Statistics and Application 统计学与应用, 2017, 6(5), 565-575
Published Online December 2017 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2017.65064

Hans汉斯

Bayesian Quantile Regression Associated with the Ordinal Data

Shaokai Yin, Litan Yan

Department of Mathematics, Donghua University, Shanghai
Email: 576985742@qq.com

Received: Dec. 7th, 2017; accepted: Dec. 22nd, 2017; published: Dec. 29th, 2017

Abstract

In this paper, we introduce an ordinal Bayesian quantile regression model associated with the ordinal data based on asymmetric Laplace distribution. We show that the posterior distributions of estimated parameters always proper when the prior distributions are given, and we also give an efficient Gibbs sampling algorithm for fitting the model to such data. To illustrate this approach, we give a simulation and a real data example.

Keywords

Bayesian Inference, Quantile Regression, Ordinal Data, Asymmetric Laplace Distribution, Gibbs Sampling

有序数据的贝叶斯分位数回归

尹绍锴，闫理坦

东华大学理学院，上海
Email: 576985742@qq.com

收稿日期：2017年12月7日；录用日期：2017年12月22日；发布日期：2017年12月29日

摘　要

对于一般的分位数回归模型，基于非对称拉普拉斯分布提出了关于有序数据的贝叶斯推理框架。指出了非对称分布的尺度参数在估计中应该被参数化。给出选择尺度参数与模型参数的先验分布的条件，其后验分布是真实概率分布，并采用吉布斯抽样法与马尔卡夫蒙特卡洛模拟方法进行参数估计。

文章引用: 尹绍锴, 闫理坦. 有序数据的贝叶斯分位数回归[J]. 统计学与应用, 2017, 6(5): 565-575.
DOI: 10.12677/sa.2017.65064

关键词

贝叶斯推断，分位数回归，有序数据，非对称拉普拉斯分布，吉布斯抽样

1. 引言

1978 年，Koenker and Bassett [1]首次提出了用分位数回归方法来描述因变量的条件分位数与自变量之间的关系。分位数回归的提出引起了广泛的关注。分位数回归被广泛应用于农业、基因芯片技术、生存研究、经济学、医疗卫生、环境科学等领域。

在分位数回归估计中，算法的实现起着至关重要的作用。分位数回归模型估计方法中较经典的方法有 Yu 和 Moyeed 在 2001 年提出的单纯形法和内点法。其他的估计方法还有参数化方法和非参数方法。

有序数据在对事物分类的同时给出了各类的顺序，其数据仍表现为“类别”，但各类之间是有序的，可以比较优劣但其数值的大小差值却无意义。

在大多数研究领域中都会用到有序数据。目前已有大量的文献对有序数据进行了研究，但是用贝叶斯方法对有序数据进行分析目前还鲜少有人提及。

2. 贝叶斯分位数回归

对于一个随机变量ξ，如果它服从参数为(μ,σ,p)的非对称拉普拉斯分布(Asymmetric Laplace distribution, ALD)，那么其密度函数可以写为

$$f(x;\mu,\sigma,p)=\frac{p(1-p)}{p}\exp\left\{-\frac{x-\mu}{\sigma}\left(p-1_{\{x\le\mu\}}\right)\right\},\ x\in R \tag{1}$$

其中$\mu\in R$为位置参数，$\sigma>0$为尺度参数，$0<p<1$为偏度参数。

引理 如果随机变量$\xi\sim ALD(\mu,\sigma,p)$则其方差$\mathrm{Var}(\xi)=\Psi(\sigma,p)\ge 8\sigma^2$。

证明 由于

$$E(x)=\int_{-\infty}^{\infty}xf(x;\mu,\sigma,p)\mathrm{d}x=\mu+\sigma\frac{1-2p}{p(1-p)}$$

且

$$\Psi(\sigma,p)=\int_{-\infty}^{\infty}\left(x-E(\xi)\right)^2 f(x;\mu,\sigma,p)\mathrm{d}x=\sigma^2\frac{1-2p+2p^2}{p^2(1-p)^2}.$$

根据函数

$$g(p)=\frac{1-2p+2p^2}{p^2(1-p)^2}=\frac{1}{p^2}+\frac{1}{(1-p)^2}$$

其中$p\in(0,1)$，经计算可得

$$\min_{p\in[0,1]}g(p)=g\left(\frac{1}{2}\right)=8$$ 证毕。

如果随机变量 $\xi \sim ALD(\mu,\sigma,p)$，则其分位数函数(分布函数的逆函数)为

$$F^{-1}(x;\mu,\sigma,p)=\begin{cases}\mu+\dfrac{\sigma}{1-p}\log\dfrac{x}{p}, & \text{if } 0\le x\le p\\ \mu-\dfrac{\sigma}{p}\log\dfrac{1-x}{1-p}, & \text{if } p<x\le 1\end{cases} \tag{2}$$

从上式中可以的到一个很重要的性质：

$$F^{-1}(x;\mu,\sigma,p)\big|_{x=p}=\mu \tag{3}$$

即服从非对称拉普拉斯分布的随机变量在概率 p 处的分位数等于位置参数 μ。

设 r_t 服从如下的线性回归模型，

$$r_t=Q_p(r_t\mid x_t,\beta)+u_t,\ t=1,2,\cdots,n \tag{4}$$

式中 $Q_p(r_t\mid x_t,\beta)$ 是在观测到 x_t 的条件下，r_t 在 p 概率水平下的分位数，β 是参数向量，误差项 $\{u_t,t=1,2,\cdots,n\}$ 相互独立且服从非对称拉普拉斯分布，即 $u_t\sim ALD(0,\sigma,p)$。依据非对称拉普拉斯分布的线性变换性质，式(4)可以等价表示为

$$r_t\sim ALD\left(Q_p(r_t\mid x_t,\beta),\sigma,p\right)$$

故 r_t 的密度函数为：

$$f\left(r_t;Q_p(r_t\mid x_t,\beta),\sigma,p\right)=\frac{p(1-p)}{p}\exp\left\{-\frac{Q_p(r_t\mid x_t,\beta)}{\sigma}\left[p-I\left(r_t\le Q_p(r_t\mid x_t,\beta)\right)\right]\right\} \tag{5}$$

$t=1,2,\cdots,n$。样本的似然函数可以表示为

$$L(r\mid\beta,\sigma)=\frac{p^n(1-p)^n}{\sigma^n}\exp\left[-\frac{1}{\sigma}\sum_{t=1}^{n}\rho_p\left(r_t-Q_p(r_t\mid x_t,\beta)\right)\right] \tag{6}$$

其中 ρ_p 为损失函数

$$\rho_p(u)=u\left(p-I(u<0)\right)>0.$$

设参数 β 和 σ 相互独立，其先验密度分别为 $f(\beta)$ 和 $f(\sigma)$。根据贝叶斯定理，参数 β,σ 的联合后验密度为

$$\pi(\beta,\sigma\mid r)\propto L(r\mid\beta,\sigma)\,f(\beta)\,f(\sigma) \tag{7}$$

在已知的文献中，如 Yu 等[2]，在估计过程中将尺度参数设定为常数 1。但在实际应用中，将尺度参数设为常数缺有欠妥当。若尺度参数定为常数 1，则随机变量的方差将不小 8。在实际的数据研究中这个约束条件的存在并不合理。事实上，尺度参数的存在使得服从非对称拉普拉斯分布的随机变量的方差能取任意正值。尺度参数影响了参数估计的质量，在实际的应用当中，它应该被当为待估参数去处理，而不是主观地被设为某个常数。

对于分位数回归中的参数，它们不存在标准的共轭先验分布。这导致在贝叶斯推理中获取参数的后验分布解析表达式有一定的困难。若选择的参数的先验分布能确保参数的后验分布为真实分布，我们就可以利用 MCMC 模拟得到参数的估计。

在 Yu 的研究中[2]，尺度参数为常数时，参数 β 的先验分布是非真实均匀分布，得到的联合后验分布为真实分布。考虑到尺度参数的参数化，在无信息的前提下，参数的先验分布存在很多种可能，而被

选择的先验分布可能会是非真实的分布。在这种情况下，参数 σ 与 β 满足什么条件会使得后验联合分布为真实分布。这对贝叶斯推理是很重要的一环。而下面的定理就可以保证后验分布是真实分布。

定理 对于任意 $\sigma>0$，参数的似然函数由式(6)给出，β 和 σ 的先验密度分别为 $f(\beta)$ 和 $f(\sigma)$。若 β 服从不真实均匀分布，$f(\sigma)$ 满足

$$\int_0^{\infty} f(\sigma)\frac{1}{\sigma^n}\mathrm{d}\sigma<\infty$$

则参数的联合后验分布 $\pi(\beta,\sigma|r)$ 是一个真实分布，即满足条件

$$0<\int_{-\infty}^{\infty}\int_0^{\infty}\pi(\beta,\sigma|r)\mathrm{d}\beta\mathrm{d}\sigma<\infty$$

证明 根据式(6)与(7)有

$$\begin{aligned}\pi(\beta,\sigma|r)&\propto L(r|\beta,\sigma)f(\beta)f(\sigma)\\&=f(\beta)f(\sigma)\frac{p^n(1-p)^n}{\sigma^n}\left[-\frac{1}{\sigma}\sum_{t=1}^{n}\rho_p\left(r_t-Q_p(r_t|x_t,\beta)\right)\right]\end{aligned}$$

可以推知

$$0<\exp\left[-\frac{1}{\sigma}\sum_{t=1}^{n}\rho_p\left(r_t-Q_p(r_t|x_t,\beta)\right)\right]<1,\ 0<\int_0^{\infty}f(\sigma)\frac{1}{\sigma^n}\mathrm{d}\sigma<\infty,$$

于是

$$0<L(r|\beta,\sigma)f(\beta)f(\sigma)<f(\beta)f(\sigma)\frac{p^n(1-p)^n}{\sigma^n}$$

非真实先验分布是在贝叶斯推断中在无信息先验情况下常采用的一种先验分布形式，参数 β 服从非真实先验分布，则其概率密度的积分是无穷大的[3]，即

$$\int_{-\infty}^{\infty}f(\beta)\mathrm{d}\beta=\infty$$

可得

$$\begin{aligned}0&=\int_{-\infty}^{\infty}\int_0^{\infty}0\mathrm{d}\beta\mathrm{d}\sigma\\&<\int_{-\infty}^{\infty}\int_0^{\infty}L(r|\beta,\sigma)f(\beta)f(\sigma)\mathrm{d}\beta\mathrm{d}\sigma\\&=\prod_{i=1}^{n}\prod_{k=1}^{K}\left[F_{AL}\left(\frac{\gamma_{p,j}-x_i'\beta_p}{\sigma}\right)-F_{AL}\left(\frac{\gamma_{p,j-1}-x_i'\beta_p}{\sigma}\right)\right]^{I(y_i=k)}\\&=p^n(1-p)^n\int_{-\infty}^{\infty}f(\beta)\mathrm{d}\beta\int_0^{\infty}f(\sigma)\frac{1}{\sigma^n}\mathrm{d}\sigma=\infty\end{aligned}$$

因此 $0<\int_{-\infty}^{\infty}\int_0^{\infty}\pi(\beta,\sigma|r)\mathrm{d}\beta\mathrm{d}\sigma<\infty$ 证毕。

3. 有序数据模型

有序数据的分位数回归模型能够用连续的潜变量 t_i 来表示

$$t_i=x_i'\beta_p+\epsilon_i,\ i=1,\cdots,n \tag{8}$$

其中，x_i 是 j 维的协向量，β_p 是 j 维的未知参数向量，误差项 $\epsilon_i\sim ALD(0,1,p)$。$n$ 表示观测值的个数。潜变量与我们观测到的响应变量 y_i 存在一定的关系，且响应变量有 K 个指标。我们用割点向量 γ_p 来反应

t_i 和 y_i 的关系：

$$\gamma_{p,k-1} < t_i < \gamma_{p,k} \Rightarrow y_i = k,\ i = 1,\cdots,n;\ k = 1,\cdots,K$$

其中 $\gamma_{p,0} = -\infty$， $\gamma_{p,K} = \infty$。一般而言 $\gamma_{p,1}$ 常被设定为 0 [4]。对于给定的数据 $y = (y_1,\cdots,y_n)'$，模型的似然函数能够用包含有未知参数 $(\beta_p,\gamma_p,\sigma)$ 的函数表示

$$\begin{aligned} f(\beta_p,\gamma_p,\sigma;y) &= \prod_{i=1}^{n}\prod_{k=1}^{K} P(y_i = k \mid \beta_p,\gamma_p,\sigma)^{I(y_i=k)} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left[F_{AL}\left(\frac{\gamma_{p,j} - x_i'\beta_p}{\sigma}\right) - F_{AL}\left(\frac{\gamma_{p,j-1} - x_i'\beta_p}{\sigma}\right)\right]^{I(y_i=k)} \end{aligned}$$

其中 $F_{AL}(\cdot)$ 表示非对称拉普拉斯分布的分布函数。 $I(y_i = k)$ 为示性函数。

为了在分位数回归中得到吉布斯算法，关于非对称拉普拉斯分布，Kozumi 和 Kobayashi 在 2011 年[5]给出了一个基于指数分布和正态分布的混合表达式。若 $\epsilon_i \sim ALD(0,1,p)$，则有

$$\epsilon_i = \theta\omega_i + \tau\sqrt{\omega_i}u_i,\ i = 1,\cdots,n \tag{9}$$

其中 ω_i 和 u_i 相互独立， $u_i \sim N(0,1)$， $\omega_i \sim E(1)$， E 表示指数分布。式(9)中的常数项 (θ,τ) 为

$$\theta = \frac{1-2p}{p(1-p)},\ \tau = \sqrt{\frac{2}{p(1-p)}}.$$

非对称拉普拉斯分布的混合表达式使得在研究过程中，可以直接利用正态分布的性质。而这将对有序数据分位数回归模型的分析起到重大的作用。

贝叶斯方法来估计分位数回归模型运用到了潜变量表达式 (8)与非对称拉普拉斯分布的混合表达式(9)， p 分位回归模型能够表示为

$$t_i = x_i'\beta_p + \theta\omega_i + \tau\sqrt{\omega_i}u_i,\ i = 1,\cdots,n \tag{10}$$

上式中可以看出潜变量的条件分布服从正态分布，即 $t_i \mid \beta_p,\omega_i \sim N(x_i'\beta_p + \theta\omega_i,\tau^2\omega_i)$。这使得我们在估计过程当中，能够利用正态分布的性质来进行研究。

考虑结果拥有三个指标的情况，回归模型能够用下式表示

$$\begin{cases} t_i = x_i'\beta_p + \sigma_p\epsilon_i = x_i'\beta_p + \theta\omega_i + \tau\sqrt{\omega_i}u_i, & i = 1,\cdots,n \\ \gamma_{k-1} < t_i \le \gamma_k\ y_i = k, & i = 1,\cdots,n \end{cases} \tag{11}$$

其中 σ_p 是 p 分位水平下的尺度参数，割点向量 $(\gamma_0,\gamma_1,\gamma_2,\gamma_3)$ 满足条件 $\gamma_0 = -\infty,\gamma_3 = \infty$，且 γ_1,γ_2 为某固定常值。

对于回归模型(11)，潜变量 t_i 的条件期望包含了尺度参数 σ_p，这导致模型不能直接利用吉布斯抽样方法进行估计，但对模型进行简单的变形之后就能消除尺度参数的影响从而能利用吉布斯抽样方法来进行估计。模型(11) 能表示为

$$t_i = x_i'\beta_p + \theta\nu_i + \tau\sqrt{\sigma_p\nu_i}u_i \tag{12}$$

其中 $\nu_i = \sigma_p\omega_i$，因此潜变量的条件分布服从正态分布，即 $t_i \mid \beta_p,\sigma_p,\nu_i \sim N(x_i'\beta_p + \theta\nu_i,\tau^2\sigma_p\nu_i)$。接下来，就是要设定参数的先验分布来获得参数的后验分布。设定参数的先验分布为：

$$\begin{cases}\beta_p \sim N\left(\beta_{p0}, B_{p0}\right) \\ \sigma_p \sim IG\left(n_0/2, d_0/2\right) \\ \nu_i \sim E\left(\sigma_p\right)\end{cases}$$

其中 IG 和 E 分别表示逆伽玛分布与指数分布。通过贝叶斯定理，参数 $\left(t,\beta_p,\nu,\sigma_p\right)$ 的联合后验密度为：

$$\begin{aligned}\pi\left(t,\beta_p,\nu,\sigma_p \mid y\right) &\propto f\left(y \mid t,\beta_p,\nu,\sigma_p\right)\pi\left(t \mid \beta_p,\nu,\sigma_p\right)\pi\left(\nu \mid \sigma_p\right)\pi\left(\beta_p\right)\pi\left(\sigma_p\right) \\ &\propto \left\{\prod_{i=1}^{n} f\left(y_i \mid t_i,\sigma_p\right)\right\}\pi\left(t \mid \beta_p,\nu,\sigma_p\right)\pi\left(\nu \mid \sigma_p\right)\pi\left(\beta_p\right)\pi\left(\sigma_p\right).\end{aligned}$$

当切割点与潜变量已知的情况下，观测值 y_i 不依赖于 $\left(\beta_p,\nu\right)$，根据式(12)，潜变量 t_i 的条件密度 $\pi\left(t \mid \beta_p,\nu,\sigma_p\right)$ 满足 $\pi\left(t \mid \beta_p,\nu,\sigma_p\right)=\prod_{i=1}^{n} N\left(x_i'\beta_p+\theta\nu_i,\tau^2\sigma_p\nu_i\right)$。待估参数的后验密度为

$$\begin{aligned}\pi\left(t,\beta_p,\nu,\sigma_p \mid y\right) \propto &\left\{\prod_{i=1}^{n}\prod_{k=1}^{3} I\left(\gamma_{k-1}<t_i\le\gamma_k\right)N\left(t_i \mid x_i'\beta_p+\theta\nu_i,\tau^2\sigma_p\nu_i\right)E\left(\nu_i \mid \sigma_p\right)\right\} \\ &\times N\left(\beta_{p0},B_{p0}\right)IG\left(n_0/2,d_0/2\right),\end{aligned}$$

据此，我们就能根据后验密度来导出我们感兴趣的参数的条件后验密度。

β_p 的条件后验密度 $\pi\left(\beta_p \mid t,\sigma_p,\nu\right)$ 正比于 $\pi\left(\beta_p\right)\times f\left(t \mid \beta_p,\sigma_p,\nu\right)$

$$\begin{aligned}\pi\left(\beta_p \mid t,\sigma_p,\nu\right) &\propto \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n}\left(\frac{t_i-x_i'\beta_p-\theta\nu_i}{\tau\sqrt{\sigma_p\nu_i}}\right)^2+\left(\beta_p-\beta_{p0}\right)'B_{p0}^{-1}\left(\beta_p-\beta_{p0}\right)\right\}\right] \\ &= \exp\left[-\frac{1}{2}\left\{\beta_p'\left(\sum_{i=1}^{n}\frac{x_ix_i'}{\tau^2\sigma_p\nu_i}\right)\beta_p-\beta_p'\left(\sum_{i=1}^{n}\frac{x_i\left(t_i-\theta\nu_i\right)}{\tau^2\sigma_p\nu_i}\right)-\left(\sum_{i=1}^{n}\frac{x_i'\left(t_i-\theta\nu_i\right)}{\tau^2\sigma_p\nu_i}\right)\right.\right. \\ &\quad \left.\left.+\left(\sum_{i=1}^{n}\frac{\left(t_i-\theta\nu_i\right)^2}{\tau^2\sigma_p\nu_i}\right)+\left(\beta_p-\beta_{p0}\right)'B_{p0}^{-1}\left(\beta_p-\beta_{p0}\right)\right\}\right] \\ &\propto \exp\left[-\frac{1}{2}\left\{\beta_p'\left(\sum_{i=1}^{n}\frac{x_ix_i'}{\tau^2\sigma_p\nu_i}+B_{p0}^{-1}\right)\beta_p-\beta_p'\left(\sum_{i=1}^{n}\frac{x_i\left(t_i-\theta\nu_i\right)}{\tau^2\sigma_p\nu_i}+B_{p0}^{-1}\right)\right.\right. \\ &\quad \left.\left.-\left(\sum_{i=1}^{n}\frac{x_i'\left(t_i-\theta\nu_i\right)}{\tau^2\sigma_p\nu_i}+\beta_p'B_{p0}^{-1}\right)\beta_p\right\}\right]\end{aligned}$$

设 $\tilde{B}_p^{-1}=\left(\sum_{i=1}^{n}\frac{x_ix_i'}{\tau^2\sigma_p\nu_i}+B_{p0}^{-1}\right)$，$\tilde{\beta}_p=\tilde{B}_p\left(\sum_{i=1}^{n}\frac{x_i\left(t_i-\theta\nu_i\right)}{\tau^2\sigma_p\nu_i}+B_{p0}^{-1}\beta_{p0}\right)$ 则有

$$\pi\left(\beta_p \mid t,\sigma_p,\nu\right) \propto \exp\left[-\frac{1}{2}\left\{\beta_p'\tilde{B}_p^{-1}\beta_p-\beta_p'\tilde{B}_p^{-1}\tilde{\beta}_p-\tilde{\beta}_p'\tilde{B}_p^{-1}\beta_p\right\}\right]$$

经过简单的变形有

$$\pi\left(\beta_p \mid t,\sigma_p,\nu\right) \propto \exp\left[-\frac{1}{2}\left\{\beta_p'\tilde{B}_p^{-1}\beta_p-\beta_p'\tilde{B}_p^{-1}\tilde{\beta}_p-\tilde{\beta}_p'\tilde{B}_p^{-1}\beta_p+\tilde{\beta}_p'\tilde{B}_p^{-1}\tilde{\beta}_p-\tilde{\beta}_p'\tilde{B}_p^{-1}\tilde{\beta}_p\right\}\right]$$

因此

$$\pi\left(\beta_p \mid t,\sigma_p,\nu\right) \propto \exp\left[-\frac{1}{2}\left\{\left(\beta_p-\tilde{\beta}_p\right)' \tilde{B}_p^{-1}\left(\beta_p-\tilde{\beta}_p\right)\right\}\right]$$

结果表明，β_p 的条件后验分布为正态分布，即 $\beta_p \mid t,\sigma_p,\nu \sim N\left(\tilde{\beta}_p,\tilde{B}_p\right)$。

尺度参数 σ_p 的条件后验密度 $\pi\left(\sigma_p \mid t,\beta_p,\nu\right)$ 正比于 $f\left(t \mid \beta_p,\nu,\sigma_p\right)\pi\left(\nu \mid \sigma_p\right)\pi\left(\sigma_p\right)$

$$\pi\left(\sigma_p \mid t,\beta_p,\nu\right) \propto \prod_{i=1}^{n}\left\{\sigma_p^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{t_i-x_i'\beta_p-\theta\nu_i}{\tau\sqrt{\sigma_p\nu_i}}\right)^2\right]\times\sigma_p^{-1}\exp\left(-\frac{\nu_i}{\sigma_p}\right)\right\}\exp\left[-\frac{d_0}{2\sigma_p}\right]\sigma_p^{-\left(\frac{n_0}{2}+1\right)}$$

$$\propto \sigma_p^{-\left(\frac{n_0}{2}++\frac{3n}{2}+1\right)}\exp\left[-\frac{1}{\sigma_p}\left\{\sum_{i=1}^{n}\frac{\left(t_i-x_i'\beta_p-\theta\nu_i\right)^2}{2\tau^2\nu_i}+\frac{d_0}{2}+\sum_{i=1}^{n}\nu_i\right\}\right]$$

其中 $\tilde{n}=n_0+3n$，$\tilde{d}=\sum_{i=1}^{n}\left(t_i-x_i'\beta-\theta\nu_i\right)^2/\tau^2\nu_i+d_0+2\sum_{i=1}^{n}\nu_i$。$\sigma_p$ 的条件后验分布为逆伽玛分布即 $\sigma_p \mid t,\beta_p,\nu \sim IG\left(\tilde{n}/2,\tilde{d}/2\right)$。

ν 的条件后验密度 $\pi\left(\nu \mid t,\beta_p,\sigma_p\right)$ 正比于 $f\left(t \mid \beta_p,\nu,\sigma_p\right)\pi\left(\nu\right)$ 且 ν_i 有如下形式

$$\pi\left(\nu_i \mid w,\beta_p,\sigma_p\right) \propto \nu_i^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{w_i-x_i'\beta_p-\theta\nu_i}{\tau\sqrt{\sigma_p\nu_i}}\right)^2-\frac{\nu_i}{\sigma_p}\right]$$

$$\propto \nu_i^{-1/2}\exp\left[-\frac{1}{2\sigma_p}\left(\frac{\left(w_i-x_i'\beta_p\right)^2+\theta^2\nu_i^2-2\theta\nu_i\left(t_i-x_i'\beta_p\right)}{\tau^2\nu_i}+2\nu_i\right)\right]$$

$$=\nu_i^{-1/2}\exp\left[-\frac{1}{2}\left\{\frac{\left(t_i-x_i'\beta_p\right)^2}{\tau^2\sigma_p}\nu_i^{-1}+\left(\frac{\theta^2}{\tau^2\sigma_p}+\frac{2}{\sigma_p}\right)\nu_i-\frac{2\theta\left(t_i-x_i'\beta_p\right)}{\tau^2\sigma_p}\right\}\right]$$

$$\propto \nu_i^{-1/2}\exp\left[-\frac{1}{2}\left\{\frac{\left(t_i-x_i'\beta_p\right)^2}{\tau^2\sigma_p}\nu_i^{-1}+\left(\frac{\theta^2}{\tau^2\sigma_p}+\frac{2}{\sigma_p}\right)\nu_i\right\}\right]$$

设 $\tilde{\lambda}_i=\frac{\left(t_i-x_i'\beta_p\right)^2}{\tau^2\sigma_p}$，$\tilde{\eta}=\left(\frac{\theta^2}{\tau^2\sigma_p}+\frac{2}{\sigma_p}\right)$ 则有

$$\pi\left(\nu_i \mid t,\beta_p,\sigma_p\right) \propto \nu_i^{-1/2}\exp\left[-\frac{1}{2}\left\{\tilde{\lambda}_i\nu_i^{-1}+\tilde{\eta}\nu_i\right\}\right]$$

ν_i 的条件后验分布为广义逆伽玛分布，即 $\nu_i \mid t,\beta_p,\sigma_p \sim GIG\left(0.5,\tilde{\lambda}_i,\tilde{\eta}\right)$。

4. 数据模拟分析

我们从如下模型获得观测值

$$\begin{cases} Y_i=\beta_0+\beta_1 x+\nu,\ i=1,2,\cdots,n, \\ x\sim unif\left(0,10\right), \\ \beta_0=1,\ \beta_1=2. \end{cases}$$

β_0 与 β_1 的先验分布为 $N\left(0,10\right)$，尺度参数 σ 的先验分布设定自由度为 3 的卡方分布。采用 Gibbs 抽样法共模拟 5000 次，为消除初值对抽样分布的影响，去掉前 2000 个抽样值。在应用 Gibbs 抽样算法对

参数进行抽样时，特定参数的 Gibbs 模拟值是否是真实后验分布的合理近似，这对参数估计值的正确性意义重大，因此在统计推断之前需要对抽样分布进行检验。

图 1 给出了样本容量为 100 时，β_0，β_1 和 σ 在 0.5 分位点下的 MCMC 抽样值轨迹和核密度图。图中显示三个参数的 3000 个抽样值序列在某一个值的上下波动，因此抽样构成的马尔科夫链收敛

图 2 给出了样本容量为 100 时，β_0，β_1 和 σ 在 0.5 分位点下抽样值的自相关图。从图中可以看出随着滞后期的增加，自相关系数趋向于零。

综上，马尔科夫链收敛且自相关系数趋于零，所以在样本容量为 100，各参数的贝叶斯中位数回归估计结果是可靠的。其他样本容量和分位点下的马尔科夫链均收敛，自相关系数也趋于零。表 1 和表 2 给出了不同样本容量下得到的 MCMC 参数估计结果。

在表 1 和表 2 中。比较 β_0，β_1 和 σ 参数化与未参数化下的估计结果，可以发现在一定的样本容量，特定的分位点下，当 σ 被参数化时，β_0 和 β_1 的抽样标准差均小于 σ 未参数化是各自的标准差。这表明尺度参数的参数化可以提高被估系数的精确度，从而提高待估系数假设检验的准确性。因此在用 Gibbs 抽样方法进行贝叶斯分位数回归估计是应该将尺度参数参数化，而不是设定为常数。此外，还可以发现，

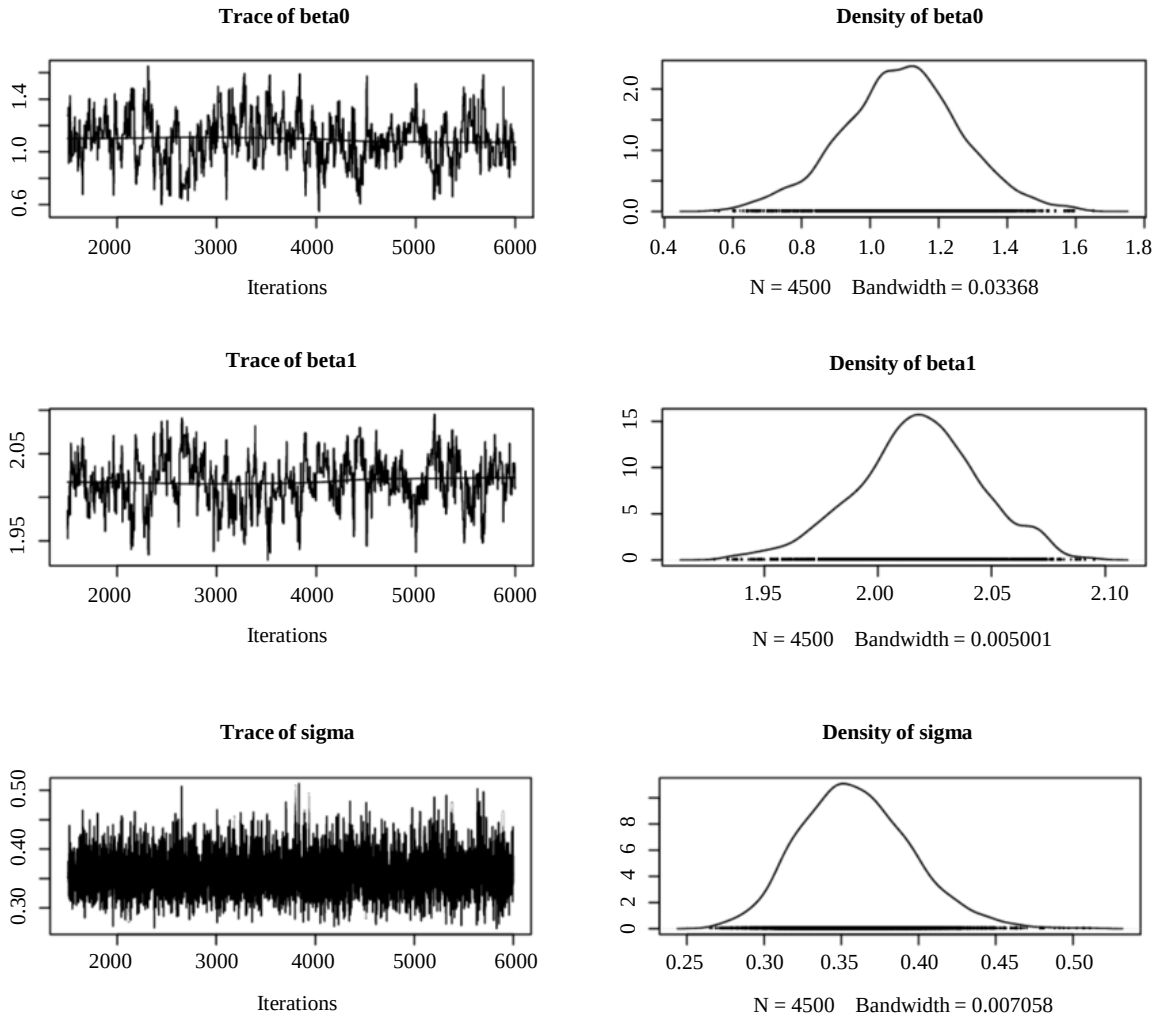

Figure 1. The MCMC trace plots and the density of β_0, β_1, σ

图 1. β_0, β_1, σ 的 MCMC 抽样值轨迹和核密度图

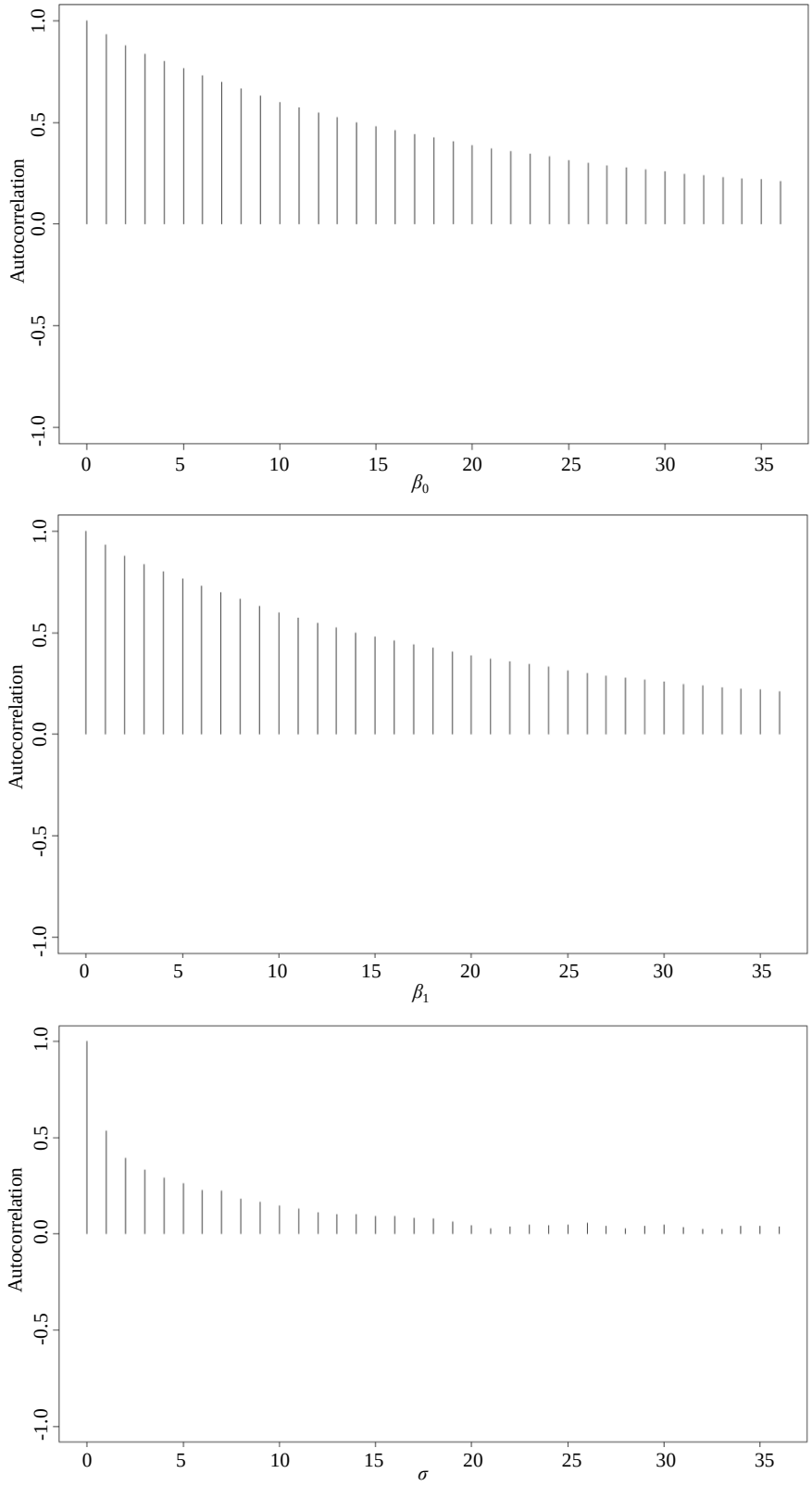

Figure 2. The autocorrelation of β_0, β_1, σ

图 2. β_0, β_1, σ 抽样值的自相关图

Table 1. Estimation of β_0 in different situation
表 1. β_0 在不同分位数水平下的估计量

样本容量	σ 未参数化			σ 参数化		
	$\beta_0(0.25)$	$\beta_0(0.5)$	$\beta_0(0.75)$	$\beta_0(0.25)$	$\beta_0(0.5)$	$\beta_0(0.75)$
	−0.388	0.583	1.223	−0.388	0.583	1.223
25	0.335	0.253	0.687	0.377	0.230	1.134
	0.281	0.267	0.282	0.221	0.227	0.224
	0.256	0.889	1.871	0.256	0.889	1.871
75	0.333	0.730	0.901	0.342	1.026	1.227
	0.249	0.271	0.225	0.171	0.233	0.148
	0.304	1.223	1.595	0.304	1.233	1.595
100	0.212	0.760	1.018	0.097	0.878	1.427
	0.291	0.202	0.235	0.139	0.166	0.134
	0.263	1.036	1.742	0.263	1.036	1.742
500	0.233	0.751	1.460	0.209	0.792	1.648
	0.133	0.132	0.139	0.083	0.089	0.095

Table 2. Estimation of β_1 in different situation
表 2. β_1 在不同分位数水平下的估计量

样本容量	σ 未参数化			σ 参数化		
	$\beta_1(0.25)$	$\beta_1(0.5)$	$\beta_1(0.75)$	$\beta_1(0.25)$	$\beta_1(0.5)$	$\beta_1(0.75)$
	2.000	2.000	2.000	2.000	2.000	2.000
25	1.718	2.025	2.036	1.832	2.045	1.989
	0.134	0.051	0.069	0.045	0.040	0.046
	2.000	2.000	2.000	2.000	2.000	2.000
50	1.939	1.982	2.138	1.962	1.958	2.101
	0.057	0.048	0.045	0.034	0.038	0.028
	2.000	2.000	2.000	2.000	2.000	2.000
75	2.000	2.042	2.068	2.035	2.040	2.017
	0.052	0.040	0.044	0.033	0.028	0.027
	2.000	2.000	2.000	2.000	2.000	2.000
1000	2.004	2.044	2.036	2.011	2.041	2.010
	0.024	0.022	0.026	0.015	0.015	0.016

注：给定样本容量，未参数化和参数化情形下的数字特征中，第一行表示参数的真值，第二行和第三行数值分别表示参数估计量分布的均值和标准差。

在给定样本容量的情况下，无论 σ 是否参数化，β_1 的标准差总是小于同一分位数水平下 β_0 的标准差。对于 β_1，无论尺度参数是否被参数化，它的精度都随着样本量的增加而提高。

参考文献 (References)

[1] Koenker, R. and Bassett, G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50.

https://doi.org/10.2307/1913643

[2] Yu, K. and Moyeed, R.A. (2001) Bayesian Quantile Regression. *Statistics and Probability Letters*, **54**, 437-447. https://doi.org/10.1016/S0167-7152(01)00124-9

[3] Zellner, A. (1971) An Introduction to Bayesian Inference in Econometrics. John Wiley Press, New York.

[4] Jeliazkov, I., Graves, J. and Kutzbach, M. (2008). Fitting and Comparison of Models for Multivariate Ordinal Outcomes. *Advances in Econometrics: Bayesian Econometrics*, **23**, 115-156. https://doi.org/10.1016/S0731-9053(08)23004-5

[5] Kozumi, H. and Kobayashi, G. (2011) Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation*, **81**, 1565-1578. https://doi.org/10.1080/00949655.2010.496117