

3.3.1.3 Distribuição hipergeométrica

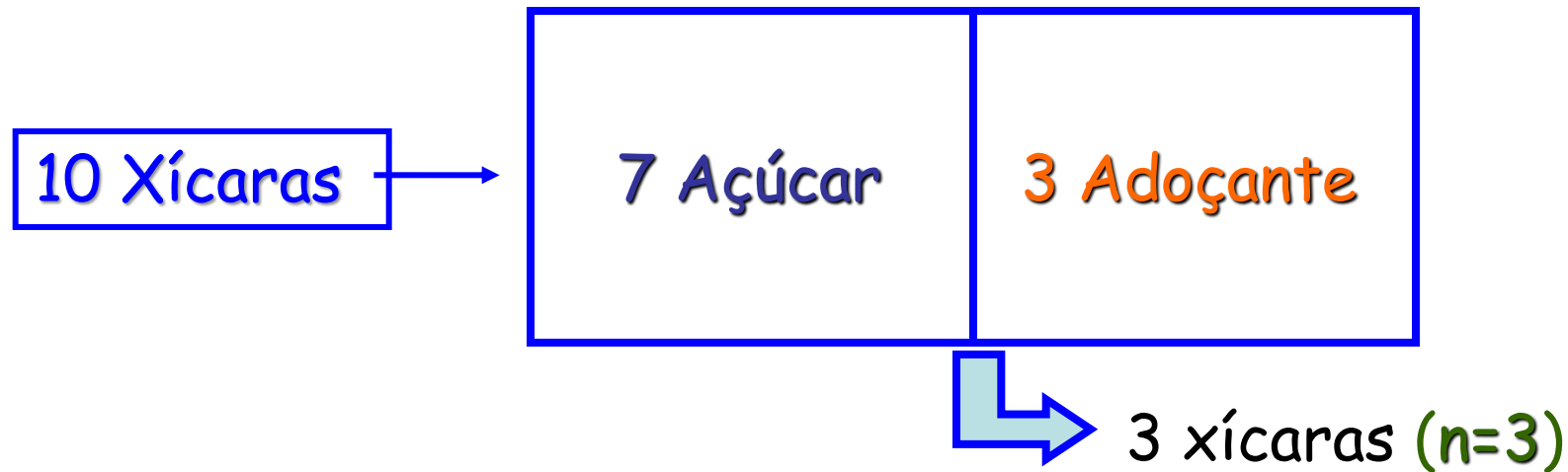
Definição: Modelo que descreve probabilisticamente os resultados de uma sequência de experimentos de Bernoulli **dependentes**. Refere-se a experimentos que se caracterizam por retiradas **sem reposição**, onde a probabilidade de sucesso **se altera** a cada retirada.

⇒ A **distribuição hipergeométrica não configura um processo de Bernoulli** porque a probabilidade de sucesso muda de um experimento para o outro

⇒ Essa distribuição é extremamente importante no contexto de **amostragem sem reposição**

Experimento: Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três com adoçante. Três pessoas se aproximam da bandeja e se servem.

Se a variável aleatória X é definida como o número de xícaras de café com açúcar, construa a distribuição de probabilidade de X .



$$S = \{A_1A_2A_3, A_1A_2A_4, \dots, A_1A_2a_1, \dots, a_1a_2a_3\} \quad \begin{array}{l} A = \text{açúcar} \\ a = \text{adoçante} \end{array}$$

$$\#S = C_{10}^3 = 120$$

X = número de xícaras de café **com açúcar**

$$S = \{A_1A_2A_3, A_1A_2A_4, \dots, A_1A_2a_1, \dots, a_1a_2a_3\}$$

X = número de xícaras de café **com açúcar**

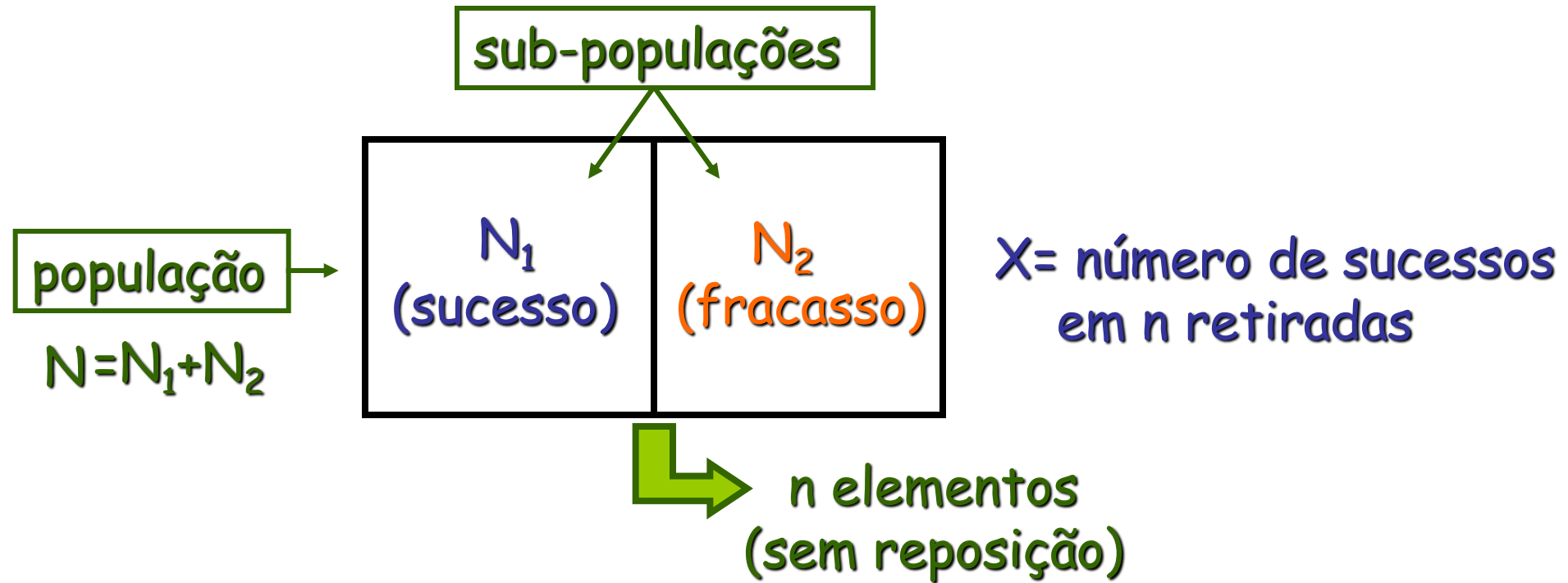
$$S_x = \{0, 1, 2, 3\}$$

$$P(X = 0) = \frac{C_7^0 C_3^3}{C_{10}^3} = \frac{1 \times 1}{120} = \frac{1}{120} = 0,008333$$

$$P(X = 1) = \frac{C_7^1 C_3^2}{C_{10}^3} = \frac{7 \times 3}{120} = \frac{21}{120} = 0,175$$

$$P(X = 2) = \frac{C_7^2 C_3^1}{C_{10}^3} = \frac{21 \times 3}{120} = \frac{63}{120} = 0,525$$

$$P(X = 3) = \frac{C_7^3 C_3^0}{C_{10}^3} = \frac{35 \times 1}{120} = \frac{35}{120} = 0,2917$$



- ⇒ Cada grupo de tamanho n formado é denominado **amostra**.
- ⇒ Todas as amostras têm a mesma chance de ocorrência.
- ⇒ Do ponto de vista probabilístico não faz diferença considerar retiradas individuais sem reposição ou retirada conjunta de grupos

Representação tabular

$X = x$	0	1	2	3	Σ
$P(X = x)$	0,00833	0,175	0,525	0,2917	1

Representação analítica

$$P(X = x) = \frac{C_7^x C_3^{3-x}}{C_{10}^3}, \text{ para } S_x = \{0, 1, 2, 3\}$$

Função de probabilidade

De modo geral, se X é uma variável que tem distribuição hipergeométrica, sua função de probabilidade será:

$$P(X = x) = \frac{C_{N_1}^x C_{N_2}^{n-x}}{C_N^n}, \text{ para } S_x = \{\max(0, n-N_2), \dots, \min(n, N_1)\}$$

Parâmetros

$$P(X = x) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}}$$

Diagram illustrating the parameters of the hypergeometric distribution. Red circles highlight the terms $\binom{N_1}{x}$, $\binom{N_2}{n-x}$, and $\binom{N}{n}$ in the formula. Red arrows point from these terms to a red-bordered box labeled "parâmetros".

A distribuição hipergeométrica tem três parâmetros:

n = número de repetições do experimento de Bernoulli

N = tamanho da população

N_1 = tamanho da subpopulação de interesse

$$X \sim \text{Hip}(n, N, N_1)$$

X tem distribuição hipergeométrica com parâmetros n , N e N_1

Medidas descritivas

♦ Média ou valor esperado

Binomial

$$E(X) = n\pi$$

$$V(X) = n\pi(1-\pi)$$

$$E(X) = \mu = \sum_{x \in S_X} x p(x)$$

Teorema: $E(X) = \mu = n \left(\frac{N_1}{N} \right)$ ← probabilidade de sucesso

♦ Variância

$$V(X) = \sigma^2 = E(X^2) - \mu^2$$

Teorema: $V(X) = \sigma^2 = n \frac{N_1}{N} \left(\frac{N_2}{N} \right) \left(\frac{N-n}{N-1} \right)$

probabilidade de fracasso
Fator de correção para populações finitas

População finita é aquela que pode ser esgotada por processo de amostragem.

⇒ Uma população será considerada finita quando tiver um **número finito** de elementos **e** a amostragem for efetuada **sem reposição**.

População infinita é aquela que não se esgota por processo de amostragem.

⇒ Uma população será considerada infinita quando tiver um **número infinito** de elementos **ou** quando amostragem for efetuada **com reposição**.

Medidas descritivas

♦ Média ou valor esperado

Teorema: $E(X) = n \frac{N_1}{N}$

A dependência não altera a média, mas altera a variância

♦ Variância

Teorema: $V(X) = n \frac{N_1}{N} \frac{N_2}{N} \left(\frac{N-n}{N-1} \right)$

Fator de correção é irrelevante para N grande

No exemplo:

$X = x$	0	1	2	3	Σ
$P(X = x)$	0,00833	0,175	0,525	0,2917	1

$$E(X) = \mu = \sum_{x \in S_X} x p(x)$$

$$E(X) = \mu = 0 \times 0,00833 + 1 \times 0,175 + 2 \times 0,525 + 3 \times 0,2917 = 2,1$$

$$V(X) = \sigma^2 = E(X^2) - \mu^2 = 4,9 - 2,1^2 = 0,49$$

$$E(X^2) = 0^2 \times 0,00833 + 1^2 \times 0,175 + 2^2 \times 0,525 + 3^2 \times 0,2917 = 4,9$$

No exemplo: $X \sim \text{Hip} (n=3, N=10, N_1=7)$

$$E(X) = n \frac{N_1}{N} = 3 \frac{7}{10} = 2,1 \text{ xícaras com açúcar}$$

Significado: Se o experimento (escolher três xícaras) for repetido um grande número de vezes, o número médio de sucessos (xícaras de café com açúcar) obtidos será 2,1.

$$V(X) = n \frac{N_1}{N} \frac{N_2}{N} \left(\frac{N-n}{N-1} \right) = 3 \frac{7}{10} \frac{3}{10} \left(\frac{10-3}{10-1} \right) = 0,49 \text{ xícaras com açúcar}^2$$

$$\sigma = \sqrt{V(X)} = \sqrt{0,49} = 0,7 \text{ xícaras com açúcar}$$

Significado: Se o experimento for repetido um grande número de vezes, a variação média do número de sucessos em relação ao valor esperado será 0,7.

3.3.1.4 Distribuição de Poisson



Maun
2. 1840. 1840.

Siméon D. Poisson
(1781 - 1840)

A distribuição de Poisson foi assim designada em homenagem ao matemático e físico francês Siméon Denis Poisson.

Modelo de descrição de acontecimentos de ocorrência rara

Exemplo clássico da **distribuição de Poisson**: número de soldados do exército da Prússia mortos anualmente por coice de cavalo, entre 1875 e 1894.

3.3.1.4 Distribuição de Poisson

Definição: descreve probabilisticamente a sequência de um **grande número** de fenômenos **independentes** entre si, cada um com probabilidade de sucesso **muito pequena**.

⇒ Ocorre naturalmente quando se deseja contar o número de eventos que ocorrem por unidade de **tempo** ou de **superfície** ou de **volume**.

⇒ Pode ser considerada como uma binomial onde o número de experimentos (**n**) é grande, **π** é pequeno (sucesso raro) e **$n\pi$** (média de sucessos) é constante.

Distribuição de Poisson ⇒ processo infinito de Bernoulli

Exemplos:

- número de peças defeituosas observadas em uma linha de produção num determinado período de tempo;
 - número de acidentes de trabalho ocorridos numa grande empresa num determinado período de tempo;
 - número de ciclones ocorridos em certa região num determinado período de tempo;
 - número de formigueiros por unidade de área em uma região;
 - número de bactérias por unidade de área em uma lâmina com extratos de uma planta;
 - número de espermatozoides inviáveis de um touro por unidade de volume de sêmen.
- ⇒ Tem inúmeras aplicações na **simulação de sistemas** modelando o número de eventos ocorridos num intervalo de tempo, quando os eventos ocorrem a uma taxa constante.

Função de probabilidade

De modo geral, se X é uma variável que tem distribuição de Poisson, sua função de probabilidade será:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \text{ para } S_x = \{0, 1, 2, \dots\}$$

espaço amostral
infinito

onde:

X : número de sucessos

$e = 2,718$ (base dos logaritmos neperianos)

λ : número médio de sucessos (sempre maior que zero)

Demonstra-se que $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ é uma função de probabilidade, provando que $\sum_{x \in S_X} p(x) = 1$. Como $S_X = \{0, 1, 2, \dots\}$, temos:

$$\begin{aligned} \sum_{x=0}^{\infty} p(x) &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) \\ &= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) \\ &= e^{-\lambda} e^{\lambda} \\ &= e^{-\lambda + \lambda} = e^0 = 1 \end{aligned}$$

Série de Taylor

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

Parâmetros

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \leftarrow \text{parâmetro}$$

A distribuição de Poisson tem apenas um parâmetro:

λ = número médio de sucessos

$$X \sim \text{Poi}(\lambda)$$

X tem distribuição de Poisson com parâmetro λ

Medidas descritivas

- ♦ **Média ou valor esperado:** $E(X) = \mu = \sum_{x \in S_X} x p(x)$

Teorema: $E(X) = \mu = \lambda$

- ♦ **Variância:** $V(X) = \sigma^2 = E(X^2) - \mu^2$

Teorema: $V(X) = \sigma^2 = \lambda$

Na Poisson média e variância são iguais!!

Medidas descritivas

♦ **Média ou valor esperado:** $E(X) = \mu = \sum_{x \in S_X} x p(x)$

$$\begin{aligned}
 E(X) &= \sum_{x \in S_X} x p(x) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \\
 &= e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda}{x} \frac{\lambda^{x-1}}{(x-1)!} \\
 &= e^{-\lambda} \sum_{x=1}^{\infty} \lambda \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}, \text{ fazendo } y = x - 1, \text{ temos} \\
 &= e^{-\lambda} \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda} \lambda e^{\lambda} = e^{-\lambda+\lambda} \lambda = e^0 \lambda = \lambda
 \end{aligned}$$

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

Teorema: $E(X) = \mu = \lambda$

♦ **Variância:** $V(X) = \sigma^2 = E(X^2) - \mu^2$

$$E(X^2) = \sum_{x \in S_X} x^2 p(x) = \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x^2 \frac{\lambda}{x} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda} \lambda \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!}, \text{ fazendo } y = x - 1, \text{ temos}$$

$$= e^{-\lambda} \lambda \sum_{y=0}^{\infty} (y+1) \frac{\lambda^y}{y!} = e^{-\lambda} \lambda \sum_{y=0}^{\infty} \left(y \frac{\lambda^y}{y!} + \frac{\lambda^y}{y!} \right) = e^{-\lambda} \lambda \left(\sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} + \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \right)$$

$$= e^{-\lambda} \lambda (\lambda e^{\lambda} + e^{\lambda}) = \lambda e^{\lambda} e^{-\lambda} \lambda + e^{\lambda} e^{-\lambda} \lambda = e^0 \lambda^2 + e^0 \lambda = \lambda^2 + \lambda$$

$$V(X) = E(X^2) - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Teorema: $V(X) = \sigma^2 = \lambda$

$$\sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} = \lambda e^{\lambda}$$

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

- ◆ Coeficiente de assimetria

Teorema: $a_3 = \frac{1}{\sqrt{\lambda}}$

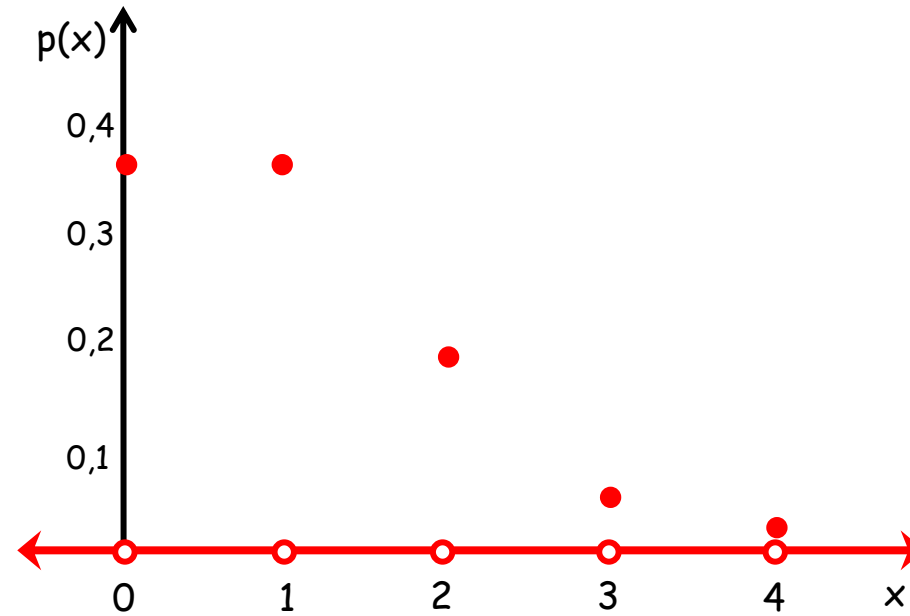
A Poisson é assimétrica positiva, tendendo para a simetria quando μ cresce.

- ◆ Coeficiente de curtose

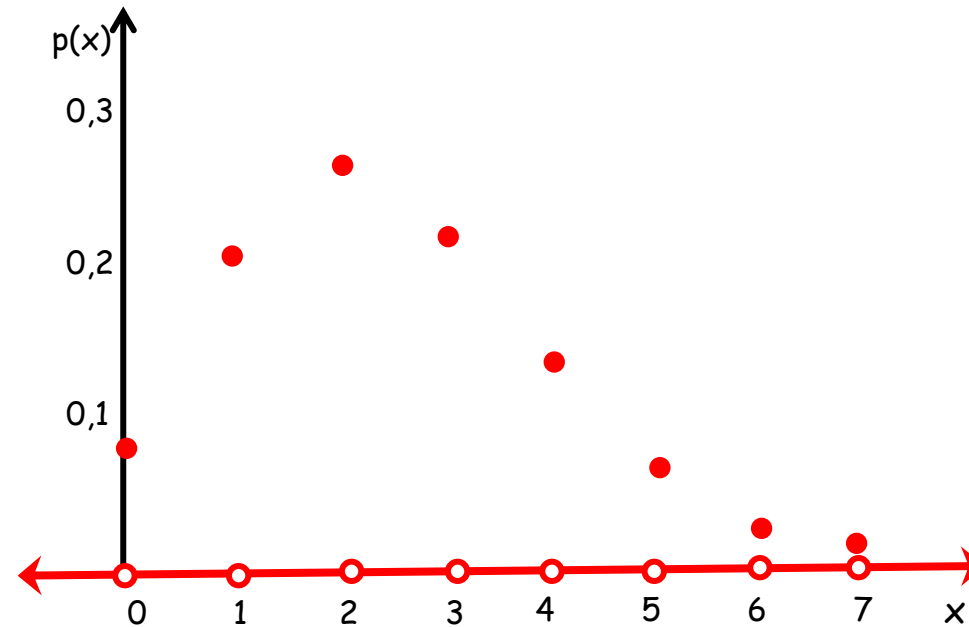
Teorema: $a_4 = \sqrt{\lambda}$

A Poisson é platicúrtica, tendendo para mesocúrtica quando μ cresce.

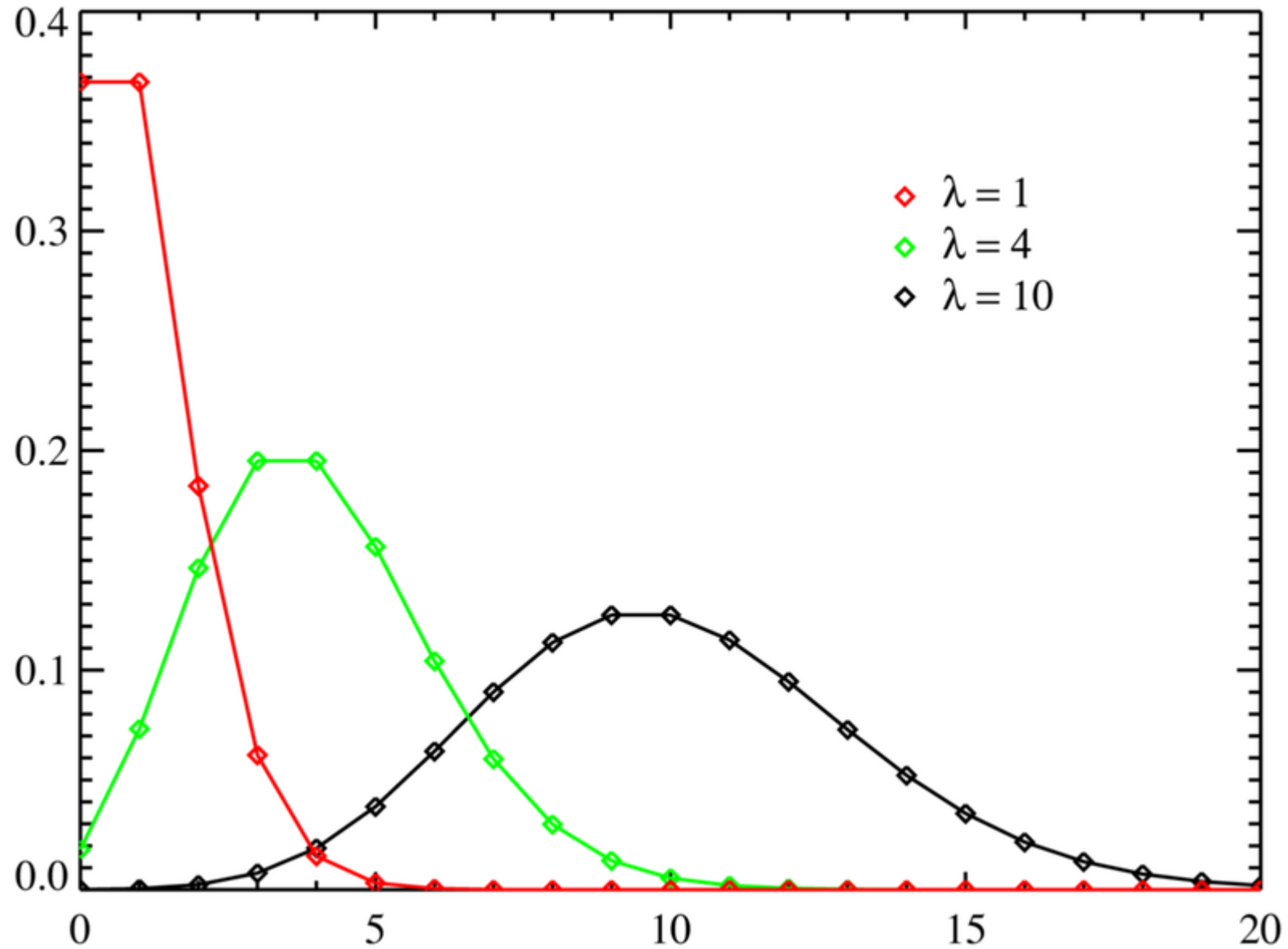
$X \sim \text{Poi} (\lambda=1)$



$X \sim \text{Poi} (\lambda=2,5)$



A Poisson tende para a simetria quando $\mu=\lambda$ cresce.



3. Distribuição hipergeométrica

Descrição probabilística de uma sequência de experimentos de Bernoulli **dependentes**.

Importante no contexto de amostragem **sem reposição**.

X = número de sucessos $S_X = \{\max(0, n - N_2), \dots, \min(n, N_1)\}$

Função de probabilidade

$S_X = \{0, 1, 2, \dots, n\}$

$$P(X = x) = \frac{C_{N_1}^x C_{N_2}^{n-x}}{C_N^n}$$

Parâmetros: n = número de repetições do experimento

N = tamanho da população

N_1 = tamanho da subpopulação de interesse

Medidas descritivas

$$E(X) = \mu = n \frac{N_1}{N}$$

$$V(X) = \sigma^2 = n \frac{N_1}{N} \frac{N_2}{N} \left(\frac{N-n}{N-1} \right)$$

4. Distribuição de Poisson

Descrição probabilística da sequência de um **grande número** de fenômenos **independentes**, com probabilidade de sucesso **muito pequena**, que ocorrem por unidade de **tempo** ou de **superfície** ou de **volume**, com média de sucessos constante.

Importante na descrição de eventos de ocorrência rara.

X = número de sucessos $S_X = \{0, 1, 2, \dots\}$ ← espaço amostral infinito

Função de probabilidade

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Parâmetro: λ = número médio de sucessos

Medidas descritivas

$$E(X) = \mu = \lambda$$

$$V(X) = \sigma^2 = \lambda$$

$$a_3 = \frac{1}{\sqrt{\lambda}}$$

$$a_4 = \sqrt{\lambda}$$

Bibliografia

BUSSAB, W.O.; MORETTIN, P.A. Estatística Básica. São Paulo: Saraiva. 2006. 526p.

FERREIRA, D.F. Estatística Básica. Lavras: Editora UFLA, 2005, 664p.

FREUND, J.E., SIMON, G.A. Estatística Aplicada. Economia, Administração e Contabilidade. 9.ed., Porto Alegre: Bookman, 2000. 404p.

MEYER, P. L. Probabilidade: aplicações à estatística. Rio de Janeiro: LTC, 1976.

MLODINOW, L. O andar do bêbado. Como o acaso determina nossas vidas. Rio de Janeiro: Editora Zahar, 2009, 264p.

SILVEIRA JÚNIOR, P., MACHADO, A.A., ZONTA, E.P., SILVA, J.B. da Curso de Estatística. v.2, Pelotas: Universidade Federal de Pelotas, 1992. 234p.

Sistema Galileu de Educação Estatística. Disponível em:
<http://www.galileu.esalq.usp.br/topico.html>