Primeira avaliação (A1)

Disciplina: Inferência Estatística
Professor: Luiz Max de Carvalho

20 de Setembro de 2021

- Por favor, entregue um único arquivo PDF;
- O tempo para realização da prova é de 4 (quatro) horas, mais vinte minutos para upload do documento para o e-class;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo, ou figura geométrica de sua preferência;
- A prova vale 80 pontos; a pontuação restante é contada como bônus.
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto.

Dicas

- Se X tem distribuição exponencial com parâmetro $\lambda > 0$, então, para $x > 0$ as funções de densidade de probabilidade e densidade acumulada são, respectivamente,

$$f_X(x) = \lambda \exp\left(-\lambda x\right),$$
$$F_X(x) = 1 - \exp(-\lambda x).$$

- Se X tem distribuição Gama com parâmetros $\alpha > 0$ e $\beta > 0$ e f.d.p.,

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x),$$

  para $x > 0$, então $W = 1/X$ tem distribuição Gama-inversa, com f.d.p.

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-(\alpha+1)} \exp(-\beta/w),$$

  para $w > 0$. Ademais, $E[W] = \beta/(\alpha-1)$ e $\operatorname{Var}(W) = \beta^2/[(\alpha-1)^2(\alpha-2)]$.

- Se $X_i \sim \text{Gama}(\alpha_i, \beta)$, com $\alpha_i > 0$ para todo i e $\beta > 0$, então $Y = \sum_{i=1}^n X_i$ tem distribuição Gama com parâmetros $\alpha_y = \sum_{i=1}^n \alpha_i$ e $\beta_y = \beta$.

- Se $X \sim \text{Gama}(\alpha, \beta)$, $Y = cX$ tem distribuição $\text{Gama}(\alpha, \beta/c)$ para $c > 0$.

1. Circling the square.

Um círculo C_r de raio r é inscrito em uma folha de papel quadrada com lado b. Suponha que desejamos estimar a área A deste círculo. Para tanto, vamos amostrar vetores aleatórios de uma distribuição uniforme definida sobre a folha de papel e, para estimar a área da circunferência, contar a proporção de vetores caindo dentro e fora de C_r e multiplicar esta proporção pela área total da folha de papel.

a) (2,5 pontos) Mostre que se X e Y são variáveis aleatórias i.i.d. com distribuição Uniforme$(0, b)$, então (X, Y) possui função de densidade de probabilidade constante sobre $(0, b) \times (0, b)$;

b) (7,5 pontos) Você deixa cair grãos de milho sobre a folha e conta quantos deles cairam dentro do círculo e fora do círculo (porém na folha). Vamos supor que este mecanismo gera observações i.i.d. uniforme sobre $(0, b)^2$. Represente os grãos que caíram sobre a folha através de $(X_1, Y_1), ..., (X_n, Y_n)$ e defina $Z_i = \mathbb{I}((X_i, Y_i) \in C_r)$, $i = 1, ..., n$ como uma variável indicadora que recebe valor 1 se o grão está dentro da circunferência. Suponha que depois de medir $\boldsymbol{Z}$ você joga fora $\boldsymbol{X}$ e $\boldsymbol{Y}$, isto é, guarda o milho no pote de novo para fazer pipoca mais tarde. Construa um modelo estatístico parametrizado pela área, A, da circunferência que reflete este experimento. Encontre uma estatística suficiente mínima para o parâmetro deste modelo.

   Dica: desenhe um diagrama e considere as áreas envolvidas (evite avaliar integrais!);

c) (5 pontos) Considere $\delta_1(\boldsymbol{Z}) = b^2 \bar{Z}_n$. Este é um estimador não enviesado da área do círculo?

d) (5 pontos) Calcule o erro quadrático médio $R(A, \delta_1)$ de δ_1 e discuta como ele se comporta em relação à quantidade de interesse. O que acontece com $R(A, \delta_1)$ quando A cresce?

Conceitos trabalhados: Modelo estatístico, suficiência, viés, EQM.
Nível de dificuldade: fácil.
Resolução:[1] Para responder a) basta notar que X e Y i.i.d implica

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{b^2}, \ x, y \in (0, b), \\ 0, \text{ caso contrário,} \end{cases}$$

que é constante em $(0, b) \times (0, b)$. Para responder b) temos que encontrar $\eta = \Pr(Z_i = 1)$ e então nosso modelo estatístico será $Z_i \sim \text{Bernoulli}(\eta(A))$, isto é, a probabilidade de $Z_i = 1$ depende de A de acordo com uma função conhecida da área A, que é a quantidade de interesse. Vamos mostrar duas maneiras de computar η com **forte** preferência pela maneira mais simples. Sabemos que para qualquer variável aleatória cuja distribuição tenha pdf f, vale

$$\Pr(X \in C) = \int_C f(x) \, dx.$$

[1] O título é um trocadilho ("Circle 'in' the square") que faz referência ao famoso problema de construir um quadrado com a mesma área de um círculo usando apenas compasso e régua: `https://en.wikipedia.org/wiki/Squaring_the_circle`.

Logo, concluímos que

$$\Pr(Z_i = 1) = \int_{C_r} f_{X,Y}(x, y)\, dxdy = \frac{\pi r^2}{b^2}.$$

isto é, a razão entre área do círculo e do quadrado. Desta forma, $\eta(A) = A/b^2$. Agora vamos discutir uma solução bem mais trabalhosa para o problema. Note que

$$\Pr(Z_i = 1) = \Pr((X_i, Y_i) \in C_r) = \Pr(X_i^2 + Y_i^2 \leq r^2).$$

Sem perda de generalidade, vamos centrar C_r em $(b/2, b/2)$. Se definirmos $U = (X_i - b/2)^2$, $V = (Y_i - b/2)^2$ e $T = U + V$, temos que

$$\Pr(Z_i = 1) = F_T(r^2).$$

Lembrando da fórmula para convolução (soma) de duas v.a.s independentes, temos

$$F_T(t) = \int_{-\infty}^{\infty} F_U(t - v) f_V(v)\, dv.$$

Para computar F_U, fazemos

$$\begin{aligned}
\Pr(U \leq u) &= \Pr((X_i - b/2)^2 \leq u), \\
&= \Pr(|X_i - b/2| \leq \sqrt{u}) = 2F_X(\sqrt{u}), \\
&= 2\frac{\sqrt{u}}{b}
\end{aligned}$$

o que também significa que $f_V(v) = 1/\sqrt{v}$ porque U e V são i.i.d. Juntando tudo, temos

$$\begin{aligned}
F_T(t) &= \int_0^{\frac{b^2}{4}} 2\frac{\sqrt{t - v}}{b\sqrt{v}}\, dv, \\
&= \Re\left(\frac{2}{b^2}\left\{\frac{b\sqrt{4t - b^2} - 4t\arctan\left(\frac{\sqrt{4t-b^2}}{b}\right)}{4} + \frac{\pi t}{2}\right\}\right), \\
&= \frac{\pi t}{b^2},
\end{aligned}$$

onde $\Re(a + bi) = a$ denota a parte real e $\arctan(x) = y$ se $\tan(y) = x$. A última linha segue do fato de que $\Re(\sqrt{4r^2 - b^2}) = 0$ para $r \in (0, b/2)$, isto é, para os círculos que nos interessam aqui. Avaliar $F_T(r^2)$ completa a computação necessária. **Eu avisei que era melhor evitar computar integrais!**
Vamos mostrar que a soma é suficiente mínima neste exemplo. Para encontrar uma estatística suficiente, vamos utilizar o teorema da fatorização[2]:

$$\begin{aligned}
f_n\left(\boldsymbol{Z} \mid \eta(A)\right) &= \prod_{i=1}^{n} \eta(A)^{z_i}\left(1 - \eta(A)\right)^{1-z_i}, \\
&= \eta(A)^S\left(1 - \eta(A)\right)^{n-S},
\end{aligned}$$

[2]Também chamado de Teorema da Fatorização de Neyman-Fisher, ou NFFT na sigla em inglês.

com $S = \sum_{i=1}^{n} z_i$. Daí, vemos que podemos fatorar a densidade condicional conjunta em $u(\eta(A)) = 1$ e $v[S, \eta(A)] = \eta(A)^S (1-\eta(A))^{n-S}$ e portanto S é suficiente para $\eta(A)$. Para mostrar que S é suficiente **mínima**, vamos considerar uma estatística suficiente $T(\boldsymbol{Z})$ de modo que possamos escrever

$$f_n(\boldsymbol{Z} \mid \eta(A)) = g[T, \eta(A)]h(\eta(A)).$$

Agora, tome $A_1 \neq A_2$ e faça

$$\frac{\eta(A_1)^S (1-\eta(A_1))^{n-S}}{\eta(A_2)^S (1-\eta(A_2))^{n-S}} = \frac{g[T, \eta(A_1)]h(\eta(A_1))}{g[T, \eta(A_2)]h(\eta(A_2))},$$
$$\left(\frac{\eta(A_1)\,[1-\eta(A_2)]}{\eta(A_2)\,[1-\eta(A_1)]}\right)^S \left(\frac{(1-\eta(A_1))}{(1-\eta(A_2))}\right)^n = \frac{g[T, \eta(A_1)]h(\eta(A_1))}{g[T, \eta(A_2)]h(\eta(A_2))},$$

de onde obtemos

$$S = f(T) = \frac{\log\left(\frac{g[T,\eta(A_1)]h(\eta(A_1))}{g[T,\eta(A_2)]h(\eta(A_2))}\right) - n\log\left(\frac{(1-\eta(A_1))}{(1-\eta(A_2))}\right)}{\log\left(\frac{\eta(A_1)[1-\eta(A_2)]}{\eta(A_2)[1-\eta(A_1)]}\right)},$$

e, com isso, mostramos que $S = f(T(\boldsymbol{Z}))$ para toda estatística suficiente T, como queríamos.
Outra forma de mostrar que S é suficiente mínima é mostrar que S é uma função bijetiva do EMV $\hat{p}$ de $p := \eta(A)$. Isso implica em $\hat{p}$ ser suficiente, o que, pelo Teorema 7.8.3 de DeGroot (Observação 10 dos slides), significa que $\hat{p}$ é suficiente mínimo e, assim, S também será suficiente mínima, por ser uma função bijetiva de uma estatística suficiente mínima. Sendo assim, vamos encontrar $\hat{p}$. Maximizar $f_n(\boldsymbol{Z} \mid p)$ é equivalente a maximizar

$$G(p) := \log f_n(\boldsymbol{Z} \mid p) = S\log p + (n-S)\log(1-p).$$

Derivando, obtemos

$$G'(p) = \frac{S}{p} + \frac{S-n}{1-p}$$

e

$$G''(p) = -\frac{S}{p^2} + \frac{S-n}{(1-p)^2}.$$

Como $S \leq n$, temos $G''(p) \leq 0$ para todo p. Portanto, $\hat{p}$ é a solução de $G'(p) = 0$, que podemos, sem muito esforço, identificar como sendo S/n. Dessa forma, temos $S = n\hat{p}$ e está mostrada a suficiência mínima de S. Agora, vamos trabalhar em c). Escrever $Z_1, \ldots Z_n \sim \text{Bernoulli}(\eta(A))$ implica

$$E[\delta_1] = \frac{b^2}{n}E\left[\sum_{i=1}^{n} Z_i\right] = \frac{b^2}{n}nE[Z_1] = b^2\eta(A) = \pi r^2.$$

Concluímos portanto que δ_1 é **não-viesado**.

De c) sabemos que

$$
\begin{aligned}
R(A, \delta_1) &= \operatorname{Var}\left(\delta_1(\boldsymbol{Z})\right), \\
&= \frac{b^4}{n^2} \operatorname{Var}(S) = \frac{b^4 n}{n^2} \operatorname{Var}(Z_1), \\
&= \frac{b^4}{n} \left(\frac{A}{b^2} \left(1 - \frac{A}{b^2} \right) \right), \\
&= \frac{A(b^2 - A)}{n},
\end{aligned}
$$

o que responde d). Note que $\operatorname{Var}(\delta_1)$ é uma função côncava e atinge seu máximo em $A = b^2/2$, que é o ponto médio da amplitude de A, $(0, b^2/4)$. A Figura 1 mostra um esboço de $R(A, \delta_1) = \operatorname{Var}(\delta_1)$. ■

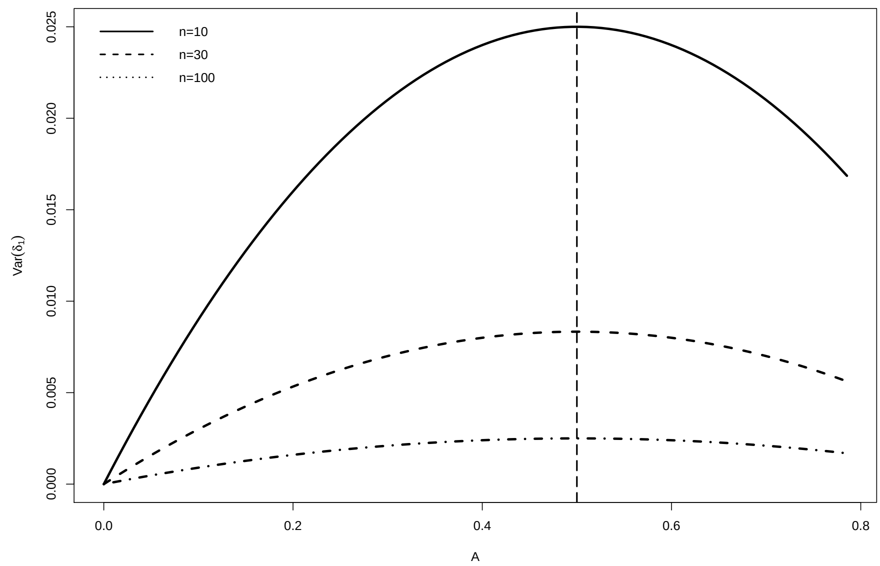

Figura 1: **Erro quadrático médio do estimador** δ_1. Mostramos a curva para $n = 10, 30$ e 100 com $b = 1$. Note que o máximo é atingido em $(b^2/4 - 0)/2 = b^2/2$.

2. The shinning.

Suponha que você é a pessoa responsável pelo controle estatístico de qualidade na fábrica de lâmpadas *LuminaEu*. Seu chefe, Astolfo, lhe envia uma planilha com os valores $X_1, X_2, \ldots, X_n$ dos tempos de falha de n lâmpadas (em dias). Você lê no manual da empresa que um modelo exponencial i.i.d. com parâmetro θ é apropriado para análise.

a) (5 pontos) Mostre que o estimador de momentos para θ coincide com o EMV neste caso;

b) (10 pontos) Discuta se o estimador do item anterior é eficiente para amostras finitas. O que acontece assintoticamente?

c) (5 pontos) Conhecendo Astolfo, no entanto, você sabe que ele não saberá interpretar quaisquer estimativas diretas da taxa θ, então decide considerar a probabilidade de excedência[3] $\alpha := \Pr(X_1 > c)$ para um certo $c > 0$. Encontre um estimador de máxima verossimilhança para α;

Conceitos trabalhados: Método dos momentos, EMV, reparametrização, invariância, eficiência.
Nível de dificuldade: fácil.
Resolução: Sabendo que $E[X_1] = 1/\theta$, o estimador de momentos pode ser obtido escrevendo $\bar{x}_n = 1/\theta$ o que nos leva a $\hat{\theta}_{MM} = n/S$, com $S = \sum_{i=1}^n X_i$. Para o encontrar o EMV, escrevemos

$$f_n(\boldsymbol{X} \mid \theta) = \theta^n \exp\left(-S\theta\right).$$

Tomando o log e diferenciando, temos

$$\lambda_n'(\boldsymbol{X} \mid \theta) = \frac{n}{\theta} - S, \tag{1}$$

$$\lambda_n''(\boldsymbol{X} \mid \theta) = -\frac{n}{\theta^2}, \tag{2}$$

que nos informam que o problema de otimização é côncavo, o que garante que nosso velho conhecido $\hat{\theta}_{EMV} = n/S = 1/\bar{x}_n$ é o único ponto de máximo e o nosso EMV para θ. Isso mostra que os estimadores coincidem.
Para entender se $\delta(X) = n/S$ é eficiente, convêm entender primeiro se é viesado. Das dicas podemos deduzir que δ tem distribuição Gamma inversa com parâmetros $\alpha_\delta = n$ e $\beta_\delta = n\theta$, o que nos leva à conclusão de que $E_\theta[\delta] = n/(n-1)\theta$ e portanto $\text{vies}(\delta) = -\theta/(n-1)$. Das dicas sabemos também que

$$\text{Var}_\theta(\delta) = \frac{n^2\theta^2}{(n-1)^2(n-2)}.$$

Agora vamos verificar se δ atinge a cota inferior de Crámer-Rao:

$$\text{Var}_\theta(\delta) \geq \frac{[m'(\theta)]^2}{nI(\theta)},$$

[3] Em inglês, *exceedance probability.*

onde $m(\theta) := E_\theta[\delta]$ e $I(\theta)$ é a informação de Fisher. Dos cálculos acima, sabemos que $[m'(\theta)]^2 = n^2/(n-1)^2$. Para computar $I(\theta)$ vamos nos aproveitar da identidade $I(\theta) = \text{Var}_\theta(\lambda'(x \mid \theta))$:

$$I(\theta) = \text{Var}_\theta\left(\frac{1}{\theta} - x\right) = \text{Var}_\theta\left(\frac{1}{\theta}\right) + \text{Var}_\theta\left(-x\right) = 0 + \frac{1}{\theta^2}.$$

Juntando tudo, temos

$$\text{Var}_\theta(\delta) \geq \frac{n\theta^2}{(n-1)^2} = V_{\text{op}},$$

o que mostra que δ está longe de ser eficiente para amostras finitas. Com efeito, $\text{Var}_\theta(\delta)/V_{\text{op}} = n/(n-2)$, o que indica que o EMV e o EMM são assintoticamente eficientes.

Das dicas, sabemos que $\alpha = 1 - F_X(c) = \exp(-\theta c)$, o que nos leva a

$$\theta = -\log(\alpha)/c. \tag{3}$$

Para responder c) e encontrar o EMV de α, temos dois caminhos: (i) lembrar da invariância do EMV, substituir o estimador do item anterior em (3) ou; (ii) reescrever a verossimilhança de acordo com o novo parâmetro (procedimento chamado de *reparametrização*) e maximizar esta nova função. Como você deve estar advinhando, aqui vamos fazer as duas coisas. Fazendo a substituição, temos $\hat{\alpha} = \exp(-\hat{\theta}c) = \exp(-nc/S)$ e, para o caminho (ii):

$$\begin{aligned} f_n(\boldsymbol{X} \mid \alpha) &= [-\log(\alpha)/c]^n \exp\left(S\log(\alpha)/c\right), \\ &\propto [-\log(\alpha)]^n \exp\left(S\log(\alpha)/c\right). \end{aligned}$$

Fazendo o procedimento usual, temos

$$\lambda_n'(\boldsymbol{X} \mid \alpha) = \frac{n}{\alpha\log(\alpha)} + \frac{S}{c\alpha}, \tag{4}$$

$$\lambda_n''(\boldsymbol{X} \mid \alpha) = -\frac{n}{\alpha^2\log(\alpha)} - \frac{n}{\alpha^2\log^2(\alpha)} - \frac{S}{c\alpha^2}. \tag{5}$$

Para começar, vemos que atacar o problema diretamente foi uma má ideia: as equações são mais difíceis de resolver e de verificar. Não obstante, podemos resolver (4) e encontrar... $\hat{\alpha} = \exp(-nc/S)$, para surpresa de 0 pessoas! Isso conclui c). ■

3. Cool and normal!

Suponha que você é a pessoa responsável por analisar a concentração de ácido em pedaços de queijo vindos da famosa fábrica de frios francesa *J'skeci*. Assumindo uma distribuição normal para as concentrações em n medições independentes de n pedaços distintos, você precisa descobrir a média μ e a variância v desta distribuição.

a) (5 pontos) Considere a priori imprópria

$$\xi(\mu, v) \propto 1/v \tag{6}$$

Mostre que a posteriori $\xi(\mu, v \mid \boldsymbol{x})$ é própria;

Dica: Procure com atenção o núcleo de distribuições conhecidas.

b) (7,5 pontos) Exiba o estimador de Bayes sob perda quadrática para v e o estimador de Bayes sob perda absoluta para μ e discuta se esses estimadores são viesados;

c) (5 pontos) Encontre uma priori conjugada para (μ, v);

d) (2,5 pontos) Mostre que a priori em (6) pode ser vista como um limite particular (dos hiperparâmetros) da priori conjugada do item anterior.

Conceitos trabalhados: Bayes, propriedade, conjugação
Nível de dificuldade: médio.

Resolução: Para começar, vamos escrever a verossimilhança:

$$f_n(\boldsymbol{x} \mid \mu, v) = \left[\frac{1}{\sqrt{2\pi v}}\right]^n \exp\left(-\frac{1}{2v}\sum_{i=1}^n (x_i - \mu)^2\right).$$

Descartando termos que não dependem dos parâmetros de interesse e utilizando a igualdade $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\mu - \bar{x}_n)^2 = s_n^2 + n(\mu - \bar{x}_n)^2$, temos

$$\begin{aligned} f_n(\boldsymbol{x} \mid \mu, v) &\propto v^{-n/2} \exp\left(-\frac{1}{2v}\left\{s_n^2 + n(\mu - \bar{x}_n)^2\right\}\right), \\ &\propto \exp\left(-\frac{n(\mu - \bar{x}_n)^2}{2v}\right) v^{-n/2} \exp\left(-\frac{s_n^2}{2v}\right). \end{aligned}$$

Daí, temos a posteriori

$$\begin{aligned} \xi(\mu, v \mid \boldsymbol{x}) &\propto f_n(\boldsymbol{x} \mid \mu, v)\xi(\mu, v), \\ &\propto \underbrace{\exp\left(-\frac{n(\mu - \bar{x}_n)^2}{2v}\right)}_{\text{Normal}(\bar{x}_n, \frac{v}{n})} \underbrace{v^{-n/2-1} \exp\left(-\frac{s_n^2}{2v}\right)}_{\text{Gama-Inversa}(\frac{n}{2}, \frac{s_n^2}{2})}, \qquad (7) \\ &= p_1(\mu \mid \boldsymbol{x}, v) p_2(v \mid \boldsymbol{x}). \end{aligned}$$

Vemos então que a posteriori se fatora em uma posteriori para μ condicional a v e aos dados, e uma posteriori marginal (em relação a μ) para v. Com isso, podemos responder b): Para v queremos a média *a posteriori*, que é o estimador de Bayes sob perda quadrática. Usando as dicas, sabemos que $E_{p_2}[v] = (s_n^2/2)/((n-2)/2) = s_n^2/(n-2)$. Já o estimador solicitado para μ é a mediana *a posteriori* de acordo com p_1, o que é simplesmente $\bar{x}_n$. Note que este estimador independe do valor de v. Agora podemos dizer que o estimador para μ é não-viesado, já que $E[\bar{X}_n] = \mu$, e que o estimador para v é viesado, já que o estimador de v não-viesado da forma $\delta_c(\boldsymbol{X}) = c\sum_{i=1}^n (X_i - \bar{X}_n)^2$ tem $c = 1/(n-1)$, como visto em aula.
Para responder c) vamos notar que precisamos encontrar $\tilde{\xi} : (-\infty, \infty) \times (0, \infty) \to (0, \infty) \in \mathcal{F}$ de modo que

$$\exp\left(-\frac{n(\mu - \bar{x}_n)^2}{2v}\right) v^{-n/2} \exp\left(-\frac{s_n^2}{2v}\right) \tilde{\xi}(\mu, v) \in \mathcal{F}.$$

As derivações acima sugerem uma estrutura condicional para $\tilde{\xi}$ da mesma forma daquela em (7): se fizermos $\tilde{\xi}(\mu, v) = \pi_1(\mu \mid v)\pi_2(v)$, podemos escrever:

$$\tilde{\xi}(\mu, v \mid \boldsymbol{x}) \propto \exp\left(-\frac{n(\mu - \bar{x}_n)^2}{2v}\right) \pi_1(\mu \mid v) \cdot v^{-n/2} \exp\left(-\frac{s_n^2}{2v}\right) \pi_2(v) \in \mathcal{F},$$

o que nos sugere o sistema de equações funcionais

$$\exp\left(-\frac{n(\mu-\bar{x}_n)^2}{2v}\right)\pi_1(\mu \mid v) = \text{Normal}(m,\tau),$$
$$v^{-n/2}\exp\left(-\frac{s_n^2}{2v}\right)\pi_2(v) = \text{Gama-inversa}(\alpha',\beta').$$

Para resolver a primeira equação, lembramos que para v fixa, podemos escrever $\pi_1(\mu, v; m, \lambda) = \text{Normal}(m, v/\lambda)$ e, completando o quadrado (duas vezes) vemos que a posteriori resultante continua na família normal (condicional a v). Também podemos ver que $\pi_2(v; \alpha_0, \beta_0) = \text{Gama-inversa}(\alpha_0, \beta_0)$ leva a uma posteriori marginal para v que permanece na família Gama-inversa, o que conclui o nosso argumento.
No que toca à questão d), vamos escrever a densidade conjunta *a priori*:

$$\tilde{\xi}(\mu, v) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}v^{-\alpha_0-1}\exp\left(-\beta_0 v\right)\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{\lambda(\mu-m)^2}{2v}\right).$$

Para obter $\tilde{\xi}(\mu, v) \propto v^{-1}$, podemos tomar $\beta_0 \to \infty$, $\alpha_0 \to 0$ pelo lado da Gama-inversa e $\lambda \to \infty$ pelo lado da normal. Desta forma, vemos que, pelo menos neste caso, uma priori imprópria pode ser vista como um limite particular de prioris próprias. ■

4. Get your ducks in a row.

Pato Donald, Huguinho, Zezinho e Luisinho estão estudando Inferência Estatística para trabalhar no *hedge fund* do Tio Patinhas. O problema em questão é a estimação do parâmetro θ de uma distribuição uniforme em $(\theta/2, 3\theta/2)$ a partir de uma amostra aleatória $X_1, X_2, \ldots, X_n$. Cada um propôs um estimador diferente para θ e seu trabalho é ajudar o Tio Patinhas a ordenar esses estimadores em ordem de qualidade.

Sejam $M := \max(X_1, X_2, \ldots, X_n)$ e $m := \min(X_1, X_2, \ldots, X_n)$.

Os estimadores escolhidos foram

1. $\delta_{\text{D}}(\boldsymbol{X}) = X_1$, para o Pato Donald;
2. $\delta_{\text{H}}(\boldsymbol{X}) = m$, para Huguinho;
3. $\delta_{\text{Z}}(\boldsymbol{X}) = M$, para Zezinho;
4. $\delta_{\text{L}}(\boldsymbol{X}) = (M+m)/2$, para Luisinho;

Para lhe ajudar na tarefa de julgar estes estimadores, Tio Patinhas enviou o seguinte conjunto de fatos úteis: para $X_1, X_2, \ldots, X_n \sim \text{Uniforme}(a, b)$, temos

$$\begin{aligned}
E[X_1] &= \frac{a+b}{2},\\
\operatorname{Var}(X_1) &= \frac{(b-a)^2}{12},\\
E[m] &= a + \frac{1}{n+1}(b-a),\\
E[M] &= b - \frac{1}{n+1}(b-a),\\
\operatorname{Var}(m) = \operatorname{Var}(M) &= \frac{n}{(n+1)^2(n+2)}(b-a)^2,\\
\operatorname{Cov}(m, M) &= \frac{(b-a)^2}{(n+1)^2(n+2)},\\
\operatorname{Corr}(m, M) &= \frac{1}{n}.
\end{aligned}$$

Os patos ainda não sabem Inferência Estatística muito bem, portanto tenha paciência com eles.

a) (2,5 pontos) Os estimadores de Huguinho e Zezinho são viesados. Mostre aos patinhos como construir versões não-viesadas, $\delta_{\text{UH}}(\boldsymbol{X})$ e $\delta_{\text{UZ}}(\boldsymbol{X})$;

b) (2,5 pontos) Discuta se algum dos estimadores do item anterior é inadmissível;

c) (2,5 pontos) Mostre que $\boldsymbol{T} = (m, M)$ é suficiente conjunta para θ;

d) (7,5 pontos) Mostre que $\delta_{\text{L}}(\boldsymbol{X}) = E[\delta_{\text{D}}(\boldsymbol{X}) \mid \boldsymbol{T}]$, isto é, que o estimador de Luisinho é o melhoramento de Rao-Blackwell do estimador do Pato Donald;

e) (5 pontos) Ordene os estimadores $\delta_{\text{D}}(\boldsymbol{X})$, $\delta_{\text{UH}}(\boldsymbol{X})$, $\delta_{\text{UZ}}(\boldsymbol{X})$ e $\delta_{\text{L}}(\boldsymbol{X})$ em termos de erro quadrático médio. Quem propôs o melhor estimador?[4]

Conceitos trabalhados: Rao-Blackwell, viés, EQM, admissibilidade.
Nível de dificuldade: médio.
Resolução: Para começar, vamos nos dar conta de que

$$\begin{aligned}
E_\theta[\delta_{\text{H}}(\boldsymbol{X})] &= \frac{\theta}{2}\frac{1}{n+1}\theta = \frac{n+3}{2(n+1)} = k_{\text{H}}\theta,\\
E_\theta[\delta_{\text{Z}}(\boldsymbol{X})] &= \frac{3}{2}\theta - \frac{1}{n+1}\theta = \frac{3n+1}{2(n+1)}\theta = k_{\text{Z}}\theta,
\end{aligned}$$

o que imediatamente sugere

$$\begin{aligned}
\delta_{\text{UH}}(\boldsymbol{X}) &= \frac{1}{k_{\text{H}}}\delta_{\text{H}}(\boldsymbol{X}),\\
\delta_{\text{UZ}}(\boldsymbol{X}) &= \frac{1}{k_{\text{Z}}}\delta_{\text{Z}}(\boldsymbol{X}),
\end{aligned}$$

[4]No caso de Huguinho e Zezinho, com a sua ajuda.

como soluções de a). Para reponder b), vamos lembrar que os estimadores são não-viesados e desta forma os EQMs são dados apenas pelas variâncias dos estimadores. Portanto:

$$R\left(\theta, \delta_{\mathrm{UH}}\right) = \operatorname{Var}\left(\delta_{\mathrm{UH}}(\boldsymbol{X})\right) = \frac{1}{k_{\mathrm{H}}^2}\operatorname{Var}\left(\delta_{\mathrm{H}}(\boldsymbol{X})\right) = \frac{4(n+1)^2}{(n+3)^2}\operatorname{Var}\left(\delta_{\mathrm{H}}(\boldsymbol{X})\right),$$
$$R\left(\theta, \delta_{\mathrm{UZ}}\right) = \operatorname{Var}\left(\delta_{\mathrm{UZ}}(\boldsymbol{X})\right) = \frac{1}{k_{\mathrm{Z}}^2}\operatorname{Var}\left(\delta_{\mathrm{Z}}(\boldsymbol{X})\right) = \frac{4(n+1)^2}{(3n+1)^2}\operatorname{Var}\left(\delta_{\mathrm{H}}(\boldsymbol{X})\right),$$

onde a última igualdade da segunda linha vem do fato de que m e M tem a mesma variância. Com isso vemos que o estimador de Zezinho domina uniformemente o estimador de Huguinho, que é inadmissível, portanto. Uma observação interessante, então, é de que, no que toca à tarefa de estimar θ, o máximo da amostra é estritamente mais informativo sobre o parâmetro que o mínimo.
Parte da questão c) foi trabalhada em aula. Em particular, podemos escrever a verossimilhança como

$$f_n(\boldsymbol{X} \mid \theta) = \begin{cases} \frac{\mathbb{I}(m/2<\theta)\mathbb{I}(2M/3<\theta)}{\theta^n}, 0 \leq x \leq \theta, \\ 0, \text{ caso contrário,} \end{cases}$$

o que mostra que $\boldsymbol{T} = (m, M)$ é suficiente[5], pelo Teorema da Fatorização.
A solução de d) passa por lembrar bem o que é o mecanismo de Rao-Blackwell para melhoramento de estimadores: dado um estimador δ de $g(\theta)$, e uma estatística suficiente T para $g(\theta)$, podemos sempre construir

$$\delta_0 := E_\theta[\delta \mid T],$$

que domina δ em termos de EQM. Com as informações das dicas da questão e do item c), fica claro que estamos procurando $E_\theta[\delta_{\mathrm{D}}(\boldsymbol{X}) \mid m, M]$. Notando que

$$f_1(x_1 \mid m, M) = \begin{cases} \frac{1}{M-m}, m \leq x_1 \leq M, \\ 0, \text{ caso contrário,} \end{cases}$$

isto é que $X_1 \mid m, M \sim \text{Uniforme}(m, M)$, obtemos

$$\begin{aligned} E_\theta[X_1 \mid m, M] &= E_\theta[\delta_{\mathrm{D}}(\boldsymbol{X}) \mid \delta_{\mathrm{H}}(\boldsymbol{X}), \delta_{\mathrm{Z}}(\boldsymbol{X})], \\ &= \frac{\delta_{\mathrm{Z}}(\boldsymbol{X}) + \delta_{\mathrm{H}}(\boldsymbol{X})}{2}, \\ &= \delta_{\mathrm{L}}(\boldsymbol{X}), \end{aligned}$$

como queríamos demonstrar. Para finalizar a questão e responder e), precisamos computar o EQM dos estimadores do Pato Donald e de Luisinho. Para o primeiro, temos:

$$\begin{aligned} R(\theta, \delta_{\mathrm{D}}(\boldsymbol{X})) &= \operatorname{Var}(X_1) + \left[E[X_1] - \theta\right]^2, \\ &= \frac{\theta^2}{12} + 0 = \frac{1}{12}\theta^2. \end{aligned}$$

[5] Além disso, $\boldsymbol{T}' = (m/2, 2M/3)$ também é suficiente para θ.

Já para o estimador de Luisinho, temos

$$
\begin{aligned}
E[\delta_{\mathrm{L}}(\boldsymbol{X})] &= \frac{E[m]+E[M]}{2} = \frac{(4n+4)\theta}{4(n+1)} = \theta, \\
\operatorname{Var}\left(\delta_{\mathrm{L}}(\boldsymbol{X})\right) &= \frac{\operatorname{Var}(m)+\operatorname{Var}(M)+2\operatorname{Cov}(m,M)}{4}, \\
&= \frac{\operatorname{Var}(M)+\operatorname{Cov}(m,M)}{2}, \\
&= \frac{\theta^2}{2(n+1)(n+2)}.
\end{aligned}
$$

Com isso, podemos computar

$$
R(\theta, \delta_{\mathrm{L}}(\boldsymbol{X})) = \frac{\theta^2}{2(n+1)(n+2)}.
$$

Agora estamos preparados para construir a nossa ordenação:

$$
\begin{aligned}
R(\theta, \delta_{\mathrm{D}}(\boldsymbol{X})) &= \frac{1}{12}\theta^2, \\
R(\theta, \delta_{\mathrm{UH}}(\boldsymbol{X})) &= \frac{4n}{(n+3)^2(n+2)}\theta^2, \\
R(\theta, \delta_{\mathrm{UZ}}(\boldsymbol{X})) &= \frac{4n}{(3n+1)^2(n+2)}\theta^2, \\
R(\theta, \delta_{\mathrm{L}}(\boldsymbol{X})) &= \frac{1}{2(n+1)(n+2)}\theta^2.
\end{aligned}
$$

De modo que **Zezinho** é o vencedor, com o estimador com menor EQM. Isso se deve em parte ao fato de que Zezinho utilizou uma estatística suficiente. Luisinho até foi esperto e usou Rao-Blackwell, mas como o estimador original (do Pato Donald) não era uma estatística suficiente, acabou conseguindo um estimador ruim. A Figura 2 mostra a distribuição dos estimadores considerados aqui. ■

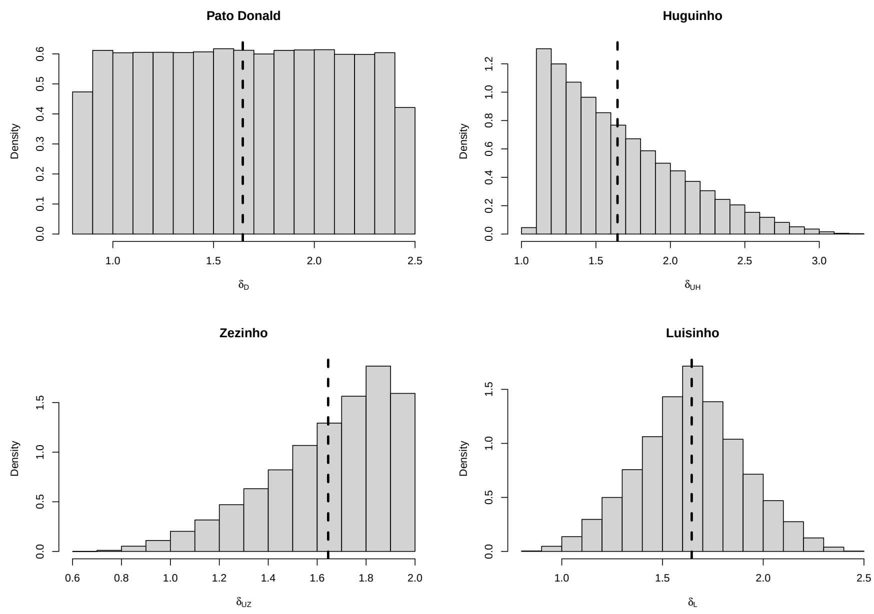

Figura 2: **Distribuição dos estimadores considerados pelos patos**. Mostramos a distribuição de cada estimador com $\theta = \pi^2/6$, $n = 3$ e $M = 10^5$ realizações.

5.Questão bônus: Boss is boss, ain't it, dad?

Considere mais uma vez o problema da questão 4. Desta vez, Tio Patinhas resolveu propor o próprio estimador, e quer mostrar que esse estimador pode ser melhor que qualquer um dos propostos anteriormente. Para isso, propõe utilizar um estimador da forma

$$\delta_{\mathrm{P}}(\boldsymbol{X}) = (1-\alpha)\delta_{\mathrm{UH}}(\boldsymbol{X}) + \alpha\delta_{\mathrm{UZ}}(\boldsymbol{X}),$$

com $\alpha \in (0, 1)$.

a) (10 pontos) Mostre que δ_{P} é não-viesado e compute seu erro quadrático médio;

   Dica: Lembre-se de que para $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y).$$

b) (10 pontos) Encontre α_{op} que faz com que δ_{P} tenha variância mínima. O estimador $\delta_{\mathrm{P}}^{\mathrm{op}}(\boldsymbol{X}) = (1-\alpha_{\mathrm{op}})\delta_{\mathrm{UH}}(\boldsymbol{X}) + \alpha_{\mathrm{op}}\delta_{\mathrm{UZ}}(\boldsymbol{X})$ domina todos aqueles derivados na questão 4? Justifique.

Conceitos trabalhados: variância mínima, combinação convexa de estimadores não-viesados.
Nível de dificuldade: médio.

Resolução: Começamos mostrando que

$$\begin{aligned} E[\delta_{\mathrm{P}}(\boldsymbol{X})] &= E[(1-\alpha)\delta_{\mathrm{UH}} + \alpha\delta_{\mathrm{UZ}}], \\ &= (1-\alpha)E[\delta_{\mathrm{UH}}] + \alpha E[\delta_{\mathrm{UZ}}], \\ &= (1-\alpha)\theta + \alpha\theta = \theta. \end{aligned}$$

Portanto,

$$\begin{aligned} R_\alpha(\theta, \delta_{\mathrm{P}}(\boldsymbol{X})) &= \mathrm{Var}\left((1-\alpha)\delta_{\mathrm{UH}} + \alpha\delta_{\mathrm{UZ}}\right), \\ &= \mathrm{Var}\left(\frac{(1-\alpha)}{k_{\mathrm{H}}}\delta_{\mathrm{H}} + \frac{\alpha}{k_{\mathrm{Z}}}\delta_{\mathrm{Z}}\right), \\ &= \frac{(1-\alpha)^2}{k_{\mathrm{H}}^2}\mathrm{Var}\,(m) + \frac{\alpha^2}{k_{\mathrm{Z}}^2}\mathrm{Var}\,(M) + 2\frac{\alpha(1-\alpha)}{k_{\mathrm{H}}k_{\mathrm{Z}}}\mathrm{Cov}(m, M), \\ &= \gamma\left(\frac{(1-\alpha)^2 n}{k_{\mathrm{H}}^2} + \frac{\alpha^2 n}{k_{\mathrm{Z}}^2} + 2\frac{\alpha(1-\alpha)}{k_{\mathrm{H}}k_{\mathrm{Z}}}\right), \end{aligned}$$

onde $\gamma = \mathrm{Cov}(m, M) = \frac{\theta^2}{(n+1)^2(n+2)}$.
Para reponder b), vamos minimizar $R_\alpha(\theta, \delta_{\mathrm{P}}(\boldsymbol{X}))$ com respeito a α:

$$\frac{d}{d\alpha}R_\alpha(\theta, \delta_{\mathrm{P}}(\boldsymbol{X})) = 2\gamma\left(\frac{\left(\left(k_{\mathrm{Z}}^2 + k_{\mathrm{H}}^2\right)n - 2k_{\mathrm{H}}k_{\mathrm{Z}}\right)\alpha - k_{\mathrm{Z}}^2 n + k_{\mathrm{H}}k_{\mathrm{Z}}}{k_{\mathrm{H}}^2 k_{\mathrm{Z}}^2}\right), \tag{8}$$

Igualando (8) a zero, obtemos

$$\begin{aligned} \alpha_{\mathrm{op}} &= \frac{k_{\mathrm{Z}}^2 n - k_{\mathrm{H}}k_{\mathrm{Z}}}{\left(k_{\mathrm{Z}}^2 + k_{\mathrm{H}}^2\right)n - 2k_{\mathrm{H}}k_{\mathrm{Z}}}, \\ &= \frac{9n+3}{10n+6}. \end{aligned}$$

Com isso, concluímos que o estimador do Tio Patinhas é melhor que os outros, finalizando a questão. Como um extra, podemos calcular, depois de alguma álgebra,

$$R_\alpha(\theta, \delta_{\mathrm{P}}(\boldsymbol{X})) = \frac{2}{(5n+3)(n+2)}\theta^2.$$

É importante notar que a melhoria não é escandalosa. Por exemplo, para $\theta = \pi^2/6$, $n = 3$, $R_\alpha(\theta, \delta_{\mathrm{UZ}}(\boldsymbol{X})) = 1.22 \times 10^{-3}$, enquanto $R_\alpha(\theta, \delta_{\mathrm{P}}(\boldsymbol{X})) = 1.10 \times 10^{-3}$.