Verossimilhança Composta: Um estudo através de simulações

Palavras-Chave: Verossimilhança aproximada, comparação de estimadores, valores aberrantes

Leonardo Mazzamboni Colussi [IMECC] - [UNICAMP]

Prof. Dr. Luiz Koodi Hotta (orientador) [IMECC] - [UNICAMP]

INTRODUÇÃO:

O método da verossimilhança composta é uma alternativa aos estimadores de máxima verossimilhança quando o modelo é complexo e/ou sua dimensão é muito alta. Dessa forma, uma alternativa é a verossimilhança composta, que é uma verossimilhança aproximada, que reduz a complexidade e os custos computacionais por meio de combinações de termos de dimensões menores.

DEFINIÇÕES:

Seja $\boldsymbol{y} = (y_1, \dots, y_m)^T$ o vetor de uma única observação amostral de uma variável aleatória de dimensão m, cuja densidade é $f(\boldsymbol{y}; \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor de parâmetros de dimensão p, tal que $\boldsymbol{\theta} \in \Theta$. Ainda, seja $\{A_1, \dots, A_K\}$ o conjunto de eventos condicionais e marginais associados às Verossimilhanças $L_K \propto f(\boldsymbol{y} \in A_K; \boldsymbol{\theta})$ (Varin e Vidoni, 2011; Varin et al., 2011).

A verossimilhança composta $\left(L_C(\boldsymbol{\theta}; \boldsymbol{y})\right)$ é dada pelo produto das verossimilhanças de um conjunto de observações de uma amostra elevada aos seus respectivos pesos $(w_i, w_i \geq 0)$, ou seja:

$$L_C(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{K} L_i(\boldsymbol{\theta}; \boldsymbol{y})^{w_i}.$$

Sendo assim, a log-verossimilhança composta $(cl(\boldsymbol{\theta}; \boldsymbol{y}))$ será:

$$cl(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{K} w_i \, l_i(\boldsymbol{\theta}; \boldsymbol{y}),$$

em que:

$$l_i(\boldsymbol{\theta}; \boldsymbol{y}) = log[L_i(\boldsymbol{\theta}; \boldsymbol{y})].$$

A verossimilhança composta pode ser dividida em dois grandes grupos: condicionais e marginais.

Condicionais:

Envolvendo blocos de observações, aplicada em estudos estratificados de caso-controle:

$$L_C(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{r=1}^{m-1}\prod_{s=r+1}^{m} f(y_r|y_r + y_s;\boldsymbol{\theta}).$$

No contexto de estudos longitudinais (Molenberghs e Verbeke, 2005) e bioinformática (Mardia et al., 2008), obteve-se a verossimilhança composta por agrupamento de densidades condicionais de pares:

$$L_C(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{r=1}^{m}\prod_{s=1}^{m} f(y_r|y_s;\boldsymbol{\theta}).$$

Alternativamente, agrupando densidades condicionais completas:

$$L_C(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{r=1}^{m} f\left(y_r|y_{(-r)};\boldsymbol{\theta}\right),$$

em que $y_{(-r)}$ representa o vetor de todas as observações exceto y_r .

Marginais:

A verossimilhança marginal composta mais simples é a usualmente conhecida, a qual assumimos independência da amostra (Chandler e Bate, 2007):

$$L(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{i=1}^{m} f(y_i;\boldsymbol{\theta}).$$

No entanto, quando os parâmetros de interesses são relacionados por meio de dependência, pode-se modelar blocos de observações, a exemplo da verossimilhança de pares:

$$L_P(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{r=1}^{m-1}\prod_{s=r+1}^{m} f(y_r, y_s;\boldsymbol{\theta}).$$

A qual obtemos um total de $\binom{m}{2}$ pares de verossimilhanças. Ainda, em alguns casos, há a possibilidade de selecionar aleatoriamente tais pares, de forma que não seja necessário a utilização de todos para a obtenção dos estimadores de máxima verossimilhança composta de pares, diminuindo ainda mais o custo computacional.

METODOLOGIA E OBJETIVO:

O método será implementado dentro do contexto de Séries Temporais Financeiras, com modelos que envolvem volatilidade. Os modelos de volatilidade devem ser capazes de reproduzir o máximo possível dos chamados fatos estilizados de séries financeiras (Cont, 2000). Em particular, será considerado o modelo cDCC (corrected dynamic conditional correlation), de Aielli, utilizando a verossimilhança composta de pares (Pakel et al., 2021).

Dessa forma, as observações das amostras serão simuladas, representando uma carteira de ativos consideravelmente grande, de forma aleatória, conforme as simulações de Monte Carlo. Para isso, será utilizado uma linguagem de programação para obter os estimadores de máxima verossimilhança e de máxima verossimilhança composta de pares, comparando-os quanto à eficiência, por meio do valor esperado, desvio padrão e erro quadrado médio. Além disso, a comparação também será feita de forma empírica, analisando a aplicação do método para um portfólio de ativos real.

Sendo assim, o objetivo é estudar a teoria inferencial do método, assim como as propriedades dos estimadores obtidos, além de demonstrar que estes são quantitativamente próximos aos de máxima verossimilhança, com menor custo computacional quando a dimensão é alta e menos sujeitos a erros numéricos. Ainda, poderá ser analisado que tais estimadores de máxima verossimilhança composta são robustos a valores aberrantes.

AGRADECIMENTOS:

Os autores agradecem os apoios da FAPESP: Leonardo (Processo Nº 2021/01516-7, bolsa de iniciação científica) e Luiz (Processo Nº 2018/04654-9, projeto temático).

BIBLIOGRAFIA

AIELLI, G.P. Dynamic conditional correlation: on properties and estimation. **Journal of Business & Economic Statistics**, v. 31, p. 282-299, 2013.

CHANDLER, R. E. e S. BATE. Inference for clustered data using the independence loglikelihood. **Biometrika,** v. 94, p. 167-183, 2007.

CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. **Quantitative Finance**, v. 1, p. 223–236, 2000.

MARDIA, K. V., G. HUGHES, C. TAYLOR e H. SINGH. A multivariate von Mises distribution with applications to bioinformatics. **Canadian Journal of Statistics**, v. 36, p. 99-109, 2008.

MOLENBERGHS, G. e G. VERBEKE. **Models for discrete longitudinal data**. Nova Iorque, Springer-Verlag, 2005

PAKEL, C., S. SHEPARD, K. SHEPARD e R. F. Engle. Fitting vast dimensional time-varying covariance models. **Journal of Busines & Economic Statistics**, v. 39, p. 652–668, 2021

VARIN, C. e P. VIDONI. Parwise likelihood inference for general state space models. **Econometric Reviews**, v. 28, p. 170-185, 2011

VARIN, C., N. Reid e D. Firth. An overview of composite likelihood methods. **Statistica Sinica**, v. 21, p. 5–42, 2011