Projeto PGM

Pedro Menezes de Araújo

Introdução

Como projeto final da matéria Modelos Gráficos Probabilísticos pretendo implementar um algoritmo MCMC para a estimação de modelos regressão quantílica com variável resposta ordinal. O trabalho será baseado no artigo *Bayesian Quantile Regression for Ordinal Models* publicado na revista *Baysian Analysis* no ano de 2016.

O artigo descreve um modelo de regressão quantílica ordinal e constrói um amostrador MCMC considerando um casos mais simples (3 categorias) e um caso geral, sendo o primeiro com Gibbs Sampler e o segundo com Gibbs com passos de Metropolis-Hastings para um parâmetro específico. No MH a proposta é feita via passeio aleatório e com base na verossimilhança marginalizada com respeito a alguns parâmetros.

Objetivos

Implementar os amostradores propostos para o modelo geral no software R. Fazer um estudo de simulação para avaliar a convergência e autocorrelação das cadeias. Avaliar a possibilidade de também implementar o modelo no Stan.

Metodologia

Um modelo quantílico para uma variável ordinal pode ser expresso através da variável latente z_i da seguinte maneira:

$$z_i = x_i'\beta_p + e_i$$

Aonde x_i' é um vetor coluna de tamanho k e β_p um vetor linha também de tamanho k com os efeitos das covariáveisβ_p no quantil p. E e_i é um vetor com resíduos seguindo uma distribuição de probabilidade skew $AL(0,1,p)$ com densidade:

$$f_p(e_i) = p(1-p) \begin{cases} \exp\{-e_i(p-1)\} & e_i < 0 \\ \exp\{-e_i p\} & e_i \geq 0 \end{cases}$$

Um resultado importante sobre a distribuição de e_i é que ela pode ser reescrita como uma mistura de uma normal com uma exponencial, tendo a forma:

$$e_i = \theta w_i + \tau\sqrt{w_i}u_i \tag{1}$$

Sendo $u_i \sim N(0,1)$, $w_i \sim \exp(1)$ e θ, τ constantes bem definidas.

A variável z_i assume valores reais e as categorias observáveis y_i se relacionam com z_i através de "pontos de cortes" $\gamma_{p,j}$. Temos que:

$$\gamma_{p,j-1} < z_i \leq \gamma_{p,j-1} \implies y_i = j$$

Em que $i = 1, \ldots, n$, $j = 1, \ldots, J$, $\gamma_{p,0} = -\infty$, $\gamma_{p,J} = \infty$ e $\gamma_{p,1}$ geralmente sendo 0 para o modelo ser identificável.

Reescrevendo z_i com o resultado em (1), encontra-se $z_i = x_i'\beta_p + \theta w_i + \tau\sqrt{w_i}u_i$, e agora $z_i \mid \beta_p, w_i \sim N(x_i'\beta_p + \theta w_i, w_i u_i)$, tal resultado garante contas mais simples na hora de encontrar as condicionais completas.

Fixando $\delta_p, j = \log(\gamma_{p,j} - \gamma_{p,j-1})$ para evitar problemas relacionados a ordenação dos parâmetros durante a amostragem e assumindo prioris:

$$\beta_p \sim N(\beta_{p0}, B_{p0})$$
$$\delta_p \sim N(\delta_{p0}, D_{p0})$$

É possível encontrar as posterioris completas para β_p, que é normal, z_i, que é normal truncada e w_i que é normal-inversa generalizada. Para δ_p é necessário um passo MH com uma proposta $\delta_p' = \delta_p + u$, $u \sim N(0, d^2\hat{D})$, onde d pode ser escolhido arbitrariamente e $\hat{D}$ é a hessiana obtida ao maximizar a log verossimilhança com respeito a δ_p. A probabilidade de aceitação é:

$$\min\left\{1, \frac{f(y \mid \beta_p, \delta_p')\pi(\beta_p, \delta')}{f(y \mid \beta_p, \delta_p)\pi(\beta_p, \delta)}\right\}$$

Referência

Rahman, Mohammad Arshad. Bayesian Quantile Regression for Ordinal Models. Bayesian Anal. 11 (2016), no. 1, 1–24. doi:10.1214/15-BA939. [https://projecteuclid.org/euclid.ba/1423083637