Exercícios de Probabilidade

Paulo Justiniano Ribeiro Jr

Versão compilada em 22 de abril de 2024 às 10:20

1. Considere o lançamento de um dado normal.
   (a) Quais os resultados possíveis?
   (b) Qual a probabilidade de sair a face 5?
   (c) Qual a probabilidade de cada possível resultado?
   (d) Qual a probabilidade de sair uma face que seja um número divisível por 3?

Solução:
Solução com elementos básicos de probabilidades:

(a) $\Omega = \{1, 2, 3, 4, 5, 6\}$
(b) $P[\{5\}] = \frac{1}{6}$
(c) Espaço amostral equiprovável

Face	1	2	3	4	5	6
Probabilidade	1/6	1/6	1/6	1/6	1/6	1/6

(d) $P[\{3, 6\}] = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0,333$

Solução alternativa utilizando conceito e notação de variável aleatória:

$$\begin{aligned} Y &: \text{face no lançamento de um dado} \\ y &\in \{1, 2, 3, 4, 5, 6\} \end{aligned}$$

Distribuição de probabilidades de Y:

Y	1	2	3	4	5	6
$P[Y = y]$	1/6	1/6	1/6	1/6	1/6	1/6

Da tabela pode-se extrair as probabilidades pedidas.

$$P[Y = 5] = 1/6$$
$$P[Y = 3 \cup Y = 6] \overset{Mut.Exc.}{=} P[Y = 3] + P[Y = 6] = \frac{2}{6} = 0,333$$

---

2. Considere o lançamento de um dado não usual, no qual a probabilidade de cada face é proporcional ao seu valor.
   (a) Quais os resultados possíveis?
   (b) Qual a probabilidade de sair a face 5?
   (c) Qual a probabilidade de cada possível resultado?
   (d) Qual a probabilidade de sair uma face que seja um número divisível por 3?

Solução:

Solução com elementos básicos de probabilidades:

(a) $\Omega = \{1, 2, 3, 4, 5, 6\}$
Cada ponto do espaço amostral tem uma probabilidade associada e as probabilidades somam 1.
Como cada uma é proporcional ao valor da face:

$$w + 2w + 3w + 4w + 5w + 6w = 21w = 1 \longrightarrow w = 1/21$$

(b) $P[\{5\}] = 5/21 = 0,238$

(c) $1/21, 2/21, 3/21, 4/21, 5/21, 6/21$ (espaço não-equiprovável)

(d) $P[\{3, 6\}] = \frac{3}{21} + \frac{6}{21} = 0,429$

Solução alternativa usando notação de variável aleatória.

$$\begin{aligned} &Y : \text{face no lançamento deste dado} \\ &y \in \{1, 2, 3, 4, 5, 6\} \end{aligned}$$

Distribuição de probabilidades de Y:

Y	1	2	3	4	5	6
$P[Y = y]$	p_1	p_2	p_3	p_4	p_5	p_6
$P[Y = y]$	w	$2w$	$3w$	$4w$	$5w$	$6w$

Como para uma distribuição discreta de probabilidades $\sum_{i=1}^{6} p_i = 1$:

$$\begin{aligned} w + 2w + 3w + 4w + 5w + 6w &= 1 \\ 21w &= 1 \\ w &= 1/21 \end{aligned}$$

A distribuição das probabilidades fica como na tabela a seguir.

Y	1	2	3	4	5	6
$P[Y = y]$	1/21	2/21	3/21	4/21	5/21	6/21

E as probabilidades pedidas são obtidas desta tabela.

$$P[Y = 5] = \frac{5}{21} = 0.238$$

$$P[Y = 3 \cup Y = 6] \overset{Mut.Exc.}{=} P[Y = 3] + P[Y = 6] = \frac{3}{21} + \frac{6}{21} = 0,429$$

---

3. Considere o lançamento de dois dados e o interesse está na soma das faces.

   (a) Quais os resultados possíveis?

   (b) Qual a probabilidade da soma ser 5?

   (c) Qual a probabilidade de cada possível resultado?

   (d) Qual a probabilidade que a soma das faces seja um número divisível por 3?

Solução:
Solução com elementos básicos de probabilidades:

(a)

$$\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,1),(2,3),(2,4),(2,5),(2,6),\\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

Este espaço é equiprovável, ou seja, cada uma dos 36 resultados possíveis tem a mesma probabilidade de ocorrência (1/36).

(b) $P[\text{soma } 5] = P[\{(1,4),(2,3),(3,2),(4,1)\}] = 4/36 = 0,111$

(c) $1/36$ (espaço equiprovável)

(d) $P[\{(1,2),(1,5),(2,1),(2,4),(3,3),(3,6),(4,2),(4,5),(5,1),(5,4),(6,3),(6,6)\}] = 12/36 = 0,333$

Solução usando a definição de uma variável aleatória:

$$Y : \text{soma das faces no lançamento de dois dados}$$
$$y \in \{2,3,4,5,6,7,8,9,10,11,12\}$$

Distribuição de probabilidades de Y:

Y	2	3	4	5	6	7	8	9	10	11	12
$P[Y=y]$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

As probabilidades pedidas, obtidas pela tabela são:

$$P[Y=5] = \frac{4}{36} = 0.111$$

$$P[Y=3 \cup Y=6 \cup Y=9 \cup Y=12] \overset{Mut.Exc.}{=} P[Y=3] + P[Y=6] + P[Y=9] + P[Y=12]$$
$$= \frac{2}{36} + \frac{5}{36} + \frac{4}{36} + \frac{1}{36} = \frac{12}{36} = 0,333$$

---

4. Um dado foi fabricado com o centro em madeira leve e cada face com uma chapa metálica porém de diferentes características (espessura/densidade) em cada face?

   (a) Quais os resultados possíveis?
   (b) Como calcular a probabilidade de sair a face 5?
   (c) Como calcular a probabilidade de cada possível resultado?
   (d) Como calcular a probabilidade de sair uma face que seja um número divisível por 3?

Solução:
A ideia aqui é discutir que problema não pode ser resolvido com as informações dadas pois parece questionável assumir que o espaço amostral seja equiprovável.
É necessário alguma informação ou suposição adicional. Por exemplo poderia-se fazer um experimento no qual se lance os dados várias vezes, para fornecer as probabilidades (estimadas) de cada face.

---

5. Voce vai a um cassino em uma mesa que tem um jogo no qual se lançam dois dados como em um problema anterior. A regra é a de que se a soma for 6, 7 ou 8 voce ganha, valor igual ao apostado, caso contrário, perde o apostado.

   (a) Qual sua opinião sobre suas chances de ganhar?
   (b) Quais os resultados possíveis?
   (c) Qual sua opinião sobre a probabilidade da soma ser 5?
   (d) Qual sua opinião sobre a probabilidade de cada possível resultado?
   (e) Qual sua opinião sobre a probabilidade de sair uma face que seja um número divisível por 3?

Solução:
A ideia aqui é discutir que o problema é resolvido com alguma suposição adicional.
É possível considerar que o espaço amostral seja equiprovável, ou seja, que os dados são *regulares* (honestos).
Alguma atribuição subjetiva de probabilidades é feita.
Com isto e revisitando o problema 3) tem-se que:

(a) Qual sua opinão sobre suas chances de ganhar?
$P[Y = 6 \cup Y = 7 \cup Y = 8] \overset{Mut.Exc.}{=} P[Y = 6] + P[Y = 7] + P[Y = 8] = \frac{5}{36} + \frac{6}{36} + \frac{5}{36} = \frac{16}{36} = 0,444$

(b) $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

(c) $P[Y = 5] = 4/36 = 1/9 = 0,111$

(d) Ver a distribuição de Y no problema 3).

(e) Neste caso voltamos ao espaço amostral original e definimos o evento e sua probabilidade:

$$\begin{aligned} &A : \text{uma face tem um número divisível por 3} \\ A =&\{(1, 3), (1, 6), (2, 3), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ &(4, 3), (4, 6), (5, 3), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\} \end{aligned}$$

Como o espaço amostral é suposto equiprovável:

$$P[A] = \frac{n(A)}{n(\Omega)} = \frac{20}{36} = \frac{5}{9} = 0,556$$

---

6. Em um teste múltipla escolha de quatro questões, deve-se marcar uma alternativa em cada questão. Cada questão possui dcinco alternativas, das quais apenas uma é correta. Qual a probabilidade de um indivíduo acertar, por mero acaso, alguma questão?

Solução:
Vamos ver uma primeira solução baseada nos conceitos fundamentais de probabilidades.
Notar que estão envolvidos os conceitos de: *eventos*, *eventos complementares* e *independência de eventos*.

$$\begin{aligned} &A_i : \text{acerta a } i\text{-ésima questão} \;\; i = 1, \ldots, 4 \\ &\forall i \;\; P(A_i) = 0,2 \quad \text{e } P(\overline{A_i}) = 0,8 \end{aligned}$$

$$\begin{aligned} P(\text{acertar alguma}) &= 1 - P(\text{errar todas}) = 1 - P(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \overline{A_4}) \overset{ind}{=} 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3}) \cdot P(\overline{A_4})) = \\ &= 1 - (0,8)^4 = 0{,}59 \end{aligned}$$

Uma solução alternativa utilizando conceitos de variáveis aleatórias e distribuição de probabilidades.
Identificamos que o interesse é no número de questões certas.
São $n = 4$ questões e cada questão tem duas possibilidades, pode estar certa ($p = 0,20$) ou errada ($p = 1-0,20 = 0,80$).
Como se assume independência e que a chance de acerto ao acaso (*chute*) é a mesma em todas questões é um caso de distribuição **Binomial**.

$$\begin{aligned} &Y : \text{número de questões certas} \\ &Y \sim \mathrm{B}(n = 4, p = 0,20) \\ P[Y = y] =& \binom{n}{y} p^y (1-p)^{n-y} = \binom{4}{y} 0,2^y (1-0,2)^{4-y} \\ P[Y > 0] =& 1 - P[Y = 0] = 1 - \binom{4}{0} 0,2^0 (1-0,2)^{4-0} = 1 - 1 \cdot 1 \cdot 0,8^4 = 0{,}59. \end{aligned}$$

---

7. Um reservatório recebe água de três fontes diferentes. A primeira tem 5% de chance de apresentar alguma contaminação, a segunda tem 6,5% e a terceira tem 12%. Qual a probabilidade do reservatório ser contaminado?

A solução utiliza conceitos de união e interseção de eventos, probabilidade complementar e faz a suposição de independência dos eventos para obter a resposta.

Solução:
Evento A : a água da primeira fonte é contaminada
Evento B : a água da segunda fonte é contaminada
Evento C : a água da terceira fonte é contaminada
Dados:

$$P[A] = 0,05 \quad ; P[\overline{A}] = 0,95$$
$$P[B] = 0,065 \quad ; P[\overline{B}] = 0,935$$
$$P[C] = 0,12 \quad ; P[\overline{C}] = 0,88$$

$$P[\text{contaminação}] = P[A \cup B \cup C] = 1 - P[\overline{A} \cap \overline{B} \cap \overline{C}] \stackrel{ind}{=} 1 - P[\overline{A}] \cdot P[\overline{B}] \cdot P[\overline{C}]$$
$$= 1 - 0.95 \cdot 0.935 \cdot 0.88 = 0{,}2183$$

---

8. Alguns biólogos fizeram estudos de laboratório sobre o comportamento de animais quando submetidos a um estímulo, o quais poderiam apresentar ou não resposta positiva. Em particular estavam interessados nas respostas positivas os estímulo. Considera-se que na população destes animais, 10% sejam sensíveis ao estímulo.

   O biologo A possuia um grupo em que 10 animais eram sensíveis e 20 eram insensíveis. Ele selecionou ao acaso 8 animais para teste.

   O biólogo B dispunha de um grande número de animais e foi testando um a um até encontrar o terceiro sensível ao estímulo.

   O biólogo C tomou fazia testes diários e encontrava uma média de 2,8 animais sensíveis a cada dia.

   O biólogo D submeteu 10 animais ao estímulo.

   O biólogo E dispunha de um grande número de animais e foi testando um a um até encontrar um sensível ao estímulo.

   (a) Qual a probabilidade do biólogo A encontrar ao menos 2 animais sensíveis?
   (b) Qual a probabilidade do biólogo B precisar testar no máximo 6 animais?
   (c) Qual a probabilidade do biólogo C encontrar menos que dois sensíveis em um determinado dia?
   (d) Qual a probabilidade do biólogo D encontrar mais que 3 animais sensíveis?
   (e) Qual a probabilidade do biólogo E precisar testar mais que 3 animais?

   Sugestão: especifique a(s) variável(eis) aleatória, sua(s) distribuição(ções) e suposições feitas.

Solução:
Os items deste problemas podem ser resolvidos de diferentes formas.
A solução que vamos propor pare este exercício utiliza as definições de diferentes tipos de *distribuições de probabilidades discretas.*
As perguntas foram formuladas de forma que em cada item é possível definir a variável aleatória e sua distribuição de probabilidades.
Desta forma ilustramos a definição de de cada distribuição e ilustramos como o uso de variáveis aleatórias auxilia obter soluções.

(a) Aqui tem-se um grupo (*finito*/pequeno) de indivíduos, cada um com uma de *duas possíveis características* e retira-se (*sem reposição*) uma parte deles. O contexto é compatível com a distribuição **hipergeométrica**.

$$X_a : \text{número de sensíveis entre os 8 selecionados}$$
$$x_a \in \{0, 1, 2, \dots 8\}$$
$$X_a \sim HG(N = 30, K = 10, n = 8)$$
$$P[X_a = x] = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{x}\binom{20}{8-x}}{\binom{30}{8}}$$
$$P[X_a \geq 2] = 1 - P[X_a \leq 1] = 1 - (P[X_a = 0] + P[X_a = 1]) = 0{,}846$$

(b) Agora o contexto é que tomamos indivíduos um a um e cada um tem *uma de duas possíveis características*, até que o *k-ésimo* (terceiro) com certa característica apareça e contamos *quantos obtidos não tinham a característica*. Mais genericamente dizemos que temos o número de "falhas" até o k-ésimo "sucesso". Isto é compatível com a distribuição **binomial negativa** (ou de *Pascal*).

$$\begin{aligned}
X_b &: \text{número de não sensíveis examinados até encontrar o terceiro sensível} \\
x_b &\in \{0, 1, \ldots\} \\
X_b &\sim BN(k = 3, p = 0,1) \\
P[X_b = x] &= \binom{x+k-1}{x} p^k (1-p)^x = \binom{x+2}{x} 0,1^3 0,9^x \\
P[X_b \leq 3] &= P[X_b = 0] + P[X_b = 1] + P[X_b = 2] + P[X_b = 3] = 0{,}016
\end{aligned}$$

(c) Neste caso queremos saber o *número de ocorrências* (sem que se saiba um limite) de algo que ocorre seguindo certa *taxa*. O contexto é o da distribuição de **Poisson**.

$$\begin{aligned}
X_c &: \text{número de sensíveis encontrados em um dia} \\
x_c &\in \{0, 1, \ldots\} \\
X_c &\sim P(\lambda = 2,8) \\
P[X_c = x] &= \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} = \frac{\mathrm{e}^{-2,8} 2,8^x}{x!} \\
P[X_c < 2] &= P[X_c = 0] + P[X_c = 1] = 0{,}231
\end{aligned}$$

(d) Aqui tem-se um número indivíduos retirados de uma população (ilimitada, muito grande ou retirados com reposição). Cada um tem ou não certo atributo com probabilidade constante de conter o atributo. O interesse é no número de indivíduos que possui o atributo. Isto caracteriza a distribuição **binomial**.

$$\begin{aligned}
X_d &: \text{número de sensíveis entre os 10 examinados} \\
x_d &\in \{0, 1, \ldots 10\} \\
X_d &\sim B(n = 10, p = 0,1) \\
P[X_d = x] &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{10}{x} 0,1^x 0,9^{10-x} \\
P[X_d > 3] &= 1 - P[X_d \leq 3] = 1 - (P[X_d = 0] + P[X_d = 1] + P[X_d = 2] + P[X_d = 3]) = 0{,}013
\end{aligned}$$

(e) O contexto é que tomamos indivíduos um a um e cada um tem *uma de duas possíveis características*, até que o *primeiro* com certa característica apareça e contamos *quantos obtidos não tinham a característica*. Mais genericamente dizemos que temos o número de "falhas" até o primeiro "sucesso". Isto é compatível com a distribuição **geométrica** que é então um caso especial da binomial negativa com $k = 1$.

$$\begin{aligned}
X_e &: \text{número de não sensíveis examinados até encontrar o primeiro sensível} \\
x_e &\in \{0, 1, \ldots\} \\
X_e &\sim G(p = 0,1) \\
P[X_e = x] &= p(1-p)^x = 0,1 \cdot 0,9^x
\end{aligned}$$

Perguntou-se a probabilidade de testar mais de três animais, ou seja, testar quatro ou mais, o que corresponde a encontrar três ou mais insensíveis antes do primeiro sensível.

$$P[X_e \geq 3] = 1 - P[X_e \leq 2] = 1 - (P[X_e = 0] + P[X_e = 1] + P[X_e = 2]) = 0{,}729$$

Soluções computacionais com o programa **R**:

```r
pa <- phyper(1, m=10, n=20, k=8, low=FALSE)
pb <- pnbinom(3, size=3, prob=0.1)
pc <- ppois(1, lambda=2.8)
pd <- pbinom(3, size=10, prob=0.1, low=FALSE)
pe <- pgeom(2,prob=0.1, low=FALSE)
```

---

9. Um site de vendas pela internet registra 40% dos acessos do estado do PR, 50% de outros estados e 10% do exterior. 20% dos acessos do PR resultam em uma compra, enquanto que os percentuais para outros estados e exterior são de 10% e 30%, respectivamente.

   (a) Qual a probabilidade de um acesso resultar em compra?

   (b) Se foi feita uma compra, qual a probabilidade de ela ter sido do exterior?

   Solução:
   Este problema é uma aplicação do Teorema de Bayes pois temos uma inversão na condicional: conhecendo as $P[\text{compra}|\text{acesso}]$ queremos calcular a $P[\text{acesso}|\text{compra}]$.
   Eventos:

$$PR: \text{acesso do PR} \quad OE: \text{acesso de outros estados} \quad EX: \text{acesso do exterior}$$

$$C: \text{compra}$$

   Probabilidades informadas:

$$P[PR] = 0,40 \quad P[OE] = 0,50 \quad P[EX] = 0,10$$

$$P[C|PR] = 0,20 \quad P[C|OE] = 0,10 \quad P[C|EX] = 0,30$$

   (a) $P[C] = P[PR \cap C] + P[OE \cap C] + P[EX \cap C] = P[PR] \cdot P[C|PR] + P[OE] \cdot P[C|OE] + P[EX] \cdot P[C|EX] = (0,40)(0,20) + (0,50)(0,10) + (0,10)(0,30) = 0{,}16$

   (b) $P[EX|C] = \frac{P[EX \cap C]}{P[C]} = \frac{P[EX] \cdot P[C|EX]}{P[C]} = \frac{(0{,}10)(0{,}30)}{(0{,}40)(0{,}20)+(0{,}50)(0{,}10)+(0{,}10)(0{,}30)} = 0{,}1875$

---

10. Considere o surgimento de um defeito na pista em um trecho de rodovia com extensão de 20 km.

   (a) Qual a probabilidade de que o defeito ocorra nos primeiros 5 km?

   (b) Qual a probabilidade de que o defeito ocorra entre os quilômetros 12 e 15?

   (c) Se o defeito ocorre na segunda metade do trecho, qual a probabilidade de seja nos últimos 3 km?

   (d) O custo de manutenção é de R\$2.000,00 se ocorre nos primeiros 5 km, de R\$5.000,00 se ocorre entre 5 e 16 kms e de R\$10.000,00 se ocorre nos ultimos 4 km. Que custo espera-se ter em 20 manutenções?

   Solução:
   O problema pode ser resolvido intuitivamente pois a probabilidade é constante em todo tracho. Entretanto, vamos apresentar a solução utilizando o conceito de distribuição de probabilidades, no caso uma uniforme contínua.

$$X: \text{local (km) onde surge o defeito na pista}$$
$$\text{supondo probabidade constante em toda extensão do trecho}$$
$$X \sim \mathrm{U}_c[0, 20]$$
$$f(x) = \frac{1}{20-0} = \frac{1}{20}\mathrm{I}_{[0,20]}(x) \qquad F(x) = \frac{x-0}{20-0} = \frac{x}{20}$$

   (a) $P[X < 5] = P[X < 5] = \frac{2}{20} = 0,25$
   Note que: $P[X < 5] = \int_0^5 f(x)\mathrm{d}x = F(5)$

   (b) $P[8 < X < 15] = \frac{7}{20} = 0,35$ Note que: $P[8 < X < 15] = \int_8^{15} f(x)\mathrm{d}x = F(15) - F(8)$

   (c) $P[X > 17|X > 10] = \frac{P[X>17 \cap X>10]}{P[X>10]} = \frac{\frac{3}{20}}{\frac{10}{20}} = \frac{3}{10} = 0,3$

   (d) Vamos definir outra variável aleatória:

$$Y \sim \text{custo do reparo}$$
$$y \in \{2.000, 5.000, 10.000\}$$

y	2.000	5.000	10.000
$P[Y = y]$	$P[X =< 5]$	$P[5 < X =< 16]$	$P[X > 16]$
$P[Y = y]$	5/20	11/20	4/20

$$Y_1 + Y_2 + \dots Y_{20} = 20 \cdot \mathrm{E}[X] = 20 \cdot \sum_{i=1}^{3} y_1 P[Y = y_i] = 2.000 \cdot \frac{5}{20} + 5.000 \cdot \frac{11}{20} + 10.000 \cdot \frac{4}{20} = 5250$$

---

11. O rendimento de uma frota de veículos de uma locadora tem a seguinte função de densidade de probabilidades. Calcule o solicitado nos itens a seguir.

$$f(y) = \begin{cases} \frac{k(y-5)}{2} & \text{se } 5 \leq y < 7 \\ \frac{k(11-y)}{4} & \text{se } 7 \leq y \leq 11 \\ 0 & \text{caso contrário} \end{cases}$$

(a) O valor de k.

(b) $P[Y < 7]$

(c) $P[Y > 10]$.

(d) $P[Y \leq 9]$

(e) $P[7,5 < Y < 9,5]$.

(f) $P[Y > 6|Y \leq 7]$.

(g) $P[Y \leq 10|Y > 8]$.

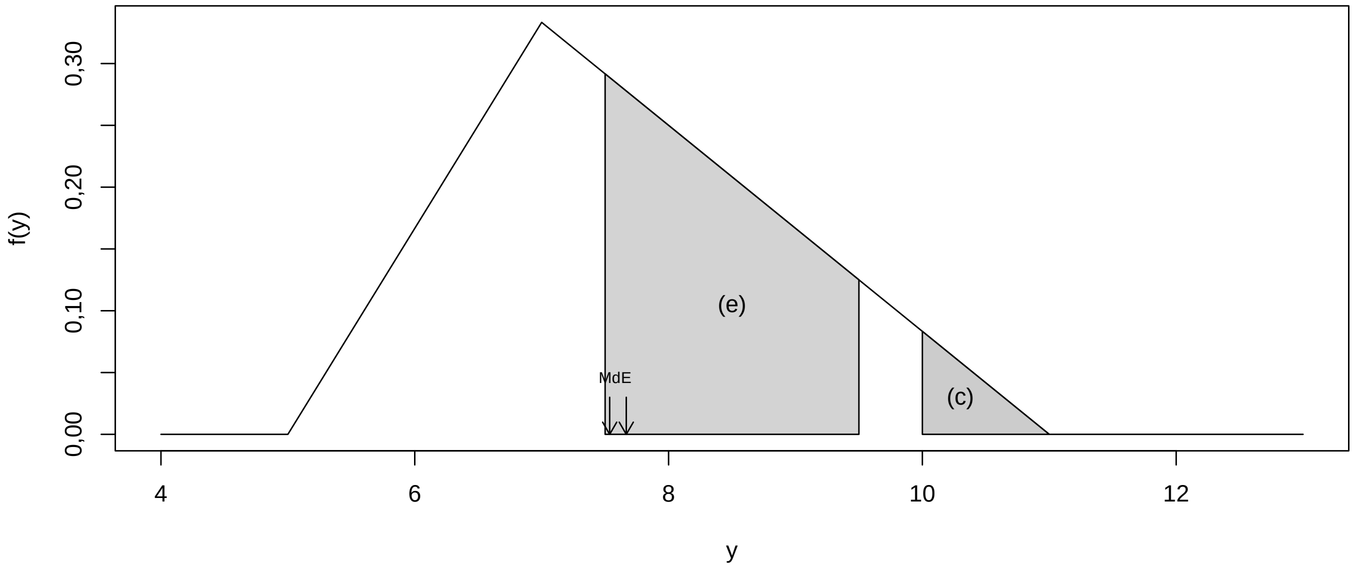
(h) O consumo médio.

(i) O consumo mediano.

Neste problema existe uma distribuição de probabilidade para o consumo que não é nenhuma das ”conhecidas”(“ou rotuladas´´).
O problema é resolvido com a definição geral de probabilidades para uma variável aleatória contínua.
Pode-se resolver analiticamente (matematicamente) ou com auxílio do computador (computacionalmente).

Solução:

Nas soluções apresentadas a seguir utiliza-se a notação matemática de integral para indicar as probabilidades, que são áreas sob a curva de $f(y)$. Entretanto, com o formado da $f(y)$ as áreas podem ser calculadas geometricamente utilizando áreas de triângulos e/ou trapézios.



(a)

$$\int_5^{11} f(y)\,\mathrm{d}y = 1$$
$$\int_5^7 \frac{k(y-5)}{2}\mathrm{d}y + \int_7^{11} \frac{k(11-y)}{4}\,\mathrm{d}y = 1$$
$$\frac{k}{2}\left[\frac{7^2-5^2}{2} - 5(7-5)\right] + \frac{k}{4}\left[11(11-7) - \frac{11^2-7^2}{2}\right] = 1$$
$$k = \frac{1}{3}$$

(b) $P[Y < 7] = \int_5^7 f(y)\,\mathrm{d}y = \frac{1/3}{2}\left[\frac{7^2-5^2}{2} - 5(7-5)\right] = 0,333$

(c) $P[X > 10] = \int_{10}^{11} f(y)\,\mathrm{d}y = \frac{1/3}{4}\left[11(11-10) - \frac{11^2-10^2}{2}\right] = 0,0417$

(d) $P[Y \leq 9] = 1 - \int_9^{11} f(y)\,\mathrm{d}y = 1 - \frac{1/3}{4}\left[11(11-9) - \frac{11^2-9^2}{2}\right] = 0,833$

(e) $P[7,5 < Y < 9,5] = \int_{7,5}^{9,5} f(y)\,\mathrm{d}y = \frac{1/3}{4}\left[11(9,5-7,5) - \frac{9,5^2-7,5^2}{2}\right] = 0,417$

(f) $P[Y > 6|Y \leq 7] = \frac{P[6<Y\leq 7]}{P[Y\leq 7]} = \frac{\int_6^7 f(y)\,\mathrm{d}y}{\int_5^7 f(y)\,\mathrm{d}y} = 0,75.$

(g) $P[Y \leq 10|Y > 8] = \frac{P[8<Y\leq 10]}{P[Y>8]} = \frac{\int_8^{10} f(y)\,\mathrm{d}y}{\int_8^{11} f(y)\,\mathrm{d}y} = 0,889.$

(h) $E[Y] = \int_5^{11} y f(y)\,\mathrm{d}y = 7,67$

(i) $\int_5^{Md} f(y)\,\mathrm{d}y = 0,5 \Rightarrow Md = 7,54$

Soluções computacionais com o programa **R**.
Neste caso computacional implementamos f(y) a utilizamos integração numérica

```
## Definindo a função
fy <- function(y){
   K <- 1/3
   fy <- numeric(length(y))
```

```
   fy[y> 5 & y<7] <- K*(y[y> 5 & y<7]-5)/2
   fy[y>= 7 & y<=11] <- K*(11-y[y>= 7 & y<=11])/4
   return(fy)
}
## Verificando se a área total sob a função é 1
integrate(fy, 0, 20)

## 1 with absolute error < 8,6e-05

##
## £P[Y < 7]£
fr.b <- integrate(fy, low=5, up=7)$value
## £P[Y > 10]£.
fr.c <- integrate(fy, low=10, up=11)$value
## £P[Y \leq 9]£
fr.d <- integrate(fy, low=5, up=9)$value
## £P[7,5 < Y < 9,5]£.
fr.e <- integrate(fy, low=7.5, up=9.5)$value
## £P[Y > 6 | Y \leq  7]£.
fr.f <- integrate(fy, low=6, up=7)$value/integrate(fy, low=5, up=7)$value
## £P[Y \leq 10 | Y >  8]£.
fr.g <- integrate(fy, low=8, up=10)$value/integrate(fy, low=8, up=11)$value
## Calculando a esperança (média)
yfy <- function(y) y * fy(y)
(fr.Ey <- integrate(yfy, 5, 11)$value)

## [1] 7,667

## Definindo uma função para calcular um quantil qualquer e obtendo o valor da mediana
qfy <- function(q){
    f <- function(x) (integrate(fy, lower=5, upper=x)$value - q)^2
    optimize(f, interval=c(5,11))$minimum
}
(fr.Md <- qfy(0.5))

## [1] 7,536
```

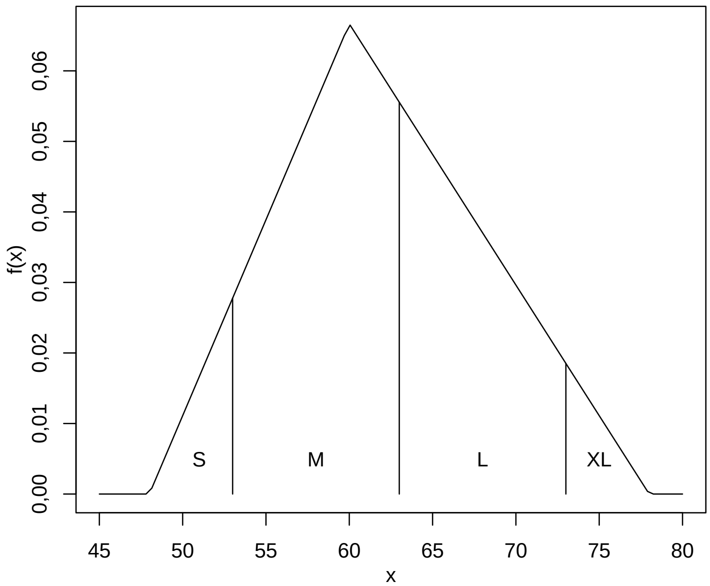
---

12. Uma determinada indústria classifica ovos como: XL acima de 73 g, L 63 a 73 g, M 53 a 63 g, S abaixo de 53 g. Suponha que um produtor produza ovos cujos tamanhos (pesos) são descritos pela seguinte função de densidade de probabilidades:

$$f(x) = \begin{cases} \frac{(x-48)}{12k} & \text{se } 48 \leq x < 60 \\ -\frac{(x-78)}{18k} & \text{se } 60 \leq x \leq 78 \\ 0 & \text{se } x < 48 \text{ ou } x > 78 \end{cases}$$

   (a) Qual o valor de k ?
   (b) Qual a proporção de ovos que deve ser produzida em cada classificação?
   (c) Se o produtor recebe R$ 0,05 por ovo S, R$ 0,10 por ovo M, R$ 0,12 por ovo L e R$ 0,18 por ovo XL, quanto deve receber em um lote de 10.000 ovos?
   (d) Qual o tamanho (peso) mediano dos ovos?
   (e) Forneça a expressão da distribuição acumulada $F(x)$.
   (f) Qual o tamanho (peso) para o qual apenas 20% dos ovos estão acima dele?

Solução:
Assim como no problema anterior aqui tem-se uma distribuição de probabilidades que não é uma das "conhecidas´´ ("rotuladas´´) e utilizam-se princípios gerais para solução.
Mas também mostra-se como a classificação dos ovos pode ser tratada como uma nova variável (categórica) definida a partir de faixas de valores da primeira.



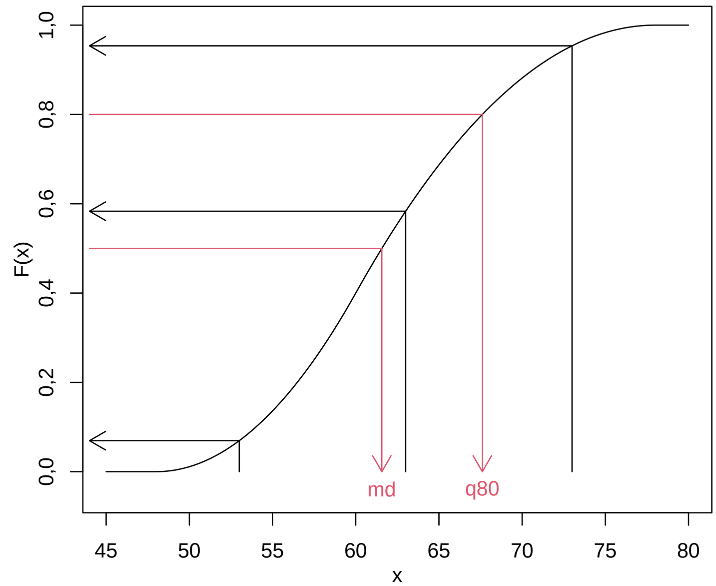


Figura 1: Funções de distribuição de probabilidades ($f(x)$) e acumulada ($F(x)$) do problema. Segmentos e setas são as soluções de alguns dos items do problema.

(a) $k = 15$

(b) Pode-se resolver de três formas diferentes: geometricamente (áreas dos polígonos indicados na figura), integrando-se $f(x)$ ou avaliando-se e fazendo as diferenças dos valores de $F(x)$ nos pontos que definem as classificações.

$$C : \text{valor por ovo}$$
$$c \in \{0,05; 0,10; 0,12; 0,18\}$$

c_i	0,05	0,10	0,12	0,18
$P[C = c_i]$	$P[X < 53] =$ 0,069	$P[53 \leq X < 63] =$ 0,514	$P[63 \leq X \leq 73] =$ 0,37	$P[X > 73] =$ 0,046

(c) $10,000 \cdot E[C] = 10.000 \cdot (0,05 \cdot 0,0694 + 0,10 \cdot 0,514 + 0,12 \cdot 0,37 + 0,18 \cdot 0,0463) = 10.000 \cdot 0,108 = 1076,4$

(d) $md : \int_{md}^{78} f(x)dx = 0,5 \longrightarrow F(md) = 0,50 \longrightarrow md = 61,6$

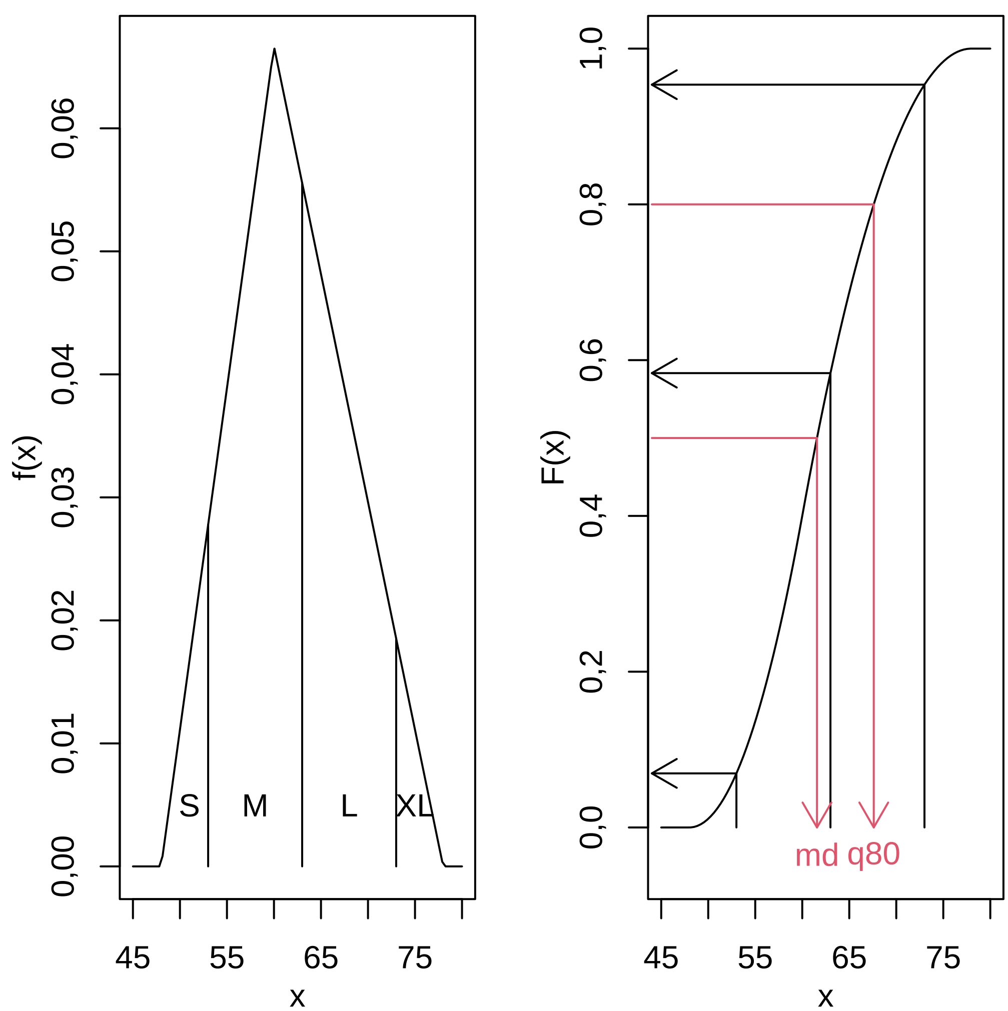
(e)
$$F(x) = \int_{48}^{x} f(x)dx = \begin{cases} 0 & \text{se } x < 48 \\ \frac{1}{180}\left[\frac{(x^2-48^2)}{2} - 48(x-48)\right] & \text{se } 48 \leq x < 60 \\ 0,4 - \frac{1}{270}\left[\frac{(x^2-60^2)}{2} - 78(x-60)\right] & \text{se } 60 \leq x < 78 \\ 1 & \text{se } x > 78 \end{cases}$$

(f) $q_{0,80} : \int_{q_{0,80}}^{78} f(x)dx = 0,20 \longrightarrow q_{0,80} = 67,6$

Soluções computacionais (linguagem R):

```r
## definindo f(x)
ddist <- function(x){
    y <- numeric(length(x))
    y[x >= 48 & x < 60] <- (x[x >= 48 & x < 60]-48)/180
    y[x >= 60 & x < 78] <- -(x[x >= 60 & x < 78]-78)/270
    return(y)
}
## definindo F(x)
pdist <- function(x){
    y <- numeric(length(x))
    ind <- x >= 48 & x < 60
    y[ind] <- ((x[ind]^2-48^2)/2 - 48*(x[ind]-48))/180
    ind <- x >= 60 & x < 78
    y[ind] <- 0.4 - ((x[ind]^2-60^2)/2 - 78*(x[ind]-60))/270
    y[x >= 78] <- 1
    return(y)
}
## definindo F^{-1}(x)
qdist <- function(q){
    uniroot(function(x) pdist(x) - q, interval=c(48,78))$root
}
## b) Proporções em cada classe
## integrando f(x)
(PrS <- integrate(ddist, 48, 53)$value)

## [1] 0,06944

(PrM <- integrate(ddist, 53, 63)$value)

## [1] 0,5139

(PrL <- integrate(ddist, 63, 73)$value)

## [1] 0,3704

(PrXL <- integrate(ddist, 73, 78)$value)

## [1] 0,0463

## utilizando F(x)
(PC <- diff(pdist(c(48,53,63,73,78))))

## [1] 0,06944 0,51389 0,37037 0,04630

## c) Valor médio por ovo
(EC <- drop(crossprod(c(0.05, 0.10, 0.12, 0.18), PC)))

## [1] 0,1076

## d) mediana
(md <- qdist(0.5))

## [1] 61,57
```

```r
## f) quantil 0,80
(q80 <- qdist(0.8))

## [1] 67,61

## Gráficos
par(mfrow=c(1,2), mar=c(3.5,3.5,1,1), mgp=c(2,1,0))
curve(ddist, from=45, to=80, ylab="f(x)")
segments(c(53,63,73), 0, c(53,63,73), ddist(c(53,63,73)))
text(c(51, 58, 68, 75), 0.005, c("S","M","L","XL"))

curve(pdist, from=45, to=80, ylab="F(x)", ylim=c(-0.05, 1))
segments(c(53,63,73), 0, c(53,63,73), pdist(c(53,63,73)))
arrows(c(53,63,73), pdist(c(53,63,73)), 44, pdist(c(53,63,73)), length=0.15)

segments(44, 0.5,md, 0.5, col=2)
arrows(md, 0.5, md,0, length=0.15, col=2)
text(md, 0, "md", pos=1, col=2)

segments(44, 0.8, q80, 0.8, col=2)
arrows(q80, 0.8, q80,0, length=0.15, col=2)
text(q80, 0, expression(q80), pos=1, col=2)
```



13. Suponha que os escores obtidos por estudantes em um teste *online* possam ser bem modelados por uma distribuição normal com média $\mu = 120$ e variância $\sigma^2 = 12^2$.

(a) Considera-se como estudante de alta performance os que atingem um escore a partir de 135. Qual o percentual esperado de estudantes de alta performance entre todos os que fazem o teste?

(b) Estudantes com escore abaixo de 100 devem se reinscrever e só podem voltar a fazer o teste após seis meses e os com escore entre 100 e 125 são convidados a refazer o teste após um mês. Quais as proporções de estudantes que deverá se reinscrever e que deverá refazer o teste após um mês?

(c) Define-se como *quartis* os escores abaixo dos quais espera-se encontrar 25, 50 e 75% dos estudantes. Quais os valores dos escores que definem os quartis?

(d) Quanto deveria ser o valor μ da média dos escores para que ao menos 30% dos escores fossem de alta performance?

(e) Há um outro teste que possui média $\mu = 125$ e variância $\sigma^2 = 6^2$. Em qual deles espera-se a maior proporção de estudantes de alta performance?

Solução:

$$X \sim \mathrm{N}(120, 12^2)$$

(a) $P[X > 135] = P[Z > \frac{135-120}{12}] = P[Z > 1{,}25] = 0{,}1056$

(b)

$$P[X < 100] = P[Z < \frac{100-120}{12}] = P[Z < \text{-}1{,}6667] = 0,0478$$
$$P[100 < X < 125] = P[\frac{100-120}{12} < Z < \frac{125-120}{12}] = P[-1,67 < Z < 0,417]$$

(c)

$$P[X < Q_1] = 0,25$$
$$z_1 = -0,674 = \frac{Q_1 - 120}{12}$$
$$Q_1 = 120 - 8,09 = 112$$

Usando o fato de que a distribuição é simétrica temos ainda que:

$$Q_2 = \mu = 120$$
$$Q_3 = 120 + 8,09 = 128$$

(d) $z = \frac{135-\mu}{15} = 0{,}524 \longrightarrow \mu = 128,7$

(e)

$$X_1 \sim \mathrm{N}(120, 12^2)$$
$$X_2 \sim \mathrm{N}(125, 6^2)$$
$$P[X_1 \geq 135] = P[Z_1 > \frac{135-120}{12}] = P[Z_1 > 1,25] = 0,106$$
$$P[X_2 \geq 135] = P[Z_2 > \frac{135-120}{12}] = P[Z_2 > 1,67] = 0,0478$$

Comandos em R para soluções:

```
(qa <- pnorm(135, mean=120, sd=12, lower=FALSE))

## [1] 0,1056

(qb <- diff(pnorm(c(-Inf, 100, 125), mean=120, sd=12)))
```

```
## [1] 0,04779 0,61375

(qc <- qnorm(c(.25, .50, .75), mean=120, sd=12))

## [1] 111,9 120,0 128,1

(qd <- 135 - 12 * round(qnorm(0.70), dig=3))

## [1] 128,7

(qez <- (135 - c(120, 125))/c(12, 6))

## [1] 1,250 1,667

(qep <- pnorm(135, m=c(120, 125), sd=c(12, 6), lower=FALSE))

## [1] 0,10565 0,04779
```
