Variáveis aleatórias discretas

- Conceito
- Função de probabilidade e função de distribuição
- Valor esperado e variância

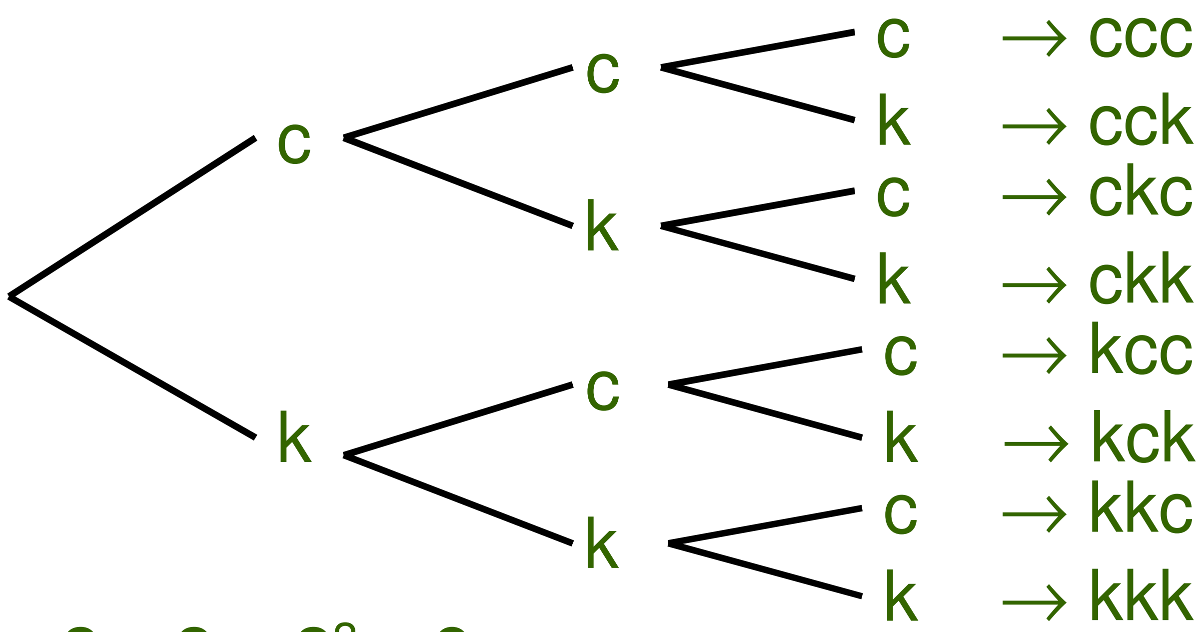
Distribuições de probabilidade

- Binomial
- Hipergeométrica
- Poisson

Variáveis aleatórias

Experimento aleatório: Lançamento de uma moeda honesta três vezes e observação das faces que ocorrem.

Diagrama em árvore



$\#S = 2 \text{ x } 2 \text{ x } 2 = 2^3 = 8$

$S = \{ccc, cck, ckc, kcc, kkc, kck, ckk, kkk\}$

⇨ Ferramental matemático se amplia consideravelmente se o espaço amostral for numérico

Experimento aleatório: Lançamento de uma moeda honesta três vezes e observação das faces que ocorrem.

$$S = \{ccc, cck, ckc, kcc, kkc, kck, ckk, kkk\}$$

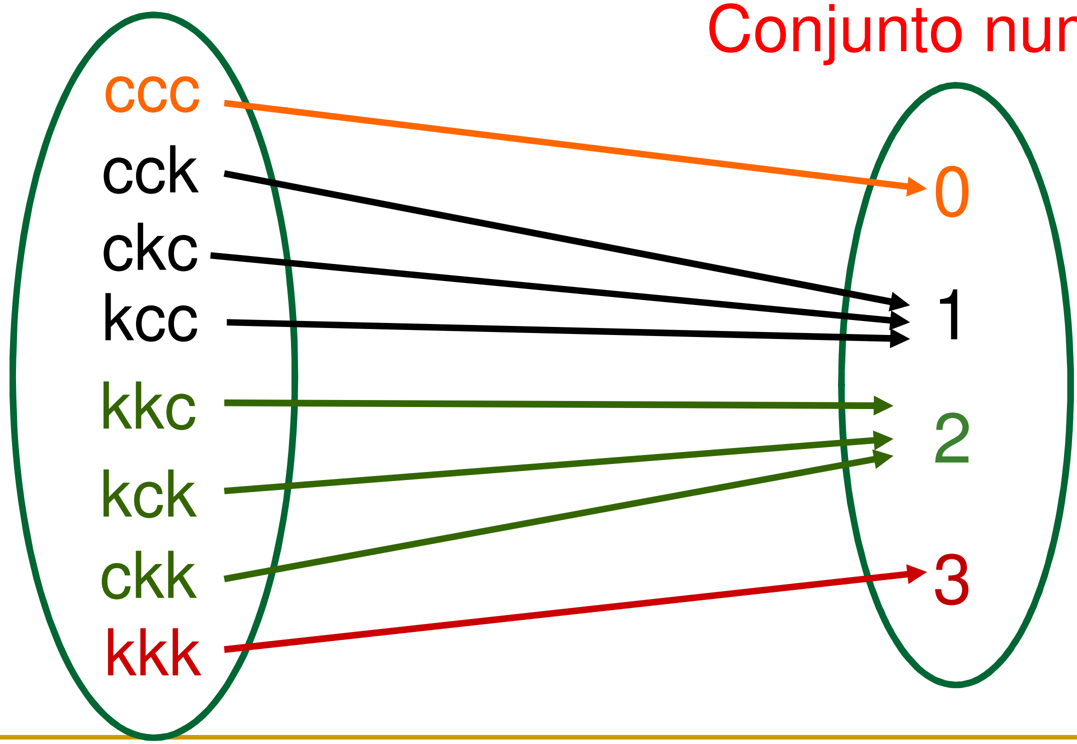
X = número de caras ocorrido nos três lançamentos

Quais são os possíveis valores de X? $X = \{0, 1, 2, 3\}$

Conjunto não numérico

Conjunto numérico

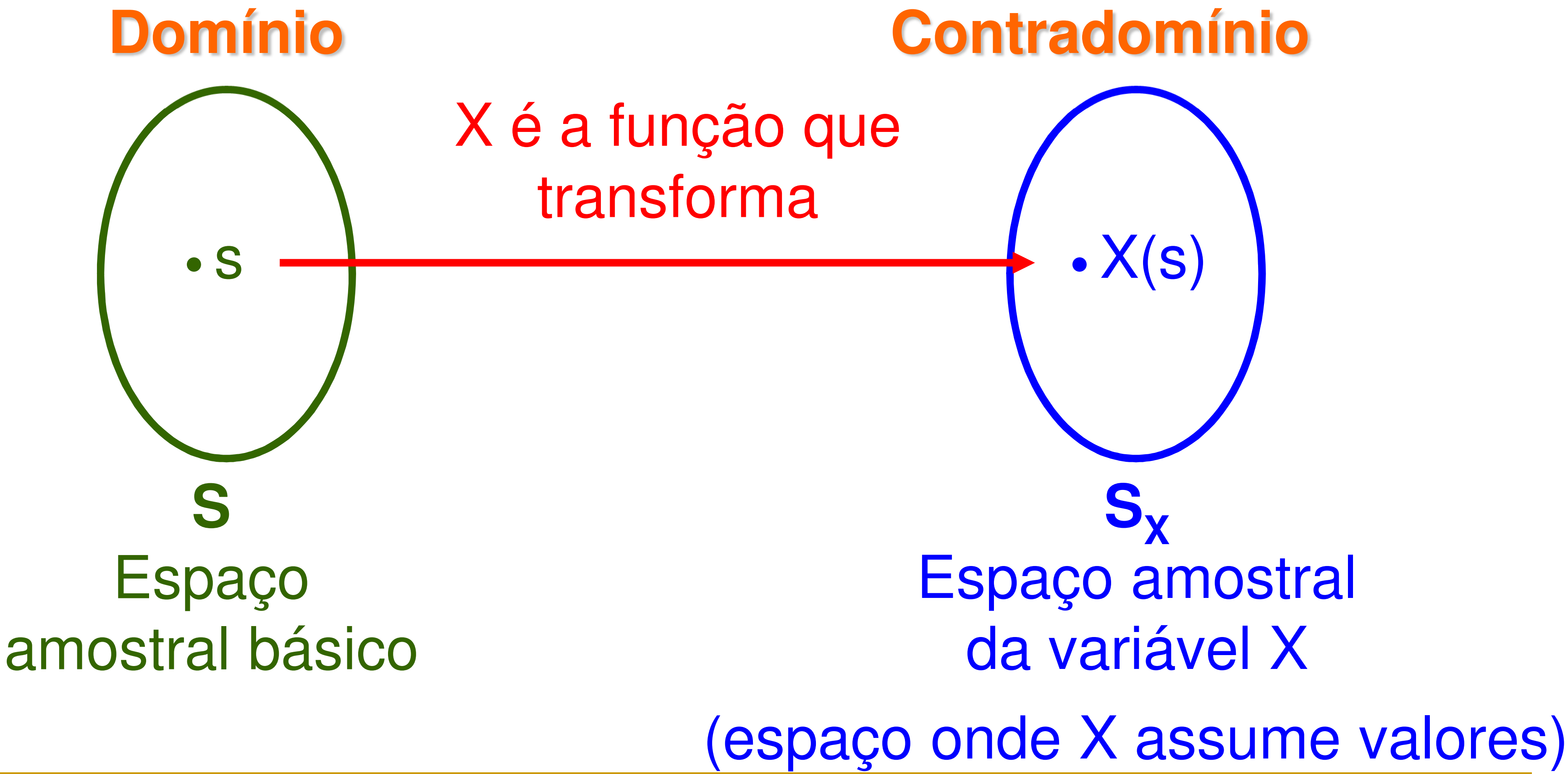
$X(ccc) = 0$
$X(cck) = 1$
$X(ckc) = 1$
$X(kcc) = 1$
$X(kkc) = 2$
$X(kck) = 2$
$X(ckk) = 2$
$X(kkk) = 3$



X é a variável que transforma um conjunto não numérico num conjunto numérico

Variável aleatória

Definição: É uma função (ou regra) que transforma um espaço amostral qualquer em um espaço amostral numérico, que será sempre um subconjunto do conjunto dos números reais.



Variáveis aleatórias $\begin{cases} \text{Discretas} \\ \\ \text{Contínuas} \end{cases}$

Variáveis aleatórias discretas

Definição: São discretas todas as variáveis cujo espaço amostral S_X é **enumerável** finito ou infinito.

Se X é uma variável aleatória discreta, então S_X é um subconjunto dos **inteiros**.

Exemplo: Lançamento de uma moeda honesta até que ocorra a face cara e observação das faces que ocorrem.

$S = \{k, ck, cck, ccck, cccck, ccccck, \ldots\}$

X = número de **coroas** até que ocorra cara

$S_X = \{0, 1, 2, 3, 4, 5, \ldots\}$

$X(k) = 0$

$X(ck) = 1$

$$S \xrightarrow{X} S_X$$

Y = número de **lançamentos** até que ocorra cara

$S_Y = \{1, 2, 3, 4, 5, 6, \ldots\}$

$Y(k) = 1$

$Y(ck) = 2$

$$S \xrightarrow{Y} S_Y$$

1. Função de probabilidade

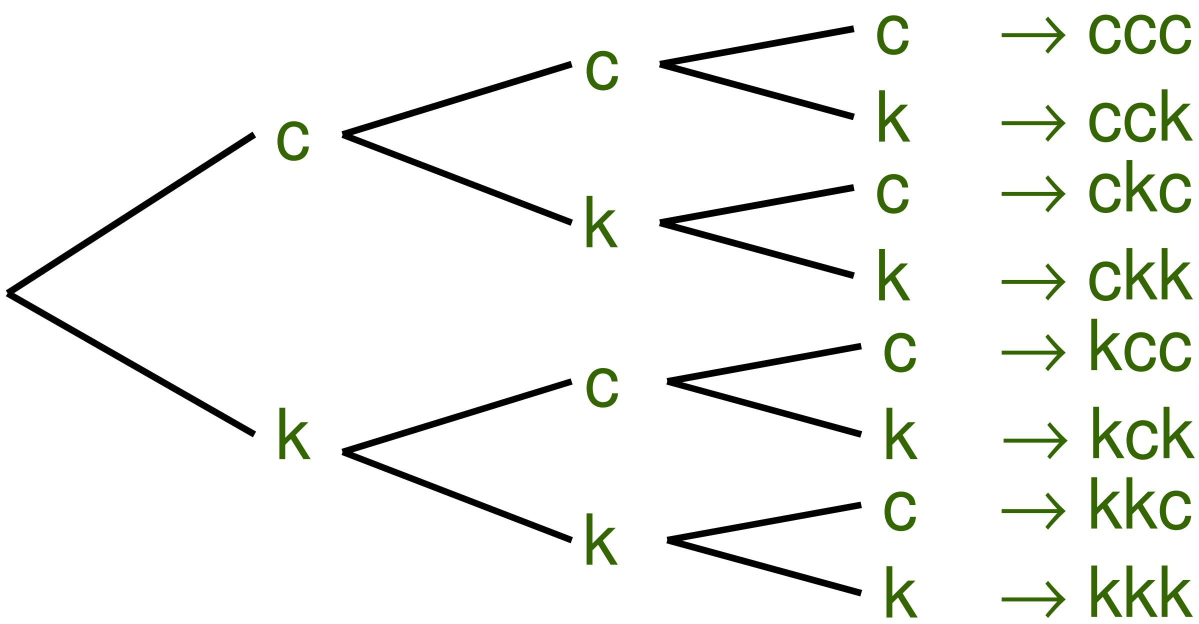
Definição: Seja X uma variável aleatória discreta e S_X o seu espaço amostral. A função de probabilidade $P(X=x)$, ou simplesmente $p(x)$, será a função que associa a cada valor de X a sua probabilidade de ocorrência, desde que atenda duas condições:

1. $p(x) \geq 0, \forall\, x \in S_X$

2. $$\sum_{x \in S_X} p(x) = 1$$

Exemplo: Lançamento de uma moeda honesta três vezes e observação das faces que ocorrem.

Diagrama em árvore



S = {ccc , cck , ckc , kcc , kkc , kck , ckk , kkk}

X = número de caras nos três lançamentos $\quad S_X = \{0,\ 1, 2, 3\}$

$$\left.\begin{aligned} p(0) &= 1/8 \\ p(1) &= 3/8 \\ p(2) &= 3/8 \\ p(3) &= 1/8 \end{aligned}\right\} \text{Primeira condição}$$

$$\frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1 \Big\} \text{Segunda condição}$$

Existem três formas de representar uma função:

- **Representação tabular:** consiste em relacionar em uma **tabela** os valores da função de probabilidade.

- **Representação gráfica:** consiste em representar **graficamente** a relação entre os valores da variável e suas probabilidades

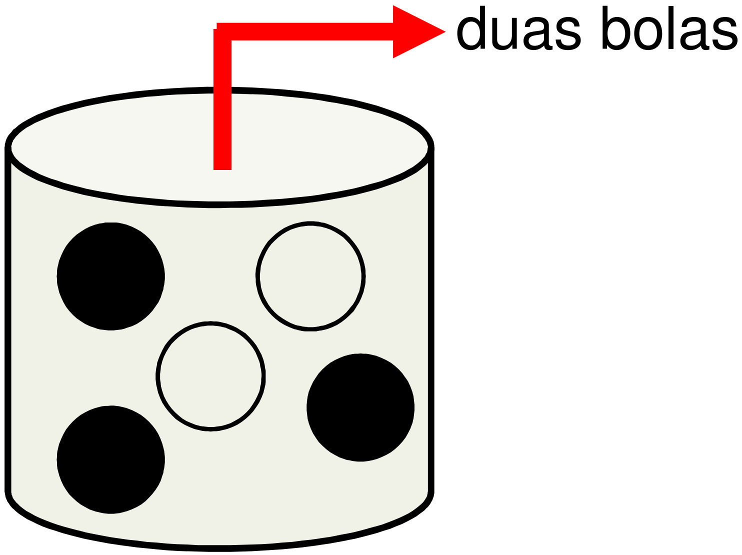
- **Representação analítica:** estabelece uma **expressão** geral para representar o valor da função de probabilidade num ponto genérico da variável

Exemplo: De uma urna com **três** bolas pretas e **duas** brancas, **retiram-se duas** bolas juntas. Se X é o número de bolas pretas retiradas, determine a função de probabilidade P(X=x).

$$\# S = C_5^2 = 10$$

$$S = \{B_1B_2,\ P_1B_1,\ P_1B_2,\ P_2B_1,\ P_2B_2,\ P_3B_1,\ P_3B_2,\ P_1P_2,\ P_1P_3,\ P_2P_3\}$$

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



$$P(X = 0) = \frac{C_3^0\ C_2^2}{C_5^2} = \frac{1}{10}$$

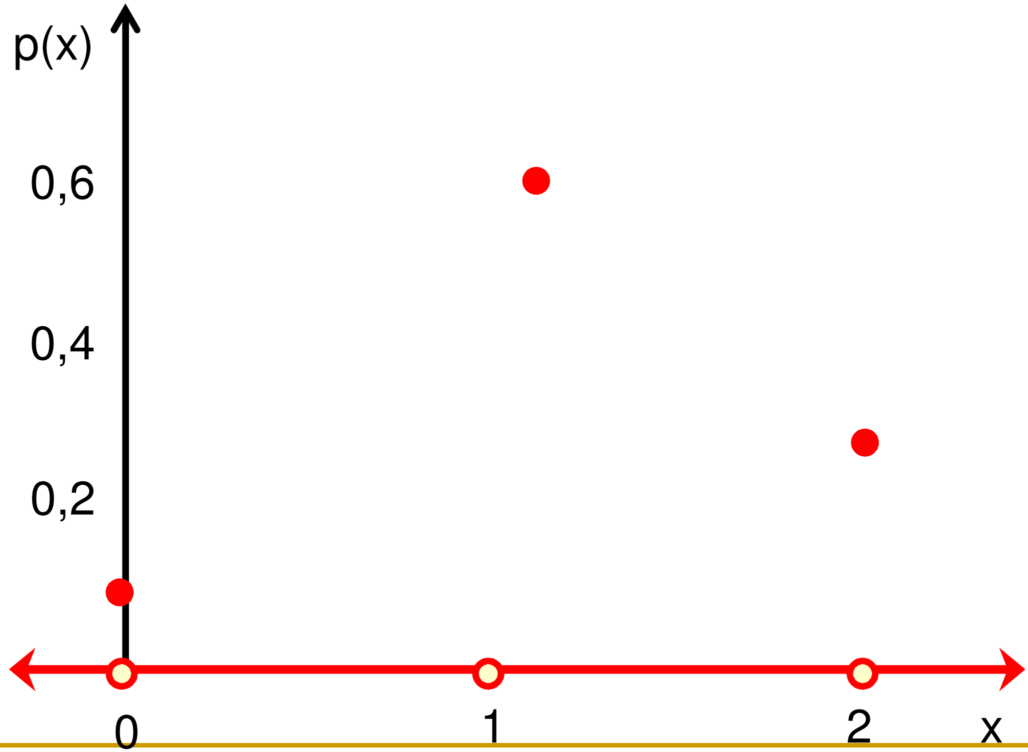
$$P(X = 1) = \frac{C_3^1\ C_2^1}{C_5^2} = \frac{6}{10}$$

$$P(X = 2) = \frac{C_3^2\ C_2^0}{C_5^2} = \frac{3}{10}$$

▫ Representação tabular

X=x	0	1	2	Σ
P(X=x)	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	1

▫ Representação gráfica



Representação analítica

$$P(X = 0) = \frac{C_3^0\ C_2^2}{C_5^2}$$

$$P(X = 1) = \frac{C_3^1\ C_2^1}{C_5^2}$$

$$P(X = 2) = \frac{C_3^2\ C_2^0}{C_5^2}$$

$$\boxed{P(X = x) = \frac{C_3^x\ C_2^{2-x}}{C_5^2}}\text{, para } S_X = \{0,\ 1,\ 2\}$$

2. Função de distribuição ou probabilidade acumulada

Definição: Seja X uma variável aleatória discreta e S_X o seu espaço amostral. A função de distribuição, denotada por $F(x)$ ou $P(X \leq x)$, é a função que associa a cada valor de X a probabilidade $P(X \leq x)$. Desta forma, temos

$$F(x) = P(X \leq x) = \sum_{X \leq x} P(X = x)$$

No exemplo:

X=x	0	1	2	Σ
P(X=x)	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	1
F(x)	$\frac{1}{10}$	$\frac{7}{10}$	1	-

$$F(x) = P(X \leq x) = \sum_{X \leq x} P(X = x)$$

$$F(0) = P(X \leq 0) = \sum_{x \leq 0} P(X = x) = P(X = 0) = \frac{1}{10}$$

$$F(1) = P(X \leq 1) = \sum_{x \leq 1} P(X = x) = P(X = 0) + P(X = 1) = \frac{1}{10} + \frac{6}{10} = \frac{7}{10}$$

$$F(2) = P(X \leq 2) = \sum_{x \leq 2} P(X = x) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \frac{1}{10} + \frac{6}{10} + \frac{3}{10} = 1$$

3. Medidas descritivas

$$S \xrightarrow{X} S_X$$

Conjunto não numérico $\xrightarrow{X}$ Conjunto numérico $\rightarrow$ Possibilita o cálculo de medidas descritivas: média, variância, etc.

No exemplo:

$$S = \{B_1B_2,\ P_1B_1,\ P_1B_2,\ P_2B_1,\ P_2B_2,\ P_3B_1,\ P_3B_2,\ P_1P_2,\ P_1P_3,\ P_2P_3\}$$

X = número de bolas pretas em duas retiradas

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

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

$X=x$	0	1	2	Σ
$P(X=x)$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	1

▫ Média ou valor esperado

Definição: Seja X uma variável aleatória discreta e S_X o seu espaço amostral. O valor médio de X, denotado por $E(X)$, ou μ_X, ou simplesmente μ, é a média dos valores de X **ponderada** pelas suas respectivas probabilidades de ocorrência. Deste modo, tem-se

$$E(X) = \mu = \frac{\sum_{x \in S_X} x\,p(x)}{\sum_{x \in S_X} p(x) = 1} = \sum_{x \in S_X} x\,p(x)$$

No exemplo:

X=x	0	1	2	Σ
P(X=x)	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	1

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

$$= 0 \times \frac{1}{10} + 1 \times \frac{6}{10} + 2 \times \frac{3}{10} = \frac{12}{10} = 1{,}2 \text{ bolas pretas}$$

Significado do valor esperado: se o experimento fosse repetido um grande número de vezes, esperaríamos que o número médio de bolas pretas escolhidas fosse 1,2.

Importante!!!

⇨ Não confundir μ_x com $\overline{X}$.

μ_x é a média de todos os valores de X (para os quais a probabilidade é conhecida)

$\overline{X}$ é a média de alguns valores de X (usualmente uma amostra de valores)

Exercício:

O tempo (em minutos) para que um operário processe certa peça é uma variável com distribuição dada na tabela abaixo.

x	2	3	4	5	6	7
p(x)	0,10	0,10	0,30	0,20	0,20	0,10

(a) Calcule o tempo médio de processamento. **(R: μ = 4,60)**

(b) Para cada peça processada o operário ganha um fixo de R$ 1,00, mas se processa a peça em menos de 6 minutos, ganha R$ 0,50 por cada minuto poupado. Por exemplo, se ele processa a peça em 5 minutos, recebe a quantia de R$ 0,50. Encontre a média de G = quantia ganha por peça (fixo + comissão).

(R: μ = 1,75)

Variância

Definição: Seja **X** uma variável aleatória discreta e S_X o seu espaço amostral. A variância de **X**, denotada por **V(X)**, ou σ_X^2, ou simplesmente σ^2, é o grau médio de dispersão dos valores de **X** em relação à sua média. Esta medida é definida como a média ou valor esperado dos quadrados dos desvios em relação à média. Deste modo, temos

$$V(X) = \sigma^2 = \boxed{E(X-\mu)^2} = \sum_{x \in S_X} (x-\mu)^2 p(x)$$

OU

$$= \boxed{E(X^2) - \mu^2}$$

onde:

$$E(X^2) = \sum_{x \in S_X} x^2 p(x)$$

$$\mu^2 = [E(X)]^2 = \left[\sum x\, p(x)\right]^2$$

No exemplo:

X=x	0	1	2	Σ
P(X=x)	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	1

$$E(X) = \mu = \frac{12}{10} = 1{,}2$$

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

$$= \sum_{x \in S_X} (x-\mu)^2 p(x)$$

$$= \left(0 - \frac{12}{10}\right)^2 \times \frac{1}{10} + \left(1 - \frac{12}{10}\right)^2 \times \frac{6}{10} + \left(2 - \frac{12}{10}\right)^2 \times \frac{3}{10} = \frac{36}{100}$$

$$V(X) = \sigma^2 = E(X^2) - \mu^2 = \frac{18}{10} - \left(\frac{12}{10}\right)^2 = \frac{36}{100} = 0{,}36 \text{ bolas pretas}^2$$

$$E(X^2) = \sum x^2 p(x)$$

$$= 0^2 \times \frac{1}{10} + 1^2 \times \frac{6}{10} + 2^2 \times \frac{3}{10} = \frac{18}{10}$$

Desvio padrão

Definição: Raiz quadrada positiva da variância.

$$\sigma = \sqrt{V(X)}$$

No exemplo:

$$\sigma = \sqrt{V(X)} = \sqrt{0{,}36} = 0{,}6 \text{ bolas pretas}$$

Significado do desvio padrão: se o experimento fosse repetido um grande número de vezes, a variação média do número de bolas pretas escolhidas em torno do valor esperado seria 0,6.

Importante!!!

- Não confundir σ^2 com s^2.

  σ^2 é a variância de **todos** os valores de X (para os quais a probabilidade é conhecida)

  s^2 é a variância de **alguns** valores de X (usualmente uma amostra de valores)

- Da mesma forma, não confundir σ com **s**.

Exercício:

Um vendedor recebe uma comissão de R$ 100,00 por uma venda. Baseado em suas experiências anteriores ele calculou a distribuição de probabilidades das vendas semanais:

x	0	1	2	3	4
p(x)	0,10	0,20	0,40	0,20	0,10

(a) Qual é o valor esperado de comissão por semana? R$ 200,00

(b) Qual é a probabilidade de ganhar pelo menos R$ 300,00 por semana? 0,30

(c) Qual o desvio padrão das vendas semanais? 1,10

Distribuições de probabilidade

O que é uma distribuição de probabilidade?

Uma distribuição de probabilidade é essencialmente um **modelo de descrição probabilística de uma população**.

população

X=x	0	1	2	Σ
P(X=x)	0,1	0,6	0,3	1

Distribuição de probabilidade

- **Parâmetros:** caracterizações numéricas que permitem a individualização de um modelo (distribuição) em determinado contexto

Distribuições discretas

1. **Distribuição de Bernoulli**
2. **Distribuição Binomial**
3. **Distribuição Hipergeométrica**
4. **Distribuição de Poisson**
5. Distribuição Multinomial
6. Distribuição Geométrica
7. Distribuição Binomial Negativa
8. Distribuição Hipergeométrica Negativa
9. Distribuição Uniforme Discreta

Distribuições de probabilidade de variáveis discretas

1. Distribuição de Bernoulli

Definição: Modelo que descreve probabilisticamente os resultados de um experimento de Bernoulli.

O experimento de Bernoulli é definido como o experimento aleatório que **possui apenas dois resultados possíveis**.

Exemplos: sexo no nascimento de um bebê, face no lançamento de uma moeda, produto perfeito ou defeituoso, satisfação ou insatisfação de um funcionário da empresa, etc.

Experimento: Um produto é avaliado quanto à qualidade

S = {perfeito, defeituoso}

Consideramos um dos resultados como sucesso:

sucesso = perfeito

fracasso = defeituoso

Se for conhecido a taxa de produtos sem defeito que é fabricada, por exemplo, 87%, concluímos que a probabilidade de o produto ser perfeito é 0,87.

O evento **{defeituoso}** é **complemento** do evento **{perfeito}**, então sua probabilidade será **1– 0,87**.

π = 0,87 = probabilidade de sucesso

1-π = 0,13 = probabilidade de fracasso

2. Distribuição binomial

Definição: Modelo que descreve probabilisticamente os resultados de uma **sequência de experimentos de Bernoulli independentes** entre si, ou seja, onde a probabilidade de sucesso é **constante** em todas as repetições do experimento.

Se $$X = Y_1 + Y_2 + \ldots + Y_n$$

onde:

Y_i ~ Ber (π) e independentes;

então, a variável X tem distribuição binomial.

Distribuição binomial ⇨ processo finito de Bernoulli

⇨ **n** experimentos de Bernoulli independentes, com probabilidade de sucesso π constante para todos eles

⇨ É importante no contexto de **amostragem com reposição**

Experimento: As peças fabricadas por uma pequena indústria podem ser consideradas perfeitas ou defeituosas. Imagine que a chance de uma peça não ter defeito algum é de 60%. Se uma peça desta fábrica é escolhida ao acaso e sua situação é registrada, temos um experimento de Bernoulli.

S = {perfeita, defeituosa}

onde:

p(perfeita) = 0,6 = π

p(defeituosa) = 1 - 0,6 = 0,4 = 1 - π

Se **três** peças são escolhidas, uma a uma, e o resultado é registrado, temos uma sequência de três experimentos de Bernoulli **independentes**, pois, a cada escolha, a probabilidade de sucesso permanecerá inalterada.

$\#S = 2^3 = 8$

P = perfeita
D = defeituosa

S = {PPP, PPD, PDP, DPP, PDD, DPD, DDP, DDD}

Sucesso = perfeita

A variável X é definida como o número de sucessos em **n** experimentos de Bernoulli independentes, com probabilidade de sucesso igual a π.

$n=3$ e $\pi=0,6$

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

Qual é a função de probabilidade P(X=x) associada a variável X?

S = {PPP, PPD, PDP, DPP, PDD, DPD, DDP, DDD}

S_X = {0,1,2,3}

P(X=x) = ?   X: nº peças perfeitas

$$P(X=0) = 0{,}4^3 = 1 \times \pi^0 \times (1-\pi)^3 = 0{,}064$$

$$P(X=1) = 3 \times 0{,}6^1 \times 0{,}4^2 = 3 \times \pi^1 \times (1-\pi)^2 = 0{,}288$$

$$P(X=2) = 3 \times 0{,}6^2 \times 0{,}4^1 = 3 \times \pi^2 \times (1-\pi)^1 = 0{,}432$$

$$P(X=3) = 0{,}6^3 = 1 \times \pi^3 \times (1-\pi)^0 = 0{,}216$$

Como podemos determinar de quantas maneiras diferentes teremos **x** sucessos e **3-x** fracassos?

$$C_3^x = \frac{3!}{x!(3-x)!}$$

Representação tabular

$X = x$	0	1	2	3	Σ
$P(X = x)$	0,064	0,288	0,432	0,216	1

Representação analítica

$$P(X = x) = \underbrace{C_3^x}_{\text{Número de casos}} \underbrace{0{,}6^x (1-0{,}6)^{3-x}}_{\text{Probabilidade de um caso}} \quad \text{, para } S_X = \{0, 1, 2, 3\}$$

Função de probabilidade

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

$$P(X = x) = C_n^x \, \pi^x (1-\pi)^{n-x} \text{, para } S_X = \{0, 1, ..., n\}$$

Parâmetros

$$P(X = x) = C_n^x \pi^x (1-\pi)^{n-x}$$

parâmetros

A distribuição binomial tem dois parâmetros:

n = número de repetições do experimento de Bernoulli

π = probabilidade de sucesso

$$X \sim \text{Bin}\,(n,\pi)$$

X tem distribuição binomial com parâmetros n e π

Medidas descritivas

- **Média ou valor esperado**

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

Teorema: $E(X) = \mu = n\pi$

- **Variância**

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

Teorema: $V(X) = \sigma^2 = n\,\pi(1-\pi)$

RESUMO - Distribuição binomial

Descrição probabilística de uma sequência de experimentos de Bernoulli independentes, ou seja, a probabilidade de sucesso é constante em todas as repetições do experimento.

Função de probabilidade

$$P(X = x) = C_n^x \, \pi^x \, (1 - \pi)^{n-x} \quad , \text{ para } S_X = \{0, 1, ..., n\}$$

Parâmetros

n = número de repetições no experimento

π = probabilidade de sucesso

Medidas descritivas

$E(X) = \mu = n\,\pi$ $\qquad$ $V(X) = \sigma^2 = n\,\pi\,(1-\pi)$

Exercício: Num determinado processo de fabricação a chance de uma peça sair defeituosa é de 10%. As peças são acondicionadas em caixas com 5 unidades cada uma.

(a) Qual a probabilidade de haver exatamente 1 peça defeituosa numa caixa? **(32,81%)**

(b) Qual a probabilidade de haver duas ou mais peças defeituosas numa caixa? **(8,14%)**

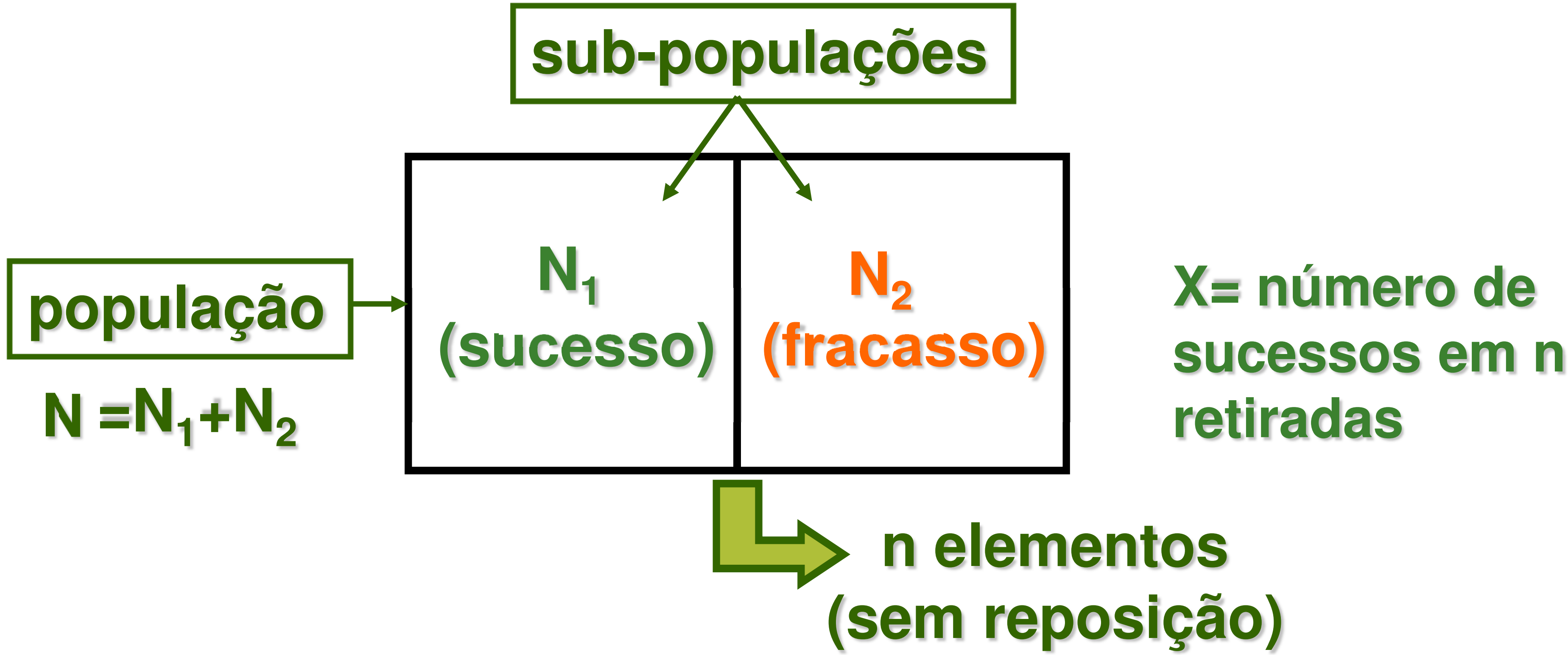
(c) Se a empresa paga uma multa de R$ 10,00 por caixa em que houver alguma peça defeituosa, qual o valor esperado da multa num total de 1000 caixas? **(R$ 4.100)**

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 se difere da Distribuição binomial** porque a probabilidade de sucesso muda de um experimento para o outro

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



- Como não há reposição, a probabilidade de sucesso (retirar elementos da sub-população de tamanho N_1) se altera a cada retirada.

- Do ponto de vista probabilístico não faz diferença considerar retiradas individuais sem reposição ou retirada conjunta de grupos

Experimento: Uma caixa contém 10 bolas coloridas: sete são verdes e três laranjas. Três bolas são retiradas da caixa, uma após a outra e sem reposição. Se a variável aleatória X é definida como o número de bolas verdes retiradas, construa a distribuição de probabilidade de X.

Bolas $\longrightarrow$

Verde $N_1=7$

Laranja $N_2=3$

$N=10$

3 bolas ($n=3$)

X = número de bolas **verdes**

$S = \{L_1L_2L_3,\ L_1L_2V_1, L_1L_2V_2,\ \ldots,\ V_5V_6V_7\}$

V = verde
L = laranja

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

X = número de bolas **verdes**

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

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

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

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

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

Representação tabular

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

Representação analítica

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

Função de probabilidade

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

$$P(X = x) = \frac{C_{N_1}^x \; C_{N_2}^{n-x}}{C_N^n} \quad \text{, para } S_X = \{0, 1, ..., n\}$$

Parâmetros

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

parâmetros

A distribuição hipergeométrica tem três parâmetros:

n = número de repetições do experimento

N = tamanho da população

N_1 = tamanho da sub-população de interesse (sucesso)

$$X \sim Hip\,(n, N, N_1)$$

X tem distribuição hipergeométrica com parâmetros n, N e N_1

Medidas descritivas

Binomial

$\mu = n\,\pi$

$\sigma^2 = n\,\pi\,(1-\pi)$

- **Média ou valor esperado**

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

Teorema: $E(X) = \mu = n\dfrac{N_1}{N}$ ← **probabilidade de sucesso**

- **Variância**

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

probabilidade de fracasso

Teorema: $V(X) = \sigma^2 = n\,\dfrac{N_1}{N}\,\dfrac{N_2}{N}\left(\dfrac{N-n}{N-1}\right)$ ← **Fator de correção**

RESUMO - Distribuição hipergeométrica

Descrição probabilística de uma sequência de experimentos de Bernoulli dependentes. Importante no contexto de amostragem sem reposição.

Função de probabilidade

$$P(X=x)=\frac{C_{N_1}^{x}\,C_{N_2}^{n-x}}{C_{N}^{n}}, \quad \text{para } S_X = \{0, 1, ..., n\}$$

Parâmetros

n = número de repetições do experimento

N = tamanho da população

N_1 = tamanho da sub-população de interesse

Medidas descritivas

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

Exercício: Pequenos motores são guardados em caixas de 50 unidades. Um inspetor de qualidade examina cada caixa testando 5 motores. Se nenhum for defeituoso, a caixa é aceita. Se pelo menos 1 for defeituoso, todos 50 são testados. Há 6 motores defeituosos numa caixa. Qual a probabilidade de que seja necessário examinar todos os motores desta caixa? **R: 0,4874**

Tem-se $N_1 = 6$, $N = 50$, $n = 5$, $P(X \geq 1) = ?$

$$P(X = 0) = \frac{C_6^0 C_{44}^5}{C_{50}^5} = \frac{1086008}{2118760} = 0{,}51257$$

$P(\text{examinar tudo}) = 1 - P(X = 0) = 0{,}48743$

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**.

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

Distribuição de Poisson ⇨ processo infinito de Bernoulli

⇨ Ocorre quando se deseja contar o número de um tipo particular de eventos que ocorrem por unidade de **tempo**, de **superfície** ou de **volume** (num espaço contínuo).

Exemplos:

- nº de peças defeituosas observadas em uma linha de produção em um dia;
- nº de acidentes de trabalho ocorridos numa grande empresa em um ano;
- nº de ciclones ocorridos em certa região em uma estação do ano;
- nº de formigueiros por km^2 em uma região;
- nº de acidentes que acontecem em 300km de uma rodovia;
- nº de carros que passam em um pedágio de uma rodovia em 30 minutos;
- nº de ligações que chegam em uma central telefônica em uma manhã.

⇨ A distribuição de Poisson tem inúmeras aplicações na **simulação de sistemas** modelando o número de eventos ocorridos num intervalo de tempo (exemplo: sistemas de filas).

Função de probabilidade

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

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

espaço amostral infinito

onde:

X: número de sucessos

e = 2,718 (base dos logaritmos neperianos)

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

Parâmetros

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

A distribuição de Poisson tem apenas um parâmetro:

λ = número médio de sucessos

$X \sim Poi(\lambda)$

X tem distribuição de Poisson com parâmetro λ

Exercício: Em uma central telefônica de uma pequena cidade do interior chegam ligações a uma taxa de 1 a cada 30 minutos. Qual a probabilidade de que no intervalo de 1 hora:

(a) Não chegue ligações? **13,53%**

(b) Chegue no máximo duas ligações? **67,67%**

(c) Chegue pelo menos duas ligações? **59,40%**

Solução: Neste caso, tem-se:

λ = 2 (taxa de ligações **por hora**)

X = nº de ligações **por hora**

Então:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \text{ para } x = 0, 1, 2, 3, ...$$

(a) $P(X = 0) = \frac{e^{-2}2^{0}}{0!} = e^{-2} = 13{,}53\%$

(b) $P(X \leq 2) = \frac{e^{-2}2^{0}}{0!} + \frac{e^{-2}2^{1}}{1!} + \frac{e^{-2}2^{2}}{2!} = 5e^{-2} = 67{,}67\%$

(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - \left[\frac{e^{-2}2^{0}}{0!} + \frac{e^{-2}2^{1}}{1!}\right] = 1 - 3e^{-2} = 59{,}40\%$

Medidas descritivas

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

Teorema: $\mathbf{E(X) = \mu = \lambda}$

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

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

Na Poisson média e variância são iguais!!

RESUMO - Distribuição de Poisson

Descrição probabilística da sequência de um grande número de fenômenos independentes, todos com probabilidade de sucesso constante e muito pequena.

Função de probabilidade

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

Parâmetro

λ = número médio de sucessos

Medidas descritivas

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

Exercício:

Um dado é formado por chapas de plástico de 10x10 cm. Em média aparecem 50 defeitos por metro quadrado de plástico, segundo uma distribuição de Poisson.

(a) Qual a probabilidade de uma determinada face apresentar exatamente 2 defeitos? **(7,58%)**

(b) Qual a probabilidade de o dado apresentar no mínimo dois defeitos? **(80,08%)**

(c) Qual a probabilidade de que pelo menos 5 faces sejam perfeitas? **(24,36%)**

(a) Qual a probabilidade de uma determinada face apresentar exatamente 2 defeitos?

Em média aparecem 50 defeitos/m^2 = (50/10000) defeitos/cm^2

Como cada face tem 10cm x 10 cm = 100 cm^2, tem-se então:

λ = (50/10000) defeitos/cm^2 x 100 cm^2 = 0,5 defeitos por face.

A probabilidade de uma face apresentar dois defeitos será:

$$P(X = 2) = \frac{e^{-0,5}(0,5)^2}{2!} = 7,58\%$$

(b) Qual a probabilidade de o dado apresentar no mínimo dois defeitos?

No dado inteiro, a área total será a = 6x100 cm^2 = 600 cm^2 e o número médio de defeitos será então:

λ = (50/10000) defeitos /cm^2 x 600 cm^2 = 3 defeitos

A probabilidade de o dado apresentar no mínimo 2 defeitos será:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots = 1 - P(X \leq 1)$$

$$= 1 - [P(X = 0) + P(X = 1)] =$$

$$= 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} \right] =$$

$$= 1 - [0{,}0498 + 0{,}1494] = 80{,}08\%$$

(c) Qual a probabilidade de que pelo menos 5 faces sejam perfeitas?

A probabilidade de uma face ser perfeita é a probabilidade de ela não apresentar defeitos, isto é:

$$P\ (X = 0) = \frac{e^{-0,5}(0,5)^{0}}{0\,!} = 60,65\%$$

Tem-se então uma binomial Y com n = 6 (número de faces do dado) e p = 60,65%(probabilidade de uma face ser perfeita)

Então a probabilidade de pelo menos 5 perfeitas, será:

$$P(Y \geq 5) = P(Y = 5) + P(Y = 6)$$

$$= \binom{6}{5}.(0,6065)^{5}.(0,3935)^{1} + \binom{6}{6}.(0,6065)^{6}.(0,3935)^{0} \quad = 24,36\%$$