



4. PRINCIPAIS MODELOS DISCRETOS

2012

Principais modelos probabilísticos discretos

4.1. Modelo Bernoulli

Muitos experimentos admitem apenas **dois** resultados.

Exemplos:

1. Uma peça é classificada como defeituosa ou não defeituosa;
2. O resultado de um exame médico para detecção de uma doença é positivo ou negativo.
3. Um entrevistado concorda ou não com uma afirmação feita;
4. No lançamento de um dado ocorre ou não face 6;
5. Um item produzido é classificado como conforme ou não conforme.

Situações com alternativas **dicotômicas** podem ser representadas genericamente por resposta do tipo **sucesso** ou **insucesso** (**fracasso** ou **falha**).

Esses experimentos recebem o nome de **ensaios de Bernoulli** e originam uma v.a. com distribuição de Bernoulli.

Distribuição de Bernoulli

X é uma v.a. que assume apenas dois valores: **1** se ocorrer **sucesso** (S) e **0** se ocorrer **fracasso** (F). Sendo p a probabilidade de sucesso, $0 < p < 1$.

$X(S) = 1$ e $X(F) = 0$. A distribuição de probabilidade é dada por

x	0	1
$P(X=x)$	$1 - p$	p

$$f(x) = P(X = x) = \begin{cases} p^x (1-p)^{1-x}, & \text{se } x = 0, 1. \\ 0, & \text{c.c.} \end{cases}$$

Notação: $X \sim \text{Bernoulli}(p)$ indica que a v.a. X tem distribuição de Bernoulli. O **parâmetro da distribuição** é p .

Se $X \sim \text{Bernoulli}(p)$, então

$$E(X) = p$$

$$\text{e } \text{Var}(X) = p(1 - p).$$

Repetições **independentes** de um ensaio de Bernoulli dão origem ao modelo binomial.

4.2. Modelo binomial

Exemplo. Uma moeda é lançada 3 vezes e a probabilidade de cara é p em cada lançamento. Determinar a distribuição de probabilidade da variável **número de caras** nos 3 lançamentos (X).

Denotemos S: **sucesso**, ocorre **cara** (c) e F: **fracasso**, ocorre **coroa** (k).

O espaço amostral para este experimento é

$$\Omega = \{FFF, FFS, FSF, SFF, FSS, SFS, SSF, SSS\}.$$

Fazemos $X_i \sim \text{Bernoulli}(p)$, $i = 1, 2, 3$. Logo, $X = X_1 + X_2 + X_3$ representa o número de caras nos 3 lançamentos.

Ω	Probabilidade	X_1	X_2	X_3	$X=X_1+X_2+X_3$
FFF	$(1-p)^3$	0	0	0	0
FFS	$(1-p)^2p$	0	0	1	1
FSF	$(1-p)^2p$	0	1	0	1
SFF	$(1-p)^2p$	1	0	0	1
FSS	$(1-p)p^2$	0	1	1	2
SFS	$(1-p)p^2$	1	0	1	2
SSF	$(1-p)p^2$	1	1	0	2
SSS	p^3	1	1	1	3

Calculamos

$$P(X = 0) = P(\{FFF\}) = (1-p)^3,$$

$$P(X = 1) = P(\{FFS, FSF, SFF\}) = 3p(1-p)^2,$$

$$P(X = 2) = P(\{FSS, SFS, SSF\}) = 3p^2(1-p) \text{ e}$$

$$P(X = 3) = P(\{SSS\}) = p^3.$$

A **distribuição de probabilidade** da v.a. X é dada por

x	0	1	2	3
$f(x) = P(X = x)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

$f(x)$ pode ser escrita como

$$f(x) = \begin{cases} \binom{3}{x} p^x (1-p)^{3-x}, & \text{se } x = 0, 1, 2, 3, \\ 0, & \text{c.c.} \end{cases}$$

$$\text{em que } \binom{3}{x} = \frac{3!}{x!(3-x)!}.$$

Distribuição binomial

Repetição de n ensaios de Bernoulli independentes, todos com a mesma probabilidade de sucesso p . A variável aleatória que conta o número de sucessos nos n ensaios de Bernoulli é denominada de variável aleatória binomial com parâmetros n e p .

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{se } x = 0, 1, \dots, n, \\ 0, & \text{c.c.,} \end{cases}$$

em que $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ representa o coeficiente binomial.

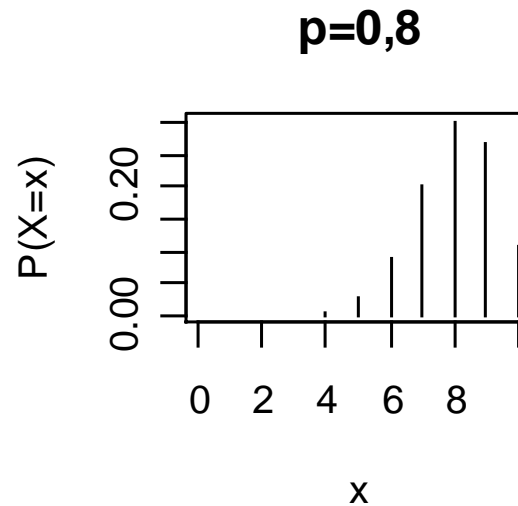
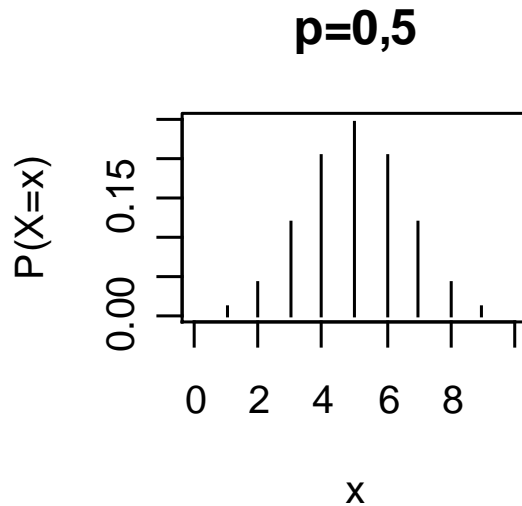
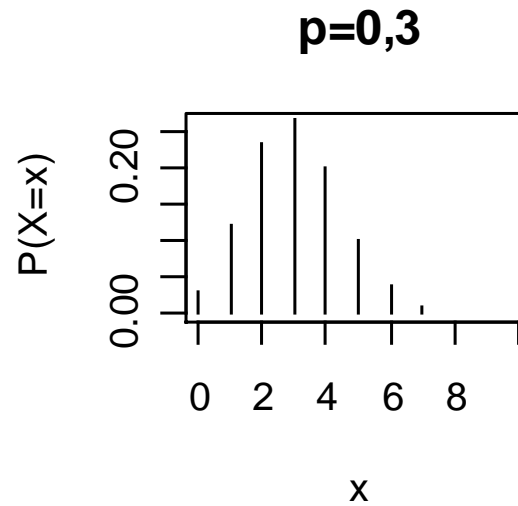
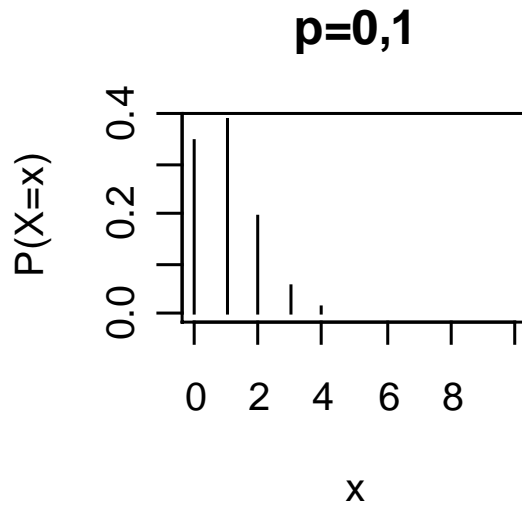
Notação: $X \sim B(n, p)$ para indicar que a v.a. X tem distribuição Binomial com parâmetros n e p .

Se $X \sim B(n, p)$, então

$E(X) = np$ e

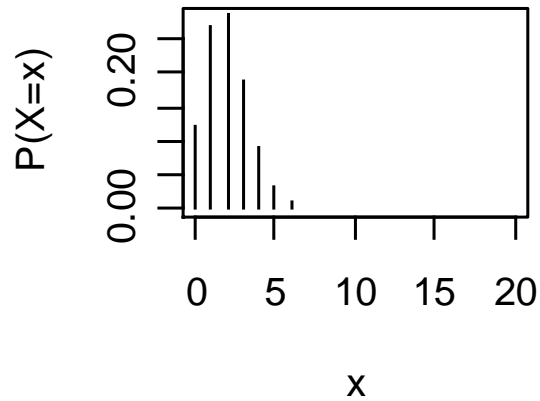
$\text{Var}(X) = np(1 - p)$.

Distribuição B($n = 10, p$)

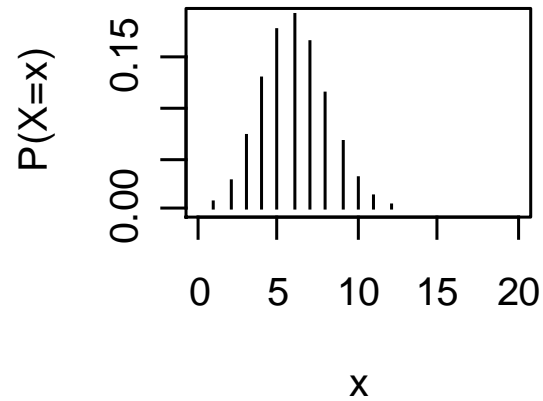


Distribuição B($n = 20$, p)

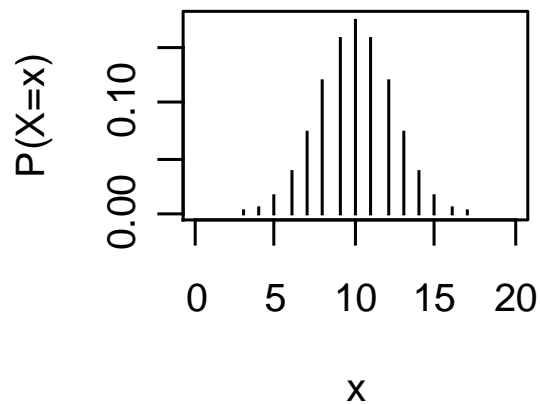
$p=0,1$



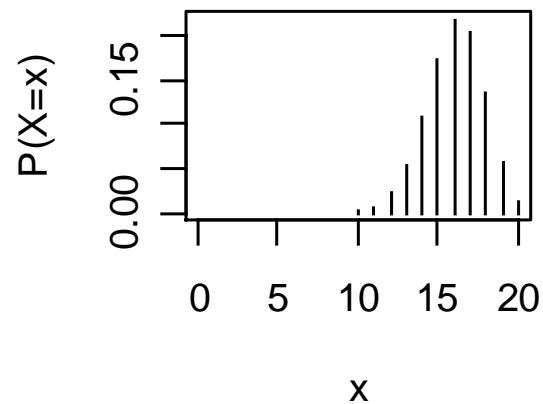
$p=0,3$



$p=0,5$

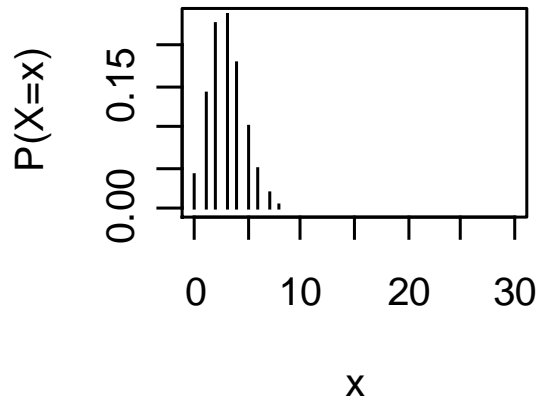


$p=0,8$

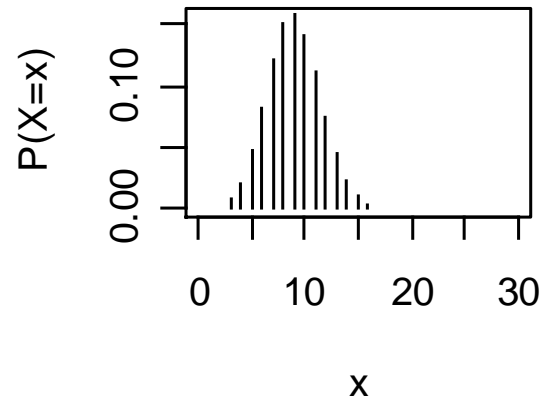


Distribuição B($n = 30, p$)

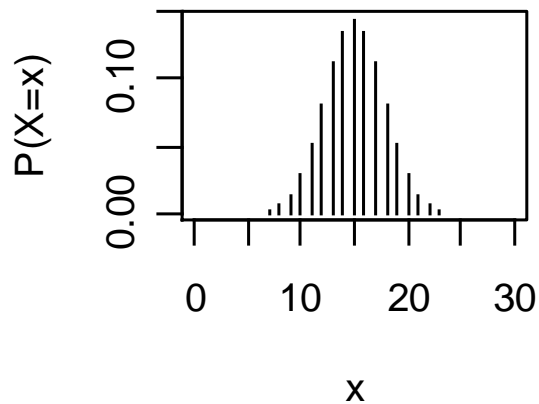
$p=0,1$



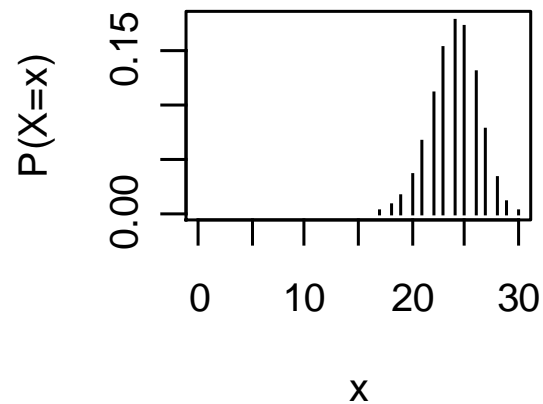
$p=0,3$



$p=0,5$



$p=0,8$



Exemplo

O professor da disciplina de Estatística elaborou uma prova de múltipla escolha, composta de 10 questões cada uma com 5 alternativas. A aprovação na disciplina requer pelo menos 6 questões corretas. Se um aluno responde a todas as questões baseado em palpite (“chute”), qual a probabilidade de ser aprovado?

Solução. X é a v.a. número de questões respondidas corretamente nas 10 questões. Eventos: S: “questão respondida corretamente” e F: “questão respondida incorretamente”.

$P(S) = 1 / 5$ e $P(F) = 4 / 5$. Logo, $X \sim B(10, p)$.

$$f(x) = \begin{cases} \binom{10}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x}, & \text{se } x = 0, 1, \dots, 10, \\ 0, & \text{c.c.} \end{cases}$$

A probabilidade de aprovação é

$$\begin{aligned} P(X \geq 6) &= 1 - P(X < 6) = 1 - P(X \leq 5) = 1 - F(5) \\ &= 1 - 0,9936306 = 0,00637. \end{aligned}$$

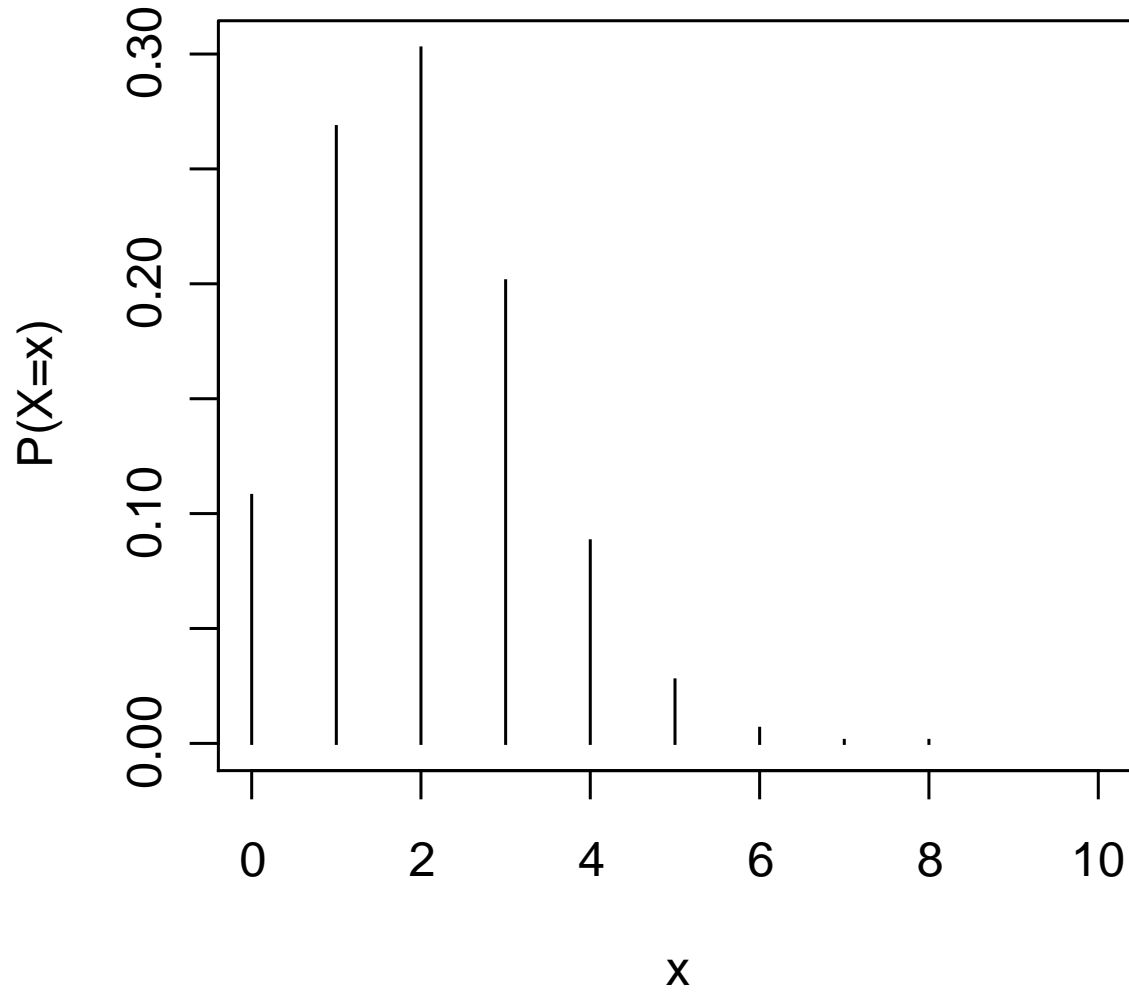
Exemplo

x	f(x)	F(x)
0	0,107374	0,10737
1	0,268435	0,37581
2	0,301990	0,67780
3	0,201327	0,87913
4	0,088080	0,96721
5	0,026424	0,99363
6	0,005505	0,99914
7	0,000786	0,99992
8	0,000074	1,00000
9	0,000004	1,00000
10	0,000000	1,00000

Em R: `dbinom(0:10,10,1/5)` e `pbinom(0:10, 10, 1/5)`.

Exemplo

B(10,p=0,20)



Exemplo

Um fabricante adquire certo tipo de componente de um fornecedor. Segundo este fornecedor, a **proporção** de componentes **defeituosos** é **2%**.

(a) O fabricante seleciona **15** componentes de um lote para inspeção. Qual a probabilidade de que seja encontrado **pelo menos um** componente **defeituoso** neste lote?

(b) O fabricante adquire **10** lotes por mês e de cada lote são selecionados **15** componentes para inspeção, como no item (a). Qual a probabilidade de que sejam encontrados **três** lotes com **pelo menos um** componente **defeituoso**?

Solução. (a) Definimos o evento sucesso (S) como “o componente selecionado é **defeituoso**”. Pelo enunciado, $P(S) = p = 0,02$. A v.a. X é definida como sendo o número de componentes defeituosos (sucessos) em $n = 15$ componentes. Supondo independência, $X \sim B(n = 15, p = 0,02)$.

Devemos calcular $P(X \geq 1)$, que é dada por

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$

$$= 1 - \binom{15}{0} \times 0,02^0 \times (1 - 0,02)^{15-0}$$

$$= 1 - 0,98^{15} = 0,261.$$

Em Excel:

$$= 1 - \text{DISTRBINOM}(0; 15; 0,02; \text{FALSO})$$

Exemplo

Solução. (b) Definimos o evento sucesso (S) como “o lote contém pelo menos um componente defeituoso”. De acordo com o item (a), $P(S) = p = 0,261$. A v.a. Y é definida como sendo o número de lotes com pelo um componente defeituoso (sucessos) em $n = 10$ lotes. Supondo independência, $Y \sim B(n = 10, p = 0,261)$.

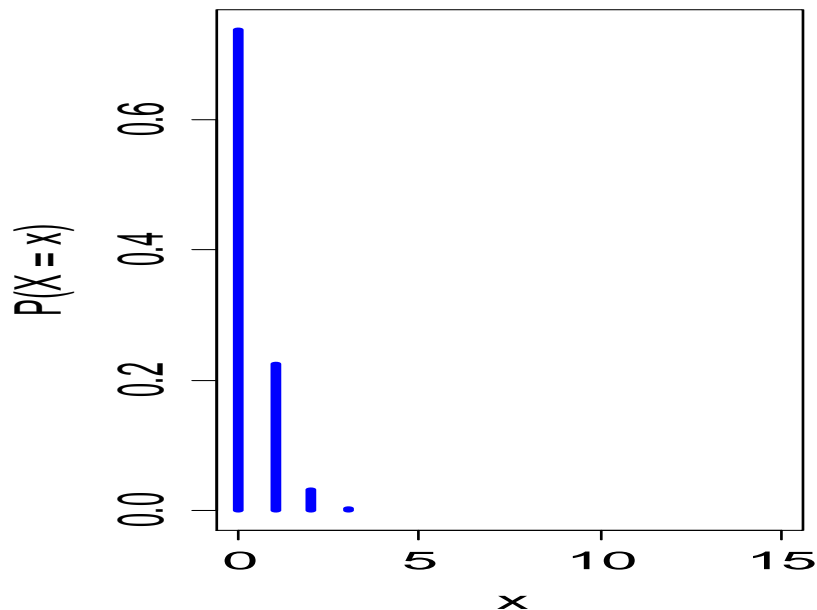
Devemos calcular $P(Y = 3)$, que é dada por

$$P(Y = 3) = \binom{10}{3} \times 0,261^3 \times (1 - 0,261)^{10-3} = 0,257.$$

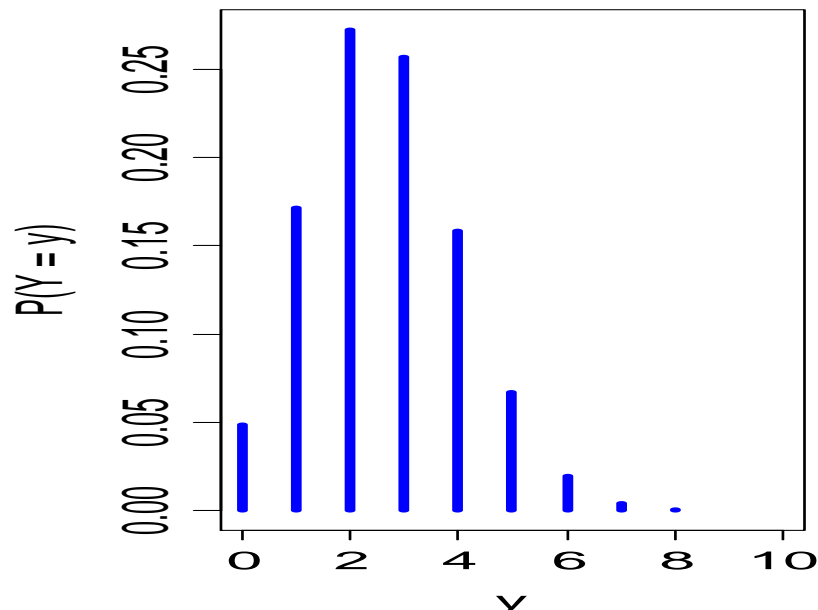
Em Excel:

= DISTRBINOM(3; 10; 0,261; FALSO)

(a)



(b)



4.3. Modelo hipergeométrico

Um conjunto de N elementos é dividido em duas classes. Uma classe com M ($M < N$) elementos (sucessos) e a outra com $N - M$ elementos (fracassos).

Por exemplo, no caso de N itens produzidos, podem ser considerados M itens defeituosos e $N - M$ itens não defeituosos.

Uma amostra de tamanho n ($n < N$) é sorteada sem reposição. A v.a. X é definida como o número de elementos com a característica de interesse (sucesso) na amostra de tamanho n .

(1) n elementos são selecionados de um conjunto de N elementos. (2) x sucessos são escolhidos de uma classe com M sucessos. (3) Finalmente, $n - x$ fracassos são escolhidos de uma classe com $N - M$ fracassos.

A função de probabilidade da v.a. X é

$$f(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, & \text{se } x = \max\{0, n - (N - M)\}, \dots, \min\{n, M\}, \\ 0, & \text{c.c.} \end{cases}$$

Notação: $X \sim H(N, M, n)$ indica que a v.a. X tem distribuição hipergeométrica com parâmetros N , M e n .

$$\text{Se } X \sim H(N, M, n), \text{ então } E(X) = n \left(\frac{M}{N} \right) \text{ e } \text{Var}(X) = n \left(\frac{M}{N} \right) \left(1 - \frac{M}{N} \right) \left(\frac{N-n}{N-1} \right).$$

Exemplo (Hines *et al.*, 2006, p. 105)

Em um departamento de inspeção de recebimento, lotes de eixo de bomba são recebidos periodicamente. Os lotes contêm 100 unidades e o seguinte plano de amostragem de aceitação é usado. Seleciona-se uma amostra de 10 unidades sem reposição. O lote é aceito se a amostra tiver, no máximo, um eixo defeituoso. Suponha que um lote seja recebido e que 5% dos itens sejam defeituosos. Qual a probabilidade de que o lote seja aceito ?

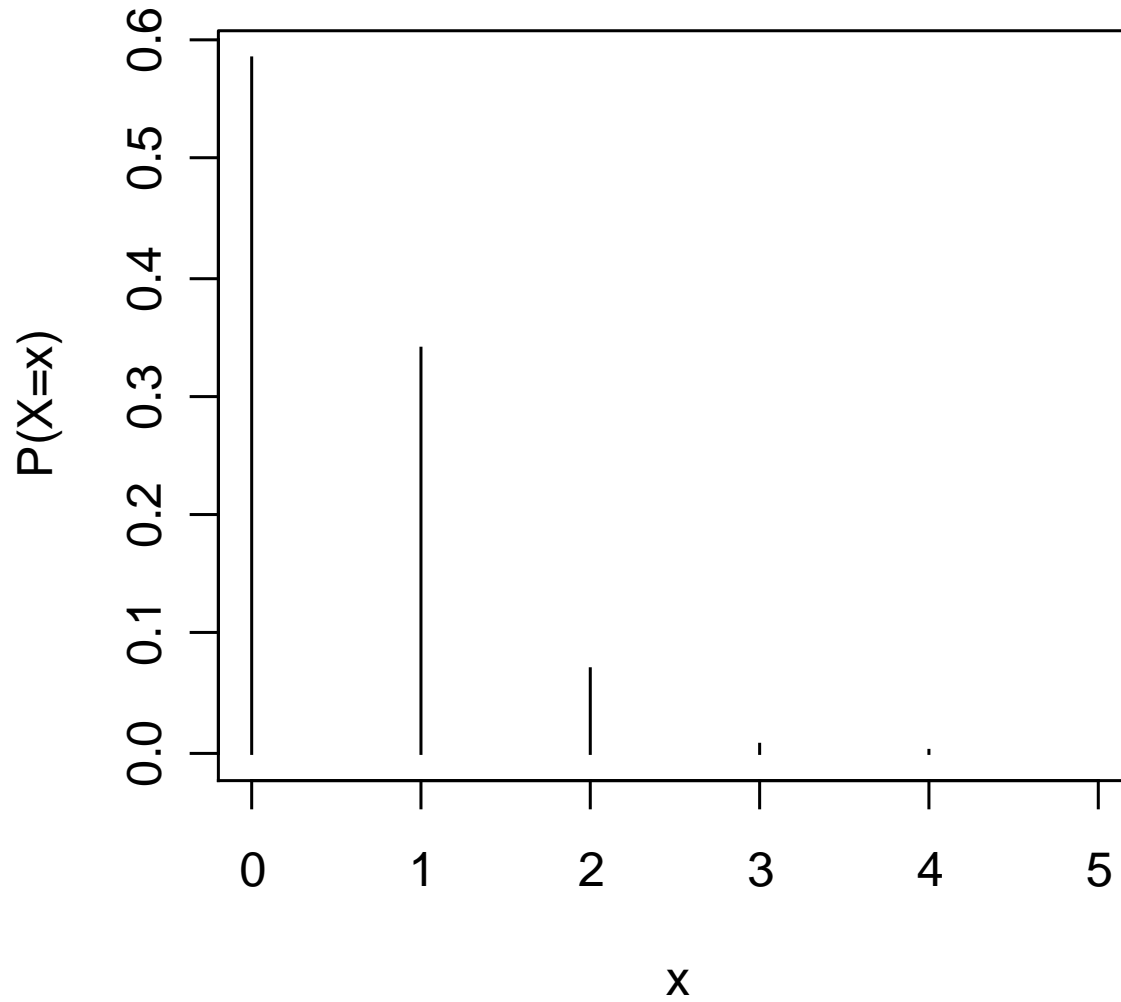
X: número de defeituosos na amostra $\Rightarrow X \sim H(N = 100, M = 5, n = 10)$.

$$P(\text{aceitar o lote}) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= \frac{\binom{5}{0} \binom{95}{10}}{\binom{100}{10}} + \frac{\binom{5}{1} \binom{95}{9}}{\binom{100}{10}} = 0,923.$$

Em R: `dhyper(0,5,95,10) + dhyper(1,5,95,10)` ou `phyper(1,5,95,10)`.

Em Excel: `=DIST.HIPERGEOM(0;10;5;100) + DIST.HIPERGEOM(1;10;5;100)`.



4.4. Modelo de Poisson

Muitos experimentos consistem em observar a ocorrência de eventos em determinada **unidade** (de tempo, volume, comprimento, área, ...)

Exemplos

1. Número de consultas a uma base de dados em **um minuto**.
2. Número de acidentes de trabalho **por semana** em uma fábrica.
3. Número de pequenas manchas por **m²** no esmaltado de uma geladeira.
4. Número de chamadas que chegam a uma central telefônica de uma empresa a cada **10 min**.
5. Número de carros que chegam ao campus entre **7:00 e 8:00h**.
6. Número de microorganismos por **cm³** de água contaminada.
7. Número de defeitos em cada **teclado** produzido por uma fábrica.

Suposições básicas

O fenômeno estudado ocorre em intervalos (de tempo, por exemplo).

O intervalo pode ser dividido em **subintervalos** com comprimentos suficientemente **pequenos** tais que

- a probabilidade de ocorrência de **mais um evento** em um subintervalo é **pequena**,
- a probabilidade de ocorrência de um evento em um subintervalo seja **a mesma** para todos os subintervalos e **proporcional** ao comprimento do subintervalo e
- a **contagem** em cada subintervalo seja independente de outros subintervalos.

Pode ser provado que a distribuição do **número** de ocorrências é **Poisson**.

Distribuição de Poisson

Uma v. a. discreta X tem distribuição de Poisson com **parâmetro** μ se sua função de probabilidade é dada por

$$f(x) = \begin{cases} \frac{e^{-\mu} \mu^x}{x!}, & \text{se } x = 0, 1, 2, \dots, \\ 0, & \text{c.c.}, \end{cases}$$

em que x é **número de eventos** em t unidades de medida,

λ é o número médio de eventos (**taxa**) em uma unidade de medida ($t = 1$) e

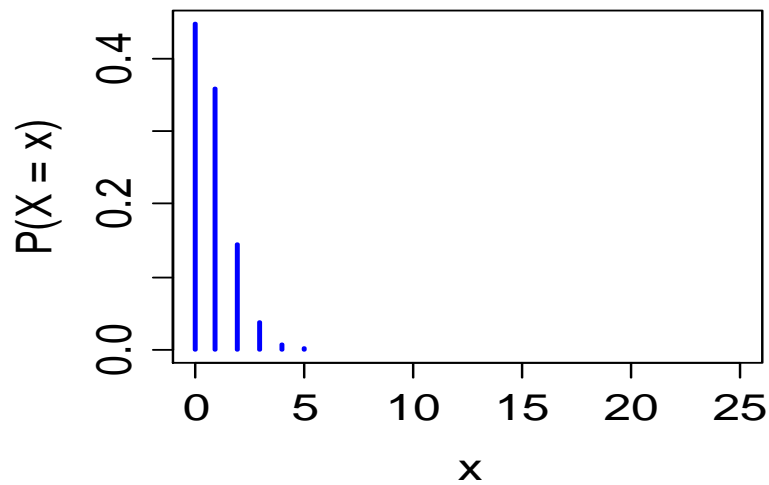
$\mu = \lambda t$ é o número médio de eventos em **t unidades** de medida.

Notação: $X \sim \text{Po}(\mu)$ indica que a v.a. X tem distribuição de Poisson com parâmetro μ .

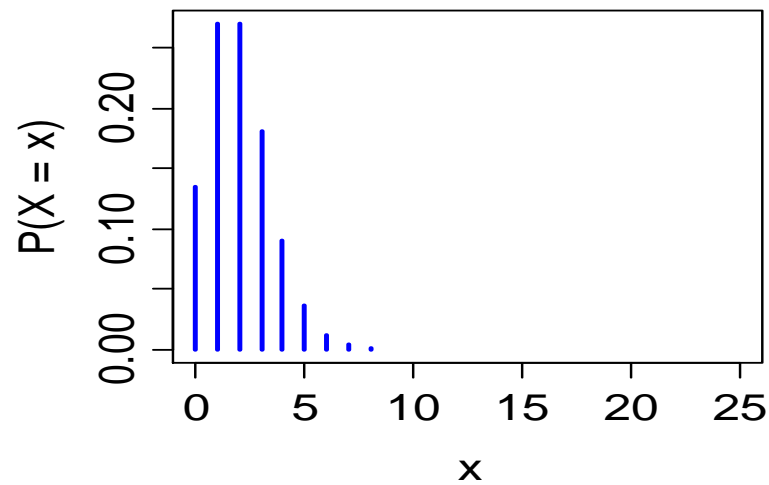
Propriedades: $E(X) = \mu$ e $\text{Var}(X) = \mu$.

Distribuição $Po(\mu)$

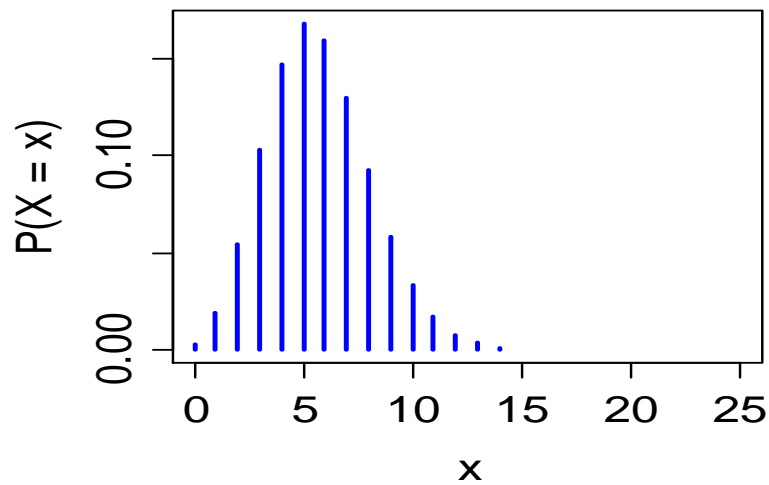
Po(0,8)



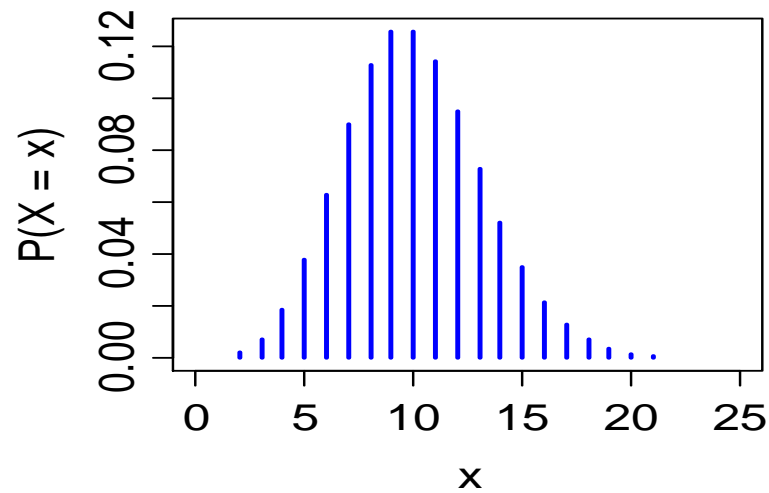
Po(2)



Po(5,7)



Po(10)



Exemplo

As chegadas a um posto de atendimento ocorrem de forma independente seguindo a distribuição de **Poisson**. Suponha que a **média** de chegadas é **3** a cada **4 minutos**. Qual é a probabilidade de que este posto receba **no máximo 2** solicitações em um intervalo de **2 minutos**?

Solução. Se X é número de chegadas a este posto **a cada 2 minutos**,

então $X \sim \text{Po}(\mu)$. Aqui, $t = 2$ min e $\lambda = \frac{3}{4} = 0,75$. Logo, $\mu = 0,75 \times 2 = 1,5$. Ou seja,

$X \sim \text{Po}(1,5)$ e

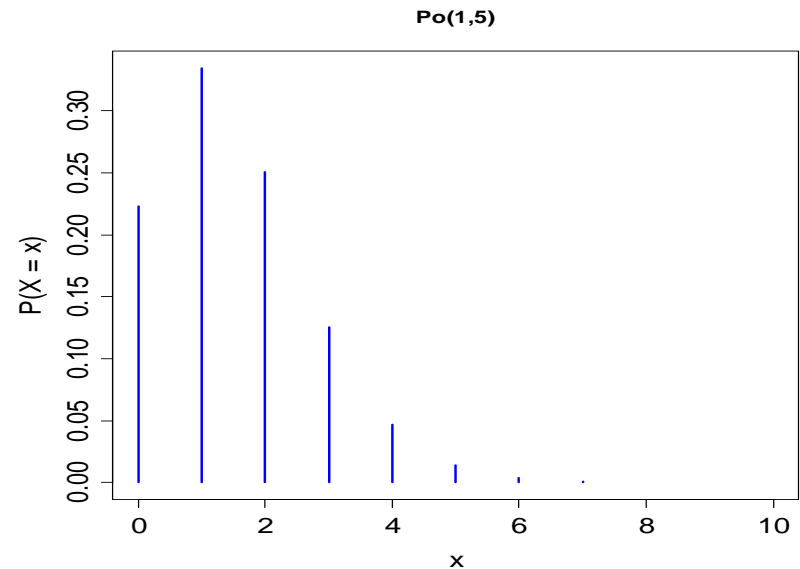
$$f(x) = \frac{e^{-1,5} 1,5^x}{x!}, \text{ se } x = 0,1,2,3,\dots$$

Calculamos

$$P(X \leq 2) = F(2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= e^{-1,5} \left(1 + 1,5 + \frac{1,5^2}{2} \right) = 0,809.$$

Em R: `ppois(2, 1.5)`; em Excel: `=POISSON(2;1,5; VERDADEIRO)`.



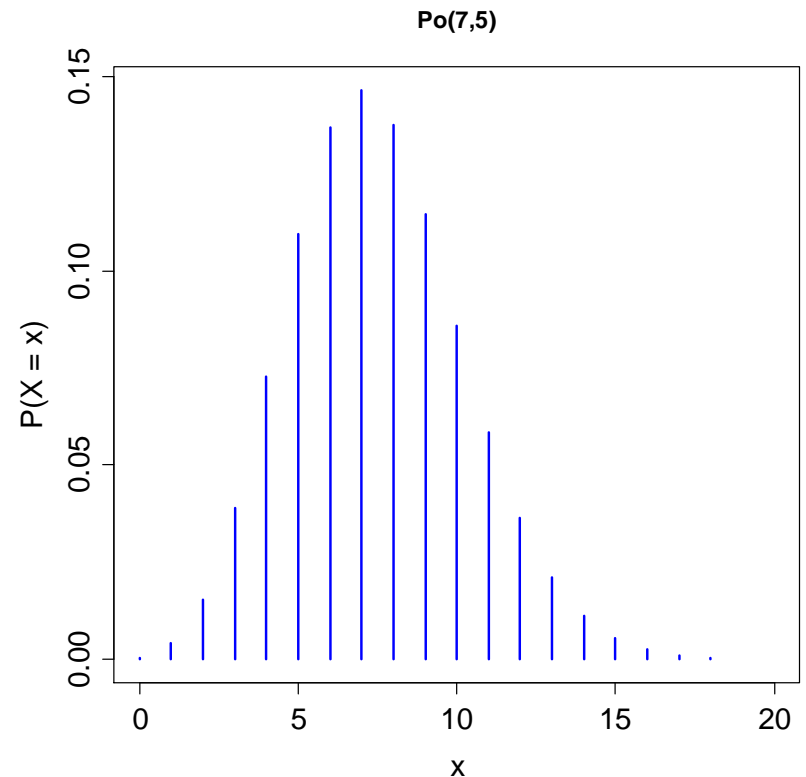
Exemplo

O número de **pedidos** de empréstimos que um banco recebe **por dia** é uma variável aleatória, sendo que **em média** são recebidos **7,5** empréstimos **por dia**. Determine as probabilidades de que, **em um dia** qualquer, o banco receba

- (a) **Exatamente 2** pedidos de empréstimo;
- (b) **No máximo 2** pedidos de empréstimo;
- (c) **No mínimo 8** pedidos de empréstimo.

Solução. Supomos que X (número de pedidos de empréstimos que o banco recebe por dia) tem distribuição **Poisson** com média $\mu = 7,5$. Logo,

$$f(x) = \frac{e^{-7,5} 7,5^x}{x!}, \text{ se } x = 0, 1, 2, \dots$$



Exemplo

x	f(x)=P(X=x)
0	0,000553
1	0,004148
2	0,015555
3	0,038889
4	0,072916
5	0,109375
6	0,136718
7	0,146484
8	0,137329
9	0,114440
10	0,085830
11	0,058521
12	0,036575
13	0,021101
14	0,011304
15	0,005652
16	0,002649
17	0,001169
18	0,000487
19	0,000192
20	0,000072
21	0,000026
22	0,000009
23	0,000003
24	0,000001
25	0,000000
26	0,000000
27	0,000000

Em R: `dpois(0:27, 7.5)`; em Excel:

Microsoft Excel - Pasta1

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela

B2 =POISSON(A2;7,5;FALSO)

	A	B	C	D	E	F
1	x	f(x)				
2	0	0,000553				
3	1	0,004148				
4	2	0,015555				
5	3	0,038889				
6	4	0,072916				
7	5	0,109375				
8	6	0,136718				
9	7	0,146484				
10	8	0,137329				
11	9	0,114440				
12	10	0,085830				
13	11	0,058521				
14	12	0,036575				
15	13	0,021101				
16	14	0,011304				
17	15	0,005652				
18	16	0,002649				
19	17	0,001169				
20	18	0,000487				
21	19	0,000192				
22	20	0,000072				
23	21	0,000026				
24	22	0,000009				
25	23	0,000003				
26	24	0,000001				
27	25	0,000000				
28	26	0,000000				
29	27	0,000000				
30						

Exemplo

Calculamos

$$(a) P(X = 2) = \frac{e^{-7,5} (7,5)^2}{2} = 0,0156,$$

$$(b) P(X \leq 2) = F(2) = P(X = 0) + P(X = 1) + P(X = 2) \\ = 0,000553 + 0,004148 + 0,015555 = 0,0203 \text{ e}$$

$$(c) P(X \geq 8) = 1 - P(X < 8) = 1 - F(7) = 1 - \sum_{x=0}^7 P(X = x) \\ = 1 - (0,000553 + \dots + 0,146484) \\ = 1 - 0,5246385 = 0,4754.$$

Exemplo

Contaminação é um problema de fabricação de discos ópticos. O número de **partículas** de contaminação que ocorrem em um disco óptico tem uma distribuição de Poisson e o número **médio** de partículas por **cm²** de superfície é **0,1**. A área da superfície do disco em estudo é **100 cm²**. Encontre a probabilidade de que 12 partículas sejam encontradas em um disco.

Solução. Se X é o número de partículas na superfície do disco, então $X \sim \text{Po}(\mu)$. Temos $t = 100 \text{ cm}^2$ e $\lambda = 0,1$ por cm^2 . Logo, $\mu = t \times \lambda = 100 \times 0,1 = 10$. Ou seja, $X \sim \text{Po}(10)$ e

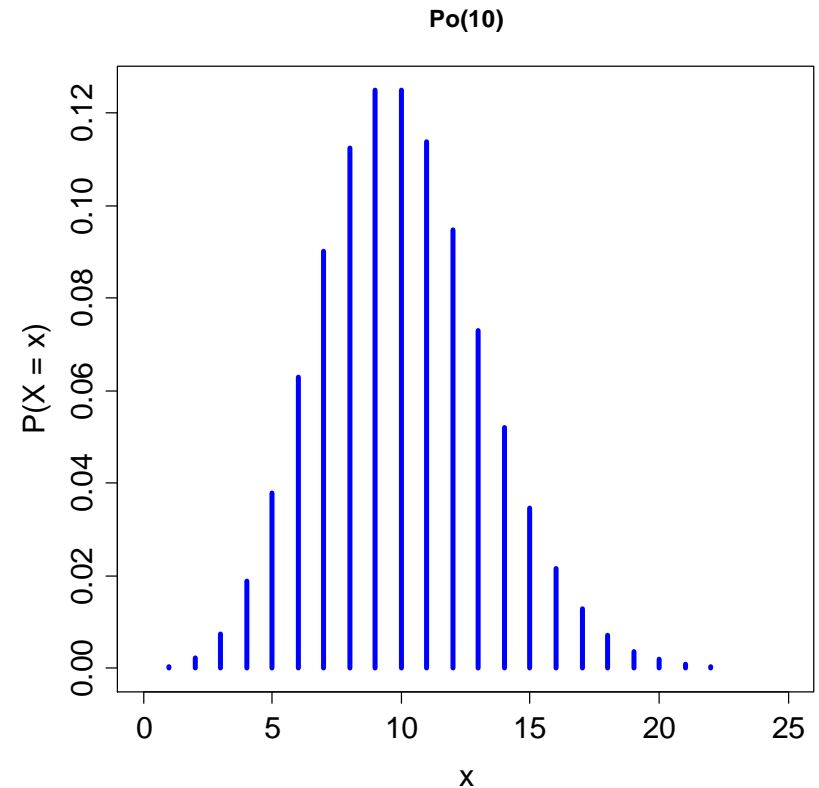
$$f(x) = \frac{e^{-10} 10^x}{x!}, \quad x = 0, 1, 2, \dots$$

Calculamos

$$P(X = 12) = \frac{e^{-10} 10^{12}}{12!} = 0,095.$$

Em R: `dpois(12,10);`

em Excel: `=POISSON(12;10;FALSO)`.



Resultado. Se X_1, \dots, X_n são variáveis aleatórias independentes com distribuição de Poisson com parâmetros μ_1, \dots, μ_n respectivamente, então a variável aleatória $Y = X_1 + \dots + X_n$ tem distribuição Poisson com parâmetro $\mu = \mu_1 + \dots + \mu_n$.

Exemplo. Em uma fábrica, dados históricos mostram que em três semanas típicas os números médios de acidentes são 2,5 na primeira semana, 2 na segunda semana e 1,5 na terceira semana. Suponha que o número de acidentes por semana segue uma distribuição de Poisson. Qual a probabilidade de que ocorram 4 acidentes em três semanas típicas?

Solução. X_i representa o número de acidentes na i -ésima semana, $i = 1, 2, 3$, com $X_i \sim \text{Po}(\mu_i)$. Supomos que X_1, X_2 e X_3 são independentes. Portanto, $Y = X_1 + X_2 + X_3$ tem distribuição Poisson com parâmetro $\mu = 2,5 + 2 + 1,5 = 6$. Calculamos

$$P(Y = 4) = \frac{6^4 e^{-6}}{4!} = 0,1339.$$

4.5. Modelo geométrico

Ensaio de **Bernoulli** são realizados de forma **independente** e cada um com probabilidade de **sucesso** igual a p .

Estamos interessados no número de ensaios que **antecedem** a ocorrência do **1º** sucesso.

A v.a. X que conta este número tem **distribuição geométrica** com parâmetro p , notando que $X \in \{0, 1, 2, \dots\}$.

Se “**S**” e “**F**” representam os eventos sucesso e fracasso e $X = x$, temos a sequência

$$\underbrace{\text{F} \quad \text{F} \quad \dots \quad \text{FS.}}_{x \text{ fracassos}}$$

Sendo assim,

$$P(X = x) = \underbrace{(1-p) \times (1-p) \times \dots \times (1-p)}_{x \text{ fracassos}} \times p.$$

Distribuição geométrica

Se ensaios de Bernoulli independentes e com probabilidade de sucesso igual a p são realizados, o número de ensaios que **antecedem** o **primeiro sucesso** tem uma distribuição geométrica com parâmetro p e sua função de probabilidade é dada por

$$f(x) = P(X = x) = (1 - p)^x p, \quad \text{se } x = 0, 1, 2, \dots \text{ e } 0 < p < 1.$$

Notação: $X \sim \text{Geo}(p)$.

Se $X \sim \text{Geo}(p)$, então

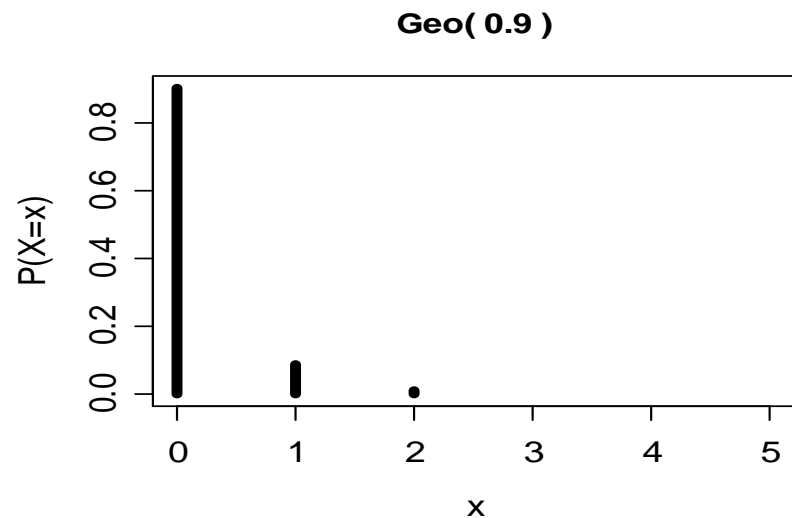
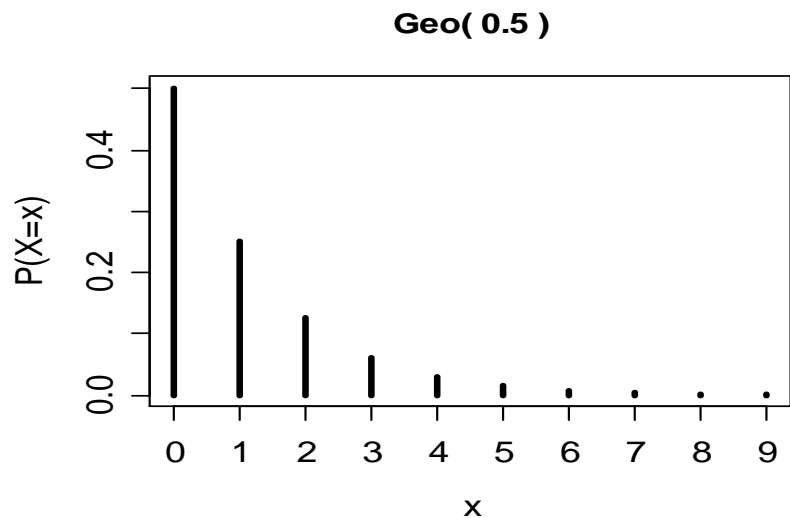
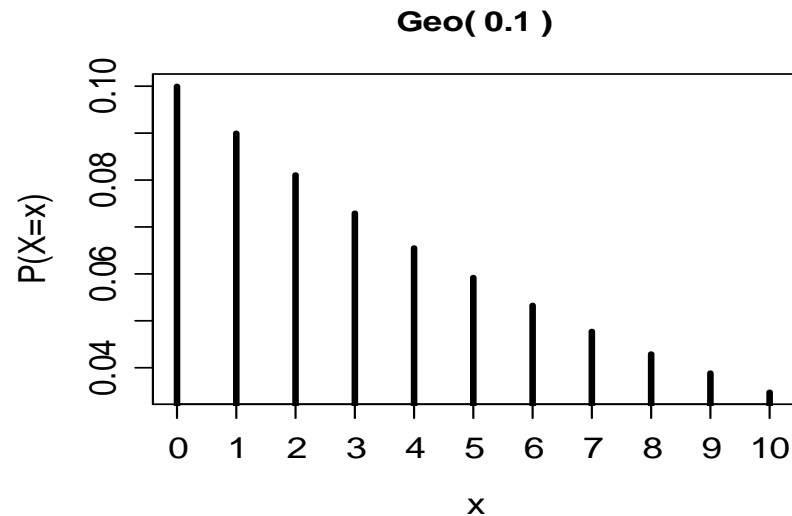
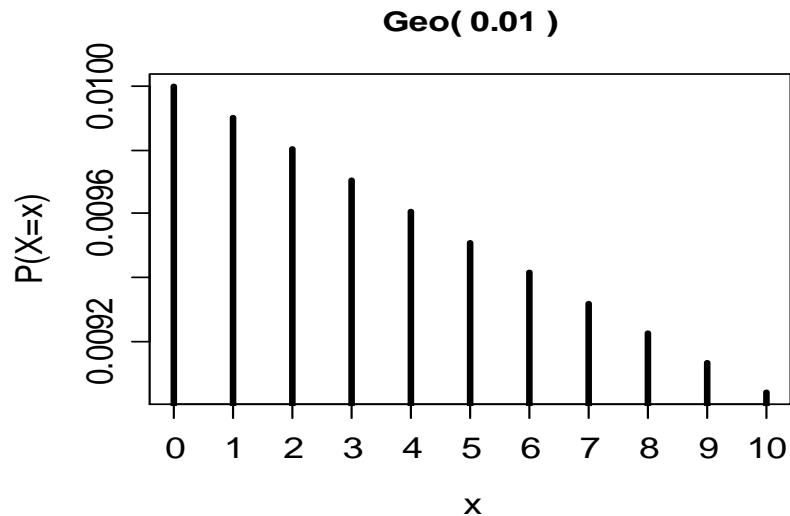
$$E(X) = (1 - p) / p \text{ e}$$

$$\text{Var}(X) = (1 - p) / p^2.$$

Propriedade: Se $X \sim \text{Geo}(p)$, então $P(X > k + m \mid X > m) = P(X > k)$.

É a **única** distribuição discreta com esta propriedade (“**falta de memória**”).

Distribuição Geo(p)



Outra definição de distribuição geométrica

Se ensaios de Bernoulli independentes e com probabilidade de sucesso igual a p são realizados, o número de ensaios Y até que ocorra o primeiro sucesso tem uma distribuição geométrica com parâmetro p e sua função de probabilidade é dada por

$$f(y) = P(Y = y) = (1 - p)^{y-1} p, \quad \text{se } y = 1, 2, \dots \text{ e } 0 < p < 1.$$

Relação entre as duas definições:

$$Y = X + 1,$$

$$E(Y) = E(X) + 1 = (1 - p) / p + 1 = 1 / p \quad \text{e}$$

$$\text{Var}(Y) = \text{Var}(X) = (1 - p) / p^2.$$

Obs. Qual a relação entre a distribuição geométrica e os álbuns de figurinhas?

Exemplo (Hines *et al.*, 2006, p. 101)

Certo experimento deve ser realizado **até que** seja obtido um resultado bem sucedido. As realizações são independentes e o **custo** de cada experimento é **\$25.000**, sendo que se o resultado for um **insucesso**, há um custo **adicional** de **\$5.000** para o preparo da próxima realização.

- (a) Obtenha o **custo esperado** do experimento.
- (b) Se o orçamento **não pode** ultrapassar \$500.000, qual a probabilidade de que este valor seja ultrapassado.

Solução. Definimos Y como sendo o número de realizações até que ocorra o primeiro resultado bem sucedido, notando que $Y \in \{1, 2, \dots\}$ e tem distribuição geométrica com parâmetro p e $f(y) = (1 - p)^{y-1}p$ (veja lâmina 29). Pelo enunciado, o custo é uma v.a., função de Y , dada por

$$C(Y) = 25000 Y + 5000 (Y - 1) = 30000 Y - 5000.$$

Usando **propriedades** do valor esperado obtemos

$$E[C(Y)] = 30000 E(Y) - 5000 = 30000 / p - 5000.$$

Se $p = 0,25$, o custo esperado vale \$115.000.

Exemplo (Hines *et al.*, 2006, p. 101)

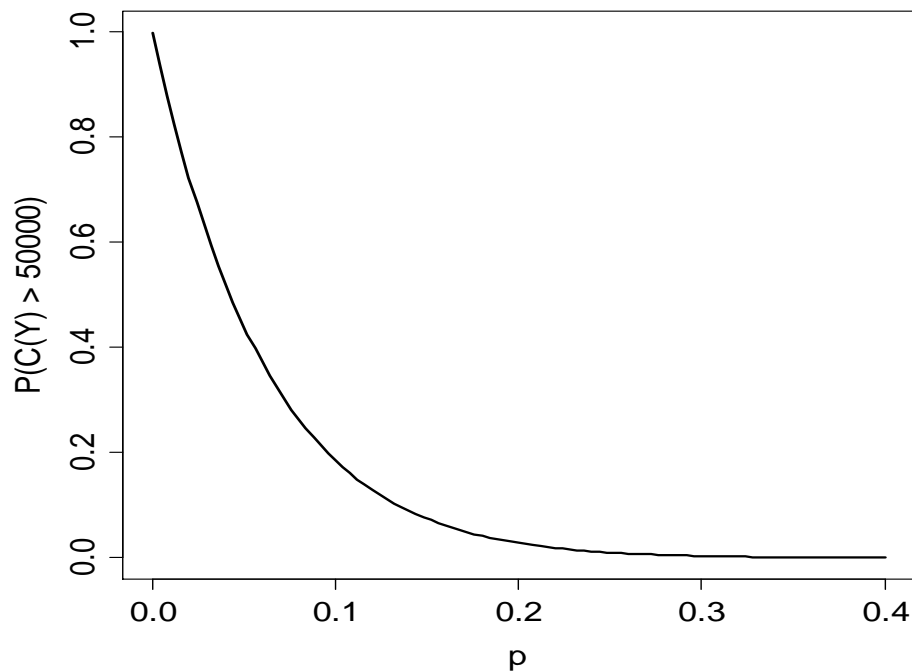
Na letra (b) devemos calcular $P(C(Y) > 500000)$. Usando a expressão de $C(Y)$,

$$\begin{aligned} P(C(Y) > 500000) &= P(30000 Y - 5000) > 500000 \\ &= P(Y > 505000 / 30000) = P(Y > 16,8) \\ &= 1 - P(Y \leq 16,8) = 1 - P(Y \leq 16) \\ &= 1 - \sum_{k=1}^{16} (1-p)^{k-1} p, \quad 0 < p < 1. \end{aligned}$$

Se $p = 0,25$,

$P(C(Y) > 500000) = 0,010$.

Em R: `1 - pgeom(15, 0.25)`.



4.6. Modelo binomial negativa

Ensaio de **Bernoulli** são realizados de forma **independente** e cada um com probabilidade de **sucesso** igual a p .

Interesse no número de ensaios que **até que** ocorram r sucessos, $r \geq 1$.

A v.a. X que conta este número tem **distribuição binomial negativa** com parâmetros r e p , notando que $X \in \{r, r+1, r+2, \dots\}$.

Se “**S**” e “**F**” representam os eventos sucesso e fracasso e $X = x$, temos sequências do tipo

$\underbrace{\text{FSF SF} \dots \text{FS}}_{r-1 \text{ sucessos em } x-1 \text{ ensaios}}$, cada uma com probabilidade = $p^r (1-p)^{x-r}$.

$$\text{Número de sequencias} = \binom{x-1}{r-1} = \frac{(x-1)!}{(r-1)!(x-r)!}.$$

Distribuição binomial negativa

Se **ensaios** de Bernoulli **independentes** e com probabilidade de sucesso igual a **p** são realizados, o número de ensaios **até que** ocorram **r** sucessos tem uma distribuição binomial negativa com parâmetros **r** e **p**. Sua função de probabilidade é dada por

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \text{ se } x = r, r+1, r+2, \dots \text{ e } 0 < p < 1.$$

Notação: $X \sim \text{BN}(r, p)$.

Se $X \sim \text{BN}(r, p)$, então

$$E(X) = r / p \text{ e}$$

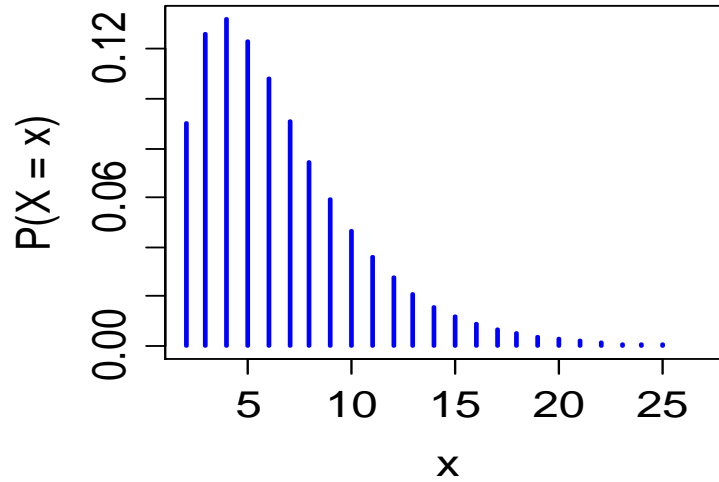
$$\text{Var}(X) = r (1 - p) / p^2.$$

Obs. (a) $r = 1$: distribuição geométrica na lâmina 29.

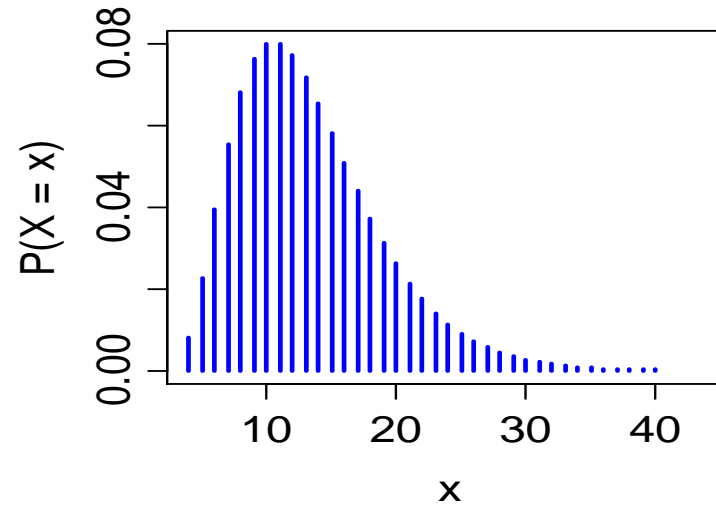
(b) Em **Excel**: função **DIST.BIN.NEG**.

Distribuição BN(r, p)

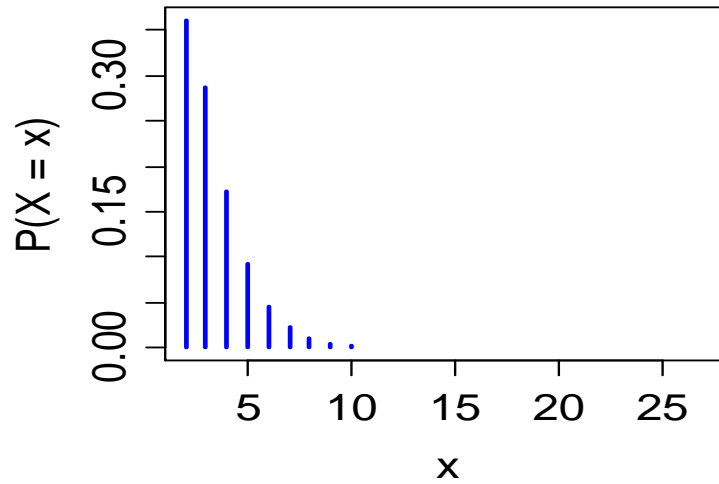
BN($r = 2, p = 0.3$)



BN($r = 4, p = 0.3$)



BN($r = 2, p = 0.6$)



BN($r = 4, p = 0.5$)

