

## Correções do Viés dos EMV via a Biblioteca *mle.tools*

André Felipe Berdusco Menezes (PIBIC/CNPq/FA/UEM), Josmar Mazucheli (Orientador), e-mail: andrefelipemaringa@gmail.com

Universidade Estadual de Maringá / Centro de Ciências Exatas / Departamento de Estatística / Maringá, PR

### Probabilidade e Estatística / Inferência Paramétrica

**Palavras-chave:** Correção de viés, Verossimilhança, Cox e Snell.

#### Resumo:

A estimação de parâmetros a partir do método da máxima verossimilhança seguramente é a técnica mais utilizada nas diversas áreas da estatística. Este método desfruta de muitas propriedades importantes. Todavia, algumas propriedades dependem essencialmente do tamanho da amostra. Devido a esta razão, existem na literatura diversas metodologias para a obtenção de estimativas próximas de serem não viesadas. Neste projeto são consideradas três metodologias usadas na correção/redução do vício dos estimadores de máxima verossimilhança. Neste relatório apresentamos uma revisão teórica dos três principais métodos de correção de vício dos estimadores de máxima verossimilhança. Além disso, exibimos a eficiência da biblioteca *mle.tools* para estimar o vício usando a fórmula de Cox-Snell e para calcular a informação de Fisher.

#### Introdução

Desde que foi proposto por Fisher em uma série de artigos durante o período de 1912 a 1934 o método da máxima verossimilhança tem sido empregado em diversos problemas no contexto da inferência estatística, muito devido as atraente propriedades que os estimadores de máxima verossimilhança (EMV) desfrutam. Por exemplo, os EMV são assintoticamente não viciados, eficientes, consistentes, invariante sob transformação e assintoticamente com distribuição Normal.

Contudo, nota-se que as propriedades estatísticas que fazem dos EMV atraentes, em sua maioria, são propriedades assintóticas, isto é, são válidas para grandes amostras. Assim, propriedades como a não tendenciosidade, podem não ser atendida para amostras pequenas ou moderadas. De fato, os EMV possuem um vício de ordem  $O(n^{-1})$ , em que  $n$  é tamanho da amostra, logo esse vício reduz a medida que a amostra aumenta.

Na literatura existem alguns métodos para correção de viés de segunda ordem. A abordagem mais usual para correção de viés foi proposta por Cox e Snell (1968) e consiste em subtrair os EMV da expressão analítica do viés obtida através de uma expansão de segunda de Taylor sob o vetor score. Uma abordagem preventiva para correção do vício dos EMV foi introduzida por Firth (1993) e consiste em uma

transformação sob o vetor escore. Além disso, outro método alternativo é a obtenção do(s) vício(s) através do método *Bootstrap* introduzido por Efron (1982).

Aplicando as abordagens mencionadas para correção de vício, em específico a metodologia corretiva de Cox-Snell, muitos pesquisadores derivaram estimadores quase não viciados para os parâmetros de várias distribuições de probabilidade. Neste projeto de iniciação científica apresentamos uma revisão teórica dos três principais métodos de correção de vício dos estimadores de máxima verossimilhança. Além disso, mostramos a eficiência da biblioteca *mle.tools* para estimar o vício usando a fórmula de Cox-Snell e para calcular a informação de Fisher esperada e observada.

### Materiais e métodos

Considere que  $l = l(\boldsymbol{\theta} | \mathbf{x})$  denote a função de log-verossimilhança de um vetor de parâmetros  $p$ -dimensional baseado em uma amostra aleatória  $\mathbf{x} = (x_1, \dots, x_n)$ .

Os cumulantes conjuntos das derivadas de  $l$  são dados por:

$$\mathbf{I}_{ij} = \mathbb{E} \left[ \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right], \quad \mathbf{I}_{ijl} = \mathbb{E} \left[ \frac{\partial^3 \ell}{\partial \theta_i \partial \theta_j \partial \theta_l} \right] \quad \text{e} \quad \mathbf{I}_{ij,l} = \mathbb{E} \left[ \left( \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right) \left( \frac{\partial \ell}{\partial \theta_l} \right) \right]$$

para  $i, j, l = 1, \dots, p$ . Cox e Snell (1968) mostraram que quando os dados são independentes, porém não necessariamente identicamente distribuídos, o vício do  $r$ -ésimo elemento do EMV de  $\boldsymbol{\theta}$ , pode ser expresso como

$$\mathcal{B}(\hat{\theta}_r) = \sum_{i=1}^p \sum_{j=1}^p \sum_{l=1}^p \mathbf{I}^{ri} \mathbf{I}^{jl} [0.5 \mathbf{I}_{ijl} + \mathbf{I}_{ij,l}] + \mathcal{O}(n^{-2}),$$

em que  $r = 1, \dots, p$ , e  $\mathbf{I}^{ij}$  denota o  $(i,j)$ -ésimo elemento da inversa da matriz informação de Fisher esperada.

Por outro lado, Firth (1993) introduziu uma correção que consiste em transformar o vetor escore  $\mathbf{U}(\boldsymbol{\theta})$  antes de obter as estimativas de máxima verossimilhança. A função escore modificada é definida como

$$\mathbf{U}^*(\boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta}) - \mathbf{I}(\boldsymbol{\theta}) \mathcal{B}(\hat{\boldsymbol{\theta}})$$

em que  $\mathbf{I}(\boldsymbol{\theta})$  é a matriz de informação de Fisher e  $\mathcal{B}(\hat{\boldsymbol{\theta}})$  é o vetor de vício de segunda ordem.

Uma abordagem alternativa a correção de vício analítica é baseada no método de reamostragem *Bootstrap* introduzido por Efron (1982). Os estimadores com correção de vício *Bootstrap* paramétrico (PBE, acrônimo em inglês) utilizam as EMV dos dados para gerar amostras pseudo aleatórias da distribuição, estimar o vício e depois subtrair o vício das estimativas obtidas.

O estimador com correção de viés *Bootstrap* é dado por:

$$\hat{\theta}_{PBE} = 2\hat{\theta} - \frac{1}{B} \sum_{j=1}^B \hat{\theta}_{(j)}.$$

## Resultados e Discussão

Importante ressaltar que dois trabalhos foram realizados durante o projeto. No primeiro apresentamos as funcionalidades e a eficiência da biblioteca *mle.tools* para obtenção do vício dos EMV, bem como para o cálculo da informação de Fisher esperada e observada. Já no segundo derivamos as correções analíticas de Cox e Snell (1968) para o vício dos parâmetros da distribuição unit-Gamma. Um estudo de simulação foi conduzido para avaliar a performance dos estimadores propostos. Já no segundo. Devido ao espaço reduzido iremos apresentar somente resultados parciais do primeiro trabalho.

A versão atual da biblioteca *mle.tools*, enviada ao CRAN em Fevereiro de 2017, tem implementada três funções --- *observed.varcov*, *expected.varcov* e *coxsnell.bc* --- que são de grande interesse na análise de dados sob o paradigma da verossimilhança. Essas função calculam respectivamente, a informação de Fisher observada, a informação de Fisher esperada e EMV com correção de vício utilizando a fórmula apresentada de Cox e Snell (1968).

Considerando os dados que descrevem os tempos entre pulsos elétricos sucessivos na superfície da fibra muscular isolada apresentados ajustamos as distribuições Weibull exponenciada, Weibull estendida de Marshall-Olkin, Weibull, Exponencial estendida de Marshall-Olkin e Exponencial. O tamanho amostral é de 799 observações. Para cada distribuição reportamos os EMV e os EMV com correção de viés.

As estimativas dos parâmetros e suas versões com correção de viés são expostas na Tabela 1. Podemos verificar que os EMV corrigidos para  $\alpha$  e  $\lambda$  das distribuições MOE-Weibull e Exp-Weibull são bastante diferentes dos EMV originais. Para as outras distribuições não é possível observar muita diferença em virtude do tamanho amostral dos dados.

Tabela 1: EMV e suas versões com correção de viés.

Distribuição	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\lambda}$
MOE-Weibull	0.3460	1.3247	0.0203	0.3283	1.3240	0.0188
Exp-Weibull	1.9396	0.7677	0.2527	1.8973	0.7625	0.2461
Weibull	–	1.0829	0.0723	–	1.0811	0.0723
MOE-Exponencial	1.1966	–	0.0998	1.1820	–	0.0994
Exponencial	–	–	0.0913	–	–	0.0912

## Conclusões

Neste projeto foi estudado em detalhes as três abordagens presente na literatura para correção de viés dos estimadores de máxima verossimilhança, uma tarefa muito usual no estudo da teoria assintótica. Considerando a distribuição unit-Gamma, pouco explorada na literatura, deduzimos correções analíticas para os vieses de seus parâmetros e mostramos por meio de um estudo de simulação Monte Carlo que os estimadores propostos são não viesados e consistentes.

Uma revisão sistemática apontou que diversos pesquisadores vem desenvolvendo estimadores corrigidos até segunda ordem utilizando a metodologia proposta por Cox e Snell (1968). Contudo, as expressões analíticas são notoriamente complicadas ou até mesmo impossíveis de serem deduzidas. Nesse contexto, foi mostrado que a biblioteca *mle.tools* pode ser uma alternativa bastante eficiente para obtenção de estimadores corrigidos. A utilidade da função *coxsnell.bc* foi testada e comparada com as expressões analíticas dos vieses para trinta e uma distribuições de probabilidade contínuas.

Por fim, é importante ressaltar que ao longo deste projeto vários artigos foram publicados em prestigiados periodicos da área de estatística.

## Agradecimentos

Agradecemos pelo suporte financeiro o CNPq, FA e UEM.

## Referências

COX, D. R.; SNELL, E. J. **A general denition of residuals**. Journal of the Royal Statistical Society, Series B., v. 30, n. 2, p. 248-275, 1968.

EFRON, B. **The Jackknife, the Bootstrap and other resampling plans**. SIAM, 1982. v. 38.

FIRTH, D. **Bias reduction of maximum likelihood estimates**. Biometrika, 1993. v. 80, n. 1, p. 27-38,