

# 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: S e 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, per-

manece. 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  $E$  um experimento nos moldes da definição 1.1.2. Suponha que  $E$  tenha um conjunto de resultados possíveis de cardinalidade dois. Um dos resultados será chamado de sucesso,  $S$ , e o outro de fracasso,  $F$ . Um sucesso ocorrerá com probabilidade  $0 < p \leq 1$  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  $E$  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 consequência 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

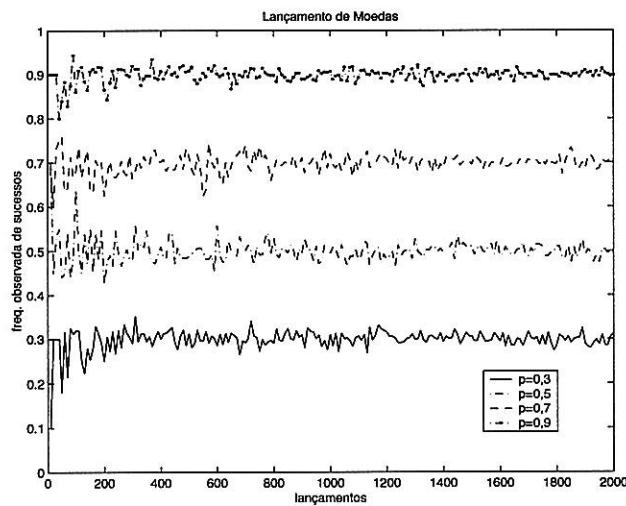
Seja  $\Upsilon$  uma seqüência de experimentos de Bernoulli independentes, no sentido da definição 4.2.1. Realizados  $n$  destes experimentos, seria de interesse saber quantos sucessos ocorreram.

**Exemplo 4.3.1 (Quão honesta é uma moeda)**

*As seguintes simulações (figura 4.1) mostram como se compõem 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 consequência direta da relação entre esperança e probabilidade, vista no capítulo 3, pág. 90.*

**Definição 4.3.1 (Distribuição Binomial)**

Figura 4.1: Honestidade de Moedas



*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.*

**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:

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

Cada uma 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,

$$\mathbb{E}(X) = np \quad (4.2)$$

$$Var(X) = np(1 - p) \quad (4.3)$$

**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  $\mathbb{E}(X(X - 1))$ :

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

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

sendo a penúltima soma o Binômio de Newton ( $p + (1-p)$ )<sup>n-2</sup>, i.e., 1.

Finalmente,

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

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 numéricico. 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:

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

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.  $\text{Bin}(n, p)$ . 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  $\text{Bin}(n, p)$  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  $k = 1, 2, \dots, n$ :

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

isto é,

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

Note que, de forma análoga, podemos concluir que:

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

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  $\mathbb{P}(X = k)$ ,  $k = 1, 2, \dots, n$ . 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 ( $m_2 \in \mathbb{N}$ ). ■

A 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

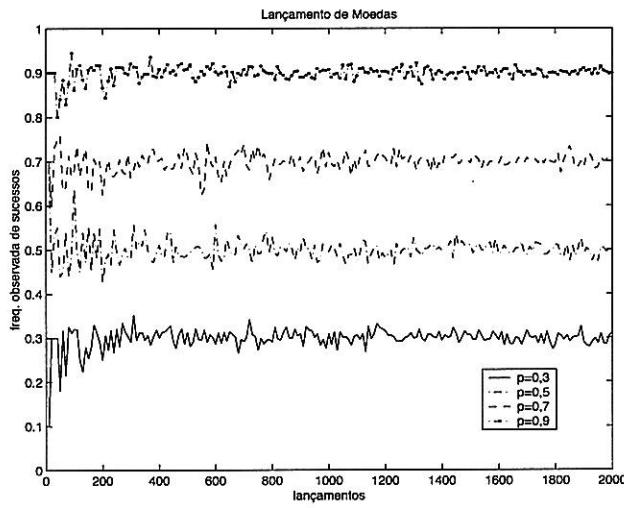
O exemplo abaixo, cuja motivação é devida a uma nota de N.L. Johnson, [21], mostra uma comparação interessante entre os conceitos de desvio médio e desvio padrão aplicados à distribuição binomial.

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

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

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

Figura 4.2: Modas da Distribuição Binomial



$$= \sum_{r=m}^n \left[ n \binom{n-1}{r-1} p^r q^{n-r+1} - n \binom{n-1}{r} p^{r+1} q^{n-r} \right]$$

$$= n \binom{n-1}{m-1} p^m q^{n-m+1}$$

$$= m \binom{n}{m} p^m q^{n-m+1} \quad (4.6)$$

(4.7)

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):

$$\sum_{r=0}^n \binom{n}{r} p^r q^{n-r} |r - np| = - \sum_{r=0}^{\lfloor np-1 \rfloor} \binom{n}{r} p^r q^{n-r} (r - np) +$$

$$+ \sum_{r=\lceil np \rceil}^n \binom{n}{r} p^r q^{n-r} (r - np). \quad (4.8)$$

*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*

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

---

<sup>1</sup>Note que aqui, a fórmula de Stirling é diferente da definida por (2.14), na página 433. 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

obtendo-se

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

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

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

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

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

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 rápida,

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

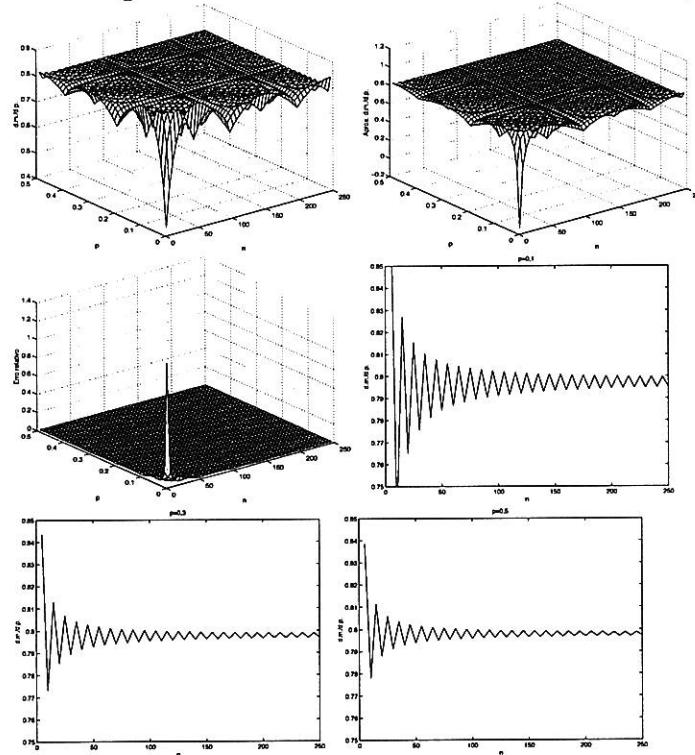
$n$	$p$	0,1 (A) (B)	0,2 (A) (B)	0,3 (A) (B)	0,4 (A) (B)	0,5 (A) (B)
10	0,7351 0,7313	0,7640 0,7630	0,7733 0,7729	0,7771 0,7768	0,7782 0,7779	
20	0,7652 0,7642	0,7807 0,7804	0,7855 0,7854	0,7874 0,7874	0,7879 0,7879	
50	0,7846	0,7909	0,7929	0,7937	0,7939	
100	0,7912	0,7949	0,7954	0,7958	0,7959	

(A) Valor exato dado pela equação 4.10

(B) Valor aproximado dado pela equação 4.11

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 quanto boa é a aproximação, funcionando muito mesmo para pequenos valores 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  $p = 0,5$  do que para  $p = 0,1$ .

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:

$$B_k = A_{k-1} \cap \{\text{Sucesso na } k\text{-ésima realização}\}, \quad (4.12)$$

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

Note-se que, dada a independência dos experimentos,

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

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  $\{\mathbb{P}(X = k)\}_{k=1}^\infty$  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,

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

**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  $Var(X)$ :

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

Portanto,

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

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)).

$$\begin{array}{c} \text{fracassos} \\ \underbrace{\text{FF} \cdots \text{F}}_{Y_1} \text{ S} \end{array} \quad \begin{array}{c} \text{fracassos} \\ \underbrace{\text{FF} \cdots \text{F}}_{Y_2} \text{ S} \end{array} \quad \dots \quad \begin{array}{c} \text{fracassos} \\ \underbrace{\text{FF} \cdots \text{F}}_{Y_r} \text{ S} \end{array} \quad (4.14)$$

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,*

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

**Demonstração:**

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

$$A = B \cap C,$$

onde  $B = \{\text{há } (r-1) \text{ sucessos nas } (n-1) \text{ primeiras tentativas}\}$  e  $C = \{\text{há sucesso na } n\text{-ésima tentativa}\}$ .  $B$  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,

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

**Demonstração:**

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

$$A = B \cap C,$$

onde  $B = \{\text{há } n \text{ fracassos nas } (n+r-1) \text{ primeiras tentativas}\}$  e  $C = \{\text{há sucesso na } (n+r)\text{-ésima tentativa}\}$ .  $B$  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:

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

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  $n = 0, 1, \dots$  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)**

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

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

**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  $Var(X) = E(X(X + 1)) - E(X) - (E(X))^2$ :

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

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

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

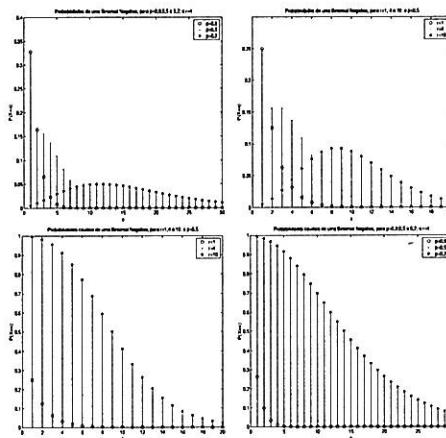
**Exemplo 4.5.1** Suponha que um dado equilibrado seja lançado até que 1 seja a face superior pela quarta vez. Seja  $X$  o número de lançamentos realizados.

$X$  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 pelo gráfico de  $1 - F(x)$  vs  $x$ , visto na figura 4.4.

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 acontecer rapidamente. Essa noção de grandeza de  $p$  depende do valor de  $r$  da

Figura 4.4: Probabilidades Caudais da Distribuição Binomial Negativa



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 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  $\Xi$  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, consequentemente, 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,  $i = 1, 2, \dots, N^3$ .

**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  $\Xi$  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,

---

<sup>3</sup>Todos os experimentos após o  $m$ -ésimo sucesso ou  $(N - m)$ -ésimo fracasso são triviais, isto é, as probabilidades de fracasso ou sucesso são respectivamente iguais a 1, em cada caso.

$$\begin{aligned}
 (i) \quad & \mathbb{P}(X = i) = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \\
 (ii) \quad & \mathbb{E}(X) = \frac{nm}{N} \\
 (iii) \quad & \text{Var}(X) = \frac{nm}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right)
 \end{aligned}$$

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 demonstrarmos 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)**

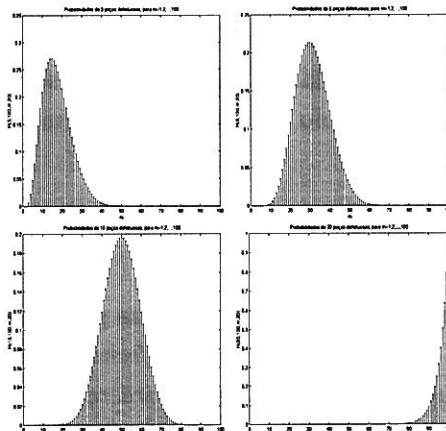
*Suponha que uma empresa A produza peças para serem vendidas à empresa B. A empresa B deseja fechar negócio com A se tiver indícios que os lotes de peças que irá receber terão proporção de peças defeituosas não superior a 5%. Sabendo-se que as peças são vendidas em lotes de cem, quantas peças de um lote 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  $h_k(100, m, 20)$  para todos os possíveis valores de  $m$  ( $1, 2, \dots, 100$ ), para cada valor de  $k$  em  $\{3, 6, 10, 20\}$ .

Figura 4.5: Probabilidades da Hipergeométrica



Como pode ser notado na figura 4.5, 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 verossímil. 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 verossímil 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 verossímil  $m = 100$ , 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 verossímil 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 4.5, 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<sup>4,5</sup>, 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  $H(N, j, n)$ .

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

---

<sup>4</sup> 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  $\hat{m}$ .

<sup>5</sup> Deve ser utilizada a convenção de que não se calculam as razões quando um dos valores for nulo.

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

ou seja,

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

se, e somente se,

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

o que significa que

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

Logo:

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

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 [25]. 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, consequentemente,  $n - i$  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:

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

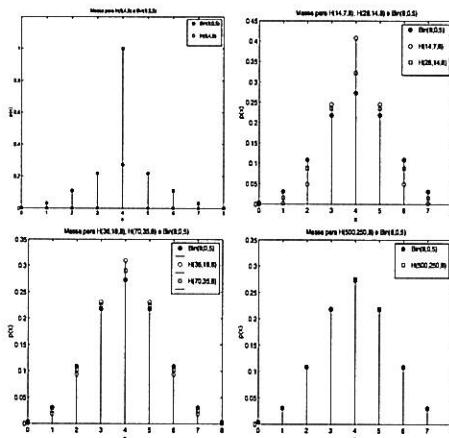
onde as passagens são feitas seguindo raciocínios análogos aos feitos para  $\mathbb{E}(X)$  e, portanto,

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

■

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  $N \rightarrow \infty$ , aquela deveria desaparecer , desde que  $m/N \rightarrow p \in [0, 1]$  e  $n/N \rightarrow 0^6$ .

Figura 4.6: A Hipergeométrica e a Binomial



Voltemos à esperança e variância de uma v.a. hipergeométrica,  $X$ .

<sup>6</sup>A condição de que  $m/N \rightarrow p \in [0, 1]$  nos diz que a proporção de defeituosas se estabiliza em  $p$  enquanto a condição de que  $n/N \rightarrow 0$  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.

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. Formalmente:

**Proposição 4.6.2 (Converg. da Hipergeométrica para a Binomial)**

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

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

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

Na figura 4.6, 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 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  $\text{Bin}(8; 0,5)$  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  $\text{Bin}(N, m/N)$  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 na figura 4.6 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  $p = \lim_{N \rightarrow \infty} \frac{m}{N}$ . Podemos escrever:

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

$$\frac{n!}{N(n-1)\cdots(N-n+1)} \asymp \binom{n}{i} \frac{m^i}{N^i} \frac{(N-m)^{n-i}}{N^{n-i}},$$

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

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

## 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<sup>7</sup>, 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  $\text{Bin}(n, p)$ , onde  $p = p_n$ . Suponha que  $\lim_{n \rightarrow \infty} np_n = \lambda \in \mathbb{R}$ . Então, para todo  $i = 1, 2, \dots,$*

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

---

<sup>7</sup>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  $n$ .

**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  $\lambda$ ,  $X \sim Po(\lambda)$ .

**Demonstração:**

$$\begin{aligned} \mathbb{P}(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &\quad \text{usando a fórmula de Stirling} \\ &\asymp \frac{\sqrt{2\pi n^{n+\frac{1}{2}}} e^{-n}}{i! \sqrt{2\pi(n-i)^{n-i+\frac{1}{2}}} e^{-(n-i)}} p^i (1-p)^{n-i} \\ &\asymp \frac{e^{-i}}{i!} \frac{n^{n+\frac{1}{2}}}{(n-i)^{n-i+\frac{1}{2}}} p^i (1-p)^{n-i} \\ &= \frac{e^{-i}}{i!} \frac{n^{n-i+\frac{1}{2}}}{(n-i)^{n-i+\frac{1}{2}}} (np)^i (1-p)^{n-i} \\ &= \frac{e^{-i}}{i!} (np)^i \left( \frac{n-np}{n-i} \right)^{n-i} \sqrt{\frac{n}{n-i}} \\ &\asymp \frac{\lambda^i e^{-\lambda}}{i!}, \end{aligned}$$

pois

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

quando  $n \rightarrow \infty$ . ■

As consequências 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,  $0,3/100000\text{hab} * \text{ano}$  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  $\lambda = 1000000 \times p = 3$ .

Em geral, podemos utilizar essa aproximação sempre que  $np$  for moderado, sendo  $n$  grande e  $p$  pequeno. O leitor pode estar questionando quanto 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  $\text{Bin}(n, p)$ .*

*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<sup>8</sup> 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 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 consequência 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, preocupar-nos-emos apenas com os bem sucedidos, ou seja, estaremos contando quantas vezes  $E$  ocorre em um determinado período de tempo. Para ter uma idéia

---

<sup>8</sup>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.

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)  $P(E \text{ ocorra em um intervalo de tempo de tamanho } h) = \lambda h + o(h)$ ;
- (2)  $P(\text{duas ou mais ocorrências de } E \text{ em um intervalo de tempo } h) = o(h)$ ; e
- (3)  $\forall t \in \mathbb{R}^+, \forall n, j_1, j_2, \dots, j_n \in \mathbb{Z}, \{I_k\}_{k=1}^n$  qualquer partição de  $[0, t]$ , sendo  $E_k$  o evento exatamente  $j_k$  dos eventos ocorrem em  $I_k$ ,  $k = 1, E_1, E_2, \dots, E_n$  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  $\lambda t$ .

#### 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  $\lambda t$ , isto é, para cada  $t > 0$ ,

$$P(N(t) = k) = \frac{\exp\{-\lambda t\} (\lambda t)^k}{k!} \mathbb{I}(k = 1, 2, \dots).$$

### Demonstração: REFAZER

$$\begin{aligned} P(N(t) = k) &= P(k \text{ intervalos contêm uma ocorrência de } E \text{ e } n - k \text{ intervalos} \\ &\quad \text{contêm 0 ocorrências}) + P(N(t) = k \text{ e pelo menos um intervalo} \\ &\quad \text{contém duas ou mais ocorrências de } E) \\ &= \mathbb{P}(A_n) + \mathbb{P}(B_n) \end{aligned}$$

Seja  $B^*$  o seguinte evento:

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

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

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

quando  $n \rightarrow \infty$ .

Por outro lado,

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

e, consequentemente,

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

Como

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

e, quando  $a_n \rightarrow a$

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

$$\mathbb{P}(A) = \frac{\exp\{-\lambda t\} (\lambda t)^k}{k!}$$

e, finalmente,

$$\begin{aligned}P(N(t) = k) &= \mathbb{P}(A) + \mathbb{P}(B) \\ &= \frac{\exp\{-\lambda t\} (\lambda t)^k}{k!}.\end{aligned}$$

■

## 4.8 Função Geratriz de Probabilidades

Existem funções fundamentais para caracterizarem-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.