

**Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Estatística**

**Introdução aos Modelos Probabilísticos Discretos:  
Binomial, Hipergeométrico, Binomial Negativo, Geométrico e Poisson**

Edna A. Reis e Ilka A. Reis

**Apostila Didática  
RTE-01/2016**

Este material é de domínio público. Pede-se apenas para citar a fonte:

Reis, E. A., Reis, I. A. (2016). *Introdução aos Modelos Probabilísticos Discretos: Binomial, Hipergeométrico, Binomial Negativo, Geométrico e Poisson*. Relatório Técnico do Departamento de Estatística da Universidade Federal de Minas Gerais.

# SUMÁRIO

<b>1. Variáveis Aleatórias e Distribuições de Probabilidades</b>	<b>03</b>
1.1. Variável Aleatória	03
1.2. Função de Probabilidade	05
1.3. Função de Distribuição Acumulada	06
1.4. Esperança e Variância	08
1.5. Propriedades de Linearidade da Esperança e da Variância	10
<b>2. Modelo Binomial</b>	<b>12</b>
2.1. Exemplo Inicial	12
2.2. Definições	12
2.3. Outros Exemplos	14
<b>3. Modelo Hipergeométrico</b>	<b>15</b>
3.1. Exemplo Inicial	15
3.2. Definições	16
3.3. Outros Exemplos	17
<b>4. Modelo Binomial Negativo</b>	<b>18</b>
4.1. Exemplo Inicial	18
4.2. Definições	18
4.3. Outros Exemplos	19
4.4. Caso Especial: O Modelo Geométrico	21
<b>5. Modelo Poisson</b>	<b>22</b>
5.1. Exemplo Inicial	22
5.2. Definições	22
5.3. Suposições	24
5.4. Outros Exemplos	24
5.5. Aproximação da Binomial pela Poisson	25
5.6. Verificação da Adequação do Modelo de Poisson	27
<b>Referências Bibliográficas</b>	<b>28</b>

# 1. VARIÁVEIS ALEATÓRIAS E DISTRIBUIÇÕES DE PROBABILIDADES

Nossa vida, pessoal ou profissional, é repleta de eventos incertos. Quantos minutos levarei para ir de casa até o trabalho ou escola? Vai chover hoje à tarde? Este medicamento terá o desejado em mim? Eliminar a incerteza de nossas vidas é tarefa impossível. Assim, devemos aprender a lidar com ela.

Quantificar a incerteza, expressando-a em um valor numérico, é uma das maneiras que o ser humano criou para lidar com ela nos eventos do dia-a-dia. A quantificação da incerteza é o que chamamos de *Probabilidade*. Assim, a Probabilidade é a medida da (in)certeza sobre a ocorrência de um evento.

Até aqui, já sabemos o que são experimentos aleatórios e aprendemos a calcular probabilidades em situações simples, que envolvem espaços amostrais cujos elementos têm todos a mesma probabilidade de ocorrer (ex: jogos de dados ou de cartas). Em geral, o interesse era calcular a probabilidade de ocorrência de um elemento do espaço amostral (ex: um ás no jogo de cartas) ou de uma combinação de eventos (ex: um ás ou um valete; um ás vermelho).

Neste capítulo, vamos aprender a transformar os elementos simbólicos de um espaço amostral em valores numéricos. Essa operação de transformação, cujo resultado chamamos de *variável aleatória*, facilita muito a resolução de problemas que envolvem incerteza. A medida dessa incerteza, ou seja, a probabilidade, será associada aos valores da variável aleatória e não mais a cada elemento ou combinação de elementos do espaço amostral.

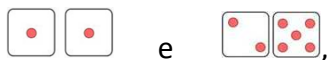
Cada *variável aleatória* tem associada a si uma *função de probabilidade*. As subseções 1.1 a 1.5 apresentarão a definição para esses dois elementos e trarão exemplos de sua aplicação na resolução de problemas. As seções 2 a 5 apresentam alguns dos modelos matemáticos para a função de probabilidade de uma variável aleatória *discreta*, como o número de filhos em uma família ou o número de votos de um candidato.

O objeto deste texto é abordar os modelos probabilísticos mais utilizados para se trabalhar com variáveis aleatórias discretas. Para detalhes e informações sobre outros modelos probabilísticos, os trabalhos de Triola (2013), Montgomery e Runger (2012) e de Walpole e colegas (2009) são excelentes fontes de consulta.

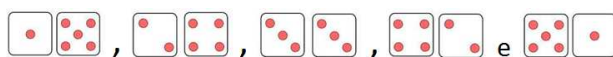
## 1.1. Variável Aleatória

**Exemplo Inicial:** vamos considerar, como exemplo, o experimento aleatório de jogar dois dados e observar as faces de cima. O espaço amostral, neste caso, é formado pelos resultados mostrados na Figura 1.1. Se os dados são honestos, cada um desses 36 resultados tem a mesma probabilidade de ocorrer (1/36).

Vamos atribuir, a cada ponto amostral, a *soma dos valores das faces*, denotada por  $X$ . Por exemplo, para os pontos amostrais



temos que  $X=2$  e  $X=7$ , respectivamente. Já os pontos amostrais



levam ao mesmo valor de  $X=6$ .

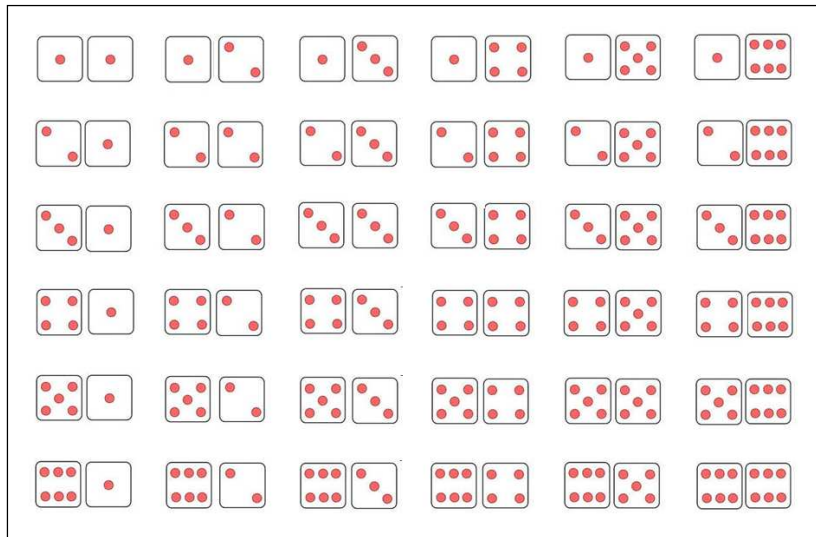


Figura 1.1: Espaço amostral no lançamento de dois dados.

A Figura 1.2 mostra os valores possíveis de  $X$  e quais elementos do espaço amostral levam a cada um desses valores. Note que cada ponto amostral leva a apenas um valor de  $X$ ; entretanto, um determinado valor de  $X$  pode vir de diferentes pontos amostrais.

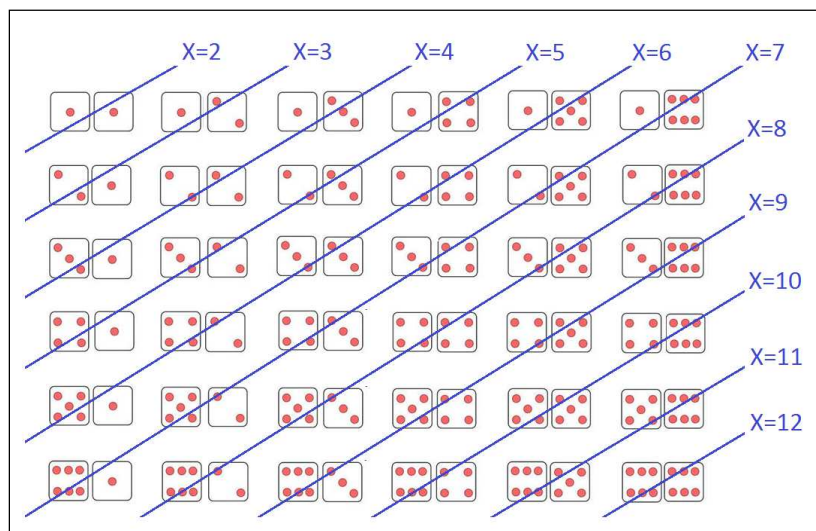


Figura 1.2: Valores possíveis da variável aleatória  $X$ : soma dos valores das faces no lançamento de dois dados.

A variável  $X$ : soma das faces de dois dados lançados é chamada de **variável aleatória**, pois associa um valor numérico a cada elemento do espaço amostral, por meio da operação matemática da soma. O adjetivo **aleatória** indica que o valor que a variável vai assumir é determinado pelo **acaso**, ou seja, seu valor somente é conhecido depois que o experimento aleatório é realizado.

**Variável Aleatória: (v.a.)** é uma **função** aplicada aos pontos do espaço amostral de um experimento aleatório e cujo resultado é um **número** real (inteiro ou não). A cada resultado possível do experimento é associado um único valor para a variável aleatória. Entretanto, mais de um resultado do experimento pode levar ao mesmo valor da variável.

Vamos adotar a seguinte notação para as variáveis aleatórias: as letras maiúsculas ( $X, Y, Z, \dots$ ) são usadas para *nomear* ou *definir* a variável aleatória, enquanto as letras minúsculas ( $x, y, z$ ) são usadas para representar os *valores* que ela pode assumir, ou seja, o *contradomínio* da variável aleatória. No exemplo do lançamento de dois dados, temos

$$X: \text{soma dos valores das faces} \quad e \quad x \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

As variáveis aleatórias podem ser **discretas** ou **contínuas**.

<b>Variável Aleatória Discreta:</b>	Assume valores em um conjunto finito ou infinito contável. Exemplos: Y: nº de filhotes nascidos vivos em uma ninhada de cinco gatos, $y \in \{0, 1, 2, 3, 4, 5\}$ ; Z: nº de colônias de bactérias por litro de leite, $z \in \{0, 1, 2, 3, \dots\}$ .
-------------------------------------	--

<b>Variável Aleatória Contínua:</b>	Assume valores em um conjunto infinito não contável. Exemplos: Y: peso de um filhote de gato, $y > 0$ ; Z: proporção de água no leite, $0 \leq z \leq 1$ .
-------------------------------------	--

Neste capítulo, iremos estudar apenas **variáveis aleatórias discretas**.

## 1.2. Função de Probabilidade

Apesar da aleatoriedade do valor da variável aleatória, podemos conhecer, de antemão, os valores ou intervalo de valores que ela **pode** assumir e qual é a **probabilidade** de cada um deles acontecer.

No exemplo do lançamento de dois dados, a probabilidade de cada um dos valores possíveis da variável aleatória  $X$  é calculada como a soma das probabilidades dos pontos amostrais associados a esses valores (que, nesse caso, são todas iguais a  $1/36$ ). Por exemplo, são quatro os resultados do lançamento de dois dados que levam à soma das faces igual a 5. Desse modo, a probabilidade de a soma das faces seja igual a 5 é  $4/36$ . Os valores das probabilidades dos demais valores de  $X$  são mostrados na Tabela 1.1.

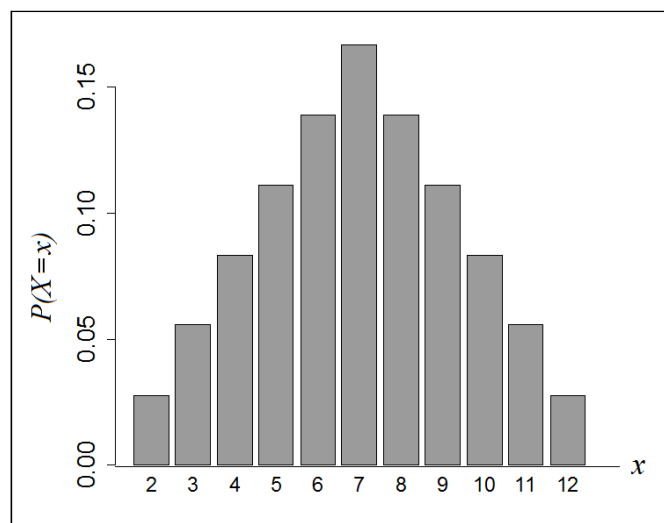
**Tabela 1.1:** Resultados possíveis da soma das faces de dois dados e suas probabilidades de ocorrência.

Valor de X	Nº de pontos amostrais	Probabilidade de X
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

O conjunto das probabilidades associadas aos valores do domínio de uma variável aleatória discreta é chamado de **função (massa) de probabilidade** dessa variável aleatória.

<b>Função (massa) de Probabilidade:</b>	É a função que associa uma <i>probabilidade</i> a cada valor possível da variável aleatória discreta. Notação: $P[X=x]$ é a probabilidade da variável aleatória discreta $X$ assumir o valor $x$ .
---	--

A função de probabilidade pode ser representada também por meio de um gráfico de barras verticais, como no caso do exemplo do lançamento de dois dados (Figura 1.3).



**Figura 1.3:** Representação gráfica da função de probabilidade da variável aleatória discreta  $X$ : soma dos valores das faces de cima no lançamento de dois dados.

Lembre-se de que a soma das probabilidades de todos os pontos do espaço amostral é igual a 1. Sabemos que cada ponto do espaço amostral leva a um único valor da variável aleatória  $X$ . Logo, a soma das probabilidades de todos os valores da variável aleatória  $X$  também é 1:

$$\sum_{\forall x} P[X = x] = 1.$$

### 1.3. Função de Distribuição Acumulada

Em alguns problemas, podemos estar interessados na probabilidade da variável aleatória discreta  $X$  assumir um valor menor ou igual a  $x$ , ou seja, em  $P[X \leq x]$ . O conjunto destas probabilidades definem a *função de distribuição (de probabilidade) acumulada de  $X$* .

<b>Função de Distribuição Acumulada:</b>	Para uma variável aleatória discreta $X$ , é definida por $F(t) \equiv P[X \leq t] = \sum_{\forall x \leq t} P[X = x], \quad -\infty < t < \infty.$
--	--

No exemplo do lançamento de dois dados, a probabilidade de a soma das faces de cima dos dois dados ser menor ou igual a 5 é calculada como:

$$P[X \leq 5] = P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5] = 1/36 + 2/36 + 3/36 + 4/36 = 10/36.$$

As demais probabilidades acumuladas para este exemplo são mostradas na Tabela 1.2.

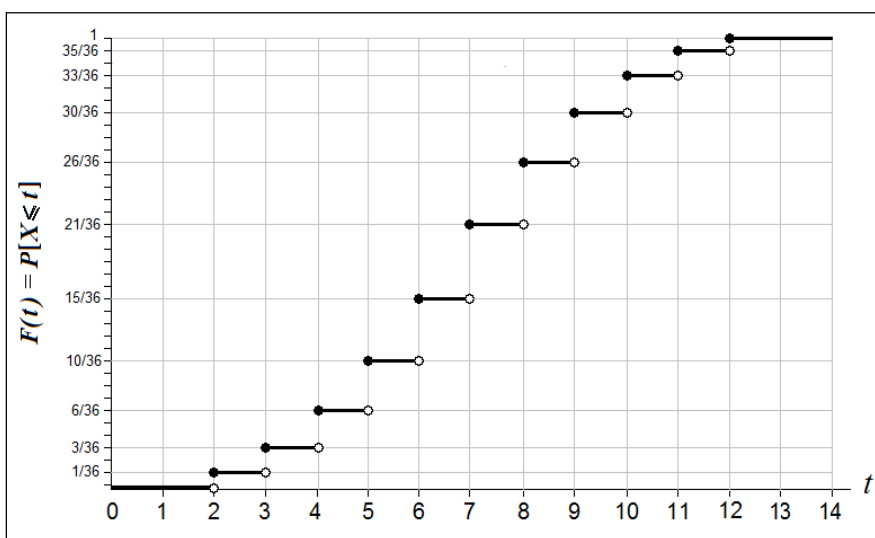
**Tabela 1.2:** Probabilidades acumuladas da soma das faces de dois dados.

x	P[X = x]	P[X ≤ x]	
2	1/36	1/36	= 1/36
3	2/36	1/36 + 2/36	= 3/36
4	3/36	1/36 + 2/36 + 3/36	= 6/36
5	4/36	1/36 + 2/36 + 3/36 + 4/36	= 10/36
6	5/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36	= 15/36
7	6/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36	= 21/36
8	5/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36	= 26/36
9	4/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36	= 30/36
10	3/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36	= 33/36
11	2/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36 + 2/36	= 35/36
12	1/36	1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36 + 2/36 + 1/36	= 1

Note que a função  $F(t)$  é definida para todos os valores de  $t$  na reta real, ou seja, para  $-\infty < t < \infty$ , e não apenas para os valores de  $X$  com probabilidade não nula. Assim, define-se, por exemplo,

$$\begin{aligned} P[X \leq 5,6] &= P[X \leq 5] = 10/36, \\ P[X \leq 7,1] &= P[X \leq 7] = 21/36, \\ P[X \leq 13] &= P[X \leq 12] = 1 \quad \text{e} \\ P[X \leq 1] &= 0. \end{aligned}$$

Desse modo, a função de distribuição acumulada  $F(t)$  é representada graficamente como uma função escada, como aquela da Figura 1.4.



**Figura 1.4:** Representação gráfica da função de distribuição acumulada da variável aleatória  $X$ : soma dos valores das faces no lançamento de dois dados.

## 1.4. Esperança e Variância

Vamos introduzir agora o conceito de duas importantes medidas para variáveis aleatórias: *esperança* e *variância*.

<b>Esperança ou Valor Esperado:</b>	Seja $X$ uma variável aleatória discreta. A <i>esperança</i> ( <i>valor esperado</i> ou <i>média</i> ) de $X$ é calculada como
-------------------------------------	--

$$E[X] = \sum_{\forall x} x \cdot P[X = x].$$

<b>Variância:</b>	Seja $X$ uma variável aleatória discreta com $E[X]=\mu$ . A <i>variância</i> de $X$ é calculada como
-------------------	--

$$Var[X] = \sum_{\forall x} (x - \mu)^2 \cdot P[X = x].$$

O *desvio-padrão* de  $X$  é calculado como  $DP[X] = \sqrt{Var[X]}$ .

Uma fórmula alternativa para o cálculo da variância é:

<b>Fórmula Alternativa Para Cálculo da Variância:</b>	$Var[X] = E[X^2] - (E[X])^2$ , com $E[X^2] = \sum_{\forall x} x^2 \cdot P[X = x]$
---	---

A esperança pode assumir qualquer valor real, enquanto a variância (e o desvio-padrão) assume apenas valores não-negativos.

**Exemplo Inicial** (*continuação*): Para a variável aleatória  $X$ : *soma das faces de dois dados lançados*, a esperança e a variância são dadas por:

$$E[X] = 2 \cdot P[X=2] + 3 \cdot P[X=3] + \dots + 12 \cdot P[X=12] = 2(1/36) + 3(2/36) + \dots + 12(1/36) = 7$$

e

$$Var[X] = (2-7)^2 \cdot P[X=2] + (3-7)^2 \cdot P[X=3] + \dots + (12-7)^2 \cdot P[X=12] = \frac{210}{36}.$$

Ou, usando a fórmula alternativa para calcular a variância, temos:

$$E[X^2] = (2^2) \cdot P[X=2] + \dots + (12^2) \cdot P[X=12] = 4(1/36) + \dots + 144(1/36) = \frac{1974}{36}$$

e, assim,

$$Var[X] = E[X^2] - (E[X])^2 = \left(\frac{1974}{36}\right) - (7)^2 = \frac{210}{36}.$$

---

A esperança indica o valor em torno do qual a variável  $X$  tende a se localizar depois de muitas repetições do experimento. Sendo a média de uma variável aleatória, a esperança indica qual é o



valor típico dessa variável. A variância, por sua vez, mede o quão espalhados (dispersos) em torno do valor da esperança estão os valores possíveis de  $X$ .

**Exemplo 1.2:** Uma mercearia compra lotes de leite de duas marcas, chamadas A e B. O lote tem cinco caixas de leite. Da observação das vendas até o momento, sabe-se que a variável aleatória  $X$ , o número de caixas vendidas por lote da marca A, tem função de probabilidade dada por

$x$	0	1	2	3	4	5
$P[X=x]$	0,01	0,08	0,23	0,34	0,26	0,08

e a variável aleatória  $Y$ , o número de caixas vendidas por lote da marca B, tem função de probabilidade dada por

$y$	0	1	2	3	4	5
$P[Y=y]$	0,00	0,01	0,05	0,20	0,41	0,33

Calcule a  $E[X]$ ,  $E[Y]$ ,  $\text{Var}[X]$  e  $\text{Var}[Y]$ .

**Solução:**

- Para a marca A:

$$E[X] = 0 \cdot P[X=0] + 1 \cdot P[X=1] + 2 \cdot P[X=2] + 3 \cdot P[X=3] + 4 \cdot P[X=4] + 5 \cdot P[X=5]$$

$$= 0(0,01) + 1(0,08) + 2(0,23) + 3(0,34) + 4(0,26) + 5(0,08) = 3.$$

Para calcular a variância, vamos usar a fórmula alternativa:

$$E[X^2] = (0^2) \cdot P[X=0] + (1^2) \cdot P[X=1] + (2^2) \cdot P[X=2] + (3^2) \cdot P[X=3] + (4^2) \cdot P[X=4] + (5^2) \cdot P[X=5]$$

$$= 0(0,01) + 1(0,08) + 4(0,23) + 9(0,34) + 16(0,26) + 25(0,08) = 51/5.$$

$$\text{e, assim, } \text{Var}[X] = E[X^2] - (E[X])^2 = \left(\frac{51}{5}\right) - (3)^2 = \frac{6}{5}.$$

- Para a marca B:

$$E[Y] = 0 \cdot P[Y=0] + 1 \cdot P[Y=1] + 2 \cdot P[Y=2] + 3 \cdot P[Y=3] + 4 \cdot P[Y=4] + 5 \cdot P[Y=5]$$

$$= 0(0,00) + 1(0,01) + 2(0,05) + 3(0,20) + 4(0,41) + 5(0,33) = 4.$$

$$E[Y^2] = (0^2) \cdot P[Y=0] + (1^2) \cdot P[Y=1] + (2^2) \cdot P[Y=2] + (3^2) \cdot P[Y=3] + (4^2) \cdot P[Y=4] + (5^2) \cdot P[Y=5]$$

$$= 0(0,00) + 1(0,01) + 4(0,05) + 9(0,20) + 16(0,41) + 25(0,33) = \frac{49}{5}.$$

$$\text{Var}[Y] = E[Y^2] - (E[Y])^2 = \left(\frac{49}{5}\right) - (4)^2 = \frac{9}{5}.$$

A marca B, além de ter, em média, um número maior de vendas, ainda tem uma variabilidade menor em torno deste valor esperado. Ou seja, as vendas da marca B seriam mais estáveis do que as da marca A. A Figura 1.5 mostra as funções de probabilidades do número de caixas vendidas por lote das duas marcas, que confirma que a marca B tem desempenho melhor, pois tem maiores probabilidades associados aos valores de vendas maiores do que 3.

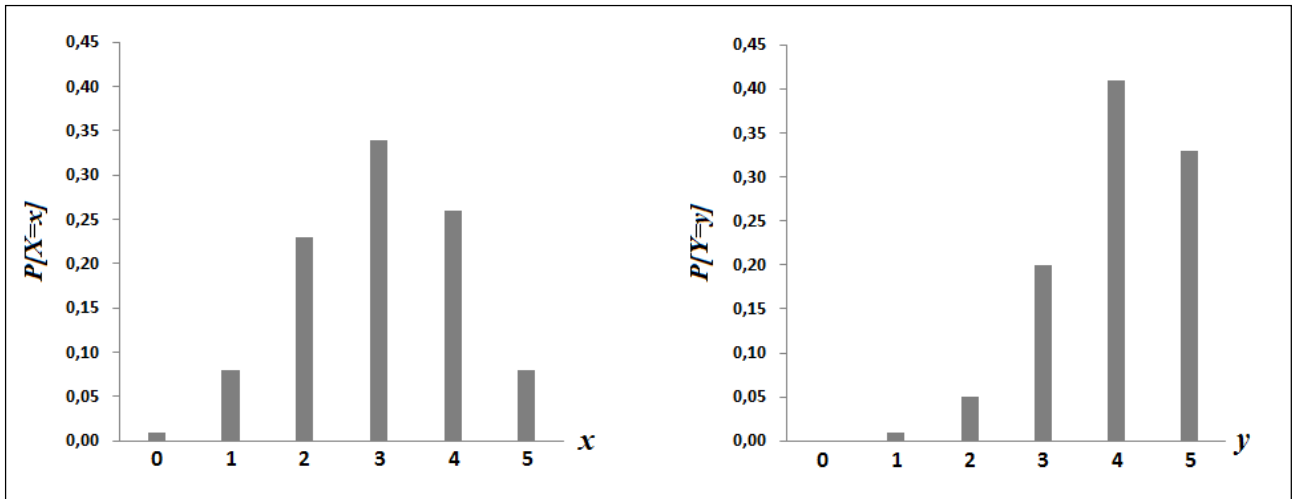


Figura 1.5: Funções de probabilidade das variáveis aleatórias X e Y do Exemplo 1.2.

### 1.5. Propriedades de Linearidade da Esperança e da Variância

A esperança e a variância têm algumas propriedades úteis nas transformações lineares de uma variável aleatória, ou seja, quando uma constante é somada à variável aleatória e/ou multiplicada à variável, criando uma nova variável aleatória.

**Propriedade 1:** Seja  $X$  uma variável aleatória com  $E[X]=\mu$  e  $\text{Var}[X]=\sigma^2$ . Se somarmos uma constante  $b$  a  $X$ , criando a nova variável aleatória  $Y=X+b$ , então

$$E[Y] = E[X] + b = \mu + b$$

e  $\text{Var}[Y] = \text{Var}[X] = \sigma^2$ .

**Propriedade 2:** Seja  $X$  uma variável aleatória com  $E[X]=\mu$  e  $\text{Var}[X]=\sigma^2$ . Se multiplicarmos  $X$  por uma constante  $a$ , criando a nova variável aleatória  $Z=aX$ , então

$$E[Z] = aE[X] = a\mu$$

e  $\text{Var}[Z] = a^2\text{Var}[X] = a^2\sigma^2$ .

Estas propriedades podem ser visualizadas na Figura 1.6. No primeiro caso, a soma da constante 4 à variável aleatória  $X$  desloca o centro (a *média*) da distribuição das probabilidades, mas não altera a dispersão (*variância*) dos valores. Já no segundo caso, a multiplicação de  $X$  pela constante 2, além de deslocar o centro da distribuição das probabilidades, também aumenta a dispersão dos valores. Se a constante multiplicada fosse tal que  $-1 < a < 1$ , a dispersão da variável resultante seria menor que a da original (Figura 1.6c).

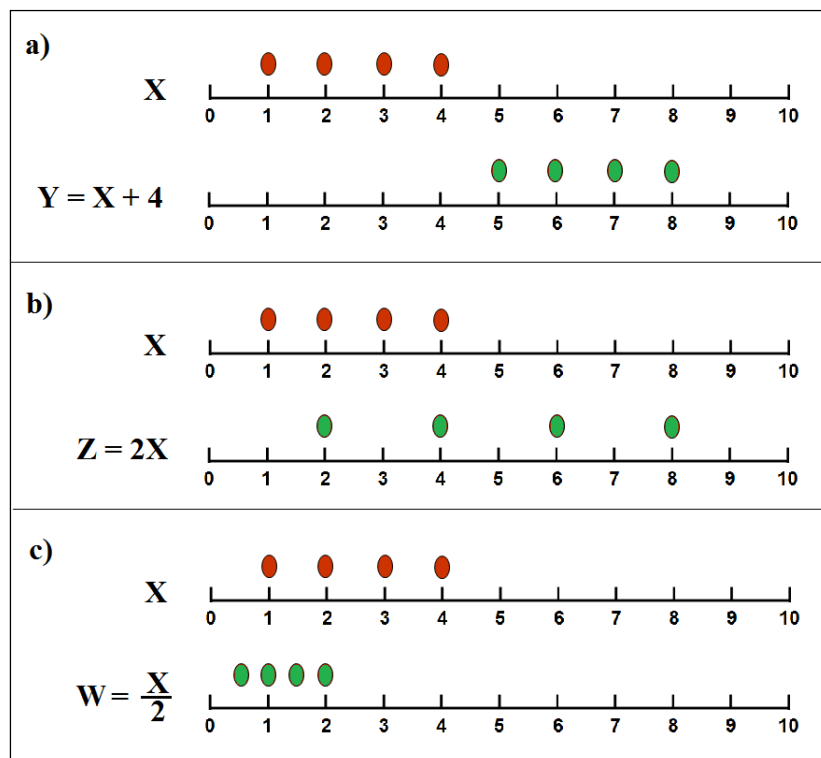


Figura 1.6: Ilustração das propriedades 1 (a) e 2 (b e c).

**Exemplo 1.3:** Uma mercearia compra um lote com cinco caixas de leite ao preço de atacado de 1,20 reais por caixa e revende cada caixa por 1,65 reais. Depois da data de validade, o leite não vendido é retirado da prateleira e a mercearia recebe, do distribuidor, um crédito igual de 0,90 reais por caixa. Sabe-se que a variável aleatória  $X$ , o número de caixas vendidas do lote, tem distribuição de probabilidade com média igual a 3 caixas e variância igual a  $6/5$ . Encontre a média e a variância do lucro do vendedor por lote.

**Solução:** O lucro do vendedor por lote ( $Y$ ) é dado pelo soma do valor que ele ganha pelas  $X$  caixas vendidas com o valor perdido pelas  $5-X$  caixas não vendidas:

$$Y = (1,65-1,20)X + (-1,20+0,90)(5-X) \Rightarrow Y = 0,75X - 1,50.$$

Logo,

$$E(Y) = (0,75)E(X) + 1,50 = (0,75)(3) + 1,50 = 0,75 \text{ reais e}$$

$$\text{Var}(Y) = (0,75)^2 \text{Var}(X) = (0,75)^2(6/5) = 0,68 \text{ reais}^2.$$

Em algumas situações, é possível identificar um padrão para a função de probabilidade de uma variável aleatória. Quando isso acontece, podemos adotar o que chamamos de **Modelo Probabilístico** para representar a função de probabilidade de uma variável aleatória. Um Modelo Probabilístico é definido por uma função de probabilidade ou por uma função de distribuição que, por sua vez, dependem de valores conhecidos, chamados **parâmetros** do modelo, e também dos valores que a variável aleatória pode assumir.

Os modelos probabilísticos são muito úteis na resolução de vários problemas práticos como dimensionamento de serviços (centrais telefônicas, portos e aeroportos), inventário da vida selvagem (método de captura-recaptura de aves e peixes), inspeções de elementos (controle de qualidade em indústrias, detecção de contrabando em alfândegas), estudo da eficácia de vacinas e muitos outros. Nas seções a seguir, vamos estudar os principais modelos probabilísticos para variáveis discretas, começando com um dos mais simples e intuitivo: o Modelo Binomial.

## 2. MODELO BINOMIAL

### 2.1. Exemplo Inicial

Uma caixa tem 3 bolas azuis e 7 bolas de outras cores. Uma bola é retirada aleatoriamente da caixa, observa-se sua cor (azul ou não azul) e ela é devolvida para a caixa. Repete-se este procedimento 6 vezes. Qual a probabilidade de 4 das 6 bolas selecionadas serem azuis?

### 2.2. Definições

Primeiramente, vamos definir o tipo de experimento aleatório no qual surge a variável aleatória Binomial.

<b>Experimento Bernoulli</b>	Um experimento aleatório de Bernoulli tem as seguintes propriedades:
<b><math>Be(p)</math></b>	- Resulta em um de dois eventos excludentes: <b>sucesso</b> ou <b>fracasso</b> ; - A <b>probabilidade de sucesso (<math>p</math>)</b> é <b>constante</b> em todas as repetições do experimento, ou seja, as repetições geram <b>resultados independentes</b> .

Alguns exemplos de experimentos que podem ser representados como de Bernoulli são:

- lançamento de uma moeda e observação da face de cima (cara ou coroa);
- lançamento de um dado e observação da face de cima (seis ou outro valor);
- nascimento de um bebê e observação do seu sexo (masculino ou feminino);
- verificação de um produto na linha de produção (conforme ou defeituoso);

A contagem dos sucessos em repetições de um experimento de Bernoulli define a variável aleatória Binomial.

<b>Variável Aleatória Binomial</b>	Seja $X$ o número de sucessos em $n$ repetições de um experimento $B(p)$ . A função de probabilidade de $X$ é dada por
<b><math>X \sim B(n;p)</math></b>	$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$

Lembre-se de que

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad n! = n(n-1)(n-2)\dots 2 \cdot 1 \quad e \quad 0! = 1.$$

**Exemplo 2.1 (continuação):** Como a bola retirada é colocada de volta na caixa para a próxima retirada, o número de bolas azuis será sempre 3 em 10; logo, a probabilidade de bola azul em qualquer retirada é sempre  $p=3/10=0,3$ . Assim,  $X$ , o número de bolas azuis em  $n=6$  retiradas, é uma v.a.  $B(6 ; 0,3)$ . A probabilidade de bola azul em 4 das 6 retiradas é dada por

$$P[X = 4] = \binom{6}{4} (0,3)^4 (1 - 0,3)^{6-4} = \frac{6!}{4!2!} \cdot (0,3)^4 (0,7)^2 = 15 \cdot (0,008) \cdot (0,49) \approx 0,06.$$

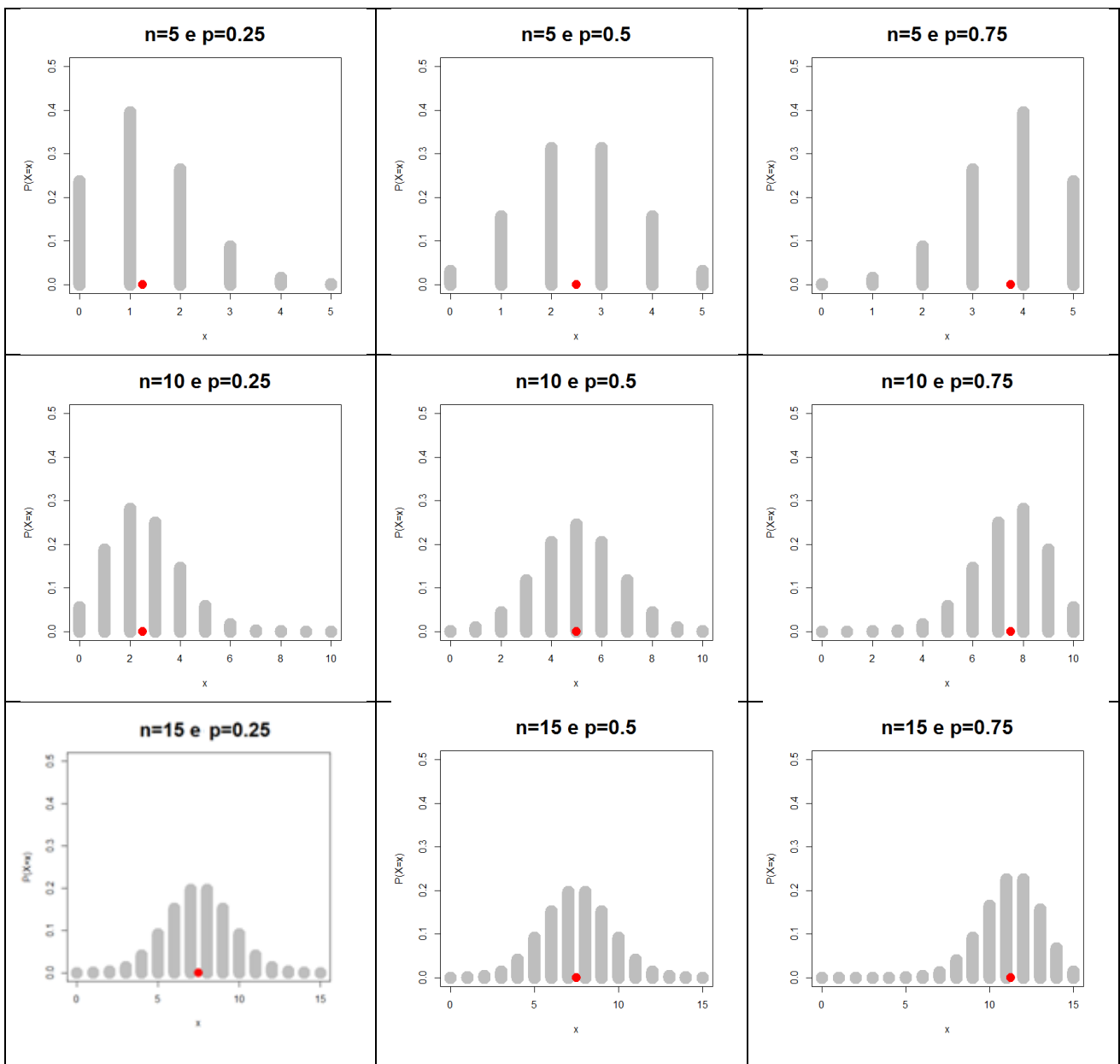
**Esperança e Variância da Binomial:**

Seja  $X \sim B(n;p)$ . A esperança e a variância de  $X$  são, respectivamente,  
 $E[X] = np$   
 e  $\text{Var}[X] = np(1-p)$ .

**Exemplo 2.1 (continuação):** O número esperado de bolas azuis em 6 retiradas é de  $6(0,3) = 1,8$ .

A Figura 2.1 mostra a função de probabilidade do modelo Binomial para diversos valores de  $n$  (nas linhas) e  $p$  (nas colunas). Observe que:

- quando  $p$  está fixo e  $n$  cresce, a esperança cresce e a variância cresce também, pois  $X$  pode assumir uma maior variedade de valores;
- quando  $n$  está fixo e  $p$  cresce, a esperança cresce. A variância, por sua vez, cresce até  $p=0,5$ , atingindo seu valor máximo ( $n/4$ ). Para valores de  $p > 0,5$ , a variância volta a diminuir.



**Figura 2.1:** Função de Probabilidade da Binomial com diferentes valores de  $n$  e  $p$  (em vermelho, a *esperança*).

## 2.3. Outros Exemplos

**Exemplo 2.2:** Uma mercearia compra um lote com cinco caixas de leite. Se a probabilidade de uma caixa ser vendida é de  $3/5$  e as caixas são vendidas de modo independente, determine a função de probabilidade do número de caixas vendidas por lote. Calcule  $E[X]$  e  $\text{Var}[X]$ .

**Solução:** Seja a variável aleatória  $X$ , o número de caixas vendidas por lote. Podemos assumir que  $X$  segue o modelo Binomial com  $n=5$  e  $p=3/5$ . Logo, a função de probabilidade de  $X$  é dada por

$$P[X = x] = \binom{5}{x} \left(\frac{3}{5}\right)^x \left(1 - \frac{3}{5}\right)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5.$$

Substituindo os valores de  $x$  na fórmula acima, obtemos:

$x$	0	1	2	3	4	5
$P[X=x]$	0,01	0,08	0,23	0,34	0,26	0,08

E ainda,  $E[X] = np = 5(3/5) = 3$  e  $\text{Var}[X] = np(1-p) = 5(3/5)(2/5) = 6/5$ .

---

**Exemplo 2.3:** No processo de fabricação de uma peça, admite-se que no máximo 5% sejam produzidas com defeito. Todos os dias, cerca de 15 unidades são selecionadas para inspeção. Se uma ou mais peças defeituosas são encontradas nesta amostra, o processo de produção é paralisado. Calcule a probabilidade de o processo ser paralisado quando:

- (a) Cinco por cento das peças inspecionadas são produzidas com defeito.
- (b) Sete por cento das peças inspecionadas são produzidas com defeito.

**Solução:** O número de ( $X$ ) tem distribuição Binomial com  $n=15$  e  $p$  igual à proporção de peças defeituosas produzidas. A função de probabilidade de  $X$  é dada por

$$P[X = x] = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, 2, \dots, 15.$$

(a) Se  $p=0,05$ , a probabilidade de o processo ser paralisado é dada por

$$\begin{aligned} P[X \geq 1] &= 1 - P[X = 0] = 1 - \binom{15}{0} (0,05)^0 (1-0,05)^{15-0} \\ &= 1 - (1 \cdot (0,95)^{15}) \approx 1 - 0,46 = 0,54. \end{aligned}$$

(b) Se  $p=0,07$ , a probabilidade de o processo ser paralisado é dada por

$$\begin{aligned} P[X \geq 1] &= 1 - P[X = 0] = 1 - \binom{15}{0} (0,07)^0 (1-0,07)^{15-0} \\ &= 1 - (1 \cdot (0,93)^{15}) \approx 1 - 0,37 = 0,63 \end{aligned}$$

Note que, quanto maior a probabilidade de uma peça do lote ser defeituosa, maior é a probabilidade de se encontrar uma ou mais peças defeituosas na amostra de 15 peças inspecionadas e, assim, maior é a probabilidade de o processo ser paralisado.

---

### 3. MODELO HIPERGEOMÉTRICO

#### 3.1. Exemplo Inicial

Suponha que uma caixa contenha 10 bolas, das quais 4 são azuis e 6 são vermelhas. Três bolas serão retiradas em sequência desta caixa e não serão devolvidas. Considere a variável aleatória  $X$ , que conta o número de bolas azuis nesta amostra de três bolas retiradas. Os valores possíveis de  $X$  são  $x = 0, 1, 2$  e  $3$ . Como podemos calcular a probabilidade de que  $X$  assumira cada um destes valores?

Primeiro, vamos determinar o tamanho do espaço amostral deste experimento. De quantas maneiras podemos retirar 3 bolas de uma caixa que contém 10 bolas? A resposta é

$$\binom{10}{3} = 120.$$

Das 4 bolas azuis que estão na caixa, de quantas maneiras podemos retirar uma quantidade  $x$  ( $x=0,1,2,3$ )? A resposta é

$$\binom{4}{x}.$$

Simultaneamente, das 6 bolas vermelhas da caixa, uma quantidade igual a

$$\binom{6}{3-x}$$

serão selecionadas. Assim, o número de maneiras possíveis de se selecionar  $x$  bolas azuis e  $3-x$  bolas vermelhas na amostra é

$$\binom{4}{x} \binom{6}{3-x}.$$

Logo, a probabilidade de  $x$  bolas azuis serem selecionadas na amostra de 3 bolas é dada por

$$P[X=x] = \frac{\binom{4}{x} \binom{6}{3-x}}{\binom{10}{3}}, \quad x=0, 1, 2 \text{ e } 3.$$

A função de probabilidade da variável  $X$  desse exemplo pode ser definida pelo **Modelo Hipergeométrico**.

### 3.2. Definições

<p><b>Variável Aleatória Hipergeométrica</b>  <math>X \sim H(N, n, k)</math></p>	<p>Seja <math>X</math> o número de sucessos em uma amostra aleatória de tamanho <math>n</math> selecionada, sem reposição, de <math>N</math> itens, dos quais <math>k</math> são sucessos. A função de probabilidade de <math>X</math> é dada por<sup>1</sup></p> $P[X = x] = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, \quad a \leq x \leq b, \quad \text{com } \begin{matrix} x_a = \max\{0, n - (N - k)\} \\ \text{e } x_b = \min\{n, k\}. \end{matrix}$
--	---

**Exemplo 3.1 (continuação):** No exemplo inicial, a variável aleatória  $X$ , o número de bolas azuis em uma amostra de  $n=3$  bolas retiradas de uma caixa com  $N=10$  bolas das quais  $k=4$  são azuis (e  $N-k=10-4=6$  bolas são vermelhas) tem probabilidades calculadas como

$$P[X = x] = \frac{\binom{4}{x} \binom{6}{3-x}}{\binom{10}{3}}, \quad x = 0, 1, 2 \text{ e } 3.$$

Temos, assim,

$$P[X=0] = \frac{\binom{4}{0} \binom{6}{3}}{\binom{10}{3}} = \frac{(1)(20)}{120} = \frac{20}{120} \approx 0,17$$

$$P[X=1] = \frac{\binom{4}{1} \binom{6}{2}}{\binom{10}{3}} = \frac{(4)(15)}{120} = \frac{60}{120} = 0,50$$

$$P[X=2] = \frac{\binom{4}{2} \binom{6}{1}}{\binom{10}{3}} = \frac{(6)(6)}{120} = \frac{36}{120} = 0,30$$

$$P[X=3] = \frac{\binom{4}{3} \binom{6}{0}}{\binom{10}{3}} = \frac{(4)(1)}{120} = \frac{4}{120} \approx 0,03.$$

Os experimentos aleatórios nos quais aplica-se o modelo Hipergeométrico são muito parecidos com aqueles que definem o modelo Binomial: conta-se o número de sucessos (bolas azuis) em um número ( $N$ ) predefinido de realizações de um experimento (retirada de uma bola da caixa). Entretanto, no modelo hipergeométrico, a probabilidade de sucesso não é constante nem há independência entre as realizações do experimento.

A maneira mais simples de diferenciar os modelos binomial e hipergeométrico é entender como as retiradas (amostragem) são feitas: *com* ou *sem reposição*. Se cada bola fosse retirada da caixa e repostada após a observação de sua cor (*amostragem com reposição*), teríamos, em todas as retiradas, 4 bolas azuis dentre a 10 bolas. Desse modo, a variável aleatória  $X$  (número de bolas azuis dentre as três selecionadas) teria função de probabilidade Binomial com  $n=3$  e  $p=4/10$ . Se a amostragem é feita sem reposição das bolas selecionadas, a probabilidade de seleção de uma bola

<sup>1</sup> A definição do menor valor que  $X$  pode assumir ( $x_a$ ) pode ser entendida da seguinte forma: se o número de bolas fracasso na urna ( $N-k$ ) é menor que o número de bolas retiradas ( $n$ ), então teremos na amostra no máximo ( $N-k$ ) bolas fracasso, e o restante da amostra ( $n-[N-k]$ ) será necessariamente completada com bolas sucesso (daí ter  $x_a = n-[N-k]$ ); caso contrário, se  $n > N-k$  (ou seja,  $n-[N-k] > 0$ ), então podemos a situação de zero bolas sucesso na amostra (todas as  $n$  bolas serem fracasso). Veja o esquema a seguir:

$$N-k < n \Leftrightarrow n > N-k \Leftrightarrow n-(N-k) > 0 \Leftrightarrow \min\{0, n-(N-k)\} = n-(N-k) \Leftrightarrow x_a = n-(N-k);$$

$$N-k > n \Leftrightarrow n < N-k \Leftrightarrow n-(N-k) < 0 \Leftrightarrow \min\{0, n-(N-k)\} = 0 \Leftrightarrow x_a = 0.$$

Já o maior valor  $X$  poderá assumir ( $x_b$ ) será o número total de sucessos na urna ( $k$ ) apenas se  $n \geq k$ ; no entanto, se  $n < k$ , ou seja, se o número de bolas retiradas ( $n$ ) for menor que o número de bolas sucesso ( $k$ ), poderemos ter na amostra no máximo  $n$  bolas sucesso.



azul em cada retirada não é constante, pois depende do resultado da seleção anterior, caracterizando a função de probabilidade Hipergeométrica.

<b>Esperança e Variância da Hipergeométrica</b>	Seja $X \sim H(N, n, k)$ . A esperança e a variância de X são, respectivamente, $E[X] = n \frac{k}{N} \quad e \quad Var[X] = n \left( \frac{N-n}{N-1} \right) \left( \frac{k}{N} \right) \left( 1 - \frac{k}{N} \right).$
---	--

**Exemplo 3.1 (continuação):** A esperança e a variância do número de bolas azuis dentre as três retiradas da caixa são dadas, respectivamente, por

$$E[X] = 3 \left( \frac{4}{10} \right) = 1,2 \quad e \quad Var[X] = 3 \left( \frac{10-3}{10-1} \right) \left( \frac{4}{10} \right) \left( 1 - \frac{4}{10} \right) = 0,56.$$

### 3.3. Outros Exemplos

**Exemplo 3.2:** Uma peça usada em motores de carros é vendida em lotes de 10 unidades. O comprador sabe que são produzidas algumas peças defeituosas, mas considera que um lote qualquer é **aceitável** se tiver, no máximo, uma peça defeituosa. A inspeção do lote pelo comprador consiste em escolher aleatoriamente 3 peças e testá-las; se nenhuma dessas 3 peças for defeituosa, o lote é considerado aceitável. Suponha que um lote a ser inspecionado tenha 2 peças defeituosas, ou seja, esse lote **não é aceitável**. Qual é a probabilidade da inspeção classificá-lo como **aceitável** ?

**Solução:** O número de peças defeituosas dentre as inspecionadas (X) tem distribuição Hipergeométrica com  $N=10$ ,  $n=3$  e  $k=2$ . A função de probabilidade de X é dada por

$$P[X=x] = \frac{\binom{2}{x} \binom{8}{3-x}}{\binom{10}{3}}, \quad x=0,1,2.$$

A probabilidade de o lote em análise ser considerado aceitável é

$$P[X=0] = \frac{\binom{2}{0} \binom{8}{3}}{\binom{10}{3}} = \frac{21}{45} = 0,467,$$

uma probabilidade muito alta de aceitar um lote “ruim”.

**Exemplo 3.3:** Um biólogo, desejando estimar a quantidade N de peixes de certa espécie em um lago, resolve seguir o seguinte procedimento (chamado de **método da captura-recaptura**): num primeiro momento (etapa da **captura**), ele vai amostrar 10 peixes desta espécie no lago e marcá-los (com tinta indelével e atóxica) e devolvê-los vivos ao lago. Após algum tempo (etapa da **recaptura**), ele vai retirar 12 peixes do lago e contar quantos deles estão marcados.

O número de peixes marcados na recaptura é a variável aleatória X, que tem distribuição Hipergeométrica com N desconhecido,  $n=12$  e  $k=10$ . A função de probabilidade de X é dada por

$$P[X = x] = \frac{\binom{10}{x} \binom{N-10}{12-x}}{\binom{N}{12}}, \quad x = 0, 1, \dots, 10.$$

Suponha que o valor observado de  $X$  seja  $x=6$ . Desse modo, a proporção de peixes marcados na amostra da recaptura é  $x/n=6/12$ . No lago todo, esta proporção é  $k/N=10/N$ . Então, uma estimativa de  $N$  pode ser obtida supondo que estas duas proporções são próximas:  $x/n \approx k/N$ , ou seja,  $N \approx (nk)/x=(12)(10)/6 = 20$  peixes.

## 4. MODELO BINOMIAL NEGATIVO

### 4.1. Exemplo Inicial

Considere que, em uma feira, 20% dos abacaxis (dentre dezenas) estão estragados. Uma pessoa quer comprar dois abacaxis (não estragados, claro). Qual a probabilidade de que ela precise analisar cinco abacaxis para concluir sua compra? E qual a probabilidade de que ela precise analisar pelo menos quatro abacaxis?

### 4.2. Definições

A contagem do número de repetições de um experimento de Bernoulli até que se obtenham  $k$  sucessos define a variável aleatória Binomial Negativa.

<b>Variável Aleatória Binomial Negativa</b>	Seja $X$ o número de repetições de um experimento $BN(k,p)$ até que ocorram $k$ sucessos. A função de probabilidade de $X$ é dada por
$X \sim BN(k,p)$	$P[X = x] = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, k+2, \dots$

**Exemplo 4.1 (continuação):** A variável aleatória  $X$  é o número de abacaxis inspecionados (repetições do experimento) até que se encontre dois ( $k=2$ ) em bom estado (sucesso); a probabilidade de um abacaxi qualquer selecionado estar em bom estado (probabilidade de sucesso) é  $p=0,8$ . Assim, a probabilidade de que a pessoa precise analisar 5 abacaxis para conseguir dois em bom estado é, pelo modelo Binomial Negativo( $k=2,p=0,8$ ),

$$P[X = 5] = \binom{5-1}{2-1} 0,8^2 (1-0,8)^{5-2} = \binom{4}{1} 0,8^2 \cdot 0,2^3 = 4 \cdot (0,64)(0,008) = 0,02048.$$

A probabilidade de que a pessoa precise analisar pelo menos 4 abacaxis é

$$P[X \geq 4] = \sum_{x=4}^{\infty} P[X = x] = 1 - P[X \leq 3] = 1 - (P[X = 2] + P[X = 3]) = 1 - (0,64 + 0,256) = 1 - 0,896 = 0,104.$$

Qual seria o “número esperado de abacaxis analisados”?

<b>Esperança e Variância da Binomial Negativa</b>	Seja $X \sim BN(k, p)$ . A esperança e a variância de X são, respectivamente, $E[X] = \frac{k}{p} \quad e \quad Var[X] = \frac{k(1-p)}{p^2}.$
---	--

**Exemplo 4.1 (continuação):** A esperança e a variância do número de abacaxis analisados é

$$E[X] = \frac{2}{0,8} = 2,5 \quad e \quad Var[X] = \frac{2(1-0,8)}{0,8^2} = 0,625.$$

Espera-se analisar entre 2 e 3 abacaxis até encontrar dois que não estejam estragados.

### 4.3. Outros Exemplos

**Exemplo 4.2:** Um cientista inocula os germes de certa doença em ratos até que dois contraíam a doença. Se probabilidade de um rato contrair a doença em cada inoculação é de 1/6, calcule a probabilidade de que oito ratos sejam necessários.

**Solução:** Podemos assumir que a variável aleatória

X: o número de ratos inoculados até que dois contraíam a doença, segue o modelo Binomial Negativo com  $k=2$  e  $p=1/6$ . Assim, a função de probabilidade de X seria dada por

$$P[X = x] = \binom{x-1}{2-1} \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{x-2}, \quad x = 2, 3, \dots$$

O número esperado de ratos a serem inoculados até que dois contraíam a doença é de  $k/p = 2/(1/6) = 12$ .

A probabilidade requerida é

$$P[X = 8] = \binom{8-1}{2-1} \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{8-2} = \binom{7}{1} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^6 \approx 0,065.$$

Pensando no planejamento do experimento, talvez a questão mais interessante fosse do tipo: qual é a probabilidade de que mais de oito ratos sejam necessários?

$$\begin{aligned} P[X > 8] &= 1 - P[X \leq 8] = 1 - \{P[X = 2] + P[X = 3] + \dots + P[X = 8]\} \\ &= 1 - \{0,028 + 0,046 + \dots + 0,065\} = 0,605 \end{aligned}$$

Ou seja, a probabilidade de que sejam necessários mais do que oito ratos é grande.

**Exemplo 4.3:** Sabe-se que, em certo aeroporto, cerca de 25% dos passageiros inspecionados têm objetos proibidos em sua bagagem. Qual é a probabilidade de que, em uma fila, 10 passageiros tenham que ser revistados até que seja encontrada uma bagagem com objetos proibidos?

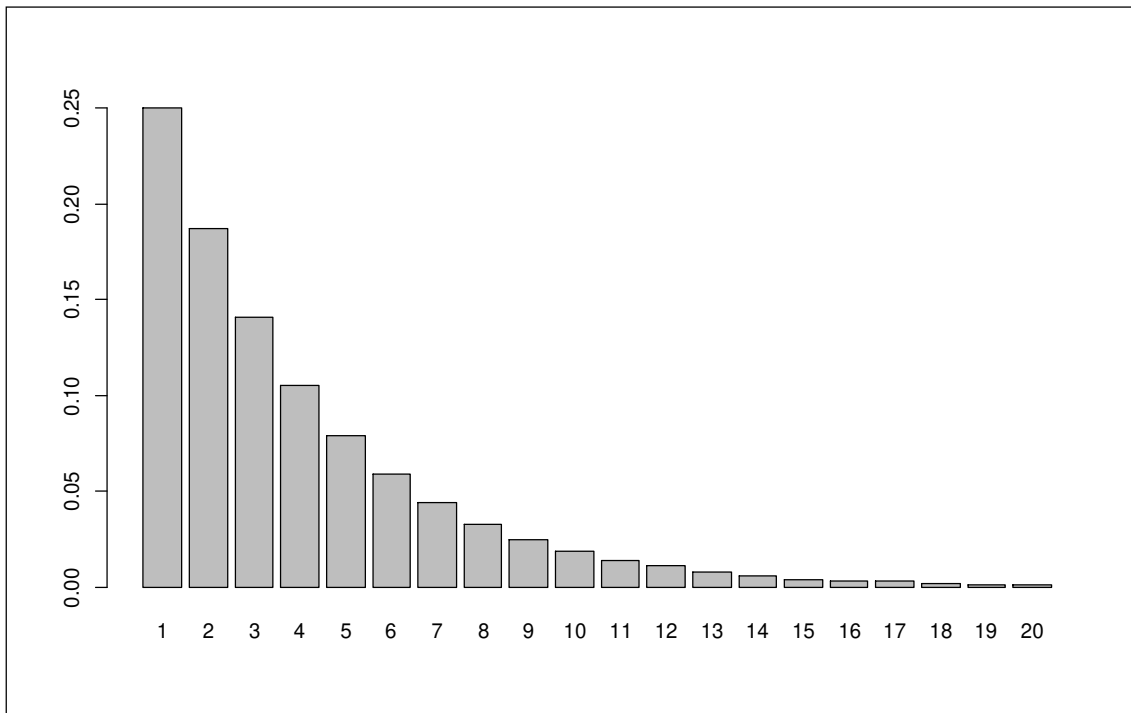
**Solução:** Seja X o número de passageiros que passam na inspeção até que seja encontrada uma bagagem com objetos proibidos. Se pudermos supor que a probabilidade de trazer objetos proibidos é a mesma para todos os passageiros e que os passageiros cometem essa infração de forma independente uns dos outros, a variável aleatória X segue o modelo Binomial Negativo com  $k=1$  e  $p=0,25$ . Assim, a função de probabilidade de X é dada por

$$P[X=x]=\binom{x-1}{1-1}(0,25)^1(0,75)^{x-1}, \quad x=1, 2, 3, \dots$$

A probabilidade de que 10 passageiros sejam inspecionados até que seja encontrada uma bagagem com objetos proibidos é

$$P[X=10]=\binom{10-1}{0}(0,25)^1(0,75)^{10-1}=1 \cdot (0,25)^1 \cdot (0,75)^9 \approx 0,02.$$

A Figura 4.1 mostra a função de probabilidade do número de passageiros que passam na inspeção até que seja encontrada uma bagagem com objetos proibidos, quando a probabilidade de que um passageiro tenha objetos escondidos seja de 0,25.



**Figura 4.1:** Função de Probabilidades da Binomial Negativa  $k=1$  e  $p=0.25$  ( $x$  variando de 1 a 20).

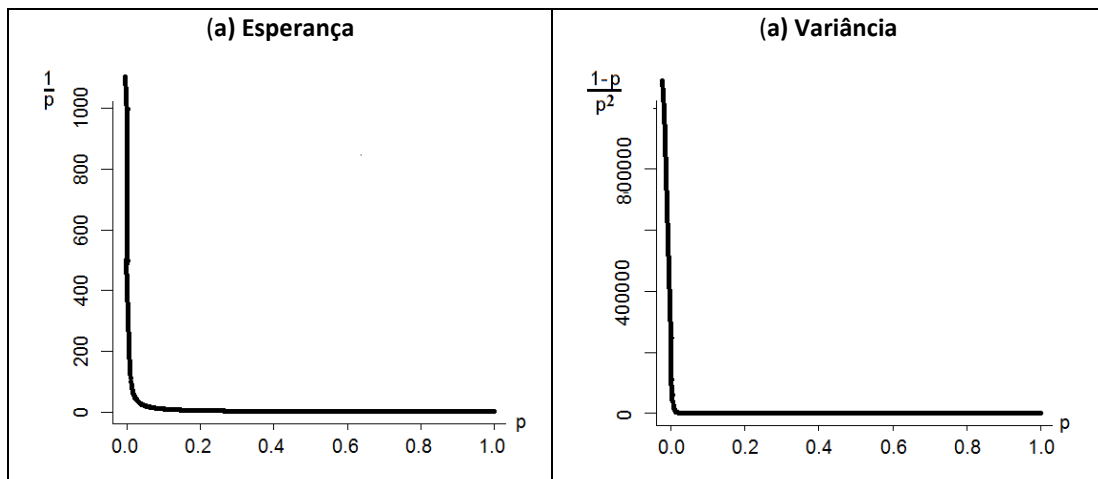
#### 4.4. Caso Especial: O Modelo Geométrico

A variável aleatória Binomial Negativa com  $k=1$  é mais conhecida como Geométrica.

<b>Variável Aleatória Geométrica</b> $X \sim G(p)$	Seja $X$ o número de repetições de um experimento $Be(p)$ até que ocorra um sucesso. A função de probabilidade de $X$ é dada por $P[X=x]=p(1-p)^{x-1}, \quad x=1, 2, 3, \dots$
---	--

<b>Esperança e Variância da Geométrica</b>	Seja $X \sim BN(k,p)$ . A esperança e a variância de $X$ são, respectivamente, $E[X]=\frac{1}{p} \quad e \quad Var[X]=\frac{1-p}{p^2}.$
--	---

Note que, quanto menor a probabilidade de sucesso, maior será o número esperado de tentativas até que se obtenha o primeiro sucesso (Figura 4.2). A variância de  $X$  também decresce rapidamente com o aumento de  $p$ .



**Figura 4.2:** Ilustração dos valores de esperança (a) e variância (b) da variável aleatória Geométrica em função da probabilidade de sucesso  $p$ .

**Exemplo 4.4:** No horário de pico, uma central telefônica está muito próxima de sua capacidade máxima e o usuário tem dificuldade para completar sua chamada. De fato, apenas 10% das chamadas são completadas neste horário. É interessante conhecer a distribuição de probabilidade do número de tentativas até que o usuário consiga completar sua chamada. Qual é a probabilidade de que sejam necessárias cinco tentativas para se conseguir completar uma chamada?

**Solução:** O número de tentativas até a chamada ( $X$ ) tem distribuição Geométrica com  $p=0,10$ . A função de probabilidade de  $X$  é dada por

$$P[X=x]=\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^{x-1}, \quad x=1, 2, 3, \dots$$

A probabilidade de que sejam necessárias cinco tentativas para completar uma chamada é

$$P[X=5]=\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^4=0,06561.$$

## 5. MODELO POISSON

### 5.1. Exemplo Inicial

Considere as seguintes variáveis definidas por uma contagem de certos tipos de evento:

- número de carros que passam por um pedágio por minuto;
- número de buracos por m<sup>2</sup> de um telhado;
- número de bactérias por litro de leite;
- número de multas por veículo;
- número de filhotes em uma ninhada.

Cada uma destas contagens pode assumir valores de zero até um valor máximo indeterminado (0,1,2,3,4,...). Essa “limitação” é definida por uma unidade de tempo (minuto), área (m<sup>2</sup>), volume (litro) ou elemento (veículo/ninhada).

Para o estudo da distribuição de probabilidade deste tipo de variável aleatória, um modelo muito utilizado é o Modelo de Poisson.

### 5.2. Definições

<b>Variável Aleatória Poisson</b> $X \sim P(\lambda)$	Se $X$ , o número de eventos ocorridos em uma unidade de tempo (espaço, volume, etc.), segue o modelo Poisson com parâmetro $\lambda$ , a função de probabilidade de $X$ é dada por $P[X=x] = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x=0, 1, 2, \dots$
--	---

<b>Esperança e Variância da Poisson</b>	Seja $X \sim P(\lambda)$ . A esperança e a variância de $X$ são, respectivamente, $E[X] = \lambda \quad e \quad Var[X] = \lambda.$
---	---

A Figura 5.1 mostra que a função de probabilidade da Poisson torna-se mais dispersa entre os valores de  $X$  à medida que se aumenta o  $\lambda$ . Note que, também, que o gráfico se torna cada vez mais simétrico à medida que  $\lambda$  aumenta<sup>2</sup>.

**Exemplo 5.2:** Suponha que se possa assumir que o número de navios petroleiros que chegam por dia a um porto tenha distribuição de probabilidade de Poisson com média igual a 10 navios-dia. As instalações do porto podem suportar até 15 navios por dia. Qual é a probabilidade de que, em um certo dia, os navios que chegarem não consigam aportar?

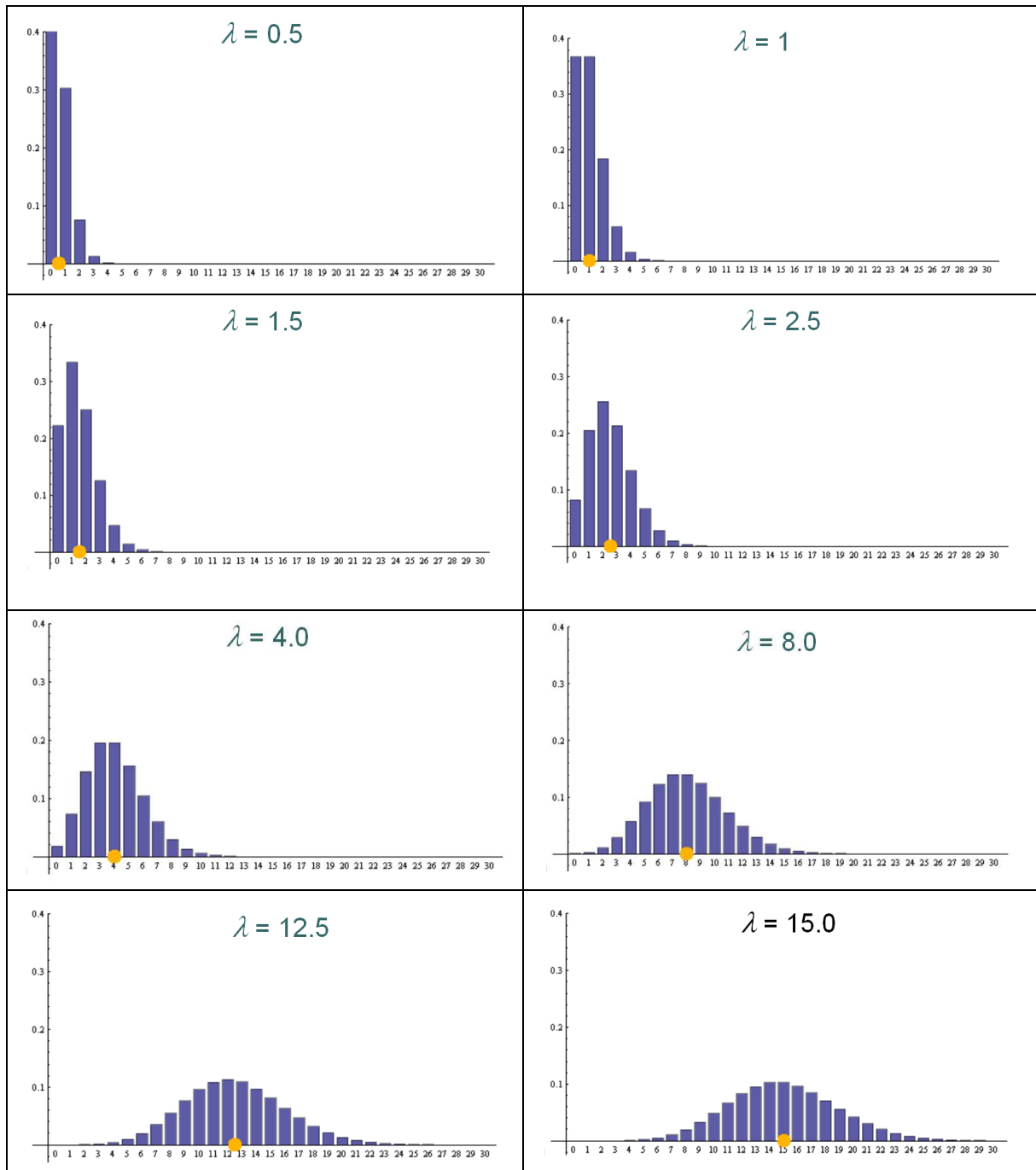
**Solução:** A função de probabilidade do número de navios petroleiros que chegam por dia ao porto ( $X$ ) é dada por

$$P[X=x] = \frac{e^{-10} 10^x}{x!}, \quad x=0, 1, 2, 3, \dots$$

A probabilidade de que cheguem mais de 15 navios em um dia é

<sup>2</sup>A distribuição Poisson se aproxima da distribuição Normal (Gaussiana) à medida que  $\lambda$  cresce.

$$P[X > 15] = 1 - P[X \leq 15] = 1 - \sum_{x=0}^{15} P[X = x] = 1 - \sum_{x=0}^{15} \frac{e^{-10} 10^x}{x!} = 1 - 0,9513 = 0,0487.$$



**Figura 5.1:** Função de Probabilidades de Poisson para vários valores de  $\lambda$ . Em amarelo, a média  $\lambda$ . Feito na plataforma WolframAlpha ([www.wolframalpha.com](http://www.wolframalpha.com)).

### 5.3. Suposições do Modelo de Poisson

A definição da variável aleatória de Poisson decorre de um experimento aleatório que atende às duas seguintes suposições:

<b>Suposição 1:</b> <b>Média constante</b>	As condições do experimento permanecem constantes no decorrer do tempo (ou no espaço), isto é, a taxa média de ocorrência de eventos, $\lambda$ , é constante ao longo do tempo (ou do espaço).
---	---

<b>Suposição 2:</b> <b>Independência</b>	Intervalos disjuntos, ou seja, sem interseção de tempo ou espaço, são independentes. Isto significa que o que ocorreu em um intervalo nada diz sobre o que vai ocorrer no outro intervalo.
---	--

Como consequência das suposições 1 e 2, temos que:

<b>Propriedade 1</b>	Se o número de eventos ocorrendo em uma unidade de tempo (ou espaço) tem distribuição de Poisson com média igual a $\lambda$ , então o número de eventos ocorrendo em $t > 0$ unidades de tempo (ou espaço) tem distribuição de Poisson com média igual a $\lambda t$ .
----------------------	---

Por exemplo, se  $X_1$ , o número de partículas emitidas por um material radioativo *por segundo*, tem distribuição Poisson com  $\lambda_1=3$ , então,  $X_2$ , o número de partículas emitidas pelo material radioativo *em dois segundos* tem distribuição Poisson com  $\lambda_2=2 \times 3=6$ .

### 5.4. Outros Exemplos

**Exemplo 5.3:** O número de carros que passam por um posto policial em uma rodovia em um intervalo de um minuto tem distribuição de Poisson com média igual a 20. Supondo que essa média se mantenha constante ao longo de 10 minutos, qual é o número médio de carros que passam pelo posto policial em 10 minutos?

**Solução:** Sejam  $X_1$  o número de carros que passam pelo posto policial em 1 minuto e  $X_2$  o número de carros que passam em 10 minutos. Se  $X_1 \sim P(20)$ , então  $X_2 \sim P(200)$ , pela *Propriedade 1*. Logo, passam, em média, 200 carros pelo posto policial em 10 minutos.

---

**Exemplo 5.4:** Assuma que o número de raios que atingem uma região por mês possa ser descrito por um modelo de Poisson com média igual a 2. Supondo que essa média se mantenha constante ao longo do ano, qual é a probabilidade de que 12 raios atinjam esta região no próximo ano?

**Solução:** Sejam  $X_1$  o número de raios que atingem a região por mês e  $X_2$  o número de raios que atingem a região por ano (12 meses). Se  $X_1 \sim P(2)$ , então  $X_2 \sim P(24)$ , pela *Propriedade 1*. A função de probabilidade de  $X_2$  é dada por



$$P[X_2 = x] = \frac{e^{-24}(24)^x}{x!}, \quad x = 0, 1, 2, 3, \dots$$

A probabilidade de que 12 raios atinjam esta região no próximo ano é

$$P[X = 12] = \frac{e^{-24}(24)^{12}}{12!} \approx 0,003.$$

## 5.5. Aproximação da Binomial pela Poisson

Apesar da definição das variáveis aleatórias Binomial e Poisson serem diferentes, as duas distribuições de probabilidade são muito semelhantes quando, na Binomial,  $n$  é muito grande e  $p$  é muito pequeno, e, na Poisson, toma-se  $\lambda=np$ .

Mais claramente, se  $X \sim B(n;p)$ , a distribuição de probabilidade de  $X$  aproxima-se da distribuição de Poisson com média  $\lambda=np$  à medida que  $n$  cresce e  $p$  decresce. Isso quer dizer que os valores de

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{e} \quad P[X = x] = \frac{e^{-np} (np)^x}{x!}$$

ficam cada vez mais próximos à medida que  $n$  cresce e  $p$  decresce.

Como exemplo, considerando  $n=100$  e  $p=0,01$ , note que a probabilidade do valor  $x=0$  calculada na Binomial(100;0.01) é

$$P[X = 0] = \binom{100}{0} (0,01)^0 (0,99)^{100} \approx 0,3660323,$$

e, na Poisson(100[0,01]=1), é

$$P[X = 0] = \frac{e^{-1} 1^0}{0!} \approx 0,3678794.$$

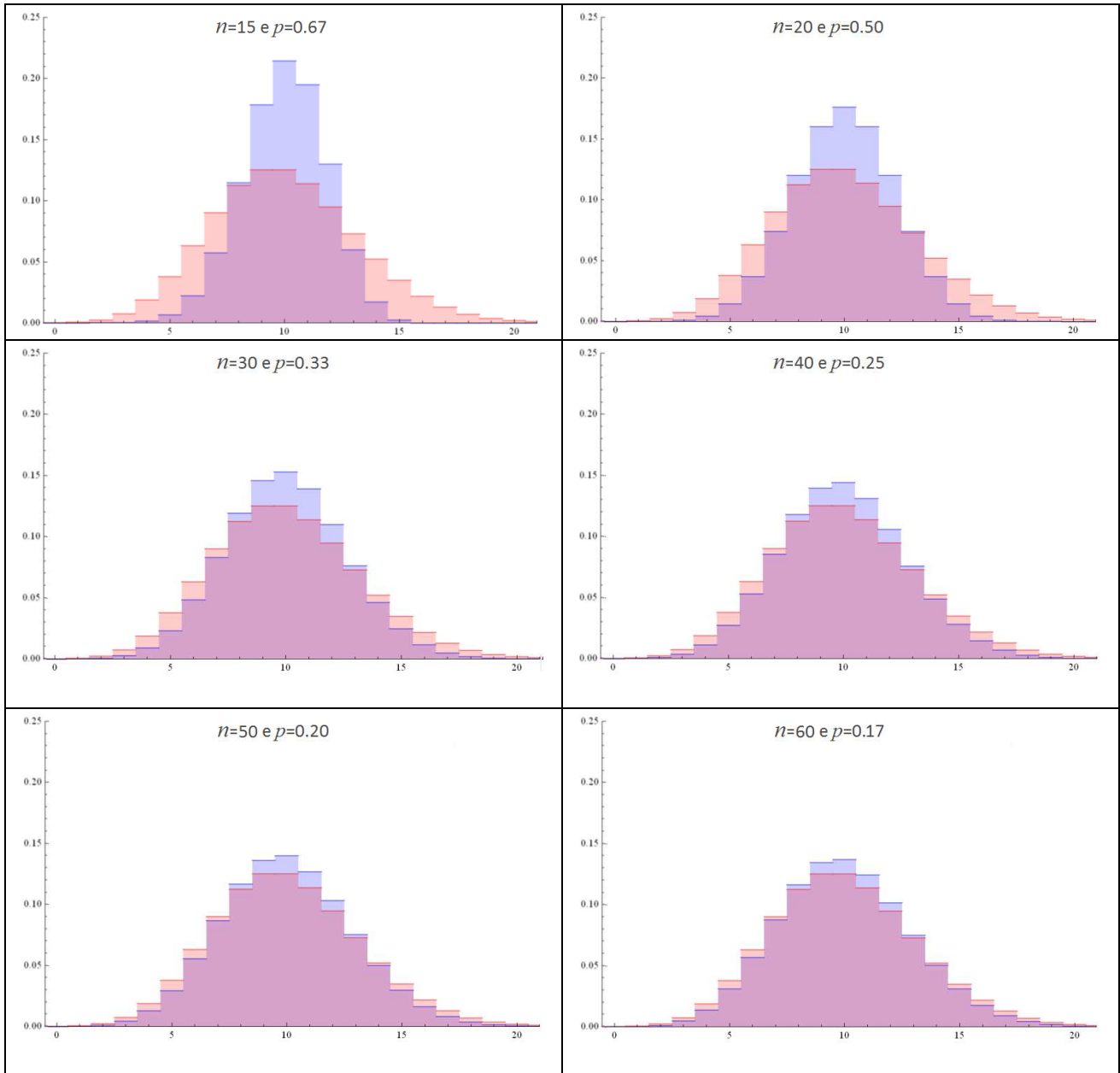
Os dois valores a partir da terceira casa decimal. Tomando  $n$  ainda maior ( $n=1000$ ) e  $p$  ainda menor ( $p=0,001$ ), as probabilidades calculadas na Binomial(1000;0,001) e na Poisson(1000[0,001]=1), respectivamente,

$$P[X = 0] = \binom{1000}{0} (0,001)^0 (0,999)^{1000} \approx 0,3676954 \quad \text{e} \quad P[X = 0] = \frac{e^{-1} 1^0}{0!} \approx 0,3678794,$$

diferem a partir da quarta casa decimal.

A Figura 5.2 mostra a aproximação entre as distribuições de probabilidade dos modelos Binomial e Poisson para diversas combinações de  $n$  e  $p$ .

**Exemplo 5.5:** O fabricante um uma bicicleta infantil recebeu diversas reclamações de um defeito no freio. Teste preliminares haviam determinado que este tipo de defeito ocorreria em 1 a cada 10 mil produtos, ou seja, com probabilidade igual a 0,0001. Decidiu-se retirar uma amostra aleatória de 200 bicicletas da produção, encontrando-se 5 delas com defeito no freio. Comente a afirmação do fabricante de que tal defeito ocorre “em 1 a cada 10 mil” usando um argumento probabilístico. Use a aproximação da Binomial pela Poisson.



**Figura 5.2:** Representação gráfica da aproximação entre as funções de probabilidades Binomial( $n,p$ ), em azul, e Poisson( $np$ ), em rosa, para valores crescentes de  $n$  e decrescentes de  $p$ . Feito na plataforma WolframAlpha ([www.wolframalpha.com](http://www.wolframalpha.com)).

**Solução:** Seja  $X$  a variável aleatória que representa o número de bicicletas com defeito dentre as 200 selecionadas. Se a alegação do fabricante está correta,  $X$  tem distribuição Binomial com  $n=200$  e  $p=0,0001$ , que pode ser aproximada pelo modelo de Poisson com  $\lambda=200(0,0001)=0,02$ . Desse modo, a probabilidade de que 5 bicicletas com defeito no freio sejam encontradas em 200 bicicletas quando a probabilidade de defeito é de 1 em 10 mil é

$$P[X = 5] = \binom{200}{5} (0,0001)^5 (0,9999)^{195} \approx \frac{e^{-0,02} (0,02)^5}{5!} = 3 \times 10^{-11},$$

uma probabilidade muito baixa. Logo, a alegação do fabricante não parece estar certa. Além disso, o número esperado de bicicletas com defeito em um lote de 200 bicicletas, caso o proprietário estivesse certo, seria  $200(0,0001) = 0,02$ , o que está bem abaixo do número de bicicletas defeituosas realmente encontradas.

## 5.6. Verificação da Adequação do Modelo de Poisson

No caso dos modelos Binomial, Hipergeométrico e Binomial Negativo, a descrição do experimento e da variável aleatória já nos leva à definição do modelo. Entretanto, no modelo Poisson, fazemos as suposições 1 e 2, mas não sabemos ao certo de elas são cumpridas e, portanto, se o modelo Poisson descreve bem a função de probabilidade da variável aleatória. Neste caso, a adequação do modelo de Poisson para descrever a função de probabilidade de uma contagem deve ser checada por meio de uma amostra aleatória de observações da variável.

**Exemplo 5.6:** Estatísticos de uma companhia telefônica estudam se o modelo de Poisson com média igual a 4,5 pode ser ajustado ao número de chamadas interestaduais que chegam, por hora, à central telefônica, durante o período noturno. Para tanto, eles registraram o número de chamadas em 650 intervalos de uma hora. O número de intervalos de uma hora em que foram recebidas  $k$  chamadas ( $k=0, 1, 2, \dots$ ) é mostrado no quadro a seguir.

No. de chamadas ( $k$ )	0	1	2	3	4	5	6	7	$\geq 8$
Frequência observada na amostra	9	38	71	115	125	106	79	50	57

Se  $X$ , o número de chamadas interestaduais por hora, tivesse distribuição Poisson ( $\lambda=4,5$ ), a probabilidade de chegarem  $k$  chamadas em uma hora, seria dada por:

$$P[X=k] = \frac{e^{-4,5}(4,5)^k}{k!}, \quad k=0, 1, 2, \dots, 7,$$

$$P[X \geq 8] = 1 - P[X \leq 7] = 1 - (P[X=0] + P[X=1] + \dots + P[X=7]).$$

Os valores destas probabilidades são dados no quadro a seguir.

No. de chamadas ( $k$ )	0	1	2	3	4	5	6	7	$\geq 8$
$P[X=k]$ se $X \sim P(4,5)$	0,011	0,050	0,112	0,169	0,190	0,171	0,128	0,082	0,087

A *frequência esperada* de  $k$  chamadas por hora ( $k=0, 1, 2, \dots$ ), no modelo Poisson ( $\lambda=4,5$ ), é obtida multiplicando a probabilidade de  $k$  chamadas pelo total  $n=650$  intervalos de uma hora. Por exemplo, ao número esperado de intervalos de uma hora com  $k=0$  chamadas é calculado como  $nP[X=0] = 650(0,011) \approx 7$ . Os demais valores são mostrados no quadro a seguir.

No. de chamadas ( $k$ )	0	1	2	3	4	5	6	7	$\geq 8$
Frequência esperada no modelo Poisson(4,5)	7	33	73	110	124	111	83	53	56

As frequências esperadas e observadas são próximas, o que levou os engenheiros à conclusão de que o modelo de Poisson ( $\lambda=4,5$ ) parece ser adequado para descrever o número de chamadas interestaduais que chegam, por hora, à central telefônica, durante o período noturno.

Esta verificação gráfica da adequação do modelo Poisson por meio de uma amostra é complementada por um procedimento de inferência estatística chamado **teste de aderência**, que não é abordado neste texto. Para maiores detalhes, ver Triola (2013).

## REFERÊNCIAS BIBLIOGRÁFICAS

- Montgomery, D.C., Runger, G.C. (2012) *Estatística Aplicada e Probabilidade para Engenheiros* (5ª edição). Editora LTC.
- Triola, M.F. (2013) *Introdução à Estatística* (11ª edição). Editora LTC.
- Walpole, R. E. e colegas (2009) *Probabilidade e Estatística* (8ª edição). Editora Pearson.