Universidade de São Paulo

Escola Superior de Agricultura “Luiz de Queiroz”

Estatística Geral

Professor

Fábio Prataviera

Variáveis aleatórias discretas

Algumas distribuições de variáveis aleatórias discretas importantes:

Distribuição Uniforme Discreta

Enquadram-se aqui as distribuições em que os possíveis valores da variável aleatória tenham todos a mesma probabilidade de ocorrência. Logo, se existem N valores possíveis, cada um terá probabilidade igual a 1/N.

$$\boxed{\mathbf{P(X = x_i) = \frac{1}{N}}}$$

Ex: Seja o lançamento de um dado e a variável aleatória X ="face superior do dado", tem-se que: ou P(X=x) = 1/6

Distribuição de Bernoulli

Imagine um experimento aleatório que podem ocorrer dois possíveis resultados, "sucesso" e "fracasso". Veja alguns exemplos:

- semente pode germinar ou não germinar;

- um cliente pode ser adimplente ou inadimplente;

- uma peça produzida por uma cia. pode ser perfeita ou defeituosa;

- um consumidor que entra numa loja pode comprar ou não comprar um produto;

- lança-se uma moeda e observa-se se o resultado é cara ou coroa;

- aplica-se doses de um inseticida e verifica-se se uma determinada praga morreu ou ão morreu

Associando-se uma variável aleatória X aos possíveis resultados do experimento, de forma que

X= 1 se o resultado for "sucesso",
X= 0 se o resultado for "fracasso".

Então, a variável aleatória X, assim definida, tem distribuição de Bernoulli, com **p** sendo a probabilidade de ocorrer **"sucesso"**, e **q = (1-p)** a probabilidade de ocorrer **"fracasso"**.

Distribuição de Bernoulli: $$P(x) = \begin{cases} p & \text{se} \quad x = 1 \\ 1\text{-}p & \text{se} \quad x = 0 \end{cases}$$

Média: $$\mu = p$$

Variância: $$\sigma^2 = pq = p(1-p)$$

Desvio padrão: $$\sigma = \sqrt{pq} = \sqrt{p(1-p)}$$

Podemos utilizar p como sendo a proporção de sucessos.

Exemplos:

- **Lançamento de uma moeda honesta:** podemos associar cara a sucesso e coroa a fracasso. Além de igualmente provável a distribuição de probabilidade é do tipo Bernoulli:

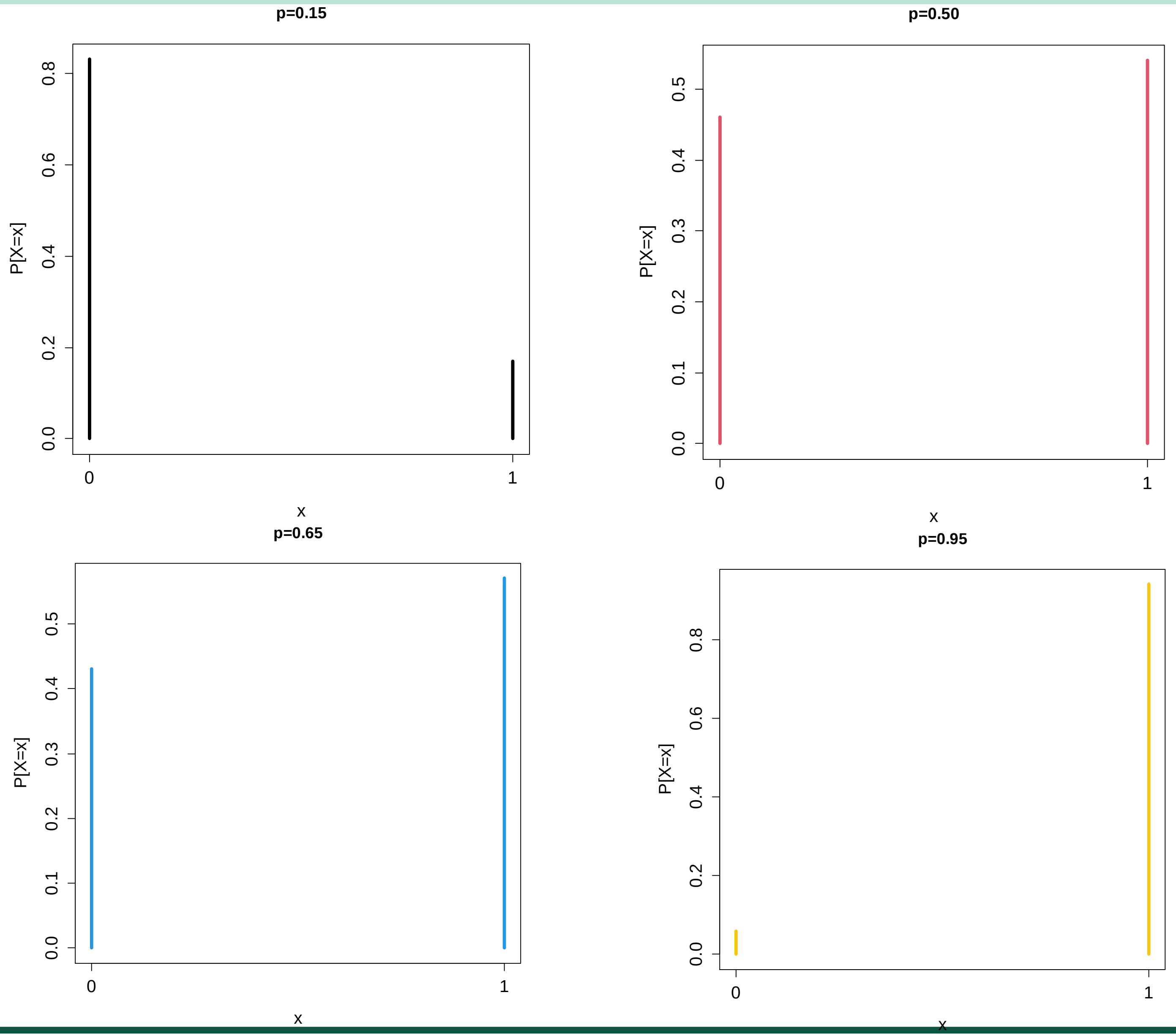
$$\textbf{Distribuição de Bernoulli:} \quad P(x) = \begin{cases} 1/2 & \text{se} \quad x = 1 \\ 1/2 & \text{se} \quad x = 0 \end{cases}$$

$$\mu = 1/2 \qquad \sigma^2 = 1/4 \qquad \sigma = 1/2$$

- A experiência tem mostrado que durante as vendas de Natal, um cliente que entra em uma determinada loja tem 60% de chance de comprar um produto qualquer. Temos, portanto, uma probabilidade de sucesso X=1 **(o cliente adquirir um produto qualquer)** de 0,6 e uma probabilidade de não adquirir X=0 um produto de q = 1-p = 0,4. Neste caso p é a proporção das vezes que um cliente compra um produto.

$$\textbf{Distribuição de Bernoulli:} \quad P(x) = \begin{cases} 0{,}6 & \text{se} \quad x = 1 \\ 0{,}4 & \text{se} \quad x = 0 \end{cases}$$

$$\mu = 0{,}6 \qquad \sigma^2 = 0{,}24 \qquad \sigma = 0{,}4898$$


p=0.15
p=0.50
p=0.65
p=0.95
P[X=x]
x
0
1
0.0
0.2
0.4
0.6
0.8
0.1
0.3
0.5

Distribuição binomial

Para que uma situação possa se enquadrar em uma distribuição binomial, deve atender às seguintes condições:

- **são realizadas n repetições (tentativas, ensaios, provas) independentes;**

- **cada tentativa é um ensaio de Bernoulli, isto é, só podem ocorrer dois resultados possíveis, com probabilidades p e q=1-p;**

- **a probabilidade p de sucesso em cada ensaio é constante.**

Se uma situação atende a todas as condições acima, então a variável aleatória

X = número de sucessos obtidos nas n ensaios terá uma distribuição binomial, com n tentativas e p (probabilidade de sucesso).

Vamos determinar a distribuição de probabilidade de X, através da probabilidade de um número genérico x de sucessos.

Suponha que ocorram sucessos (1) apenas nas x primeiras provas, e fracassos (0) nas $n - x$ provas restantes

$$\underbrace{1, 1, 1, \ldots, 1}_{x}, \underbrace{0, 0, 0, \ldots, 0}_{n-x}$$

Como as provas são independentes, a probabilidade de ocorrência de x sucessos em n tentativas é uma extensão do modelo de Bernoulli para n ensaios, ou seja,

$$\underbrace{p \cdot p \cdot p \cdots p}_{x} \cdot \underbrace{(1-p) \cdot (1-p) \cdot (1-p) \cdots (1-p)}_{n-x} = p^x (1-p)^{n-x}$$

Porém, o evento: "x sucessos em n provas" pode ocorrer de diferentes maneiras (ordens) distintas, todas com a mesma probabilidade.

Como o número de ordens é o número de combinações de n elementos tomados x a x, então a probabilidade de ocorrerem x sucessos em n provas de Bernoulli será então a distribuição binomial, dada por

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

onde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

é o **coeficiente binomial**, que dá o número total de combinações possíveis de n elementos, com x sucessos.

Simbolicamente: $X \sim B(n,p)$ A variável aleatória x tem distribuição binomial com n ensaios e uma probabilidade p de sucesso

p : probabilidade de sucesso

q = 1- p : probabilidade de não sucesso

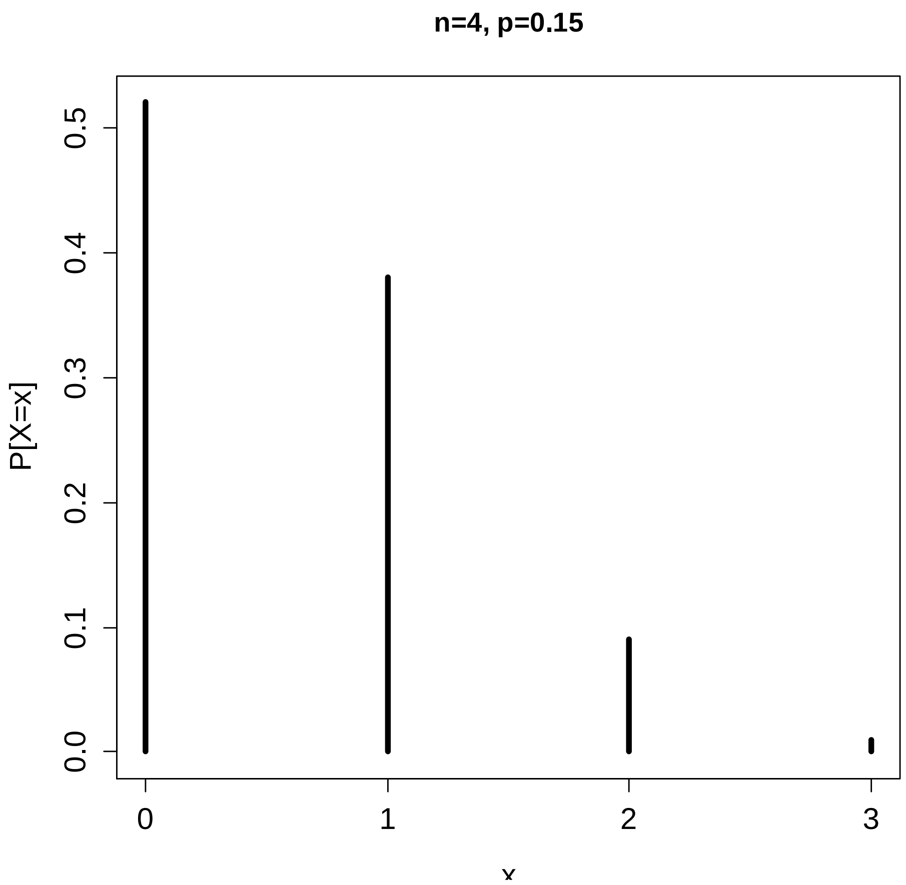
A distribuição de probabilidade da variável aleatória X = x sucessos, em n ensaios, é dada pela **distribuição Binomial**:

Distribuição Binomial: $$P_n(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

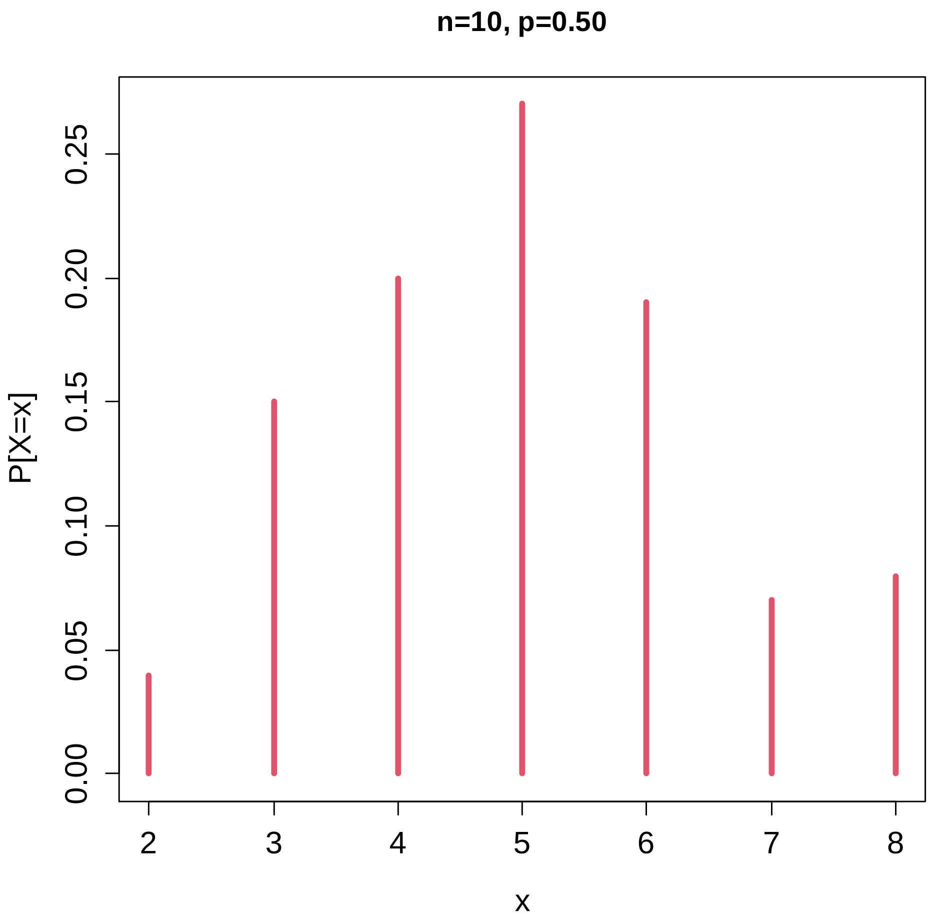
média: $\mu = np$

variância: $\sigma^2 = npq$

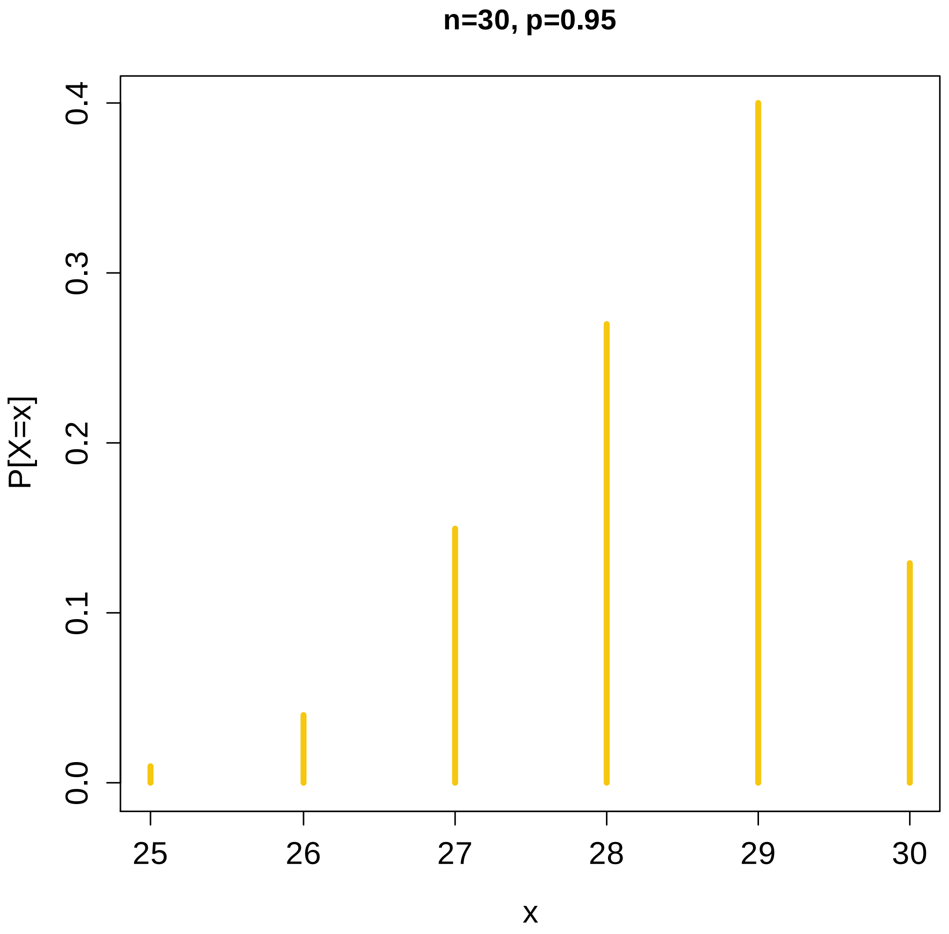
desvio padrão: $\sigma = \sqrt{npq}$

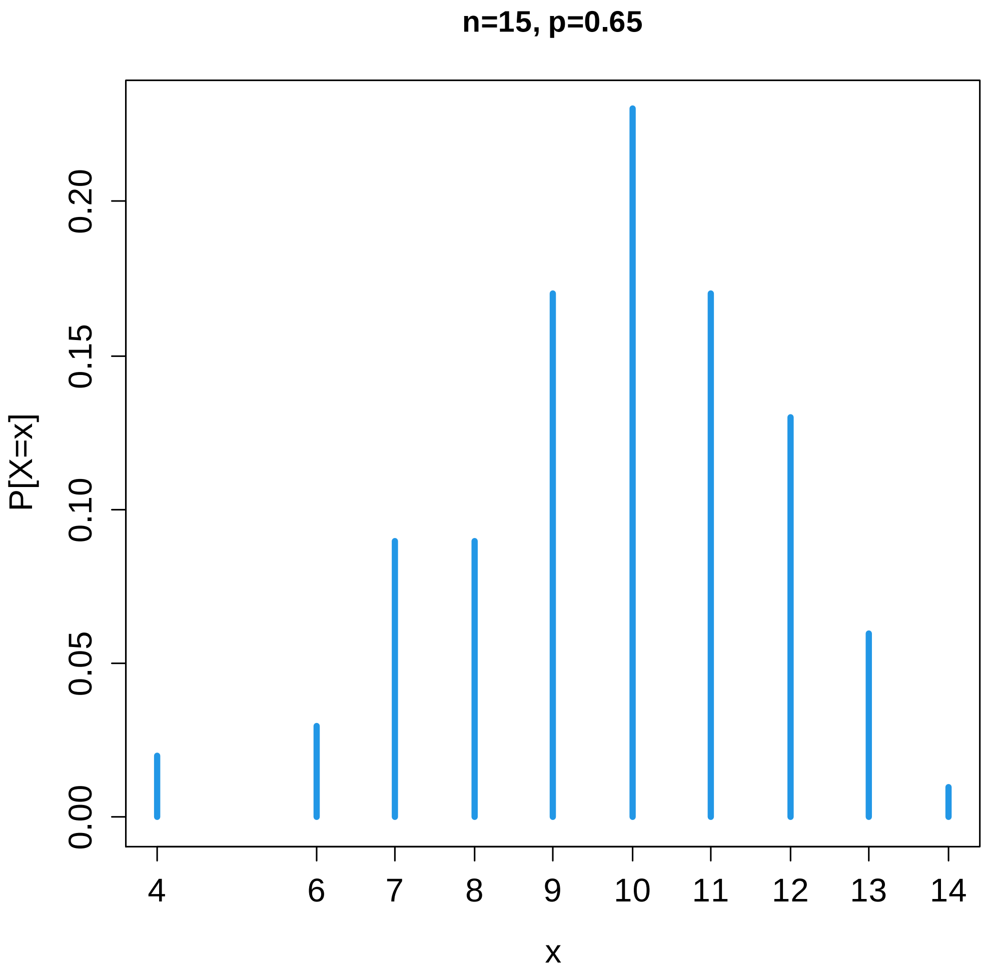


n=4, p=0.15
P[X=x]
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
x



n=10, p=0.50
P[X=x]
0.00
0.05
0.10
0.15
0.20
0.25
2
3
4
5
6
7
8
x


n=30, p=0.95
P[X=x]
0.0
0.1
0.2
0.3
0.4
25
26
27
28
29
30
x


n=15, p=0.65
P[X=x]
0.00
0.05
0.10
0.15
0.20
4
6
7
8
9
10
11
12
13
14
x

- Probabilidade de se obter valores maiores ou iguais ao valor observado

$$P_n(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} (p)^i (1-p)^{n-i}$$

- Probabilidade de se obter valores menores ou iguais ao valor observado

$$P_n(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} (p)^i (1-p)^{n-i}$$

Exemplos

(1) No lançamento de uma moeda honesta 5 vezes sucessivamente:

(a) Qual a probabilidade de ocorrer 3 caras?
(b) Qual a probabilidade de ocorrer menos de 3 caras?
(c) Qual a probabilidade de ocorrer no máximo 3 caras?
(d) Qual a probabilidade de ocorrer no mínimo 3 caras?

Sucesso é representado pelo resultado cara.

Como a moeda é honesta p=0.5 e q=1-p=0.5 número de lançamentos da moeda: n=5 a probabilidade de ocorrência de k caras em n lançamentos é dado pela distribuição binomial

$$\boxed{P_n(X=x)=\binom{n}{x}p^x q^{n-x}=\frac{n!}{x!(n-x)!}p^x q^{n-x}}$$

Temos:

$$(a)\, P_5(X=3)=\binom{5}{3}(0.5)^3(0.5)^{5-3}=\frac{5!}{3!(5-3)!}(0.5)^3(0.5)^2$$

$$(b)\, P_5(X<3)=P(0)+P(1)+P(2)=\binom{5}{0}(0.5)^0(0.5)^{5-0}+\binom{5}{1}(0.5)^1(0.5)^{5-1}+\binom{5}{2}(0.5)^2(0.5)^{5-2}$$

$$(c)\, P_5(X\leq 3)=P(0)+P(1)+P(2)+P(3)=\binom{5}{0}(0.5)^0(0.5)^{5-0}+\binom{5}{1}(0.5)^1(0.5)^{5-1}+\binom{5}{2}(0.5)^2(0.5)^{5-2}+\binom{5}{3}(0.5)^3(0.5)^{5-3}$$

$$(d)\, P_5(X\geq 3)=P(3)+P(4)+P(5)\quad \left(\text{ou } P_5(X\geq 3)=1-P_5(X<3)\right)$$

Completem as contas !!!

(2) Em um determinado processo de fabricação, 10% das peças produzidas são consideradas defeituosas. As peças são armazenadas em caixas com cinco unidades cada uma. Considere que cada peça tem a mesma probabilidade de ser defeituosa (como se houvesse repetição no experimento de retirar uma peça).

(a) Qual a probabilidade de haver exatamente três peças defeituosas numa caixa?
(b) Qual a probabilidade de haver duas ou mais peças defeituosas em uma caixa?
(c) Qual a probabilidade de uma caixa não apresentar peça defeituosa?
(d) Supondo que a empresa pague uma multa de R$ 10,00 por caixa que apresente peças defeituosas, qual o valor esperado desta multa em um lote de 1.000 caixas?

p = 0,1 é a probabilidade da peça ser defeituosa
q = 1-p = 0,9 é probabilidade da peça não ser defeituosa
Numa caixa temos 5 peças: n = 5
O problema é descrito por uma distribuição binomial onde
X=x é o número de peças defeituosas

$$Temos:$$

$$(a)\, \mathrm{P}_5(\mathrm{X}=3)=\binom{5}{3}(0.1)^3(0.9)^{5-3}=\frac{5!}{3!(5-3)!}(0.1)^3(0.9)^2$$

$$(b)\, \mathrm{P}_5(\mathrm{X}\geq 2)=1-\mathrm{P}_5(\mathrm{X}<2)=1-(P(0)+P(1))=1-\binom{5}{0}(0.1)^0(0.9)^{5-0}-\binom{5}{1}(0.1)^1(0.9)^{5-1}$$

$$(c)\, \mathrm{P}_5(\mathrm{X}=0)=\binom{5}{0}(0.1)^0(0.9)^{5-0}=\frac{5!}{0!(5-0)!}(0.9)^5=(0.9)^5=0{,}59049$$

(d) Neste item, temos uma nova variável aleatória Y = número de caixas com peças defeituosas . Então temos que calcular o a probabilidade de uma caixa ter peça defeituosa:

P(uma caixa ter peça defeituosa) =1-P(uma caixa não ter peça defeituosa)
= 1 - P(X=0) = 0, 4095

Então a nova variável Y: número de caixas com peças defeituosas, em um lote de 1.000 caixas, segue uma distribuição binomial com n = 1.000 e p = 0, 4095 e, portanto o valor esperado de Y será:

E(Y) = np = 1000.0,4095 = 409,5 caixas.

E a multa esperada:

Multa Esperada = 409,5. R$ 10,00 = R$ 4.095,00

Distribuição Hipergeométrica

- Relacionada com o número de sucessos em uma amostra contendo n observações – população finita de tamanho N
- Amostra retirada sem reposição
- Os resultados das observações são dependentes
- A probabilidade de sucesso varia de um experimento para outro
- probabilidade da variável aleatória X=k sucessos em uma amostra retirada de uma população com “r” sucessos

Fórmula da Distribuição Hipergeométrica

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

Em que

N = tamanho da população

r = número de sucessos na população

N – r = número de insucessos na população

n = tamanho da amostra

x = número de sucessos na amostra

n – x = número de insucessos na amostra

Propriedades da Distribuição Hipergeométrica

A média (valor esperado de X) da distribuição hipergeométrica é

$$\mu = np$$

onde $p = r/N$ *é a proporção de sucessos na população*

e o desvio-padrão

$$\sigma = \sqrt{\frac{nr(N\text{-}r)}{N^2}} . \sqrt{\frac{N-n}{N-1}} = \sqrt{np(1\text{-}p)} . \sqrt{\frac{N-n}{N-1}}$$

onde $\sqrt{\frac{N-n}{N-1}}$ *é chamado fator de "correção de população finita",*

que resulta a mostragem sem reposição, de uma população finita.

Quando N é grande comparado com n, a distribuição hipergeométrica de parâmetros n, N, p *se aproxima da distribuição binomil com parâmetros* $n,\ p$

Utilizando a Distribuição Hipergeométrica

Exemplo: Um departamento possui 10 computadores, dos quais 3 são verificados. 4 desses computadores possuem software ilegal. Qual a probabilidade de que 2 dos 3 computadores verificados tenham o softwrare ilegal?

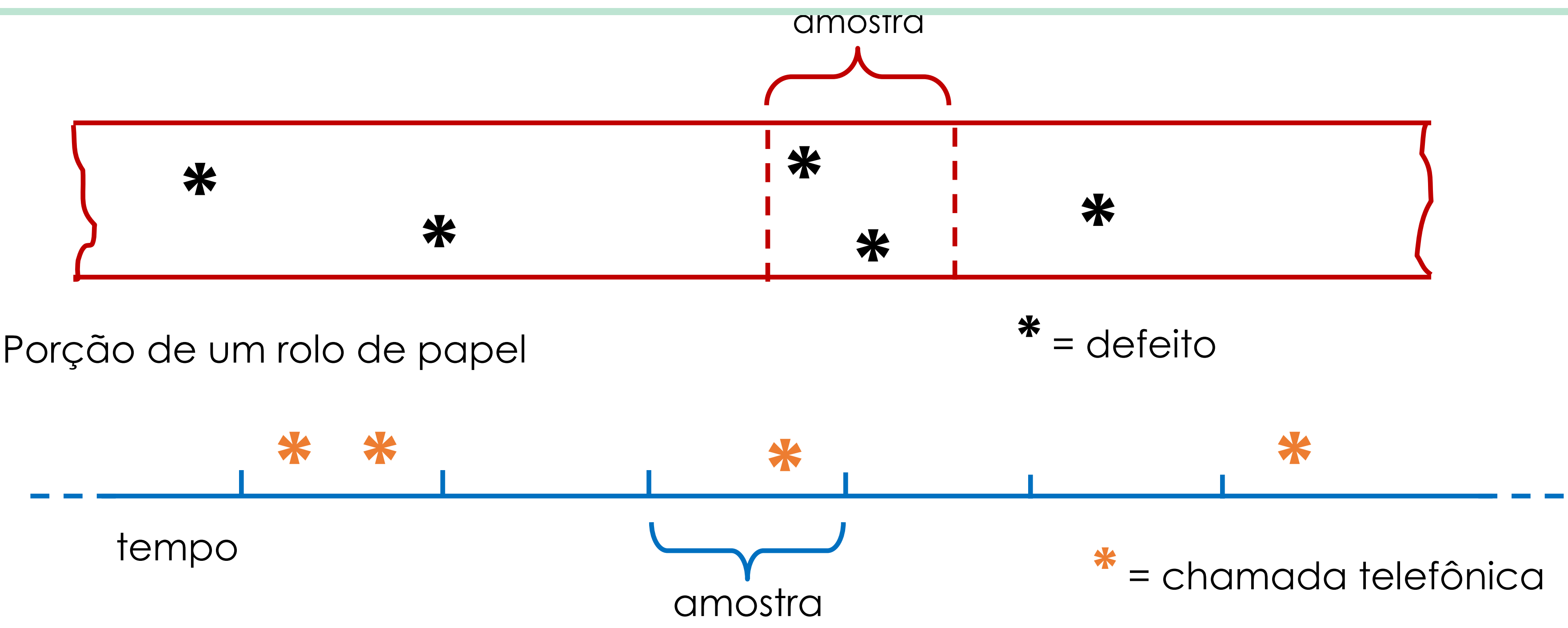
N = 10 n = 3 r = 4 x = 2

$$P(X=2)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=\frac{(6)(6)}{120}=0{,}3$$

A probabilidade de que 2 dos 3 computadores checados tenham o software ilegal é 0,30 ou 30%

Distribuição de Poisson

- Aplicada quando:
  - Você estiver interessado em contar o número de vezes em que um evento específico ocorre em um determinado intervalo (de tempo, de comprrimento, de área, volume,...)
  - A probabilidade de que um evento específico ocorra em um determinado intervalo é a mesma para todas os intervalos
  - O número de eventos que ocorrem em um determinado intervalo é independente do número de eventos que ocorrem em outros intervalos
  - A probabilidade de que dois ou mais eventos venham a ocorrer em um determinado intervalo se aproxima de zero à medida que o intervalo se torna menor.



A distribuição de Poisson é útil para descrever o número de ocorrências num Intervalo contínuo (tempo, distância).

Ex: Defeitos por centímetro quadrado, acidentes por dia, número de indivíduos por quadrante de 1 m²,...

A variáve aleatória X = número de ocorrências é discreta, mas a unidade de medida é contínua. Não é possível contar o número de não ocorrências (número de defeitos que não ocorreram por centímetro quadrado, número de chamadas telefônicas que não foram feitas, etc,...)

- Chamadas telefônicas por unidade de tempo;
- Defeitos por unidade de área;
- Acidentes por unidade de tempo;
- Chegada de clientes a um supermercado por unidade de tempo;
- Número de glóbulos sangüíneos visíveis ao microscópio por unidade de área;
- Número de partículas emitidas por uma fonte de material radioativo por unidade de tempo.

Fórmula da Distribuição de Poisson

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

onde:

x = número de ocorrências (0, 1, 2, 3, 4, 5,..., ∞)

μ = λ t =número médio (valor esperado) de ocorrências no intervalo t

λ = taxa média por unidade (de tempo, de distância, de área…)

t = número de unidades (de tempo, distância, área, volume,…)

e = 2,71828… (base dos logaritmos naturais)

A distribuição de Poisson pode ser um modelo probabilístico razoável de processos descritos na página anterior.

Probabilidade de k ocorrências: $P(X = x) = \dfrac{e^{-\mu}\mu x}{x!}$

Valor esperado (média): $\mu = \lambda t$

Variância: $\sigma^2 = \mu$

Desvio padrão: $\sigma = \sqrt{\mu}$

Exemplo: encontre $P(X = 2)$ se $\mu = 0{,}50$

$$P(X=2) = \frac{e^{-\mu}\mu^2}{k!} = \frac{e^{-0{,}50}(0{,}50)^2}{2!} = 0{,}0758$$

Exemplo: Trens chegam numa estação a uma razão de 3 trêns/hora. Observando a chegada de trens durante meia hora, qual a probabilidade de (a) não chegar nenhum trem (b) chegar 1 trem (c) chegar menos de 2 trens. Suponha que a chegada de trens possa ser descrtia por um processo de Poisson.

Temos: *Taxa de chegada:* $\lambda = 3$ *trens/hora*

intervalo de tempo: $t = 0{,}5\ h$

média de chegadas: $\mu = \lambda t = 3 \ .\ 0{,}5 \ = 1{,}5$

$$P(X = x) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-1,5}(1{,}5)^{x}}{x!}$$

$$\textbf{(a) } \mathbf{P(X=0)} = \frac{\mathbf{e^{-1,5}(1{,}5)^{0}}}{\mathbf{0!}} = \mathbf{0{,}223}$$

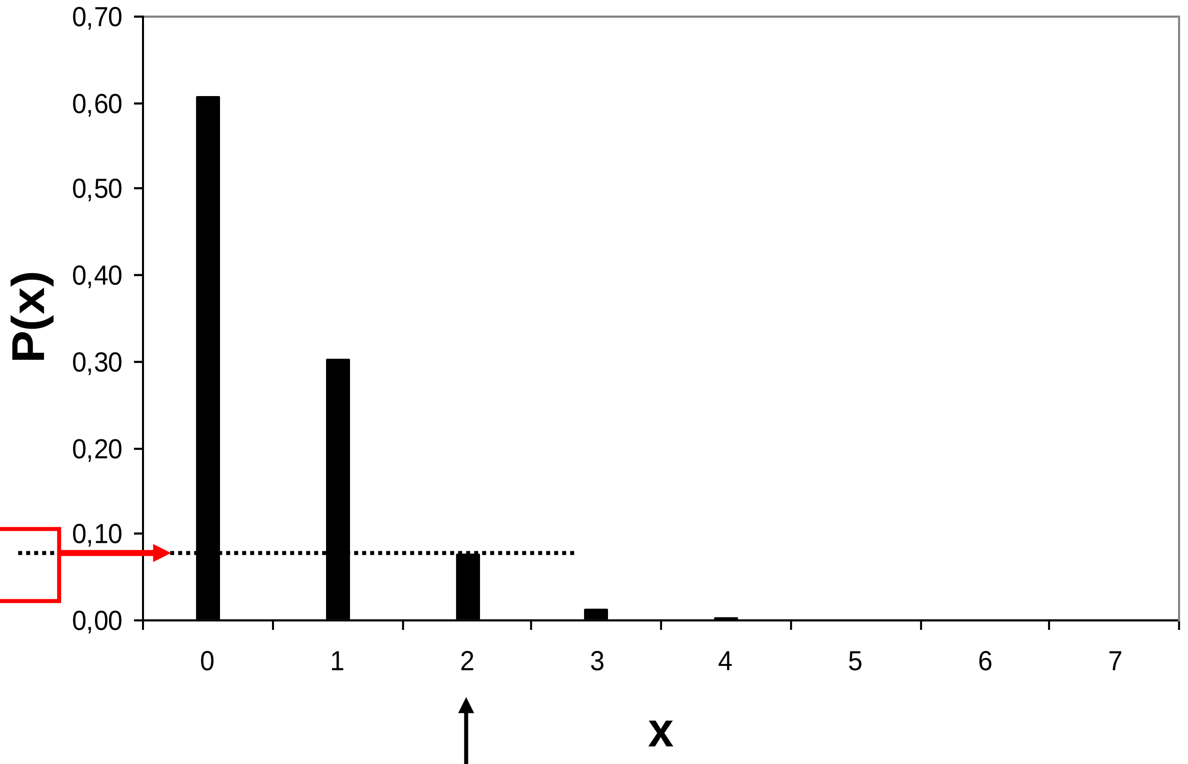
$$\textbf{(b) } \mathbf{P(X=1)} = \frac{\mathbf{e^{-1,5}(1{,}5)^{1}}}{\mathbf{1!}} = \mathbf{0{,}334}$$

$$\textbf{(c) } \mathbf{P(X<2) = P(0) + P(1) = 0{,}557}$$

Gráfico das Probabilidades de Poisson

μ = 0,50

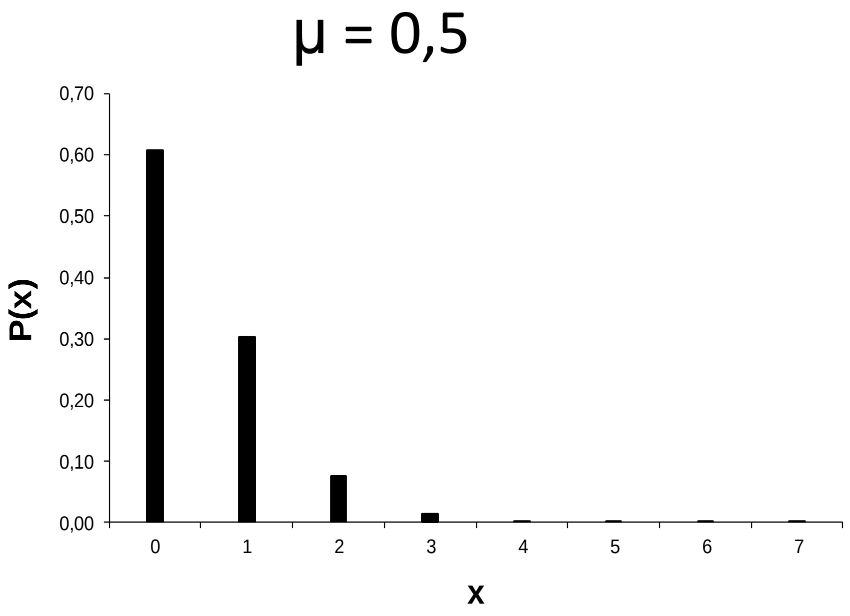
X	μ = 0,50
0	0,6065
1	0,3033
2	0,0758
3	0,0126
4	0,0016
5	0,0002
6	0.000013
7	0.00000094

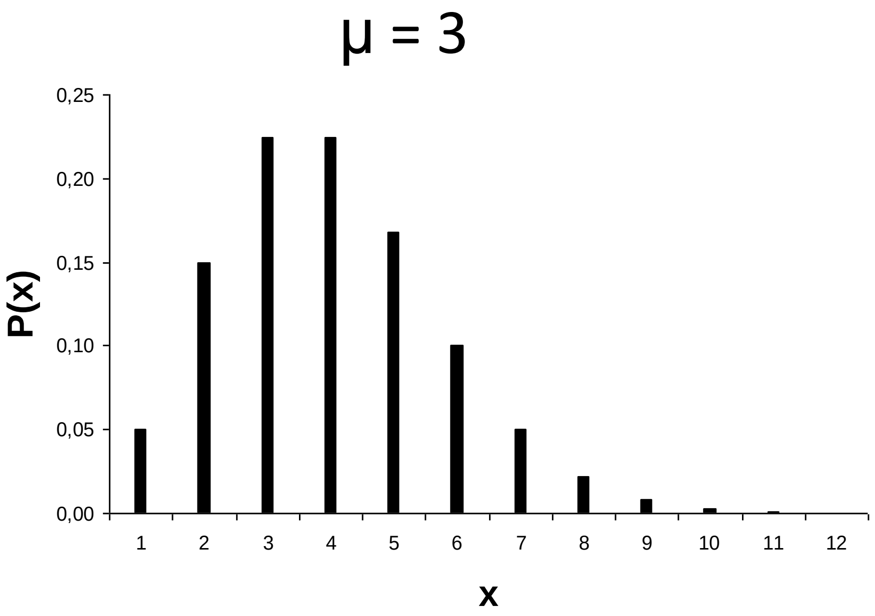


P(X = 2) = 0,0758

Forma da Distribuição de Poisson

- A forma da Distribuição de Poisson depende do parâmetro μ :





Comparação entre as distribuições Binomial e de Poisson:

	Binomial	Poisson
Resultados possíveis	Inteiros de 0 a n	Inteiros de 0 a $+\infty$
Observações	Contagem de sucessos ou falhas	Contagem de sucessos somente
parâmetros	n e p	μ

Aproximação da binomial pela poisson

Para um número muito grande de repetiçõe e probabilidade de sucesso pequena podemos calcular, ou seja, se **n** é grande e p pequeno de modo que **np ≤ 7**, podemos calcular a probabilidade de sucessos aproximando a distribuição binomial pela distribuição de Poisson com **μ = np**.

Exemplo: Determinar a probabilidade de haver **4** peças defeituosa numa amostra de **300** peças, extraida de um grande lote onde há **2%** de defeituosas:

temos uma distribuição binomial com $n = 300$ *e* $p = 0{,}02$ *. Assim,*

$$P(X = 4) = \frac{300!}{4!(300\text{-}4)!}(0{,}02)^4.(0{,}98)^{296} \quad \textit{dureza fazer esse calculo!!!}$$

mas temos uma situação em que n *é grande e* p *pequeno, portanto podemos aproximar a distribuição binomial por uma distribuição de Poisson com* $\mu = np = 300.0{,}02 = 6.$

$$P(X = 4) = \frac{\mu^x e^{-\mu}}{x!} = \frac{6^4 e^{-6}}{4!} = 0{,}135 \quad \textit{(mais fácil de calcular)}$$

Tabela resumo

Modelo	$P(X = x)$	parâmetros	$E(X)$	$Var(X)$
Bernoulli	$p^x(1-p)^{1-x}, x = 0,1$	p	p	$p(1\text{-}p)$
Binomial	$\binom{n}{x} p^x(1-p)^{n-p}, x = 0,1,2\ldots,n$	$n, \quad p$	np	$np(1\text{-}p)$
Poisson	$\frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \ldots, \infty$	μ	μ	μ
Hipergeométrica	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots$	N, r, n	$\frac{nr}{N}$	$\frac{nr}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Função de distribuição acumulada ou função de distribuição

Definição. Dada a variável aleatória *X*, chamaremos de função de distribuição acumulada (f.d.a.), ou simplesmente função de distribuição (f.d.) ***F(x)*** à função

$$F(x) = P(X \leq x).$$

Observe que o domínio de F é todo o conjunto dos números reais, ao passo que o contradomínio é o intervalo [0,1].

Se X toma um número finito de valores $x_1, x_2, \dots, x_n$ então a função de distribuição será dada por

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leqq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leqq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \leqq x < \infty \end{cases}$$

Exemplo

Seja o lançamento de uma duas vezes. Apresentar o espaço amostral. Defina a variável aleatória X como sendo o número de caras, C. Apresentar a função distribuição e o gráfico associado. Encontrar a função de distribuição acumulada e representar graficamente.

Observação: Uma variável aleatória que toma um número finito ou infinito contável de valores, é denominada variável aleatória discreta

Exemplo

Seja o lançamento de uma duas vezes. Apresentar o espaço amostral. Defina a variável aleatória X como sendo o número de caras, C. Apresentar a função distribuição e o gráfico associado. Encontrar a função de distribuição acumulada e representar graficamente.

CC	CS	SC	SS
2	1	1	0

$$P(CC) = \frac{1}{4} \qquad P(CS) = \frac{1}{4} \qquad P(SC) = \frac{1}{4} \qquad P(SS) = \frac{1}{4}$$

$$P(X=0) = P(SS) = \frac{1}{4}$$

$$P(X=1) = P(CS \cup SC) = P(CS) + P(SC) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X=2) = P(CC) = \frac{1}{4}$$

Observação: Uma variável aleatória que toma um número finito ou infinito contável de valores, é denominada variável aleatória discreta

Exemplo

Seja o lançamento de uma duas vezes. Apresentar o espaço amostral. Defina a variável aleatória X como sendo o número de caras, C. Apresentar a função distribuição e o gráfico associado. Encontrar a função de distribuição acumulada e representar graficamente.

CC	CS	SC	SS
2	1	1	0

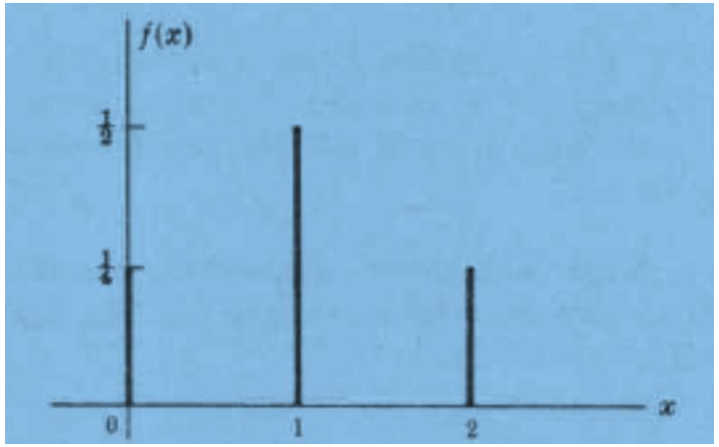
$$P(CC) = \frac{1}{4} \qquad P(CS) = \frac{1}{4} \qquad P(SC) = \frac{1}{4} \qquad P(SS) = \frac{1}{4}$$

$$P(X=0) = P(SS) = \frac{1}{4}$$

$$P(X=1) = P(CS \cup SC) = P(CS) + P(SC) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X=2) = P(CC) = \frac{1}{4}$$

x	0	1	2
$f(x)$	1/4	1/2	1/4

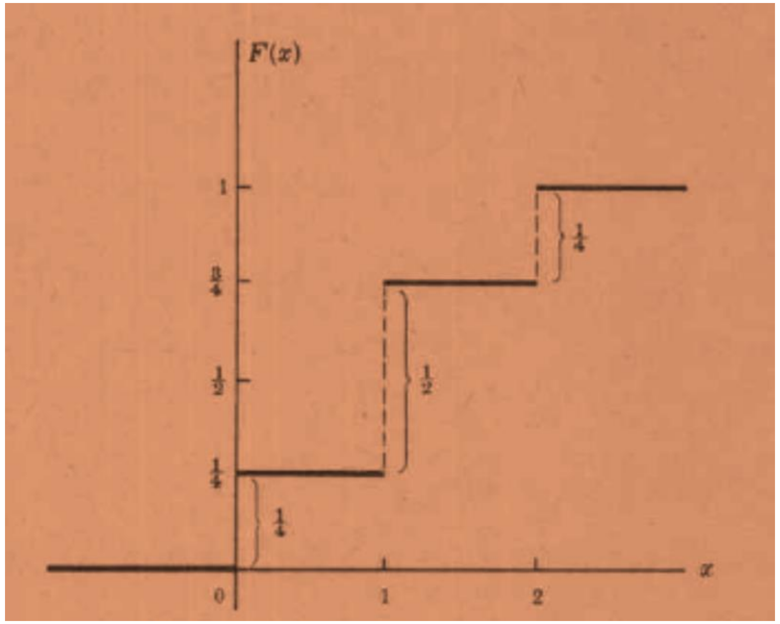


Observação: Uma variável aleatória que toma um número finito ou infinito contável de valores, é denominada variável aleatória discreta

Exemplo

Seja o lançamento de uma duas vezes. Apresentar o espaço amostral. Defina a variável aleatória X como sendo o número de caras, C. Apresentar a função distribuição e o gráfico associado. <u>Encontrar a função de distribuição acumulada e representar graficamente.</u>

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \leqq x < 1 \\ \frac{3}{4} & 1 \leqq x < 2 \\ 1 & 2 \leqq x < \infty \end{cases}$$



Observação: Uma variável aleatória que toma um número finito ou infinito contável de valores, é denominada variável aleatória discreta

Próximas aulas

- Distribuições contínuas;
- Noções de amostragem;
- Distribuições amostrais.