# Correções de Bartlett para modelos $\beta$ARFIMA

Matias Segelis Vieira e Guilherme Pumi (orientador)

## Introdução

Ao definirmos novos estimadores para um parâmetro de interesse, comumente estes satisfazem propriedades assintóticas interessantes, como consistência, ausência de vício e algum tipo de teorema central do limite. Porém, em pequenas amostras, alguns estimadores, em específico, aqueles baseados na teoria de máxima verossimilhança, apresentam vício. Existem várias técnicas para corrigir este vício em pequenas amostras. As mais conhecidas e utilizadas estão o ajuste de Skovgaard (Skovgaard, 2001), a correção de Cox-Snell (Cox and Snell, 1968) e as correções de Bartlett (Bartlett, 1937), sendo a última a qual estamos interessados neste trabalho, pois possui acurácia de mais alta ordem e pode ser facilmente implementada em uma linguagem matricial de programação, como apresenta Cordeiro, 1993.

Neste trabalho estamos interessados em estudar a correção de Bartlett e Skovgaard no contexto dos modelos $\beta$ARFIMA de Pumi et al. (2017). Os modelos $\beta$ARFIMA são uma generalização do já bem conhecido modelo $\beta$ARMA de Rocha e Cribari-Neto (2009), que são modelos dinâmicos para séries temporais limitadas. A ideia é substituir a estrutura de dependência do tipo ARMA, que modela a média condicional do processo, por uma estrutura ARFIMA, permitindo que a média condicional do processo apresente longa dependência. Embora tendo motivações semelhantes, a teoria assintótica dos modelos $\beta$ARFIMA é mais geral, sendo baseada na teoria de verossimilhança parcial que generaliza a abordagem de verossimilhança condicional presente na teoria assintótica dos modelos $\beta$ARMA. Neste trabalho temos por objetivo de estudar as correções de Bartlett (delineada neste pôster) e de Skovgaard (trabalho em andamento) para modelos $\beta$ARFIMA.

## Modelos dinâmicos para séries temporais limitadas

Neste trabalho estamos interessados em modelos do tipo *observation driven* em que a distribuição condicional da série temporal dado passado segue uma distribuição beta. Mais especificamente, seja $\{y_t\}_{t=1}^{\infty}$ um processo estocástico de interesse tomando valores (ou apropriadamente reescalado) no intervalo $(0,1)$ e seja $x_t'$ um vetor $l$-dimensional de covariáveis exógenas no tempo $t$. Seja $\mathscr{F}_t$ a sigma-álgebra gerada pela informação conhecida até o tempo $t$. A classe de processos que estamos interessados neste trabalho apresenta distribuição condicional dada pela seguinte parameterização da distribuição beta (conforme Ferrari e Cribari-Neto, 2004)

$$f(y_t;\mu_t,\nu|\mathscr{F}_{t-1}) = \frac{\gamma(\nu)}{\gamma(\nu\mu_t)\gamma(\nu(1-\mu_t))}\,y_t^{\nu\mu_t-1}(1-y_t)^{\nu(1-\mu_t)-1}, \tag{1}$$

para todo $0<y_t<1$, onde $0<\mu_t<1$ e $\nu>0$. Observe que $\mathbb{E}(y_t|\mathscr{F}_{t-1})=\mu_t$ e $\mathrm{Var}(y_t|\mathscr{F}_{t-1})=\frac{\mu_t(1-\mu_t)}{1+\nu}$. Neste caso o parâmetro $\nu$ é um parâmetro de precisão do modelo. O segundo passo é parametrizar a média condicional $\mu_t$. Sendo $g:(0,1)\to\mathbb{R}$ uma função de ligação, consideramos a seguinte especificação para $\mu_t$

$$g(\mu_t) = x_{t-1}'\boldsymbol{\beta} + \tau_t, \tag{2}$$

onde $\boldsymbol{\beta}'=(\beta_1,\ldots,\beta_l)$ são os coeficientes relacionados com as covariáveis e $\tau_t$ é um termo aditivo responsável por modelar a correlação serial presente em $\mu_t$. A especificação de $\tau_t$ depende da estrutura de dependência que se quer estudar. Na ausência de $\tau_t$, obtemos o modelo de regressão Beta de Ferrari e Cribari-Neto (2004). Se $\tau_t$ segue um modelos ARMA, temos o modelo $\beta$ARMA de Rocha e Cribari-Neto (2009). Quando $\tau_t$ segue um ARFIMA, temos os modelos $\beta$ARFIMA de Pumi et al. (2017).

## Modelos ARFIMA$(p,d,q)$

Lembramos que um processo estacionário $X_t$ é dito ser um processo ARFIMA$(p,d,q)$ se for solução estacionária de

$$\phi(\mathcal{B})X_t = \theta(\mathcal{B})(1-\mathcal{B})^{-d}\varepsilon_t$$

onde $\mathcal{B}$ é o operador de diferença dado por

$$\mathcal{B}^j(X_t) = X_{t-j}, \quad j\in\mathbb{Z},$$

$\phi(\mathcal{B}) = 1-\phi_1\mathcal{B}-\cdots-\phi_p\mathcal{B}^p$ e $\theta(\mathcal{B}) = 1+\theta_1\mathcal{B}+\cdots+\theta_q\mathcal{B}^q$ são os operadores autoregressivo e de média-móvel, respectivamente; $\phi(\mathcal{B})$ e $\theta(\mathcal{B})$ não têm raízes em comum, $(1-\mathcal{B})^{-d}$ é o operador de diferenciação fracionária, definido pela expansão binomial

$$(1-\mathcal{B})^{-d} = \sum_{j=0}^{\infty}\pi_j\mathcal{B}^j = \pi(\mathcal{B}), \qquad \text{onde} \qquad \pi_j(\mathcal{B}) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)},$$

para $d<\frac{1}{2}$, $d\neq 0,-1,-2,\cdots$, e $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ é um processo ruído branco.

## Correção de Bartlett

A ideia de melhorar a aproximação $\chi^2$ à distribuição sob hipótese nula da estatística $w$ do teste da Razão da log-Verossimilhança, multiplicando $w$ por um escalar, foi primeiramente introduzida por *Bartlett (1937)* e colocada numa estrutura geral por *Lawley (1956)*. Estritamente falando, a correção de Bartlett é uma transformação escalar aplicada na estatística do Teste da Razão de Verossimilhança que resulta numa estatística aperfeiçoada, com distribuição $\chi^2$ sob hipótese nula de ordem $O^{n^{-1}}$. Isto representa uma melhora clara em relação a estatística original, no sentido que a RV tem distribuição $\chi^2$ sob a hip´ótese nula, mas apenas de ordem $O(1)$.

Na tentativa de corrigir a estatística acima mencionada, Bartlett argumentou o seguinte, supondo sob hipótese nula simples ou composta:

$$\mathbb{E}(w) = pc + O(n^{-3/2}) \tag{3}$$

Onde $p$ é a diferença de dimensões dos espaços paramétricos sob as hipóteses nula e alternativa, $n$ é o tamanho amostral e $c$ uma constante igual a um mais o termo de ordem $O^{n^{-1}}$ que pode ser estimado consistentemente sob $H_0$. Pode-se mostrar que o valor esperado da estatística $w^*=w/c$ é mais próximo ao da distribuição $\chi^2_p$ que o valor esperado de $w$. A constante $c$ é amplamente conhecida como a correção de Bartlett.

## Formulação geral

Lawley (1956) obteve os momentos de certas derivadas da função de log-verossimilhança para proporcionar uma fórmula geral para a correção $c$ de Bartlett e para mostrar que todos os cumulantes da estatística ajustada $w^*$ concordam com os da distribuição $\chi^2$ de referência com erros da ordem $n^{-2}$. Hayakawa (1977) obteve a expansão assintótica da estatistica $w$ para ordem $n^{-1}$ sob a hipótese nula. Por fim, *Cordeiro (1993)* obteve uma forma matricial geral para as Correções de Bartlett para testar qualquer hipótese nula composta.

Denota-se a função de log-verossimilhança dos dados completos por $l=l(\beta)$ que depende dos paramêtros desconhecidos $\beta=(\beta_1,\ldots,\beta_p)^T$. O autor utiliza a seguinte notação tensorial para os cumulantes mistos das derivadas da log-verossimilhança:

$$k_{rs} = \mathbb{E}(\partial^2 l/\partial\beta_r\partial\beta_s), \quad k_{rst} = \mathbb{E}(\partial^3 l/\partial\beta_r\partial\beta_s\partial\beta_t), \quad k_{rs}^{(t)} = \partial k_{rs}/\partial\beta_t$$

e assim por diante. A matriz de informação total é $K=\{-k_{rs}\}$ e $K^{-1}=\{-k^{rs}\}$ sua inversa.

Os valores de $l$ no verdadeiro parâmetro e na estimativa de máxima verossimilhança $\hat{\beta}$ são $l_0$ e $\widehat{l_p}$, respectivamente. Lawley mostrou que $2\mathbb{E}(\widehat{l_p}-l_0) = p+\varepsilon_p+O(n^{-2})$ onde o termo de correção $\varepsilon_p$ de ordem $\frac{1}{n}$ é dado por:

$$\varepsilon_p = \sum_{\beta}(l_{rstu}-l_{rstuvw}), \qquad l_{rstu} = k^{rs}k^{tu}\left\{\frac{1}{4}k_{rstu}-k_{rst}^{(u)}+k_{rt}^{(su)}\right\}$$

e

$$l_{rstuvw} = k^{rs}k^{tu}k^{vw}\{\frac{1}{6}k_{rtv}k_{suw}+\frac{1}{4}k_{rtu}k_{rvw}-k_{rtv}k_{sw}^{(u)}-k_{rtu}k_{sw}^{(v)}+k_{rt}^{(v)}k_{sw}^{(u)}+k_{rt}^{(u)}k_{sw}^{(v)}\}.$$

A formulação matricial a seguir, dada por Cordeiro, visa facilitar as computações numéricas e algébricas das correções de Bartlett. Para expressar $\varepsilon_p$ em forma matricial é útil definir as seguintes matrizes $p\times p$ $A^{(tu)}$, $P^{(t)}$ e $Q^{(u)}$ para $t=1,\ldots,p$. Usando notação matricial, pode-se escrever:

$$\sum_{\beta} l_{rstu} = tr(K^{-1}L),$$

onde $L$ é uma matriz $p\times p$ definida pelo $(r,s)$-ésimo elemento igual ao $tr(K^{-1}A^{(rs)})$. Tem-se também:

$$\sum_{\beta}k^{rs}k^{tu}k^{vw}\left\{\frac{1}{6}k_{rtv}k_{suw}-k_{rtv}k_{sw}^{(u)}+k_{rt}^{(v)}k_{sw}^{(u)}\right\}$$

$$= -\frac{1}{6}\mathrm{tr}(K^{-1}M_1)+tr(K^{-1}M_2)-tr(K^{-1}M_3),$$

onde $M_1$, $M_2$ e $M_3$ são matrizes $p\times p$ definidas pelos $(r,s)$-ésimos elementos $tr(K^{-1}P^rK^{-1}P^s)$, $tr(K^{-1}P^rK^{-1}Q^{(s)T})$ e $tr(K^{-1}Q^rK^{-1}Q^s)$. Adicionalmente,

$$\sum_{\beta}k^{rs}k^{tu}k^{vw}\left\{\frac{1}{4}k_{rtu}k_{svw}-k_{rtu}k_{sw}^{(b)}+k_{rt}^{(u)}k_{sw}^{(v)}\right\} =$$

$$= -\frac{1}{4}tr(K^{-1}N_1)+tr(K^{-1}N_2)-tr(K^{-1}N_3),$$

onde $N_1$, $N_2$ e $N_3$ são matrizes $p\times p$ definidas analogamente por

$$N_1 = \{tr(P^rK^{-1})tr(P^sK^{-1})\}, \qquad N_2 = \{tr(P^rK^{-1})tr(Q^sK^{-1})\}$$

e

$$N_3 = \{tr(Q^rK^{-1})tr(Q^sK^{-1})\}.$$

Das equações acima, obtemos a redução

$$\varepsilon_p = tr[K^{-1}(L-M-N)], \tag{4}$$

onde $M=-\frac{1}{6}M_1+M_2-M_3$ e $N=-\frac{1}{4}N_1+N_2-N3$.

Na prática, (4) pode ser utilizada sem muita dificuldade. Em geral, estamos interessados em testar uma hipótese nula composta $H_0:\beta_2=\beta_2^{(0)}$ contra uma alternativa composta $H_1:H_0$ falsa. O critério da razão de verossimilhança para esse teste é dado por $w=2\{l(\widehat{\beta}_1,\widehat{\beta}_2)-l(\beta_1,\beta_2^{(0)})\}$. Assim, a correção de Bartlett pode ser escrita como:

$$c = 1+\frac{\varepsilon_p-\varepsilon_q}{p-q},$$

onde ambos $\varepsilon_p$ e $\varepsilon_q$ podem ser obtidos de (4). A distribuição de $w$ geralmente tem um erro de ordem $n^{-1}$ em relação à distribuição $\chi^2_{p-q}$, mas a correção $c$ de Bartlett torna o termo $n^{-1}$ igual a zero e a distribuição de $w^*$ se aproxima muito mais da distribuição $\chi^2_{p-q}$ do que a distribuição de $w$.

## References


- Bartlett, M.S. (1937). "Properties of Sufficiency and Statistical Tests". Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 160(901), 268–282
- Cordeiro, M.C. (1993). " General matrix formulae for computing Bartlett Corrections". Statistics and Probability Letters, 16(1), 11–18.
- Ferrari, S.L.P e Cribari-Neto, F. (2004) "Beta Regression for Modelling Rates and Proportions". Journal of Applied Statistics, 21(7), 799–815.
- Hayakawa, T. (1977). "The likelihood ratio criterion and the asymptotic expansion of its distribution". Annals of the Institute of Statistical Mathematics, 29(1), 359–378.
- Lawley, D. N. (1956). "A general method for approximating to the distribution of likelihood ratio criteria". Biometrika, 43(3/4), 295–303.
- Pumi, G.; Valk, M.; Bisognin, C. e Bayer, F.M. (2017). "Beta autoregressive fractionally integrated moving average models". Artigo submetido (Journal of Statistical Planning and Inference).
- Rocha, A.V. e Cribari-Neto, F. (2009). "Beta autoregressive moving average models". Test, 18(3), 529–545.
- Skovgaard, I. M. (2001). "Likelihood asymptotics". Scandinavian Journal of Statistics, 28(1), 3–32.


Universidade Federal Rio Grande do Sul - Porto Alegre - RS.

Email: matias.vieira@ufrgs.br