

Variáveis aleatórias discretas

Wagner H. Bonat
Elias T. Krainski
Fernando P. Mayer

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação



Sumário

- 1 Introdução
- 2 Função de Probabilidade
- 3 Distribuições discretas de probabilidade
- 4 Exercícios recomendados

Resumo

- Fenômeno aleatório: situação ou acontecimento cujos resultados não podem ser previstos com certeza.
- Espaço amostral: conjunto de todos os resultados possíveis de um fenômeno aleatório, denotado por Ω .
- Eventos: subconjuntos de Ω , denotado por A, B, \dots
- Conjunto vazio: conjunto sem eventos, denotado por \emptyset .
- União $A \cup B$: ocorrência de pelo menos um dos eventos A ou B .
- Intersecção $A \cap B$: ocorrência simultânea de A e B .
- Eventos disjuntos ou mutuamente exclusivos: $A \cap B = \emptyset$.
- Eventos complementares: $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$.

Definição e Notação

Definição

- **Variável aleatória** - Descrição numérica do resultado de um fenômeno aleatório.

Notação

- X denota a variável aleatória.
- x denota os valores realizados da v.a.
- Probabilidade de X assumir o valor x é denotada $P(X = x)$.

Exemplo

- X : Número de alunos em uma sala de aula.
- Uma possível realização $x = 50$.

Tipos de dados, espaço amostral e v.a

- Tipos de dados

- 1 Dados na reta real, $\Omega = \mathbb{R}$.
- 2 Dados estritamente positivos, $\Omega = \mathbb{R}_+$.
- 3 Dados positivos com zeros, $\Omega = \mathbb{R}_0 = [0, \infty)$.
- 4 Proporções, $\Omega = (0, 1)$.
- 5 Direções, $\Omega = [0, 2\pi)$.
- 6 Contagens, $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$.
- 7 Binomial, $\Omega = \{0, 1, 2, \dots, m\}$.

- Tipos de espaço amostral

- 1 **Espaço amostral Discreto**: Contêm apenas um número finito ou contável de elementos.
- 2 **Espaço amostral Contínuo**: Contêm um número infinito de elementos.

- Tipos de variáveis aleatórias

- 1 Variável aleatória é **contínua** se seu espaço amostral é contínuo.
- 2 Variável aleatória é **dicreta** se seu espaço amostral é discreto.

Descrição probabilística de v.a

- Dada a realização de um experimento aleatório qualquer, com um certo espaço de probabilidade, desejamos estudar a **estrutura probabilística** de quantidades associadas à esse experimento.
- Note que antes da realização de um experimento, **não sabemos seu resultado**, entretanto seu espaço amostral pode ser estabelecido com precisão.
- Podemos atribuir probabilidades aos *eventos* deste espaço amostral, dando origem a idéia de **distribuição de probabilidade**.
- Em geral vamos distinguir a distribuição de probabilidade de v.a's discretas e contínuas.

Sumário

- 1 Introdução
- 2 Função de Probabilidade
- 3 Distribuições discretas de probabilidade
- 4 Exercícios recomendados

Definição

A **função de probabilidade** (fp) da v.a discreta X , que assume os valores $x_1, x_2, \dots, x_n, \dots$, é a função que atribui probabilidades a cada um dos possíveis valores: $\{[x_i, p(x_i)], i = 1, 2, \dots\}$, ou seja,

$$P[X = x_i] = p(x_i) = p_i, \quad i = 1, 2, \dots$$

com as seguintes propriedades:

- 1 A probabilidade de cada valor deve estar entre 0 e 1

$$0 \leq p(x_i) \leq 1, \quad i = 1, 2, \dots$$

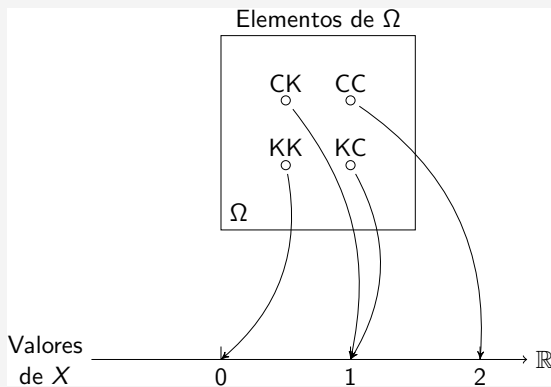
- 2 A soma de todas as probabilidades é igual a 1

$$\sum_i p(x_i) = 1.$$

Exemplo 3.3

Experimento

Lançamento de duas moedas. $X =$ número de resultados cara (C).



Exemplo 3.3

Podemos montar uma tabela de distribuição de frequência para a variável aleatória $X =$ número de resultados cara (C).

X	Frequência (f_i)	Frequência relativa (fr_i)
0	1	1/4
1	2	2/4
2	1	1/4
Total	4	1

Assim podemos associar a cada valor de X sua **probabilidade** correspondente, como resultado das **frequências relativas**.

$$P[X = 0] = 1/4$$

$$P[X = 1] = 2/4 = 1/2$$

$$P[X = 2] = 1/4$$

Exemplo 3.3

Dessa forma, a **distribuição de probabilidade** da variável aleatória $X =$ número de resultados cara (C) é a tabela:

X	$P[X = x_i] = p(x_i)$
0	1/4
1	1/2
2	1/4
Total	1

Repare que as propriedades da função de probabilidade estão satisfeitas:

- 1 As probabilidades $p(x_i)$ estão entre 0 e 1.
- 2 A soma de todas as probabilidades $p(x_i)$ é 1.

Exemplo 3.1

Com dados do último censo, a assistente social de um Centro de Saúde constatou que para as famílias da região, 20% não tem filhos, 30% tem um filho, 35% tem dois, e as restantes se dividem igualmente entre três, quatro ou cinco filhos.

Descreva a função de probabilidade da v.a N definida como número de filhos.

Exemplo 3.2

Na construção de um certo prédio, as fundações devem atingir 15 metros de profundidade e, para cada 5 metros de estacas colocadas, o operador anota se houve alteração no ritmo de perfuração previamente estabelecido. Essa alteração é resultado de mudanças para mais ou para menos, na resistência do subsolo.

Nos dois casos, medidas corretivas serão necessárias, encarecendo o custo da obra. Com base em avaliações geológicas, admite-se que a probabilidade de ocorrência de alterações é de 0.1 para cada 5 metros.

O custo básico inicial é de 100 UPC (unidade padrão de construção) e será acrescida de $50k$, com k representando o número de alterações observadas. Como se comporta a v.a custo das obras de fundações?

Função de distribuição de probabilidade

- Em muitas situações, é útil calcularmos a probabilidade **acumulada** até um certo valor.
- Definimos a **função de distribuição** ou **função acumulada de probabilidade** de uma v.a X pela expressão:

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

para qualquer número real x .

Exemplo 3.5

Uma população de 1000 crianças foi analisada num estudo para determinar a efetividade de uma vacina contra um tipo de alergia.

No estudo, as crianças recebiam uma dose da vacina e, após um mês, passavam por um novo teste. Caso ainda tivessem tido alguma reação alérgica, recebiam outra dose da vacina. Ao fim de 5 doses todas as crianças foram consideradas imunizadas.

Os resultados estão na tabela a seguir.

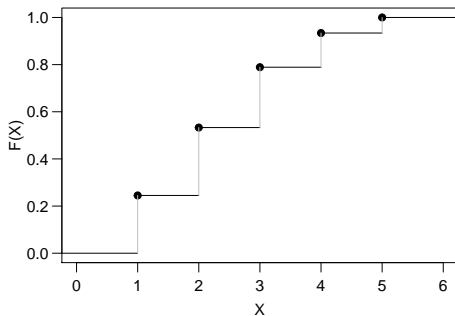
	1	2	3	4	5
Freq.	245	288	256	145	66

Para uma criança sorteado ao acaso qual a probabilidade dela ter recebido 2 doses? E até 2 doses?

Exemplo 3.5

Tabela de frequência:

	n_i	f_i	f_{ac}
1	245	0.245	0.245
2	288	0.288	0.533
3	256	0.256	0.789
4	145	0.145	0.934
5	66	0.066	1.000
Sum	1000	1.000	

Gráfico de $F(X)$:

Exemplo 3.5

Assim,

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0,533.$$

Note que podemos escrever

$$F(x) = P(X \leq x) = 0,533 \quad \text{para} \quad 2 \leq x < 3.$$

Função de distribuição

$$F(x) = \begin{cases} 0 & \text{se } x < 1 \\ 0,245 & \text{se } 1 \leq x < 2 \\ 0,533 & \text{se } 2 \leq x < 3 \\ 0,789 & \text{se } 3 \leq x < 4 \\ 0,934 & \text{se } 4 \leq x < 5 \\ 1 & \text{se } x \geq 5 \end{cases}$$

Exemplo 3.6

Num estudo sobre a incidência de câncer foi registrado, para cada paciente com esse diagnóstico, o número de casos de câncer em parentes próximos (pais, irmãos, tios, filhos, primos e sobrinhos). A distribuição de frequência para 26 pacientes é a seguinte:

	0	1	2	3	4	5
n_i	4	4	6	6	2	4

Estudos anteriores assumem que a incidência de câncer em parentes próximos pode ser teoricamente modelada pela seguinte função discreta de probabilidade:

	1	2	3	4	5	6
p_i	0.1	0.1	0.3	0.3	0.1	0.1

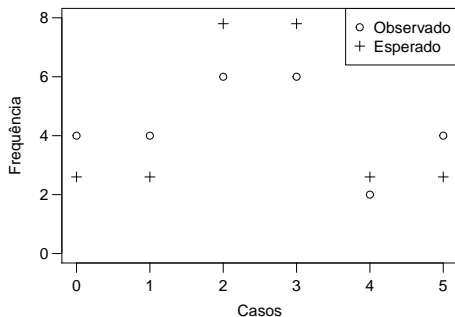
Os dados observados concordam com o modelo teórico?

Exemplo 3.6

O número de observações de incidência esperado seguindo o modelo teórico é calculado como

$$e_j = n \times p_j.$$

	n_j	e_j
0	4	2.6
1	4	2.6
2	6	7.8
3	6	7.8
4	2	2.6
5	4	2.6
Sum	26	26.0



Sumário

- 1 Introdução
- 2 Função de Probabilidade
- 3 Distribuições discretas de probabilidade**
- 4 Exercícios recomendados

Modelo Uniforme Discreto

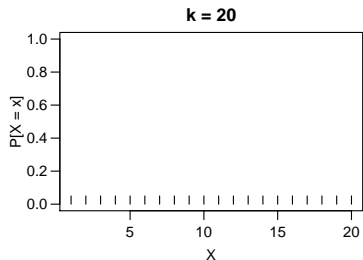
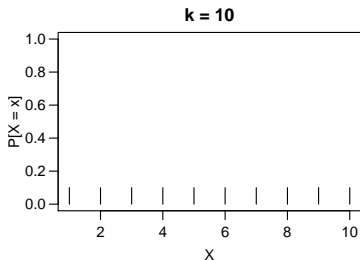
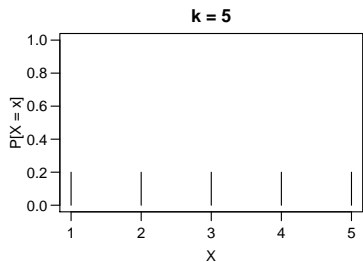
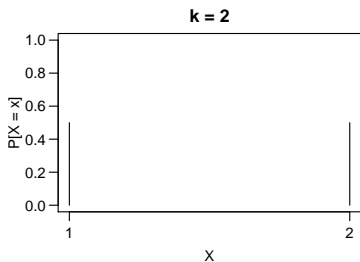
Definição: Seja X uma v.a assumindo valores $1, 2, \dots, k$. Dizemos que X segue o modelo **Uniforme Discreto** se atribui a mesma probabilidade $1/k$ a cada um desses k valores.

Então, sua função de probabilidade é dada por

$$P[X = j] = \frac{1}{k}, \quad j = 1, 2, \dots, k.$$

Notação: $X \sim U_D[1, k]$

Modelo Uniforme Discreto



Exemplo 3.7

Uma rifa tem 100 bilhetes numerados de 1 a 100. Tenho 5 bilhetes consecutivos numerados de 21 a 25 e meu colega tem outros 5 bilhetes, com os números 1, 11, 29, 68 e 93. Quem tem maior possibilidade de ser sorteado?

Modelo Bernoulli

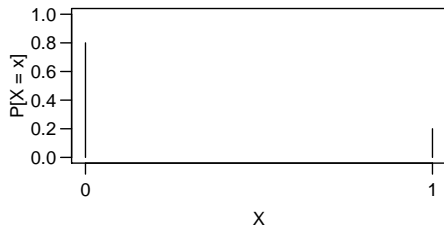
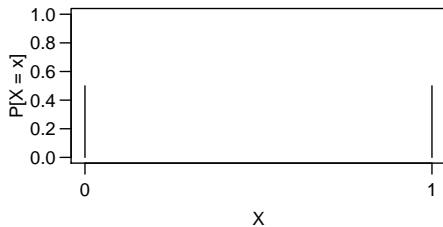
Definição: Uma variável aleatória X segue o modelo Bernoulli se assume apenas os valores 0 (“fracasso”) ou 1 (“sucesso”). Sua função de probabilidade é dada por

$$P[X = x] = p^x(1 - p)^{1-x}, \quad x = 0, 1$$

onde o parâmetro $0 \leq p \leq 1$ é a probabilidade de sucesso.

Notação: $X \sim \text{Ber}(p)$

Modelo Bernoulli

p = 0.2**p = 0.5****p = 0.7****p = 0.9**

Exemplo

No lançamento de uma moeda, considere cara como o evento de sucesso. Qual a probabilidade de sair cara, sendo que $p = 1/2$?

$$X = \begin{cases} 1, & \text{cara} \\ 0, & \text{coroa} \end{cases}$$

Temos que

X	$P[X = x]$	$p = 1/2$
0	$1 - p$	$1/2$
1	p	$1/2$

Modelo Binomial

Definição: Seja um experimento realizado dentro das seguintes condições:

- 1 São realizados n “ensaios” de Bernoulli independentes.
- 2 Cada ensaio só pode ter dois resultados possíveis: “sucesso” ou “fracasso”.
- 3 A probabilidade p de sucesso em cada ensaio é constante.

Vamos associar a v.a X o número de sucessos em n ensaios de Bernoulli. Portanto X poderá assumir os valores $0, 1, \dots, n$.

Modelo Binomial

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

$$\underbrace{1, 1, 1, \dots, 1}_x, \underbrace{0, 0, 0, \dots, 0}_{n-x}.$$

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,

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

Modelo Binomial

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

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

onde

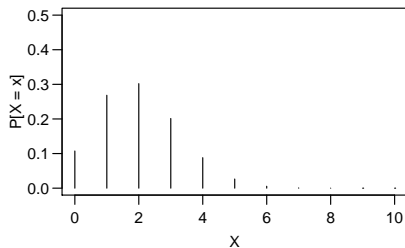
$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

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

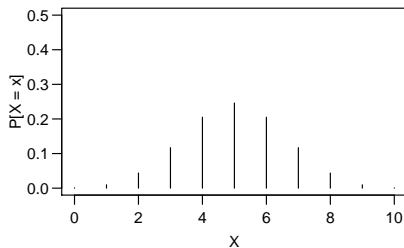
Notação: $X \sim B(n, p)$

Modelo Binomial

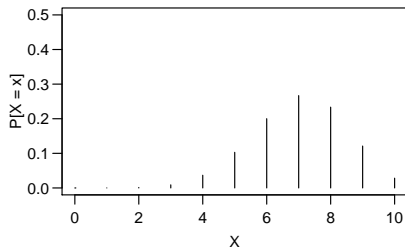
$n = 10, p = 0.2$



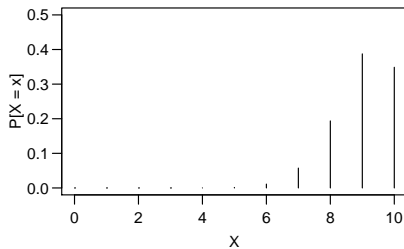
$n = 10, p = 0.5$



$n = 10, p = 0.7$



$n = 10, p = 0.9$



Modelo Binomial

- As probabilidades são completamente caracterizadas pela informação dos **parâmetros**.
- Por exemplo, para calcular a probabilidade de 3 sucessos de uma $B(12, 0.4)$ fazemos

$$P[X = 3] = \binom{12}{3} 0.4^3 0.6^9 = \frac{12!}{3!9!} 0.4^3 0.6^9 = 0.142$$

Diversos programas computacionais fazem esse cálculo facilmente. Por exemplo, com o R

```
round(dbinom(3, size = 12, prob = 0.4), 3)
```

```
[1] 0.142
```

Exemplo 3.9

O escore de um teste internacional de proficiência na língua inglesa varia de 0 a 700 pontos, com mais pontos indicando um melhor desempenho. Informações, coletadas durante vários anos, permite estabelecer o seguinte modelo para o desempenho no teste:

Pontos	[0, 200)	[200, 300)	[300, 400)	[400, 500)	[500, 600)	[600, 700)
p_i	0.06	0.15	0.16	0.25	0.28	0.10

Várias universidades americanas, exigem um escore mínimo de 600 pontos para aceitar candidatos de países de língua não inglesa. De um grande grupo de estudantes brasileiros que prestaram o último exame, escolhemos ao acaso 20 deles. Qual é a probabilidade de no máximo 3 atenderem ao requisito mínimo?

Exemplo 3.10

Um veterinário está estudando o índice de natalidade em porcos sujeitos à inseminação artificial. Para tal, coletou informações sobre a variável número de filhotes nascidos vivos em cada uma das 100 inseminações realizadas com o mesmo reprodutor. A tabela a seguir apresenta os resultados.

Número	1	2	3	4	5	6	7	8	9
Freq. Obs.	1	6	7	23	26	21	12	3	1

O veterinário informa que 11 ou mais filhotes nascidos vivos é uma ocorrência muito rara e pode ser desprezada em termos do modelo. Ele sugeriu que a variável N : número de filhotes nascidos vivos, poderia ser ajustada pelo modelo Binomial com parâmetros $n = 10$ e $p = 0.5$. O que você acha da sugestão do veterinário?

Exemplo 3.10

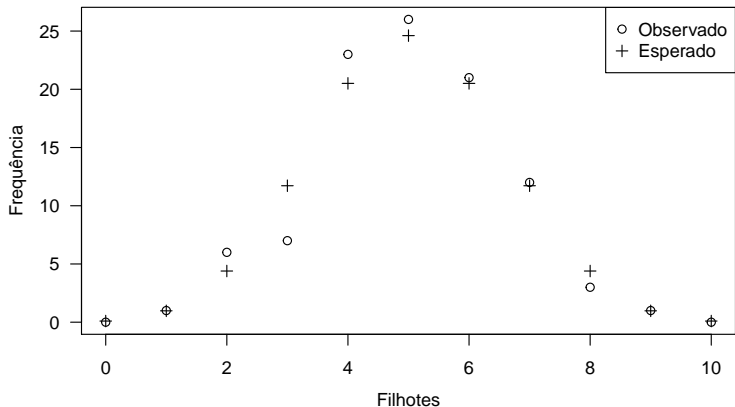
Usando o R temos:

```
obs <- c(0, 1, 6, 7, 23, 26, 21, 12, 3, 1, 0)
prob <- dbinom(0:10, size = 10, prob = 0.5)
esp <- prob * 100
cbind("Obs" = obs, "p" = round(prob, 2), "Esp" = round(esp, 2))
```

#		Obs	p	Esp
#	[1,]	0	0.00	0.10
#	[2,]	1	0.01	0.98
#	[3,]	6	0.04	4.39
#	[4,]	7	0.12	11.72
#	[5,]	23	0.21	20.51
#	[6,]	26	0.25	24.61
#	[7,]	21	0.21	20.51
#	[8,]	12	0.12	11.72
#	[9,]	3	0.04	4.39
#	[10,]	1	0.01	0.98
#	[11,]	0	0.00	0.10

Exemplo 3.10

```
plot(0:10, obs, xlab = "Filhotes", ylab = "Frequência")
points(0:10, esp, pch = 3)
legend("topright", legend = c("Observado", "Esperado"), pch = c(1, 3))
```



Modelo Geométrico

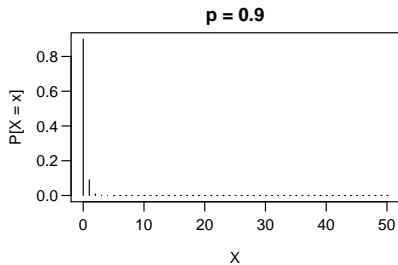
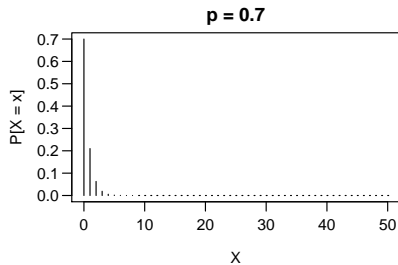
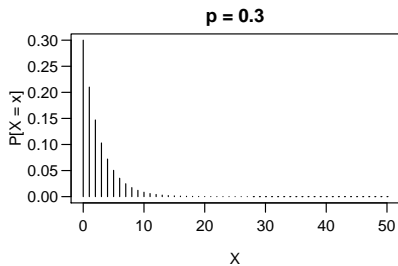
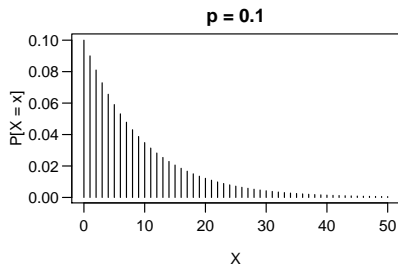
Definição: Considere o número (k) de ensaios Bernoulli **que precedem o primeiro sucesso**. Nesse caso, dizemos que a v.a X tem distribuição Geométrica de parâmetro p , e sua função de probabilidade tem a forma

$$P(X = k) = p(1 - p)^k, \quad k = 0, 1, 2, \dots$$

onde $0 \leq p \leq 1$ é a probabilidade de sucesso.

Notação: $X \sim G(p)$.

Modelo Geométrico



Exemplo 3.11

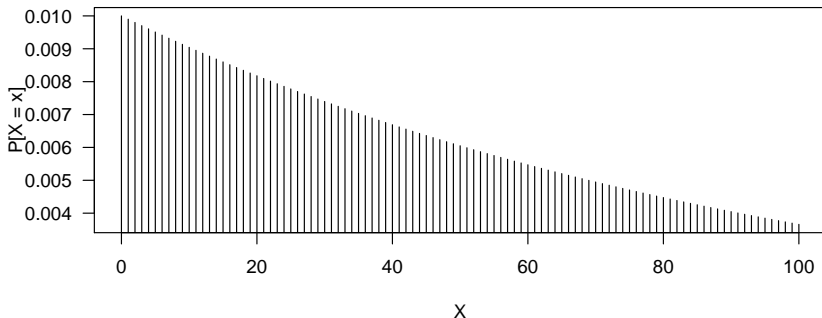
Uma linha de produção está sendo analisada para efeito de controle da qualidade das peças produzidas. Tendo em vista o alto padrão requerido, a produção é interrompida para regulagem toda vez que uma peça defeituosa é observada. Se 0.01 é a probabilidade da peça ser defeituosa, estude o comportamento da variável Q , quantidade de peças boas produzidas antes da primeira defeituosa.

Exemplo 3.11

$$P(Q = k) = 0.01 \times 0.99^k, \quad k = 0, 1, 2, \dots$$

Usando o R temos

```
x = c(0:100)
dx = dgeom(x = x, prob = 0.01)
plot(x, dx, type = "h", xlab = "X", ylab = "P[X = x]")
```



Modelo Poisson

Definição: Seja um experimento realizado nas seguintes condições:

- 1 As ocorrências são independentes.
- 2 As ocorrências são aleatórias.
- 3 A variável aleatória X é o número de ocorrências de um evento **ao longo de algum intervalo** (de tempo ou espaço).

Denominamos esse experimento de **processo de Poisson**.

Vamos associar a v.a X o número de ocorrências em um intervalo. Portanto X poderá assumir os valores $0, 1, \dots$ (sem limite superior).

Modelo Poisson

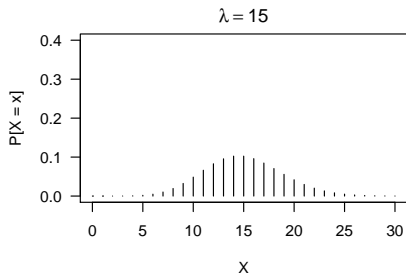
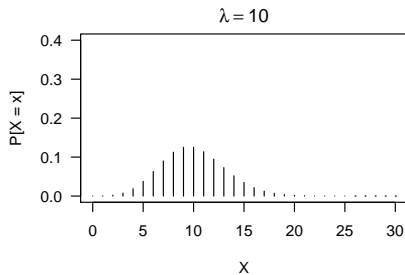
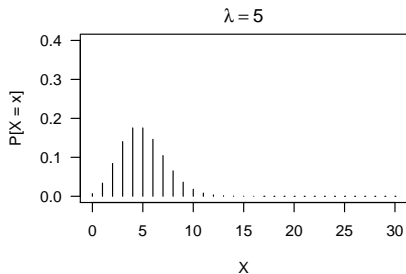
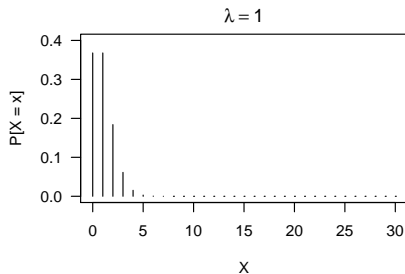
Uma v.a X segue o modelo de Poisson se surge a partir de um processo de Poisson, e sua **função de probabilidade** for dada por

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

onde o parâmetro $\lambda > 0$ é a taxa média de ocorrências em um intervalo de tempo ou espaço.

Notação: $X \sim P(\lambda)$.

Modelo Poisson



Exemplo 3.12

A emissão de partículas radioativas têm sido modelada através de uma distribuição de Poisson, com o valor do parâmetro dependendo da fonte utilizada.

Suponha que o número de partículas alfa, emitidas por minuto, seja uma variável aleatória seguindo o modelo Poisson com parâmetro 5, isto é, a taxa média de ocorrência é de 5 emissões a cada minuto.

Calcule a probabilidade de haver mais de 2 emissões em um minuto.

Exemplo 3.12

Note que $P(A > 2) = 1 - P(A \leq 2)$, portanto

$$\begin{aligned} P(A > 2) &= \sum_{a=3}^{\infty} P(A = a) = 1 - P(A \leq 2) \\ &= 1 - \sum_{a=0}^2 \frac{e^{-5} 5^a}{a!} = 0.875 \end{aligned}$$

Em R temos

```
1 - ppois(2, lambda = 5)
```

```
# [1] 0.875348
```

Exemplo 3.13

Engenheiros da companhia telefônica estudam se o modelo de Poisson com taxa de ocorrência de 4.5 chamadas por hora pode ser ajustado ao número N de chamadas interestaduais que chegam por hora, a uma central telefônica, durante o período noturno. Os dados coletados referentes a 650 períodos de uma hora, estão apresentados abaixo. Analise se esta suposição é razoável.

Chamadas	0	1	2	3	4	5	6	7	8+
Freq. Obs.	9	38	71	115	125	106	79	50	57

Exemplo 3.13

Usando o R temos

```
obs <- c(9, 38, 71, 115, 125, 106, 79, 50, 57)
prob <- c(dpois(0:7, lambda = 4.5), 1 - ppois(7, lambda = 4.5))
esp <- prob * 650
cbind("Obs" = obs, "p" = round(prob, 2), "Esp" = round(esp, 2))
```

#		Obs	p	Esp
#	[1,]	9	0.01	7.22
#	[2,]	38	0.05	32.49
#	[3,]	71	0.11	73.11
#	[4,]	115	0.17	109.67
#	[5,]	125	0.19	123.37
#	[6,]	106	0.17	111.04
#	[7,]	79	0.13	83.28
#	[8,]	50	0.08	53.54
#	[9,]	57	0.09	56.28

Modelo hipergeométrico

Definição: Considere um conjunto de n objetos dos quais m são do tipo I e $n - m$ são do tipo II . Para um sorteio de r objetos $r < n$, feito ao acaso e **sem reposição**, defina X como o número de objetos de tipo I selecionados.

Diremos que a v.a X segue o modelo Hipergeométrico e sua função de probabilidade é dada por

$$P[X = k] = \frac{\binom{m}{k} \binom{n-m}{r-k}}{\binom{n}{r}},$$

onde $k = \max \{0, r - (n - m)\} \leq k \leq \min \{r, m\}$.

Notação: $X \sim \text{HG}(m, n, r)$.

Exemplo

Considere que, num lote de 20 peças, existam 4 defeituosas. Selecionando-se 5 dessas peças, sem reposição, qual seria a probabilidade de 2 defeituosas terem sido escolhidas?

Exemplo

Pelo enunciado, sabemos que $n = 20$, $m = 4$, $r = 5$. Podemos calcular a probabilidade de $X = 2$ como

$$P[X = 2] = \frac{\binom{m}{k} \binom{n-m}{r-k}}{\binom{n}{r}} = \frac{\binom{4}{2} \binom{20-4}{5-2}}{\binom{20}{5}} = 0.217$$

Usando o R temos

```
## m = defeituosas, n = não defeituosas, k = tam. amostra
dhyper(x = 2, m = 4, n = 16, k = 5)
```

```
# [1] 0.2167183
```

Modelo hipergeométrico: Captura e Recaptura

Para determinar o tamanho de uma população, podemos usar o modelo hipergeométrico o processo resumido é o seguinte:

- Captura e marcação de m animais.
- Após um tempo, amostram-se r animais (sem reposição).
- O número de animais marcados é contado.

Sendo X a v.a “número de animais marcados”, então $X \sim \text{HG}(m, N, r)$, onde N é o tamanho da população (desconhecido).

Supondo que a amostra forneça k_0 animais marcados, podemos ter uma estimativa de N com

$$N_{est} = m \frac{r}{k_0}$$

Modelo hipergeométrico - Captura e Recaptura

Suponha que você marcou 200 indivíduos de uma espécie de peixe em um lago. Algum tempo depois você capturou 56 indivíduos, dos quais 9 estavam marcados. Qual o tamanho estimado da população?

$$\begin{aligned}N_{est} &= m \frac{r}{k_0} \\ &= 200 \cdot \frac{56}{9} \\ &= 1244,44 \approx 1244\end{aligned}$$

Sumário

- 1 Introdução
- 2 Função de Probabilidade
- 3 Distribuições discretas de probabilidade
- 4 Exercícios recomendados

Exercícios recomendados

- Seção 3.1 - 3, 5 e 6.
- Seção 3.2 - 2, 3, 5 e 6.
- Seção 3.3 - 1, 2, 3, 4, 5 e 6.
- Extras: Seção 3.4 - 1, 9, 10, 14, 16, 19, 21, e 26.