Trabalho III: o comportamento assintótico de estimadores eficientes.

Disciplina: Inferência Estatística
Professor: Luiz Max de Carvalho

17 de Setembro de 2021

Data de Entrega: 13 de Outubro de 2021.

Orientações

- Enuncie e prove (ou indique onde se pode encontrar a demonstração) de todos os resultados não triviais necessários aos argumentos apresentados;
- Lembre-se de adicionar corretamente as referências bibliográficas que utilizar e referenciá-las no texto;
- Equações e outras expressões matemáticas também recebem pontuação;
- Você pode utilizar figuras, tabelas e diagramas para melhor ilustrar suas respostas;
- Indique com precisão os números de versão para quaisquer software ou linguagem de programação que venha a utilizar para responder às questões[1];
- Este trabalho é longo. Sugiro fortemente começar a fazer assim que possível.

Introdução

Como aprendemos até agora, existem vários critérios de otimalidade para a construção e avaliação de estimadores. Um conceito fundamental é o de variância mínima, ou eficiência, uma propriedade de estimadores não-viesados.

Na vida real, no entanto, mesmo um estimador viesado ou ineficiente pode ser útil. Um dos aspectos que buscamos estudar é o comportamento *assintótico* de estimadores, isto é, o que acontece quando o tamanho de amostra, n, tende ao infinito. No que se segue, vamos estudar alguns resultados interessantes sobre o comportamento assintótico de estimadores e, utilizando simulações, investigar o seu comportamento empírico.

[1]Não precisa detalhar o que foi usado para preparar o documento com a respostas. Recomendo a utilização do ambiente LaTeX, mas fique à vontade para utilizar outras ferramentas.

Questões

Parte I: lidando com estimadores eficientes

1. Considere um modelo estatístico paramétrico $f(x \mid \theta)$ com f duas vezes diferenciável com respeito a θ e suporte independente de θ. Seja $\boldsymbol{X}$ uma amostra aleatória de tamanho n e $\delta(\boldsymbol{X})$ um estimador **eficiente** de $g(\theta)$. Defina $E[\delta(\boldsymbol{X})] = m(\theta)$ de modo que $m'(\theta) := \frac{d}{d\theta}m(\theta) \neq 0$ para todo $\theta \in \Omega$. Mostre que a distribuição de

$$\frac{\sqrt{nI(\theta)}}{m'(\theta)}\left[\delta(\boldsymbol{X}) - m(\theta)\right],$$

é normal padrão, onde $I(\theta)$ é a informação de Fisher.

Dica: Lembre-se do método Delta.

2. Seja $\boldsymbol{X}$ uma amostra aleatória de tamanho n de uma distribuição Poisson com taxa μ. Mostre que o EMV para μ é eficiente.

3. Tomando $\mu_0 = 0.5$, simule 100.000 conjuntos de dados de $n = 10$ observações com distribuição Poisson(μ_0). Para cada simulação $\boldsymbol{X}^{(m)}$, $m = 1, 2, \ldots, 10^5$, compute o EMV, $\hat{\mu}^{(m)}$. Agora, compute a fração de simulações para as quais $\hat{\mu}^{(m)} \leq 0.55$. Esta é uma estimativa da função de distribuição empírica[2] de $\hat{\mu}$, $\hat{F}(0.55)$. Compare $\hat{F}(0.55)$ com a aproximação assintótica derivada no item 1. Repita o experimento para $n = 30$ e $n = 100$. Para ajudar, aqui vai uma tabela a ser preenchida (utilize 3 dígitos de significância):

Tamanho de amostra (n)	CDF empírica	Aproximação normal
10	0.xxx	0.xxx
30	0.xxx	0.xxx
100	0.xxx	0.xxx

4. Discuta o quão boa a aproximação normal é, e se a qualidade melhora à medida que n cresce.

Parte II: condições menos que ideais

Agora vamos lidar com uma situação onde o estimador em questão não é eficiente. O EMV, por exemplo, nem sempre é eficiente, mas podemos enunciar um resultado parecido com o da seção anterior. Sob condições de regularidade, temos que

$$\sqrt{nI(\theta)}\left[\delta_{\text{EMV}} - \theta\right],$$

tem distribuição normal padrão, isto é, que o EMV é *assintoticamente* eficiente.

5. Tome $\boldsymbol{X}$ uma amostra aleatória de tamanho n de uma distribuição exponencial com taxa θ. Mostre que o EMV para θ é viesado e ineficiente.

[2]ECDF, na sigla em inglês.

6. Mostre que
$$\delta_{\text{EMV}}(\boldsymbol{X}) \sim \text{Gama-inversa}(n, n\theta).$$

7. Nesta situação, portanto, sabemos a função de distribuição do EMV exatamente. Vamos compará-la com a sua aproximação normal. Tomando $\theta = 2$, preencha a tabela a seguir:

Tamanho de amostra (n)	CDF exata	Aproximação normal
10	0.xxx	0.xxx
30	0.xxx	0.xxx
100	0.xxx	0.xxx

Dica: Se não quiser utilizar pacotes especializados para computar a CDF exata, não precisa. Basta lembrar que se $X \sim \text{Gama}(\alpha, \beta)$, então $Y = 1/X$ tem distribuição Gama-inversa(α, β), de modo que você consegue deduzir $\Pr(Y \leq y)$ a partir da função de distribuição de X, que está disponível em quase todos os pacotes estatísticos modernos.