Primeira avaliação (A1)

Disciplina: Inferência Estatística (MSc)
Instrutor: Luiz Max Carvalho
Monitor: Lucas Moschen

27 de Julho de 2022.

- O tempo para realização da prova é de 3 horas e 55 minutos;
- Leia a prova toda com calma antes de começar a responder;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo ou figura geométrica de sua preferência;
- A prova vale 100 pontos. A pontuação restante é contada como bônus;
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;
- Como convenção adotamos $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_+ = (0, \infty)$ e $\mathbb{N} = \{1, 2, \ldots\}$.

1. Uniformly cool.

Seja $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma distribuição uniforme em $(0, \theta]$ com densidade comum com respeito a Lebesgue

$$f_\theta(x) = \frac{1}{\theta}\mathbb{I}(0 < x \leq \theta).$$

Defina $m := \min(\boldsymbol{X}_n)$ e $M := \max(\boldsymbol{X}_n)$.

a) (20 pontos) Encontre o estimador de máxima verossimilhança para θ e mostre que ele é suficiente e completo. Discuta se este estimador é viesado e se é consistente.

b) (10 pontos) Mostre que $U = \frac{m}{M}$ é ancilar.

   Dica: Para uma amostra aleatória de tamanho n com cdf comum F e $a < b$, vale

$$\Pr(m \leq a, M \leq b) = [F(a)]^n - [F(b) - F(a)]^n.$$

c) (10 pontos) Calcule $E_\theta[U]$.

   Dica: Compute $E_\theta[m]$ e lembre-se de que se X e Y são v.a.s independentes, $E[XY] = E[X]E[Y]$.

Conceitos trabalhados: estimador de máxima verossimilhança; suficiência; ancilaridade; Teorema de Basu.
Nível de dificuldade: médio.
Resolução: Primeiro vamos escrever a função de densidade de probabilidade conjunta da amostra observada:

$$\begin{aligned}
f_\theta(\boldsymbol{x}_n) &= \prod_{i=1}^n f_\theta(x_i) \quad \text{(i.i.d.)}, \\
&= \prod_{i=1}^n \frac{1}{\theta}\mathbb{I}(x_i > 0)\mathbb{I}(x_i \leq \theta), \\
&= \frac{1}{\theta^n}\prod_{i=1}^n \mathbb{I}\left(\min_i x_i > 0\right)\mathbb{I}\left(\max_i x_i \leq \theta\right).
\end{aligned}$$

Assim, fazendo $L_n(\theta) = f_\theta(x_n)$ a nossa função de verossimilhança, vemos (Figura 1) que esta função é descrescente no intervalo $[\max_i x_i, \infty)$ e, portanto,

$$\sup_{\theta \in (0,\infty)} L_n(\theta) = L_n(\max_i x_i) = \frac{1}{(\max_i x_i)^n}.$$

Desta forma, dizemos que $\hat{\theta}_{\text{EMV}} = M$. Pelo teorema da fatorização de Neyman-Fisher, sabemos que todo EMV é suficiente, o que também pode ser verificado diretamente por inspeção da função de verossimilhança. Para verificar completude, precisamos verificar que para g mensurável vale

$$E_\theta[g(M)] = 0 \implies g(x) = 0 \quad \forall x \in E,$$

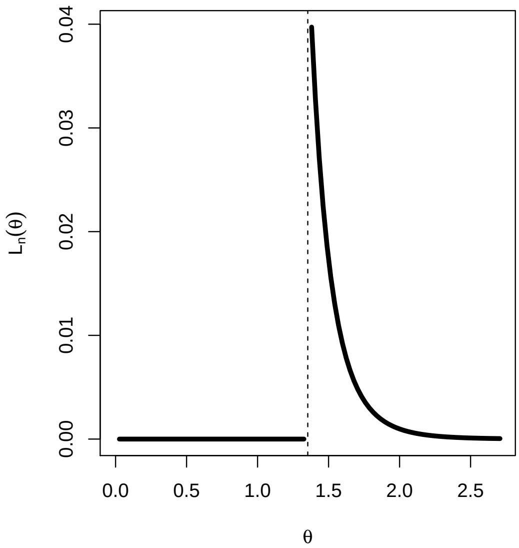

Figura 1: **Função de verossimilhança para uma uniforme em** $(0, \theta]$. Aqui mostramos a verossimilhança para $n = 10$. Como podemos notar, a função é zero até $M = \max_i X_i$ (linha vertical pontilhada). A partir deste ponto a função é decrescente e $\propto \theta^{-n}$.

com $P_\theta(M \in E) = 1$ para todo $\theta \in (0, \infty)$. Primeiro, note que

$$F_M(y) = P_\theta(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) = \prod_{i=1}^{n} P_\theta(X_i \leq y) = \left[\frac{y}{\theta}\right]^n.$$

Assim,

$$f_M(y) = \frac{dF_M(y)}{dy} = \frac{n}{\theta}\left[\frac{y}{\theta}\right]^{n-1}.$$

Assim temos que, se M é completa,

$$E_\theta[g(M)] = \frac{n}{\theta^n}\int_0^\theta g(y)y^{n-1}\,dy = 0.$$

Defina $K : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{0\}$,

$$K(\theta) = \int_0^\theta g(y)y^{n-1}\,dy.$$

Pela regra de Leibniz, temos

$$\frac{d}{d\theta}K(\theta) = g(\theta)\theta^n.$$

Ora, se $\frac{d}{dx}K(x) = 0$ para todo x, há de ser o caso de que $g(x) = 0$ para todo x – a menos, talvez, de um conjunto de medida nula –, o que é exatamente o que queríamos demonstrar. Para entender o viés deste estimador, vamos calcular $E_\theta[M]$.

$$E_\theta[M] = \frac{n}{\theta^n}\int_0^\theta y y^{n-1}\,dy = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$

Outro caminho é utilizar a fórmula de cauda para a esperança (Lo, 2018), $E_\theta[M] = \int_0^\infty \Pr(M > y)\,dy = \int_0^\infty [1 - F_M(y)]\,dy$:

$$\begin{aligned}
E_\theta[M] &= \int_0^\infty \left[1 - \left[\frac{y}{\theta}\right]^n\right] dy,\\
&= \int_0^\theta \left[1 - \left[\frac{y}{\theta}\right]^n\right] dy + \cancelto{0}{\int_\theta^\infty \left[1 - \left[\frac{y}{\theta}\right]^n\right] dy},\\
&= \int_0^\theta 1\,dy - \frac{1}{\theta^n}\int_0^\theta y^n\,dy\\
&= \theta - \frac{\theta^{n+1}}{(n+1)\theta^n} = \frac{n}{n+1}\theta.
\end{aligned}$$

Para finalizar, vamos pensar sobre a consistência deste estimador. Tudo que precisamos é calcular, para $\epsilon > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P_\theta\left(\mid M - \theta \mid > \epsilon\right) = \lim_{n\to\infty} P_\theta\left(\theta - \epsilon \le M \le \theta + \epsilon\right),\\
&= \lim_{n\to\infty} \left\{P_\theta\left(M \le \theta\right) - P_\theta\left(M \le \theta - \epsilon\right)\right\},\\
&= \lim_{n\to\infty} 1 - \left[\frac{\theta - \epsilon}{\theta}\right]^n = 1,
\end{aligned}$$

portanto $M \xrightarrow{p} \theta$. Outra maneira de ver esse resultado é notar que M é assintoticamente não-viesado e que $\lim_{n\to\infty} \operatorname{Var}_\theta(M) = 0$. Ora, por Chebychev, temos que para $\epsilon > 0$

$$P_\theta\left(|M - \theta| \ge \epsilon\right) \le \frac{E_\theta[(M-\theta)^2]}{\epsilon^2}.$$

Como o lado direito vai pra zero com $n \to \infty$, temos o resultado[1]. Agora vamos responder b) mostrando que a distribuição de $U = m/M$ não depende de θ. Em primeiro lugar, vamos usar a dica para encontrar

$$P_\theta(m \le a, M \le b) = \left[\frac{b}{\theta}\right]^n - \left[\frac{b}{\theta} - \frac{a}{\theta}\right]^n.$$

Diferenciando, encontramos

$$f_\theta(a,b) = \frac{n(n-1)}{\theta^n}[b-a]^{n-2}, 0 < a < b \le \theta.$$

Para encontrar a distribuição de U precisamos considerar transformações das variáveis aleatórias m e M. Como U é uma função de $\mathbb{R}_+^2$ em $\mathbb{R}_+$, vamos

[1]Ou seja, convergência quadrática implica convergência em probabilidade – como visto em aula e Keener (2010), prop 8.3. Isto por sua vez significa que não-viesamento assintótico + variância assintótica zero $\implies$ consistência. Ver `https://stats.stackexchange.com/a/500762/97431`

“suplementar” o mapa com uma coordenada a mais, fazendo assim $(m, M) \mapsto (U, M)$. A matriz jacobiana da transformação é

$$J = \begin{bmatrix} M & U \\ 0 & 1 \end{bmatrix},$$

e temos que $m = UM$, $f_{U,M}(u, b) = f_\theta(a, b)|J|$, ou seja,

$$f_{U,M}(u, b) = \frac{n(n-1)}{\theta^n}[b - ub]^{n-2}b = \frac{n(n-1)}{\theta^n}b^{n-1}[1-u]^{n-2}.$$

Agora falta integrar com respeito a M para obter

$$\begin{aligned} f_U(u) &= \int_0^\theta f_{U,M}(u, b)\, db, \\ &= \frac{n(n-1)}{\theta^n}[1-u]^{n-2}\int_0^\theta b^{n-1}\, db, \\ &= (n-1)[1-u]^{n-2}, \end{aligned}$$

o que indica que, de fato, a densidade marginal de U não depende de θ e portanto esta estatística é ancilar. Lucas nos aponta que uma maneira mais simples de ver o problema é notar que $Y_i = X_i/\theta$ tem distribuição uniforme em $(0, 1)$, e assim $U = M_y/m_y = (M/\theta)/(m/\theta)$ é a razão de duas quantidades que não dependem de θ. Variáveis aleatórias como M/θ e m/θ são chamadas quantidades pivotais ou pivôs, e serão úteis mais à frente no curso. Para concluir a questão, vamos responder c). Primeiro, vamos seguir a dica e computar $E_\theta[m]$:

$$P_\theta(m \leq a) = 1 - \prod_{i=1}^n P_\theta(X_i > a) = 1 - \left[1 - \frac{a}{\theta}\right]^n.$$

Utilizando a fórmula de cauda e a substituição $u = 1 - x/\theta$, temos

$$E_\theta[m] = \int_0^\theta \left[1 - \frac{x}{\theta}\right]^n dx = \frac{\theta}{n+1}.$$

Como M é completa e suficiente e U é ancilar, temos que elas são independentes pelo Teorema de Basu, e portanto,

$$E_\theta\left[\frac{m}{M}M\right] = E_\theta[m] = \frac{\theta}{n+1} = E_\theta[U]E_\theta[M] = E_\theta[U]\frac{n}{n+1}\theta,$$

o que nos leva a concluir que

$$E_\theta[U] = \frac{1}{n}.$$

Esta quantidade é independente de θ, o que era de se esperar dada a sua ancilaridade. Podemos também notar que f_U é a densidade de uma distribuição beta com parâmetros $\alpha = 1$ e $\beta = n - 1$ e então computar $\alpha/(\alpha + \beta) = 1/n$.

Discussão: Nesta questão respondemos algumas perguntas simples sobre um modelo que foi bastante estudado durante as aulas. A questão c) mostra a utilidade do teorema de Basu, que permite intuir a resposta sem necessariamente fazer a integração de f_U, que, se não soubéssemos que é uma densidade beta, daria um certo trabalho para integrar. ■

2. So many dice...

Seja $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma distribuição binomial com número de tentativas $N \in \mathbb{N}$ conhecido e probabilidade de sucesso $\theta \in (0, 1)$ desconhecida. Considere estimar a probabilidade de exatamente um sucesso:

$$g(\theta) := P_\theta(X = 1) = N\theta(1-\theta)^{N-1}.$$

a) (10 pontos) Mostre que $\delta_1(\boldsymbol{X}_n) = \mathbb{I}(X_1 = 1)$ é um estimador não-viesado de $g(\theta)$;

b) (10 pontos) Encontre uma estatística suficiente e completa, $T(\boldsymbol{X}_n)$.

c) (10 pontos) Argumente que δ_1 não é suficiente e que portanto pode ser melhorado usando T. Mostre como fazer isso e obter um estimador $\delta_2(\boldsymbol{X}_n)$.

d) (10 pontos) Exiba a fórmula para δ_2 e mostre que ele é ENVVUM.

Conceitos trabalhados: Rao-Blackwell; Estimador não-viesado de variância uniformemente mínima (ENVVUM); Lehmann-Scheffé.
Nível de dificuldade: fácil.
Resolução: A resolução aqui vai seguir bem de perto a apresentação do exemplo 7.3.24 de Casella and Berger (2002). Para começar, vamos notar que para calcular o viés de δ_1 só precisamos da distribuição de X_1 e que $E_\theta[\delta_1]$ vale

$$\begin{aligned}
E_\theta[\delta_1] &= \sum_{x_1=0}^{n} \mathbb{I}(x_1 = 1)\binom{N}{x_1}\theta^{x_1}(1-\theta)^{N-x_1}, \\
&= \binom{N}{1}\theta^1(1-\theta)^{N-1}, \\
&= g(\theta).
\end{aligned}$$

Easy-peasy. Agora, vamos pensar em encontrar T suficiente e completa. Primeiro, vamos escrever a verossimilhança:

$$\begin{aligned}
L_n(\theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n) &= \prod_{i=1}^{n} P_\theta(X_i = x_i), \\
&= \prod_{i=1}^{n}\binom{N}{x_i}(1-\theta)^N\left(\frac{\theta}{1-\theta}\right)^{x_i}, \\
&\underset{\theta}{\propto} (1-\theta)^{nN}\left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i},
\end{aligned}$$

o que imediatamente sugere que $T(\boldsymbol{X}_n) = \sum_{i=1}^n X_i$ é suficiente pelo Teorema da fatorização. Vamos derivar a distribuição de T para em seguida verificar se ela é completa. Se X é v.a. binomial com parâmetros $k \in \mathbb{N}$ e $p \in (0, 1)$, então a sua função geradora de probabilidades (PGF) vale

$$\Pi_X(s) = (1 - p + ps)^k.$$

Para encontrar a PGF da soma de n v.a.s independentes[2], podemos fazer

$$\Pi_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^{n}\Pi_{X_i}(s),$$

[2]Note que nada foi dito sobre identidade de distribuição.

o que nos leva a concluir que

$$\begin{aligned}\Pi_T(s) &= \prod_{i=1}^{n}(1-\theta+\theta s)^N,\\ &= (1-\theta+\theta s)^{nN},\end{aligned}$$

o que sugere que T tem distribuição binomial com parâmetros nN e θ. Assim, precisamos verificar que para g mensurável, temos que

$$E_\theta[g(T)] = \sum_{t=0}^{nN} g(t)\binom{nN}{t}(1-\theta)^{nN}\left(\frac{\theta}{1-\theta}\right)^t = 0 \tag{1}$$

implica que $g(t) = 0$ para $t = 0, 1, \ldots, nN$. Como visto em aula em algumas oportunidades, a soma em (1) é um polinômio em $\theta \in (0,1)$, o que implica que para que esta expressão seja indenticamente zero para todo θ os coeficientes do polinômio em questão precisam ser zero, o que ocorre precisamente quando $g(t) = 0$ para $t \in \{0, 1, \ldots, nN\}$, que é o conjunto de medida 1 relevante. Todo esse desenvolvimento poderia ser pulado se notássemos que a família binomial é uma família exponencial com espaço paramétric que contém um intervalo aberto de $\mathbb{R}^s$. Isto significa que T é suficiente e completa – ver Teorema 2.74 de Schervish (1995). Respondida a questão b), vamos agora melhorar δ_1 usando T. Isso vai ser legal, já que como δ_1 é não-viesado, vamos obter δ_2 também não-viesado! Está claro que queremos

$$\delta_2(\boldsymbol{X}_n) = E_\theta[\delta_1 \mid T] = E[\delta_1 \mid T],$$

onde a última igualdade segue do fato de que T é suficiente. Agora, procedemos aos cálculos, tomando $T(\boldsymbol{X}_n) = t$,

$$\begin{aligned}\delta_2(\boldsymbol{X}_n) &= E[\delta_1 \mid T = t] = \Pr(X_1 = 1 \mid T = t),\\ &= \frac{P_\theta(X_1 = 1, T = t)}{P_\theta(T = t)},\\ &= \frac{P_\theta(X_1 = 1)P_\theta(\sum_{i=2}^n X_i = t-1)}{P_\theta(T = t)},\\ &= \frac{\left\{\binom{N}{1}\theta^1(1-\theta)^{N-1}\right\}\left\{\binom{(n-1)N}{t-1}\theta^{t-1}(1-\theta)^{(n-1)N-t+1}\right\}}{\binom{nN}{t}\theta^t(1-\theta)^{nN-t}}.\end{aligned}$$

Depois de alguma álgebra, encontramos

$$\delta_2(\boldsymbol{X}_n) = \frac{N\binom{(n-1)N}{T-1}}{\binom{nN}{T}},$$

que naturalmente não depende de θ, já que T é suficiente para θ. A Figura 2 mostra a distribuição de δ_1 e δ_2. Para terminar de responder à questão d), notamos que como T é completa e suficiente, o Teorema de Lehmann-Scheffé não só garante que δ_2 é o ENVVUM como também que é único!

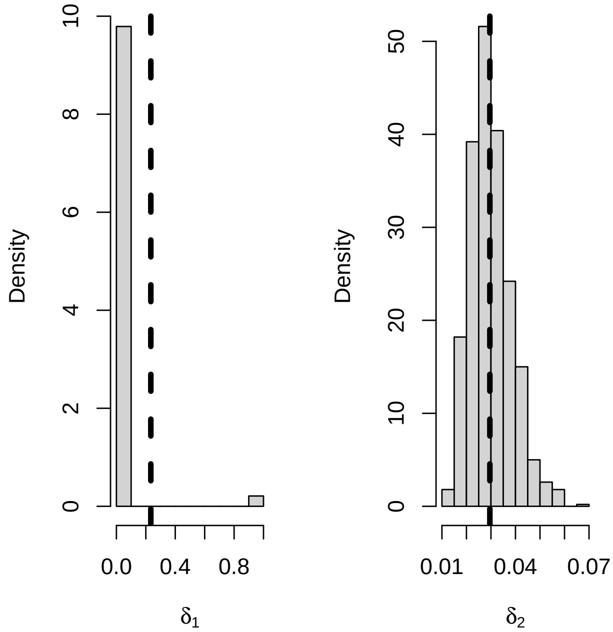

Figura 2: **Distribuição dos estimadores δ_1 e δ_2 no exemplo da binomial**. Aqui $N = 20$, $n = 50$ e $\theta_0 = .234$, o que dá $g(\theta_0) \approx 0.0295$. Resultados obtidos com cem mil replicatas de Monte Carlo.

Discussão: Nesta questão vimos como melhorar um estimador não-viesado de uma quantidade complicada através do mecanismo de Rao-Blackwell. Utilizando Lehmann-Scheffé, conseguimos fazer uma afirmação interessante sobre as propriedades do estimador melhorado, δ_2. Agora, uma observação adicional: note que sem acesso a Lehmann-Scheffé, poderíamos tentar demonstrar que[3]

$$\text{Var}_\theta(\delta_2) = \frac{[g'(\theta)]^2}{nI(\theta)},$$

isto é, que δ_2 atinge a cota inferior de Cramér-Rao (CLRB). Além de parecer um problema difícil de resolver em forma fechada, há também o fato de que a variância de δ_2 é proxima da CLRB, mas não parece atingí-la (ver Tabela 1). Isso acontece porque um ENVVUM não necessariamente atinge a cota inferior de Cramér-Rao. E este parece ser o caso aqui. ■

[3] Aqui nós "corrigimos" a expressão para levar em conta que estamos estimando $g(\theta)$ e não θ.

Tabela 1: **Simulação da Binomial com** $N = 20$ **e** $n = 50$. Resultados obtidos com cem mil replicatas de Monte Carlo.

θ_0	$\text{Var}_\theta(\delta_1)$	$\text{Var}_\theta(\delta_2)$	Cramér-Rao
0.1	1.97E-01	8.06E-04	8.11E-04
0.2	5.38E-02	1.90E-04	1.86E-04
0.3	7.26E-03	5.84E-06	5.56E-06
0.4	4.29E-04	5.26E-08	4.85E-08
0.5	1.99E-05	1.36E-10	1.17E-10
0.6	0.0	6.85E-14	5.48E-14
0.7	0.0	2.94E-18	2.13E-18
0.8	0.0	1.96E-24	9.89E-25
0.9	0.0	4.84E-35	1.04E-35

3. Exponentially efficient.

Dizermos que uma família de distribuições de probabilidade forma uma família exponencial regular de dimensão s se:

1. O suporte da distribuição, $E = \{x : f_\theta(x) > 0\}$, não depende de θ;
2. O espaço paramétrico Θ é um subconjunto aberto de $\mathbb{R}^s$;
3. Para todo vetor $a = (a_1, \ldots, a_s)$ diferente de $\mathbf{0} = (0, \ldots, 0)$ vale

$$P_\theta \left(a \cdot T(X) = a_s\right) < 1,$$

   para todo θ, ou seja, as estatísticas suficientes não são linearmente dependentes com probabilidade positiva.

Suponha que $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ formam uma amostra aleatória de uma distribuição P_θ pertencente a uma família exponencial regular de dimensão $s = 1$ parametrizada em função da média, isto é, $\theta = E[X]$.

a) (10 pontos) Encontre uma forma funcional para o estimador de máxima verossimilhança de θ em função da estatística suficiente, $T(\boldsymbol{X}_n)$. Discuta as condições para que este estimador seja único.

b) (10 pontos) Dizemos que um estimador δ de $g(\theta)$ é *eficiente* se a sua variância é igual à cota inferior de Cramér-Rao para todo $\theta \in \Theta$. Discuta se o estimador obtido no item anterior é eficiente.

   Dica: Compute a informação de Fisher, $I_X(\theta)$, e lembre-se de ajeitar a parametrização.

Conceitos trabalhados: suficiência; família exponencial; reparametrização; Informação de Fisher.
Nível de dificuldade: médio.
Resolução: A discussão aqui vai seguir de perto o começo da seção 8.3 de Keener (2010). Primeiro, vamos expressar a densidade de P_θ na forma canônica,

$$f_\eta(x) = h(x) \exp\left[\eta T(x) - A(\eta)\right],$$

e, portanto, a função *score* é

$$\begin{aligned}\kappa(\eta) &= \frac{\partial}{\partial \eta}\left\{\eta T(x) - A(\eta)\right\},\\ &= \frac{\partial}{\partial \eta}\eta T(x) - \frac{\partial}{\partial \eta}A(\eta).\end{aligned}$$

Agora, note que para uma amostra aleatória, temos

$$\begin{aligned}\kappa_n(\eta) &= \frac{\partial}{\partial \eta}\eta\sum_{i=1}^{n} T(x_i) - \frac{\partial}{\partial \eta}nA(\eta),\\ &= \sum_{i=1}^{n} T(x_i) - nE_\eta[T].\end{aligned}$$

Fazendo $E_\eta[T] = m(\eta)$, vemos que

$$m(\hat{\eta}) = \frac{1}{n}\sum_{i=1}^{n} T(x_i) =: \bar{T}_n.$$

Supondo que conseguimos inverter m, temos

$$\hat{\eta} = m^{-1}(\bar{T}_n).$$

Assim, concluimos que $\hat{\theta}_{\text{EMV}} = \bar{T}_n$. Ou seja, numa família exponencial regular, o estimador de máxima verossimilhança da média é a média amostral da estatística suficiente aplicada na amostra. Note que

$$\frac{\partial}{\partial \eta}\kappa_n(\eta) = -\frac{\partial^2}{\partial \eta^2}nA(\eta) = -n\operatorname{Var}_\eta(T),$$

que é negativo quase certamente. Dessa forma, dada a convexidade do espaço paramétrico natural, sabemos que $\hat{\eta}$ é único, e portanto $\hat{\theta}_{\text{EMV}}$ também o é. Sabemos também que $\hat{\theta}_{\text{EMV}}$ é não-viesado, já que

$$\begin{aligned}E_\eta[\hat{\theta}_{\text{EMV}}] &= \frac{1}{n}\sum_{i=1}^{n} E_\eta[T(X_i)],\\ &= \frac{1}{n}n\theta = \theta.\end{aligned}$$

Promissor no que toca à chance de ser eficiente! Agora vamos verificar se este estimador atinge a cota inferior de Cramér-Rao. Para isso, vamos seguir a dica e calcular a informação de Fisher. Primeiro, a forma canônica:

$$\begin{aligned}I_n(\eta) &= \operatorname{Var}_\eta(\kappa_n(\eta)),\\ &= E_\theta[\{\kappa_n(\eta)\}^2],\\ &= E_\theta\left[\left\{\sum_{i=1}^{n} T(x_i) - n\theta\right\}^2\right],\\ &= n\operatorname{Var}_\eta[T(X)]\end{aligned}$$

Onde a segunda linha segue do fato de que a função *score* tem esperança zero e quarta envolve abrir o quadrado e agrupar, lembrando que θ é a média de X. Agora, pela regra da cadeia, temos

$$\begin{aligned} I_n(\theta) &= \left[\frac{d}{d\theta}h^{-1}(\theta)\right]^2 I_n(h^{-1}(\theta)), \\ &= \frac{1}{n\,\mathrm{Var}_\theta[T(X)]}, \end{aligned}$$

Onde a segunda linha segue do fato de que

$$\mathrm{Var}_\theta(T(X)) = \frac{\frac{\partial^2}{\partial\theta^2}A(\eta(\theta))}{[\eta'(\theta)]^2} - \frac{\frac{\partial}{\partial\theta}A(\eta(\theta))\eta''(\theta)}{[\eta'(\theta)]^3}.$$

Mas $n\,\mathrm{Var}_\eta[T(X)]$ é a variância de $\hat{\theta}_{\mathrm{EMV}}$ e portanto concluímos que ele atinge a CRLB e portanto é eficiente!

Discussão: nesta questão exploramos algumas propriedades adicionais da família exponencial, em particular a parametrização de média de famílias regulares. Vimos que há uma forma funcional simples para o EMV da média. Durante toda a derivação era preciso saber bem as identidades diferenciais da família exponencial e as propriedades da informação de Fisher. ■

4. (Bônus) I'm never enough for you!

A mediana amostral é definida como

$$m(\boldsymbol{X}_n) = \begin{cases} X_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar,} \\ \frac{X_{(n/2)}+X_{(n/2+1)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Tome $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma distribuição qualquer, P_θ.

a) (20 pontos) Mostre que para $n > 2$, $m(\boldsymbol{X}_n)$ **nunca** é estatística suficiente.

Conceitos trabalhados: estatísticas de ordem; suficiência; Pitman-Koopman-Darmois; família exponencial.
Nível de dificuldade: difícil.
Resolução: Quando $n = 1, 2$, média e mediana coincidem, e, portanto, a afirmação não pode valer, visto que sabemos de várias situações onde a média é, de fato, suficiente. Para $n \geq 3$, vamos começar com o caso mais difícil, que é quando o suporte de P_θ depende de θ. Nesta situação, podemos escrever

$$f_\theta(x) = h(x)t(\theta)\mathbb{I}(x \in A_\theta),$$

onde $A_\theta \in \mathcal{X}$ é o suporte e t é uma função só de $\theta \in \Theta$. A densidade conjunta da amostra vale, portanto,

$$\begin{aligned} f_\theta(\boldsymbol{x}_n) &= \prod_{i=1}^n h(x_i)[t(\theta)]^n \prod_{i=1}^n \mathbb{I}(x_i \in A_\theta), \\ &= h_n(\boldsymbol{x}_n)[t(\theta)]^n \mathbb{I}(x_i \in A_\theta). \end{aligned}$$

Pelo Teorema da Fatorização, temos que

$$\mathbb{I}\left(m(\boldsymbol{x}_n) \in B_\theta^n\right) = \mathbb{I}\left(m(\boldsymbol{x}_n) \in B_\theta^n\right),$$

onde $B_\theta^n \subseteq \mathcal{X}$ é justamente o conjunto cuja indicatriz coincide com $\prod_{i=1}^n \mathbb{I}(x_i \in A_\theta)$ para as medianas de todas as amostras $\boldsymbol{x}_n$ possíveis. Note que se a mediana é de fato suficiente, $\mathbb{I}\left(m(\boldsymbol{x}_n) \in B_\theta^n\right)$ depende de θ apenas através de B_θ^n. Considere adicionar uma nova observação x_{n+1} à amostra e note que sempre é possível fazê-lo de modo a deixar $m(\boldsymbol{x}_n)$ intacta. Claramente,

$$\mathbb{I}\left(m(\boldsymbol{x}_{n+1}) \in B_\theta^{n+1}\right) = \mathbb{I}\left(m(\boldsymbol{x}_n) \in B_\theta^n\right) \cdot \mathbb{I}(x_{n+1} \in A_\theta),$$

mas x_{n+1} pode estar dentro ou fora do suporte, o que leva a uma contradição! Suponha agora que $E = \{x : f_\theta(x) > 0\}$ é independente de θ. Neste caso, podemos invocar o Teorema de Pitman-Koopman-Darmois (Darmois, 1935; Pitman, 1936; Koopman, 1936), que diz que se uma família de distribuições de probabilidade tem suporte independente de θ **e** tem estatística suficiente de dimensão finita, então esta é uma família exponencial. Ora, se P_θ pertence à família exponencial, a densidade conjunta de $\boldsymbol{X}_n = \boldsymbol{x}_n$ pode ser escrita como

$$\begin{aligned} f_n(\boldsymbol{x}_n) = \prod_{i=1}^n f_\theta(x_i) &= \prod_{i=1}^n h(x_i) \exp\left[\sum_{j=1}^s \eta_j(\theta)T_j(x_i) - B(\theta)\right], \\ &= \exp\left[\sum_{j=1}^s \eta_j(\theta) \sum_{i=1}^n T_j(x_i) - nB(\theta)\right] \prod_{i=1}^n h(x_i), \end{aligned}$$

o que sugere que $T(\boldsymbol{X}_n) = (\sum_{i=1}^n T_1(x_i), \ldots, \sum_{i=1}^n T_s(x_i))$ é estatística suficiente **mínima** para θ. Sendo assim, a mediana deveria ser função de $T(\boldsymbol{X}_n)$, mas modificar um dos extremos em $x_{(1)}, \ldots, x_{(n)}$ certamente muda $T(\boldsymbol{x}_n)$, mas não muda $m(\boldsymbol{x}_n)$, o que é um absurdo. Concluímos, assim, que a mediana também não pode ser suficiente neste caso. $\blacksquare$

Bibliografia

Casella, G. and Berger, R. (2002). *Statistical inference*. Duxbury advanced series, second edition.

Darmois, G. (1935). Sur les lois de probabilitéa estimation exhaustive. *CR Acad. Sci. Paris*, 260(1265):85.

Keener, R. W. (2010). *Theoretical Statistics: Topics for a core course*. Springer.

Koopman, B. O. (1936). On distributions admitting a sufficient statistic. *Transactions of the American Mathematical society*, 39(3):399–409.

Lo, A. (2018). Demystifying the integrated tail probability expectation formula. *The American Statistician*.

Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. In *Mathematical Proceedings of the cambridge Philosophical society*, volume 32, pages 567–579. Cambridge University Press.

Schervish, M. J. (1995). *Theory of Statistics*. Springer Science & Business Media.