Primeira avaliação (A1)

Disciplina: Inferência Estatística (MSc)
Instrutor: Luiz Max Carvalho
Monitor: Isaque Pim

02 de Outubro de 2023.

- 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;
- (Mestrado) A prova vale 100 pontos. A pontuação restante é contada como bônus;
- (Mestrado) Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;

1. Miss Independent

Seja $X_1, \dots, X_n$ variáveis aleatórias i.i.d absolutamente contínuas com densidade comum

$$f_\theta(x) = \theta e^{-\theta x}\mathbb{I}\{x > 0\},$$

onde $\theta > 0$ é desconhecido.

a) (5 pontos) Encontre a densidade de θX_i;

b) (15 pontos) Seja, $X_{(1)} \leq \cdots \leq \dots X_{(n)}$ as estatísticas de ordem e $\bar{X} = (X_1 + \cdots + X_n)/n$ a média amostral. Mostre que $\bar{X}$ e $X_{(1)}/X_{(n)}$ são independentes.

Conceitos trabalhados: suficiência mínima; ancilaridade; Teorema de Basu.
Nível de dificuldade: fácil.
Resolução: A densidade de $Y_i = h(X_i) = \theta X_i$ é $f_Y(y) = f_\theta(y/\theta)\left|\frac{\partial}{\partial y}h(y)\right| = f_\theta(y/\theta)/\theta$:

$$f_Y(y) = \exp(-y)\mathbb{I}\{y > 0\},$$

o que sugere que $Y_i \sim$ exponencial(1) e que Y_i é quantidade pivotal[1]. Primeiro, vamos mostrar que $\bar{X}_n$ é completa. Para isso, vamos estudar $S_n = \sum_{i=1}^n X_i$ e notar $\bar{X}_n$ é transformação bijetora de S_n. Primeiro, vamos encontrar a distribuição de S_n através da sua função geradora de momentos.

$$\begin{aligned} M_S(t) &= \prod_{i=1}^n \left\{\frac{1}{1 - t/\theta}\right\}, \\ &= \left(\frac{1}{1 - t/\theta}\right)^n, \end{aligned}$$

que é a MGF de uma Erlang[2] com parâmetros n e θ, cuja densidade[3] vale:

$$f_S(s) = \frac{\theta^n}{(n-1)!}s^{n-1}\exp(-\theta s)\mathbb{I}\{s > 0\}.$$

Agora, vamos considerar uma função mensurável $g : \mathbb{R}_+ \to \mathbb{R}$ de modo que

$$E_\theta[g(S_n)] = \int_0^\infty g(s) f_S(s)\,ds = 0,$$

de tal sorte que

$$\int_0^\infty g(s)s^{n-1}\exp(-\theta s)\,ds = 0. \tag{1}$$

Daqui, podemos fazer dois argumentos:

[1] O conceito será visto em detalhes mais à frente no curso.
[2] Agner Krarup Erlang (1878-1929) foi um matemático, estatístico e engenheiro dinamarquês. E sim, isso é mesmo que dizer que S_n tem distribuição Gama com forma n e taxa θ.
[3] Com respeito a Lebesgue em $\mathcal{B}(\mathbb{R})$.

- Note que (1) parece bastante a transformada de Laplace de g, $\mathcal{L}\{g\}(\theta)$; particular, se (1) vale, vale também que $\mathcal{L}\{g\}(\theta) = 0$ para todo $\theta \in \Omega$. Mas sabemos que $\mathcal{L}\{f\}(t) = 0$ para todo t se e somente se $f(x) = 0$ para todo x. É possível construir um argumento similar usando a transformada de Mellin[4].

- Um argumento mais estatístico é notar que a Erlang pertence à família exponencial e que o espaço paramétrico é um aberto de $\mathbb{R}$, o que implica que $T(s) = s$ é completa[5] – ver Teorema 3.19 de Keener (2010).

De posse da informação de que S_n é completa, concluímos que $\bar{X}_n$ é completa. Ademais, $\frac{X_{(1)}}{X_{(n)}} = \frac{\theta X_{(1)}}{\theta X_{(n)}}$ é ancilar, já que é a razão entre duas variáveis aleatórias cuja distribuição não depende de θ. Pelo Teorema de Basu (Keener, 2010, Teorema 3.2.1), temos que $\bar{X}_n \perp\!\!\!\perp \frac{X_{(1)}}{X_{(n)}}$. ■

Comentário: Uma questão bem simples, trabalhando os conceitos de forma bem transparente. O item a) é para ajudar a ver que a razão do mínimo e do máximo precisa ser ancilar. O resto foi trabalhado em aula.

2. You can't truncate sufficiency

Tome $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição normal com média $\mu \in \mathbb{R}$ e variância $\sigma^2 \in \mathbb{R}_+$. Sabemos que média amostral $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ e a variância amostral $\bar{V}_n = (n-1)^{-1} \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ são estatísticas suficientes para a distribuição normal com $\theta = (\mu, \sigma^2)$.

Em algumas situações, essas observações só podem acontecer dentro de uma região $A \in \mathcal{B}$. Por regularidade, assuma que $k(\theta) = P_\theta(X \in A) > 0$. Nesse caso, a distribuição apropriada para a nova variável aleatória restrita X^* é a distribuição condicional

$$P_\theta(X^* \in B) = P_\theta(X \in B | X \in A), \quad B \in \mathcal{B}.$$

Está distribuição é chamada de *truncamento* de X em um conjunto A.

a) (10 pontos) Mostre que a média e a variância amostrais continuam sendo estatísticas suficientes para uma distribuição normal truncada em um intervalo $A = (a, b)$ com $-\infty \le a < b \le \infty$.

b) Seja $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ uma família de distribuições dominada por uma medida σ-finita μ. Seja A um boreliano. Considere P_θ^* o truncamento de P_θ em A e $\mathcal{P}^* = \{P_\theta^* : \theta \in \Omega\}$. Mostre que:

   - (10 pontos)* Se T é suficiente para $\mathcal{P}$, então é suficiente para $\mathcal{P}^*$;
   - (10 pontos)* Se T é completa para $\mathcal{P}$, então é completa para $\mathcal{P}^*$.

Conceitos trabalhados: suficiência; completude.
Nível de dificuldade: Médio.
Resolução:

[4] `https://hal.science/hal-03152634v1/file/MELLIN.pdf`
[5] Com n conhecido e θ desconhecido, é bom que se diga.

a) Seja $f(x|\mu, \sigma^2)$ a densidade da normal de média μ e variância σ^2 e $F(x|\mu, \sigma^2)$ sua função de distribuição acumulada. Seja $f^*(x|\mu, \sigma^2)$ a densidade do truncamento desta normal. Para simplificar a notação tome $\theta = (\mu, \sigma^2)$. É de amplo conhecimento que a normal truncada em (a, b) possui densidade igual a

$$f^*_\theta(x) = \frac{f_\theta(x)}{F_\theta(b) - F_\theta(a)}, \quad a < x < b,$$

e vale 0 fora do do intervalo. Como a média e a variância amostral são suficientes para a normal não truncada, basta usar a mesma fatoração dada pelo teorema da fatoração e ver que ambas continuam sendo suficientes para a normal truncada (a densidade é a mesma a menos de uma constante e de uma variação no domínio que independe do parâmetro).

b) Agora em um esquema mais geral temos que

$$P_\theta(E|A) = \frac{P_\theta(E \cap A)}{P_\theta(A)}$$

O denominador é uma constante em θ:

$$k(\theta) = \int_A dP_\theta.$$

Considere a indicadora do conjunto A $\mathbb{1}_A$. Por Radon-Nikodym temos

$$\frac{dP^*}{d\mu} = \frac{dP^*}{dP}\frac{dP}{d\mu} = \frac{1}{k(\theta)}\mathbb{1}_A f_\theta(x)$$

Note que $P_\theta(E \cap A) \leq P_\theta(E)$, então o truncamento mantém a medida de probabilidade dominada por μ. Por fim, a dependência no dado da densidade se manteve, e o teorema da fatorização vai garantir que T continuará suficiente. Para a completude, se T é completa para a distribuição não truncada, então

$$\mathbb{E}[\mathbb{1}_A g(T)] = 0 \iff \mathbb{1}_A g(T) = 0,$$

e pelo resultado anterior é fácil ver que na distribuição truncada

$$\mathbb{E}^*[g(T)] = \frac{1}{k(\theta)}\mathbb{E}[\mathbb{1}_A g(T)].$$

Fica claro então que a completude também se preserva.

■

Comentário: Não é necessário que a família seja dominada para que os resultados desta questão valham. A família ser dominada facilita no sentido de nos dar uma densidade de probabilidade e possibilitar invocar o teorema da fatorização. Você conseguiria mostrar os resultados dessa questão sem assumir a dominação da família?

3. For the non-canonical folk (20 pontos)

Para uma família exponencial no formato canônico, mostramos em aula que $\mathbb{E}[T_j] = \partial A(\nu)/\partial \nu_j$. Isso pode ser escrito em formato vetorial como $\mathbb{E}[T] = \nabla A(\nu)$. Deduza uma fórmula diferencial análoga para $\mathbb{E}_\theta[T]$ para uma família exponencial s-paramétrica que não está em formato canônico. Assuma que o novo espaço paramétrico, Ω, tem dimensão s.

Dica: Diferenciar sobre o sinal da integral resulta em um sistema linear de equações. Escreva essas equações em formato matricial.

Conceitos trabalhados: família exponencial .

Nível de dificuldade: Fácil .

Resolução: Para uma família exponencial s-paramétrica em formato canônico temos a seguinte identidade

$$e^{A(\nu)} = \int \exp(\langle \nu, T(x)\rangle) \cdot h(x) d\mu(x).$$

Aqui μ é uma medida em $\mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}$ uma função mensurável não-negativa, $T = (T_1, \ldots T_s)$ um vetor de funções mensuráveis do $\mathbb{R}^n$ em $\mathbb{R}$ e $\nu \in N \subseteq \mathbb{R}^s$ parâmetros de um espaço paramétrico adequado. Reparametrizando a distribuição substituindo ν por uma função $\nu : \Omega \to N$, de um novo espaço de parâmetros $\Omega \subset \mathbb{R}^s$ obtemos a nova identidade

$$e^{B(\theta)} = \int \exp(\langle \nu(\theta), T(x)\rangle) \cdot h(x) d\mu(x),$$

onde $B(\theta) = A(\nu(\theta))$. Agora, seguindo a dica e diferenciando a identidade sob a integral com respeito a θ obtemos

$$e^{B(\theta)} \frac{\partial B(\theta)}{\partial \theta_i} = \int \sum_{j=1}^{s} \frac{\partial \nu_j(\theta)}{\partial \theta_i} T_j(x) \exp(\langle \nu(\theta), T(x)\rangle) \cdot h(x) d\mu(x).$$

Podemos manipular os fatores que não dependem de x na integração. Dividindo ambos os lados por $e^{B(\theta)}$ e movendo o somatório obtemos

$$\frac{\partial B(\theta)}{\partial \theta_i} = \sum_{j=1}^{s} \frac{\partial \nu_j(\theta)}{\partial \theta_i} \mathbb{E}_\theta[T_j(x)].$$

Podemos compactar a notação para

$$\nabla B(\theta) = (D\nu(\theta))^T \cdot \mathbb{E}_\theta[T(x)]$$

Supondo a invertibilidade de $D\nu(\theta)$, obtemos a fórmula diferencial para $\mathbb{E}_\theta[T(x)]$ como sendo

$$\mathbb{E}_\theta[T(x)] = \left(D\nu(\theta)^T\right)^{-1} \nabla B(\theta).$$

■

Comentário: Note que com uma regra da cadeia multidimensional o resultado também sairia, e quase que de imediato.

4. Pfizer Inc.

Modelos de *resposta à dosagem* são de grande interesse da indústria farmacêutica. Uma possível aplicação seria na área de avaliação de risco de substâncias tóxicas, onde o realizador do experimento está interessado em estimar a probabilidade de alguma resposta (e.g, aparição de um tumor) em função da dosagem de toxina[6] aplicada. Nesta questão vamos estudar as propriedades de suficiência mínima destes modelos.

Para isso, considere que vamos aplicar diferentes volumes de toxinas em ratos. $0 < d_1, < d_2 < \cdots < d_m$ foram as dosagens escolhidas para este estudo. Para cada dosagem, foi escolhida uma população de n_i ratos. Os n_i's são fixos. Seja $x = (x_1, \ldots, x_m)$ o número de respostas à aplicação das doses, respectivamente. Isto é, para o i-ésimo grupo, de n_i ratos, x_i tiveram a reação desejada. Seja $p_\theta(d)$ a probabilidade de resposta a uma dose $d > 0$, onde θ é p-dimensional com $(p \leq m)$ em um espaço de parâmetros Θ. A distribuição de probabilidade de x é dada por

$$f_M(x|\theta) = \prod_{i=1}^{m} \binom{n_i}{x_i} p_\theta(d_i)^{x_i} [1 - p_\theta(d_i)]^{n_i - x_i},$$

onde M representa uma escolha particular de modelo de resposta à dose (podemos escolher diferentes funções para modelar p_θ). Seja $\mathcal{F}_M$ a família de distribuições induzida por M.

a) (5 pontos) Mostre que $T(x) = x$ é suficiente para $\mathcal{F}_M$ para todo M.

b) (15 pontos) Mostre o seguinte resultado preliminar. Seja $\mathcal{P}$ uma família finita de distribuições com densidades p_i, $i = 0, 1, \ldots, k$, todas tendo o mesmo suporte. Então, a estatística

$$T(X) = \left(\frac{p_1(X)}{p_0(X)}, \frac{p_1(X)}{p_0(X)}, \ldots, \frac{p_k(X)}{p_0(X)} \right)$$

é suficiente mínima.

Dica: Mostre primeiro que se U é uma estatística suficiente para $\mathcal{P}$, então a razão $\frac{p_i(x)}{p_j(x)}$ é uma função apenas de $U(x)$.

c) (10 pontos) Seja A uma matriz $m \times m$ com seus elementos definidos por

$$A_{i,j} = \ln \left(\frac{p_{\theta_j}(d_i)[1 - p_{\theta_0}(d_i)]}{p_{\theta_0}(d_i)[1 - p_{\theta_j}(d_i)]} \right).$$

Mostre que se existirem vetores $(\theta_0, \theta_1, \ldots, \theta_m)$ tais que a matriz A é invertível, então $T(x) = x$ é suficiente mínima.

Conceitos trabalhados: suficiência; suficiência mínima .

Nível de dificuldade: Médio.

Resolução:

[6] Why did the statistician bring a toxin to the data analysis meeting? Because they wanted to add a little "poisonal" touch to their regression analysis...

a) Imediato pelo Teorema da Fatorização. A função de densidade já é função de x.

b) Tome o par de densidade p_i e p_j. Seguindo a dica, se U é suficiente para a família $\mathcal{P}$, então pelo teorema da fatorização existem funções g_i, g_j e h mensuráveis tais que

$$\begin{aligned} p_i(x) &= g_i(U(x))h(x), \\ p_j(x) &= g_j(U(x))h(x). \end{aligned}$$

Tomando a razão de ambas as equações obtemos que p_i/p_j depende de x pela estatística suficiente $U(x)$. Sabendo deste resultado, fica claro que a estatística $T(X)$ do enunciado é função de qualquer estatística suficiente por ser um vetor de razões das densidades em $\mathcal{P}$. Basta então mostrar que $T(X)$ é suficiente e ela será suficiente mínima por definição. Para cada i temos que

$$p_i(x) = (T(x))_i \cdot p_0(x).$$

Aqui p_0 está atuando como a função livre de parâmetro fixa. Para recuperar a densidade zero podemos definir a função depende do parâmetro para ser 1 em $i = 0$. $T(X)$ então é suficiente, e por ser função de toda estatística suficiente é suficiente mínima.

c) Se A é invertível, então A é uma aplicação linear bijetiva. Sabemos que suficiência mínima é preservada por bijeções. Defina $T(x)$ como no item anterior para as densidades $p_i = f_M(x|\theta_i)$. Como estamos num espaço finito de densidades, $T(x)$ é suficiente mínima. Vamos agora analisar a i-ésima linha do produto Ax.

$$\begin{aligned} (Ax)_i &= \sum_{j=1}^{m} \ln\left(\frac{p_{\theta_j}(d_i)[1-p_{\theta_0}(d_i)]}{p_{\theta_0}(d_i)[1-p_{\theta_j}(d_i)]}\right) \cdot x_j, \\ &= \sum_{j=1}^{m} \ln\left(\frac{p_{\theta_j}^{x_j}(d_i)[1-p_{\theta_0}(d_i)]^{x_j}}{p_{\theta_0}^{x_j}(d_i)[1-p_{\theta_j}(d_i)]^{x_j}}\right), \\ &= \ln\left(\prod_{j=1}^{m} \frac{p_{\theta_j}^{x_j}(d_i)[1-p_{\theta_0}(d_i)]^{x_j}}{p_{\theta_0}^{x_j}(d_i)[1-p_{\theta_j}(d_i)]^{x_j}}\right), \\ &= \ln\left(\prod_{j=1}^{m} \frac{p_{\theta_j}^{x_j}(d_i)[1-p_{\theta_j}(d_i)]^{-x_j}}{p_{\theta_0}^{x_j}(d_i)[1-p_{\theta_0}(d_i)]^{-x_j}}\right), \\ &= \ln\left(\frac{f_M(x|\theta_j)}{f_M(x|\theta_0)}\right). \end{aligned}$$

como o logaritmo é bijetivo no domínio em questão e a composição de funções bijetivas é bijetiva, fica claro que x é estatística minimamente suficiente pois é uma bijeção da estatística minimamente suficiente encontrada no item anterior.

■

Comentário: A questão foi inspirada em Messig and Strawderman (1993). Em particular, o artigo em questão ilustra a existência de uma estatística completa suficiente para uma família exponencial curvada (Keener, 2010, Cap 5), em particular uma estatística que é de dimensão maior que o parâmetro. Um ponto interessante a se aprender na questão é sobre a definição de estatística suficiente mínima. A dica dada no item b) induz a conclusão rápida de que $T(X)$ é função de toda estatística suficiente, mas até mesmo constantes satisfazem isso. Deve ser mostrado também que $T(X)$ é de fato uma estatística suficiente.

Bibliografia

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

Messig, M. A. and Strawderman, W. E. (1993). Minimal sufficiency and completeness for dichotomous quantal response models. *The Annals of Statistics*, 21(4):2149–2157.