha que a probabilidade de sucesso em cada um dos experimentos de Bernoulli seja p, $0 \leq p \leq 1$. Defina variáveis X, Y e W como segue:*

> *X assume valores 0 e 1 se houver, respectivamente, fracasso ou sucesso no primeiro experimento de Bernoulli;*
>
> *Y representa o número de sucessos nos n primeiros experimentos de Bernoulli realizados; e*

W é o número de experimentos necessários para que o primeiro sucesso ocorra.

Ache as distribuições de probabilidade de cada uma das v.a.'s acima definidas. Ache também as respectivas médias, modas (caso existam), medianas e variâncias.

Exercício 4.5 *Assuma as mesmas condições do experimento Π definido no exercício 4.4. Seja X o máximo entre as duas seguintes quantidades: número de sucessos nos n primeiros experimentos de Bernoulli e o número de sucessos nos n experimentos entre o $(n+1)$-ésimo e o $(2n)$-ésimo experimento de Bernoulli. Ache a distribuição de probabilidade de X, sua média, moda (caso exista), mediana e variância.*

Exercício 4.6 *Uma companhia de seguros vendeu apólices a dezoito pessoas, todas de mesma idade e condições de saúde. De acordo com as tábuas atuariais, a probabilidade de que uma pessoa, nas condições dos segurados, sobreviva dez anos depois de comprada a apólice é 0,9. Calcule o número esperado de sobreviventes aos dez anos de contrato e sua variância. Qual a probabilidade de que todas as pessoas sobrevivam. Quais as condições que devem ser satisfeitas para que possamos aplicar um modelo dado no texto? Discuta a aplicabilidade de um tal modelo.*

Exercício 4.7 *Uma prova consiste de 25 perguntas de múltipla escolha, com cinco possíveis respostas, sendo apenas uma delas correta. A nota é igual a $4/10$ vezes o número de respostas corretas. Um certo aluno se decide pelo seguinte esquema de responder a cada questão: joga-se um dado equilibrado e a face superior será a resposta. Caso saia 6, joga-se o dado novamente, até que se obtenha um número entre 1 e 5. Seja X a nota tirada por esse aluno.*

(i) Qual a distribuição de X? Defina Y como função de X, de forma que Y tenha uma distribuição binomial. Assinale as condições necessárias para sua resposta e quão razoáveis são na prática;

(ii) Calcule $\mathbb{P}(3 < Y \leq 6)$, $\mathbb{P}(4 \leq Y < 8)$, $\mathbb{P}(1 < Y < 4)$, $\mathbb{P}(Y \geq 6)$, $\mathbb{P}(Y \geq$

10), $\mathbb{P}(Y > 4)$, $\mathbb{P}(Y \leq 20)$, $\mathbb{P}(14 \leq Y < 22)$, $\mathbb{P}(Y \geq 1)$ e $\mathbb{P}(2 \leq Y \leq 4)$; e

(iii) Quais hipóteses de sua análise seriam violadas se o aluno utilizasse raciocínio e conhecimento para responder às perguntas?

Exercício 4.8 *Suponha as hipóteses do exercício 4.7. Assuma que o aluno tenha k ($k > 0$) minutos para responder às 25 perguntas. Gastam-se vinte segundos para cada lançamento do dado e preenchimento de cada resposta. Suponha que o aluno desconheça quanto tempo tem, ou seja, os lançamentos podem ser considerados independentes. Sejam W número de respostas a que o aluno consegue responder em tempo e Z a nota que o aluno obtém:*

(a) Como você pode utilizar uma distribuição binomial negativa para calcular probabilidades para W?

(b) Quais os valores de k para os quais $\mathbb{P}(W = 25) > 0,9$?

(c) Relacione k, as respectivas formas das distribuições de W e X (definida no exercício 4.7).

(d) Relacione as respectivas distribuições de X, Y (definidas no exercício 4.7), W e Z.

Exercício 4.9 *Seja X uma v.a. cuja função de probabilidade é dada por:*

$$\mathbb{P}(X = x_j) = p_j,$$

para $j = 1, 2, \ldots, r$ tal que $r \in \mathbb{N}^*$, $\sum_{i=1}^{r} p_j = 1$ e $(x_1, x_2, \ldots, x_r) \in \mathbb{R}^r$.

São feitas n repetições independentes de um experimento cujos possíveis resultados são completamente explicados por X. Prove que, para cada $j = 1, 2, \ldots, r$

$$\mathbb{E}(\textit{Frequência do valor } x_j \textit{ nas } n \textit{ repetições}) = np_j.$$

Obs.: A esperança acima definida é denominada frequência esperada de x_j.

Exercício 4.10
(a) Calcule a média, moda e mediana para uma $Bin(n, 1/2)$, $n = 30, 40, 60$. *Analise-as, com os respectivos gráficos;*

(b) Calcule a média, moda e mediana para a $Bin(n, 1/2)$, *n=30, 40, 60 e 100, utilizando-se dos respectivos gráficos. Justifique;*

(c) Verifique que, nos casos acima, se a distribuição for simétrica, média e mediana coincidem;

(d) Dê um contraexemplo de uma distribuição qualquer simétrica cuja moda não coincida com a média (que é igual à mediana). Qual(is) das três medidas seria(m) mais significativa(s) nesse caso? Justifique.

Exercício 4.11 *Sejam* $X \sim Bin(n,p)$ *e* $Y \sim Bin(n, 1-p)$, *onde* $n \in \mathbb{N}$ *e* $0 \le p \le 1$. *Prove, formalmente, que, para todo* $0 \le j \le n$,

$$\mathbb{P}(Y = j) = \mathbb{P}(X = n - j).$$

Interprete esse resultado com ajuda de gráficos, para $p = 2/10$, $p = 3/10$ *e* $p = 4/10$.

Exercício 4.12
(a) Uma pessoa pretende vender um lote de dezoito camisas que contém oito defeituosas. Um potencial comprador faz o seguinte teste: toma uma amostra de seis camisas e rejeita o lote se achar pelo menos três camisas defeituosas. Qual a probabilidade de o lote ser aceito?

(b) Suponha que uma empresa receba determinada peça em lotes de dez. Seu objetivo é rejeitar 70% dos lotes que tenham cinco ou mais peças defeituosas. Proponha uma regra de decisão, baseada em uma amostra de tamanho três, que satisfaça as exigências do comprador.

Exercício 4.13 *Uma urna contém bolas vermelhas (V) e bolas brancas (B). São extraídas cinco bolas. Calcule a probabilidade de extraírem-se duas bolas*