CE-009: Introdução a Estatística - Turma A
Avaliações Periódicas (1^o semestre 2017)

Avaliação 01

1. Um determinado exame tem probabilidade de 0,85 de detectar uma doença quando esta está presente enquanto que um segundo tipo de exame tem probabilidade de 0,70. A doença é considerada detectada se algum dos exames é positivo. Considere que um material com a doença vai ser testado por ambos exames.
   (a) Descreva os eventos relevantes com uma notação apropriada.
   (b) Forneça o espaço amostral na forma de um conjunto e aponte suas características (finito ou infinito, enumerável ou não enumerável, equiprovável ou não).
   (c) Defina em notação o evento "a doença é detectada" e forneça o conjunto que define este evento.
   (d) Qual a probabilidade da doença ser detectada?
   (e) Qual a suposição feita no cálculo da probabilidade do anterior?
   (f) Se três materiais com a doença forem testados com ambos exames, qual a probabilidade de que todos deem (falso) "negativo".

   Solução:

   (a) **Notação**:

   $$\begin{aligned} A &: \text{doença detectada no primeiro exame} \\ B &: \text{doença detectada no segundo exame} \\ D &: \text{a doença é detectada} \\ \overline{D} &: \text{a doença não é detectada} \\ P[A] &= 0,85 \\ P[B] &= 0,70 \end{aligned}$$

   (b) $\Omega = \{(AB), (A\overline{B}), (\overline{A}B), (\overline{AB})\}$.
   O espaço amostral é *finito*, *enumerável* e *não equiprovável*.
   (c) O evento "a doença é detectada" é definido pelo conjunto $D = \{(AB), (A\overline{B}), (\overline{A}B)\}$.
   (d) $P[D] = P[A \cup B] = P[A] + P[B] - P[A \cap B] = P[A] + P[B] - P[A] \cdot P[B] = 0,85 + 0,70 - 0,85 \cdot 0,70 = 0.955$
   (e) Independência entre os resultados dos dois exames.
   (f) Supondo independência entre materiais e exames:
   $P[\overline{D}]^3 = (1 - P[D])^3 = (1 - P[A \cup B])^3 = 0.00009113$

---

2. Um material genético de feijão em desenvolvimento foi testado quanto à resistência a duas doenças que afetam comumente a cultura. Os resultados de 100 exames são resumidos na tabela a seguir.

resistência	resistência à doença A	
à doença B	alta	baixa
alta	80	9
baixa	6	5

   Denote por A o evento *o material tem alta resistência à doença A* e por B o evento *o material alta resistência à doença B*.
   (a) Obtenha: $P[A]$, $P[A \cap B]$, $P[A^c]$, $P[A^c \cap B^c]$, $P[A^c \cup B]$.
   (b) Obtenha: $P[A|B]$, $P[B|A]$, $P[A|B^c]$, $P[B^c|A]$, $P[B|A^c]$.
   (c) Se um material é selecionado ao acaso qual a probabilidade de ter:
      - alta resistência a A e baixa a B?
      - alta resistência a B e baixa a A?
   (d) os eventos alta resistência a ambas doenças são mutuamente exclusivos? (justifique)
   (e) os eventos alta resistência a ambas doenças são independentes? (justifique)

Solução:

```
> m <- matrix(c(80,6,9,5),ncol=2,dimnames=list(c("A","A^c"),c("B","B^c")))
> m

     B B^c
A   80   9
A^c  6   5

> mp <- prop.table(m)
> mp

       B  B^c
A   0.80 0.09
A^c 0.06 0.05
```

(a)
- $P[A] = 0.86$
- $P[A \cap B] = 0.8$
- $P[A^c] = 0.14$
- $P[A^c \cap B^c] = 0.05$
- $P[A^c \cup B] = P[A^c] + P[B] - P[A^c \cap B] = 0.94$

(b)
- $P[A|B] = \frac{P[A \cap B]}{P[B]} = 0.9$
- $P[B|A] = \frac{P[A \cap B]}{P[A]} = 0.93$
- $P[A|B^c] = \frac{P[A \cap B^c]}{P[B^c]} = 0.55$
- $P[B^c|A] = \frac{P[B^c \cap A]}{P[A]} = 0.43$
- $P[B|A^c] = \frac{P[A^c \cap B]}{P[A^c]} = 0.64$

(c)
- $P[B \cap A^c]$ =0.06
- $P[A \cap B^c]$ =0.09

(d) Não, pois $P[A \cap B] \neq 0$, isto é, os eventos ter alta resistência a ambas doenças possuem intersecção, por isso não são mutuamente exclusivos. No contexto do exemplo, isto significa, por exemplo, que é possível ter resistência a ambas doenças ao mesmo tempo.

(e) $P[A \cap B] \neq P[A] \cdot P[B]$, isto é, o produto das marginais difere dos valores observados, por isso sabemos que os eventos não são independentes. No contexto do exemplo, as chances de ter resistência a uma doença são diferentes quando se tem ou não resistência à outra. Dizendo de outra forma, a probabilidade *marginal* (em todo universo) de se ter uma das doenças é diferente da *condicional* (no subgrupo que tem a outra doença), i.e. $P[A] \neq P[A|B]$.

```
> addmargins(mp)

       B  B^c  Sum
A   0.80 0.09 0.89
A^c 0.06 0.05 0.11
Sum 0.86 0.14 1.00

> outer(rowSums(mp), colSums(mp), "*")

         B    B^c
A   0.7654 0.1246
A^c 0.0946 0.0154

> ## note tb a ordem entre os termos!
> #outer(apply(m/sum(m),2,sum),apply(m/sum(m),1,sum),"*")
```

Avaliação 02

1. Em um lote estão misturadas 20 sementes de uma determinada cultivar (A) de uva com 5 de outra (B). Não é possível distinguir facilmente as sementes, o que só pode ser feito em um exame detalhado. Considere as diferentes possibilidades abaixo e calcule as probabilidades solicitadas.

   (a) Se forem retiradas ao acaso de uma só vez quatro sementes para inspeção, qual a probabilidade de obter ao menos uma de B?

   (b) Se forem retiradas ao acaso e inspecionadas as sementes uma a uma, retornando a semente retirada ao lote, qual a probabilidade de que sejam retiradas três ou mais de A antes de se retirar a primeira de B ?

   (c) Finalmente, considere que as sementes serão retiradas ao acaso e inspecionadas uma a uma, retornando a semente retirada ao lote antes da próxima retirada e serã feitas exatamente quatro retiradas, Qual a probabilidade de obter ao menos uma de B?

Solução:

(a)

$$\begin{aligned}
&X : \text{Número sementes do tipo } B \text{ entre quatro retiradas}\\
&x \in \{0, 1, 2, 3, 4\}\\
&X \sim \mathrm{HG(N = 25, K = 5, n = 4)}\\
P[X = x] &= \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} == \frac{\binom{5}{x}\binom{25-5}{4-x}}{\binom{25}{n}}\\
P[X \geq 1] &= P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = 1 - P[X = 0] = 1 - \frac{\binom{5}{0}\binom{20}{4}}{\binom{24}{4}} = 0.617
\end{aligned}$$

(b)

$$\begin{aligned}
&X : \text{Número de sementes retiradas de } A \text{ até retirar a primeira de } B\\
&X \sim \mathrm{G(p = 5/25 = 0,2)}\\
&x \in \{0, 1, 2, 3, 4, 5, \ldots\}\\
P[X = x] &= (1-p)^x \cdot p = (0,8)^x 0,2\\
P[X \geq 3] &= P[X = 3] + P[X = 4] + P[X = 5] + \ldots = 1 - P[X < 3] = 1 - \{P[X = 0] + P[X = 1] + P[X = 2]\} =\\
&= 1 - \{(0,8)^0 0,2 + (0,8)^1 0,2 + (0,8)^2 0,2\} = 0.422
\end{aligned}$$

(c)

$$\begin{aligned}
&X : \text{Número sementes do tipo } B \text{ entre quatro retiradas (com reposição)}\\
&x \in \{0, 1, 2, 3, 4\}\\
&X \sim \mathrm{B(n = 4, p = 0,20)}\\
P[X = x] &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{4}{x} 0,2^x 0,8^{4-x}\\
P[X \geq 1] &= P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = 1 - P[X = 0] = 1 - \binom{4}{0} 0,2^0 0,8^4 = 0.59
\end{aligned}$$

Soluções computacionais com o programa R:

```
> (pa <- phyper(0, m=5, n=20, k=4, lower=F))
[1] 0.617
> (pb <- pgeom(2, p=0.25, lower=F))
[1] 0.4219
> (pc <- pbinom(0, size=4, prob=0.2, lower=F))
[1] 0.5904
```

---

2. Registros mostram que em determinado bairro são registradas, em média, 1,7 violações de residências por dia. Assuma uma distribuição de probabilidades possivelmente adequada e calcule as probabilidades de que hajam
   (a) três ou mais ocorrências em um dia;
   (b) nenhuma ocorrência em um dia;
   (c) entre uma e quatro ocorrências em um dia.

Solução:

$$\begin{aligned}
&X : \text{Número diário de violações}\\
&x \in \{0, 1, 2, 3, 4, \ldots\}\\
&X \sim \mathrm{P}(\lambda = 1,7)\\
P[X = x] &= \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1.7}1.7^x}{x!}
\end{aligned}$$

(a) $P[X \geq 3] = P[X = 3] + P[X = 4] + P[X = 5] \ldots = 1 - \{P[X = 0] + P[X = 1] + P[X = 2]\} = 1 - \{\frac{e^{-1.7}1.7^0}{0!} + \frac{e^{-1.7}1.7^1}{1!} + \frac{e^{-1.7}1.7^2}{2!}\} = 1 - e^{-1.7}\{1 + 1.7^1 + \frac{1.7^2}{2!}\} = 0.243$

(b) $P[X = 0] = \frac{e^{-1.7}1.7^0}{0!} = 0.183$

(c) $P[1 \leq X \leq 4] = P[X=1]+P[X=2]+P[X=3]+P[X=4] \ldots == \{\frac{e^{-1.7}1.7^1}{1!} + \frac{e^{-1.7}1.7^2}{2} + \frac{e^{-1.7}1.7^3}{3!} + \frac{e^{-1.7}1.7^4}{4!}\} = 1 - e^{-1.7}\{1.7^1 + \frac{1.7^2}{2} + \frac{1.7^3}{6} + \frac{1.7^4}{24}\} = 0.788$

Soluções computacionais com o programa R:

```
> (pa <- ppois(2, lam=1.7, lower=F))
[1] 0.2428
> (pb <- dpois(0, lam=1.7))
[1] 0.1827
> (pc <- diff(ppois(c(0,4), lam=1.7)))
[1] 0.7877
```

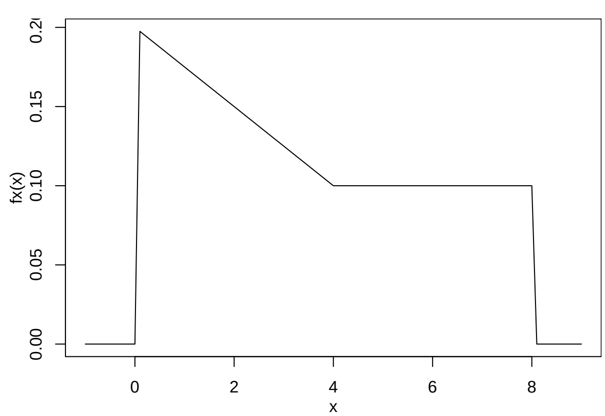
Avaliação 03

1. Uma função de densidade de probabilidade (f.d.p.) tem a expressão:

$$f(x) = \begin{cases} -0,025x + 0,2 & \text{se } 0 < x \leq 4 \\ 0,1 & \text{se } 4 < x \leq 8 \\ 0 & \text{para outros valores de x} \end{cases}$$

(a) Mostre que $f(x)$ é uma f.d.p. válida.
(b) Calcule $P[X < 3]$
(c) Calcule $P[X > 7]$
(d) Calcule $P[X < 6,5]$
(e) Calcule $P[X > 2]$
(f) Calcule $P[3 < X < 7]$
(g) Calcule $P[X < 7|X > 4]$
(h) Obtenha a tal que $P[X < a] = 0,80$
(i) Obtenha a mediana de X
(j) Estime uma valor (aproximado) para a média ($E[X]$). Este valor deve ser menor, maior ou igual à mediana? Justifique.
(k) Obtenha os quartis da distribuição de X

Solução:



(a) i. Pelo gráfico da função é possível veficar que $f(x) \geq 0 \; \forall x$,

ii. A área sob a função deve ser igual a 1. Isto pode ser verificado geometricamente:

$$A = \frac{(0,2+0,1) \cdot 4}{2} + 0,1 \cdot 4 = 1.$$

Alternativamente pode-se integrar a função no intervalo e verificar que

$$\int_0^8 f(x)\,\mathrm{d}x = \ldots = 1$$

(b) $P[X < 3] = 0.488$
(c) $P[X > 7] = 0.1$

(d) $P[X < 6,5] = 0.85$

(e) $P[X > 2] = 0.65$

(f) $P[3 < X < 7] = 0.413$

(g) $P[X < 7|X > 4] = 0.75$

(h) $P[X < a] = 0,80 \longrightarrow a = 6$

(i) $md(X) = 3.1$

(j) $E[X] = 3.47$. Maior que a mediana pois a distribuição é assimétrica (com maior cauda à direita)

(k) $Q_1(X) = 1.37$ e $Q_3(X) = 5.5$

Soluções computacionais com o programa R:

```
> fx <- function(x){
+     y <- numeric(length(x))
+     y[x > 0 & x <= 4] <- -0.025*x[x > 0 & x <= 4] + 0.2
+     y[x > 4 & x <= 8] <- 0.1
+     return(y)
+ }
> (ita <- integrate(fx, 0, 8)$value)

[1] 1

> (itb <- integrate(fx, 0, 3)$value)

[1] 0.4875

> (itc <- integrate(fx, 7, 8)$value)

[1] 0.1

> (itd <- integrate(fx, 0, 6.5)$value)

[1] 0.85

> (ite <- integrate(fx, 2, 8)$value)

[1] 0.65

> (itf <- integrate(fx, 3, 7)$value)

[1] 0.4125

> (itg <- integrate(fx, 4, 7)$value/integrate(fx, 4, 8)$value)

[1] 0.75

> Qx <- function(x, quantil) (integrate(fx, 0, x)$value - quantil)^2
> (q80 <- optimize(Qx, c(0,8), quantil=0.80)$min)

[1] 6

> (md <- optimize(Qx, c(0,8), quantil=0.5)$min)

[1] 3.101

> Ex.f <- function(x) ifelse(x > 0 & x <= 8, x*fx(x), 0)
> (Ex <- integrate(Ex.f, 0, 8)$value)

[1] 3.467

> (q1 <- optimize(Qx, c(0,8), quantil=0.25)$min)

[1] 1.367

> (q3 <- optimize(Qx, c(0,8), quantil=0.75)$min)

[1] 5.5
```