

Application of ABC Algorithm in Parameter Estimation of MA(q) Model

Linghui Li, Nijie Tao, Huizeng Zhang*

Department of Mathematics, School of Science, Hangzhou Normal University (HZNU), Hangzhou Zhejiang
Email: *zhanghz789@163.com

Received: Feb. 11th, 2019; accepted: Feb. 26th, 2019; published: Mar. 5th, 2019

Abstract

In this paper, we provide the parameter estimation of the MA(q) model by using the approximate Bayesian computation. In the approximate Bayesian computation, a way to improve the sampling efficiency of the algorithm is choosing the statistics of low dimension, which contains as much information of parameters as possible. As the variance of white noise is known, we use the sample autocorrelation function of the first q order as the statistics. Through the numerical simulations and comparison, effects of choosing the sample auto-correlation functions are much better than the sample auto-covariance function, which greatly improves the sampling efficiency. When the variance of white noise is unknown, we use the sample autocorrelation function of the first q order and the sample variance as statistics to realize a two-step approximation Bayesian parameter estimation method. As a result of numerical simulation, this method greatly improves the accuracy of estimation.

Keywords

MA(q) Model, ABC Algorithm, Maximum Likelihood Estimation

ABC算法在MA(q)模型参数估计中的应用

李玲慧, 陶倪杰, 张慧增*

杭州师范大学理学院数学系, 浙江 杭州
Email: *zhanghz789@163.com

*通讯作者。

摘要

本文利用近似贝叶斯计算给出了MA(q)模型的参数估计。在近似贝叶斯计算中，提高算法的抽样效率在于选取低维的包含参数信息尽可能多的统计量。在白噪声方差已知的情况下，本文采用前 q 阶样本自相关函数作为统计量。通过数值模拟进行比较，选用样本自相关函数远好于样本自协方差函数，大大提升了抽样的效率。在白噪声方差未知的情况下，采用前 q 阶样本自相关函数与样本方差作为统计量，分两步对参数进行近似贝叶斯估计。从数值模拟的效果来看，该方法大大提高了估计的精确度。

关键词

MA(q)模型，ABC算法，极大似然估计

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

时间序列 MA(q)模型常见的参数估计方法为极大似然估计，对于该估计常用的数值方法有格点搜索、最陡爬坡法、Newton-Raphson 方法和 Davidon-Fletcher-Powell 方法等，这些数值方法在高阶情况下计算过程较为复杂。在贝叶斯统计的框架下，利用 MCMC 方法对参数的后验分布抽样进行参数估计是近年来一种流行的方法[1] [2]。但无论是极大似然估计还是 MCMC 方法都要涉及对似然函数的计算，当样本量很大和阶数较高时，会增加计算的复杂性和计算量。本文采用操作简单又易于实现的近似贝叶斯计算 (Approximate Bayesian Computation) 来估计 MA(q)模型的参数。

近似贝叶斯计算(简称 ABC)是利用拒绝法对参数的近似后验分布进行抽样，从而给出参数估计的一种算法[3]。1984年由Rubin首次给出近似贝叶斯基本思想[4]，1997年Tavare等学者首次把该算法应用在DNA数据的统计推断中[5]，1999年Pritchard等学者对该算法加以推广[6]，并利用推广的ABC算法研究了人类Y染色体的变异。自2002年尤其是近些年来，近似贝叶斯计算作为一种复杂模型参数估计的方法广泛应用于种群遗传学[7]、系统生物学、流行病学和系统地理学等多个领域。提高ABC算法的抽样效率在于选取低维的参数充分统计量。在复杂模型中，找到参数的充分统计量一般是很难做到的，常见的做法是选取低维的包含参数信息尽可能多的统计量。Jean-Michel Marin等学者在文献[8]中利用ABC方法对MA(q)模型参数进行了估计。该文献假设模型的白噪声方差是已知的，把前 q 阶样本自协方差函数作为参数的统计量对后验分布进行抽样和参数估计。本文采用前 q 阶样本自相关函数作为参数统计量，通过数值模拟比较，选用样本自相关函数作为统计量大大提升了抽样的效率。此外，本文在白噪声方差未知的情况下，提出用前 q 阶样本自相关函数与样本方差作为参数的统计量，采用两步估计法，即先估计模型系数，再估计白噪声方差的ABC算法对模型参数进行估计。从数值模拟的效果来看，该方法大大提高了估计的精确度。

2. MA(q)模型

定义 1: 设 $\{x_k : k = 0, \pm 1, \pm 2, \dots\}$ 为一时间序列，如果

$$x_k = \mu + u_k - \theta_1 u_{k-1} - \theta_2 u_{k-2} - \cdots - \theta_q u_{k-q}, k = 0, \pm 1, \pm 2, \cdots$$

则称该时间序列为 q 阶移动平均模型, 简记为 $MA(q)$, 其中 $\theta_q \neq 0, \theta = \{\theta_1, \theta_2, \dots, \theta_q\}$ 称为模型系数, $\{u_k : k = 0, \pm 1, \pm 2, \dots\}$ 为高斯白噪声序列, 且 $u_k \sim N(0, \sigma^2)$, σ^2 称为白噪声方差。当 $\mu = 0$ 时, 称序列为中心化的 $MA(q)$ 模型。对于非中心化的 $MA(q)$ 模型只需做简单平移 $y_i = x_i - \mu$ 即可转化为中心化 $MA(q)$ 模型。而这种转化不会影响序列之间的相关关系, 因此本文只讨论中心化的 $MA(q)$ 模型, 即满足下面结构的模型

$$x_k = u_k - \theta_1 u_{k-1} - \theta_2 u_{k-2} - \cdots - \theta_q u_{k-q}, k = 0, \pm 1, \pm 2, \cdots$$

令

$$\Theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \cdots - \theta_q x^q$$

如果没有特殊说明, $MA(q)$ 模型都是指中心化的模型。称 $\Theta(x) = 0$ 为模型 $MA(q)$ 的特征方程。如果 $\Theta(x) = 0$ 的根在单位圆外, 则模型 $MA(q)$ 可逆。

命题 1 模型 $MA(q)$ 的特征方程 $\Theta(x) = 0$ 根在单位圆外当且仅当该模型可逆。

命题 2 $MA(q)$ 模型的自相关系数 ρ_k 与 $\theta_i (1 \leq i \leq q)$ 之间具有如下的关系:

$$\rho_k = \begin{cases} 1, & k = 0 \\ -\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i} \\ \frac{i-1}{1 + \theta_1^2 + \cdots + \theta_q^2}, & 1 \leq k \leq q \\ 0, & k > q \end{cases}$$

命题 1 和 **命题 2** 都是众所周知的性质, 这里不再给出证明。

命题 3 若 $MA(q)$ 过程可逆, 则模型系数 θ 满足: $|\theta_i| < C_q^i, i = 1, 2, \dots, q$ 。

证: 令特征多项式

$$\Theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \cdots - \theta_q x^q = (1 - \lambda_1 x)(1 - \lambda_2 x)(1 - \lambda_3 x) \cdots (1 - \lambda_q x)$$

上面式子的右边项展开得

$$1 - \left(\sum_{i=1}^q \lambda_i \right) x + \left(\sum_{i < j} \lambda_i \lambda_j \right) x^2 - \left(\sum_{i < j < k} \lambda_i \lambda_j \lambda_k \right) x^3 + \cdots + (-1)^q \left(\prod_{i=1}^q \lambda_i \right) x^q,$$

与左边式子相比较得,

$$\theta_1 = \sum_{i=1}^q \lambda_i, \theta_2 = -\sum_{i < j} \lambda_i \lambda_j, \theta_3 = \sum_{i < j < k} \lambda_i \lambda_j \lambda_k, \dots, \theta_q = (-1)^q \left(\prod_{i=1}^q \lambda_i \right)$$

由 $MA(q)$ 过程可逆可知, 方程的根 $\theta_i = \frac{1}{\lambda_i}$ 在单位圆外, $i = 1, 2, \dots, q$, 因此 $|\lambda_i| < 1, i = 1, 2, \dots, q$ 。因此 $|\theta_1| < q = C_q^1, |\theta_2| < C_q^2, |\theta_3| < C_q^3, \dots, |\theta_q| < C_q^q$, 即 $|\theta_i| < C_q^i (1 \leq i \leq q)$ 。

3. 近似贝叶斯计算(ABC)

为了便于说明近似贝叶斯计算方法, 首先介绍一下贝叶斯模型。假设 $X = (X_1, X_2, \dots, X_n)^T$ 的密度 $p_{X|\theta}(x|\theta)$, 其中 θ 为模型的参数, $x = (x_1, x_2, \dots, x_n)^T$ 。 θ 的先验密度为 $p_\theta(\theta)$, 则在 $X = x$ 的条件下, θ

的后验密度为

$$p_{\theta|x}(\theta|x) = \frac{p_{x|\theta}(x|\theta)p_{\theta}(\theta)}{\int p_{x|\theta}(x|t)p_{\theta}(t)dt}$$

从上面公式可以看出, 对于后验密度的计算, 需要涉及到似然函数 $p_{x|\theta}(x|\theta)$ 以及 $\int p_{x|\theta}(x|t)p_{\theta}(t)dt$ 的计算。在复杂的模型中, 这是十分困难的事情。近似贝叶斯计算可以避开这两个问题的计算。下面我们给出简单的介绍。

近似贝叶斯计算, 简称 ABC, 是一种近似后验分布的抽样模拟方法。该方法最显著的特点就是容易实现, 不需要计算似然函数 $p_{x|\theta}(x|\theta)$ 。一般来说当模型比较复杂的时候, 似然函数以及后验密度很难进行计算, 不会有一个具体的表达式。为了估计复杂统计模型的参数, 可以利用 ABC 算法模拟后验样本, 用这些样本替代似然函数的计算, 从而得到模型参数的估计。

我们首先给出基本的 ABC 算法。

算法 1 基本 ABC 算法

输入: 样本数据 $x^* \in R^n$; θ 的先验分布 $p_{\theta}(\cdot)$; 似然函数 $p_{x|\theta}(\cdot|\theta)$; R^n 的距离 $\rho(\cdot, \cdot)$; 容忍度 $\delta > 0$ 。

输出: 近似后验分布 $p_{\theta|x}(\theta|x^*)$ 的样本 $\theta_1, \theta_2, \dots, \theta_n$ 。

算法步骤:

for $i = 1$ to n do

repeat

 根据先验分布 $p_{\theta}(\cdot)$ 生成 θ'

 根据似然函数 $p_{x|\theta}(\cdot|\theta')$ 生成 x

until $\rho(x, x^*) \leq \delta$

$\theta_i = \theta'$

end for

基本 ABC 算法的缺点是当维数 n 比较大时, θ 被拒绝的概率就很高, 这样计算成本就会大大增加。为此我们选用参数 θ 的低维充分统计量 $S(x) \in R^d$ 替代 x (其中 $d \ll n$), 即下面改进的 ABC 算法。

算法 2 改进的 ABC 算法

输入: 样本数据 $x^* \in R^n$; θ 的先验分布 $p_{\theta}(\cdot)$; 似然函数 $p_{x|\theta}(\cdot|\theta)$; R^d 的距离 $d(\cdot, \cdot)$; 容忍度 $\delta > 0$ 。

输出: 近似后验分布 $p_{\theta|x}(\theta|x^*)$ 的样本 $\theta_1, \theta_2, \dots, \theta_n$ 。

算法步骤:

for $i = 1$ to n do

repeat

 根据先验分布 $p_{\theta}(\cdot)$ 生产 θ'

 根据似然函数 $p_{x|\theta}(\cdot|\theta')$ 生产 x

until $d(S(x), S(x^*)) \leq \delta$

$\theta_i = \theta'$

end for

可以证明, 当 $\delta \rightarrow 0$, 所得到的样本就是后验证分布的样本。事实上, 对于复杂的模型, 找到低维

的充分参数统计量是十分困难，因此尽可能寻找包含参数 θ 信息尽可能的统计量。

4. MA(q)模型参数的 ABC 估计

令 $\{x_k : k = 0, \pm 1, \pm 2, \dots\}$ 为 MA(q)模型，即

$$x_k = \mu + u_k - \theta_1 u_{k-1} - \theta_2 u_{k-2} - \dots - \theta_q u_{k-q}, k = 0, \pm 1, \pm 2, \dots$$

$\theta_q \neq 0$ ， $\theta = \{\theta_1, \theta_2, \dots, \theta_q\}$ 称为模型系数， $\{u_k : k = 0, \pm 1, \pm 2, \dots\}$ 为高斯白噪声序列，且 $u_k \sim N(0, \sigma^2)$ ， σ 称为白噪声标准差。

为了保证模型系数估计的唯一性，我们假设模型 MA(q)可逆的。令

$D = \{\theta \in R^q : 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0 \text{ 的根在圆外}\}$ 。由命题 1 可知，模型 MA(q)可逆当且仅当模型系数

$\theta \in D$ 。本文假定 θ 的先验分布 $p_\theta(\cdot)$ 为 D 的均匀分布， σ 的先验分布为逆伽马分布，即， $\sigma = \frac{1}{\tau}$ ，

$\tau \sim \Gamma(\alpha, \beta)$ ， α, β 为伽马分布的参数且已知， θ, τ 相对独立。

在 ABC 算法中，首先要随机生成一个服从 D 上的均匀分布 θ ，当 $q=1$ 时， $D=(0,1)$ ，即 $\theta \sim U(0,1)$ ，

当 $q=2$ 时， $D = \{\theta = (\theta_1, \theta_2) : -2 < \theta_1 < 2, \theta_1 + \theta_2 < 1, \theta_2 - \theta_1 < 1\}$ 。当 $q \geq 3$ 时，

$D = \{\theta \in R^q : 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0 \text{ 的根在圆外}\}$ 很难找到具体的不等式给出 D 的表示。根据命题 3，

若模型 MA(q)可逆，则模型系数 $\theta \in \Delta = \{\theta \in R^q : |\theta_i| < C_q^i, i = 1, 2, \dots, q\}$ 。因此我们首先模拟一个服从 Δ 上的

的均匀分布的随机向量 θ ，然后再判断 θ 是否满足 $1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0$ 的根再圆外。如果在圆外，

便接受它，否则拒绝它。这样便得到一个模拟服从 D 上均匀分布随机向量的算法：

算法 3

输入：模型阶数 $q \geq 2$

输出：生成服从 D 上均匀分布的随机向量 θ

算法步骤：

repeat

 随机生成服从 Δ 上均匀分布的随机变量 θ'

until θ' 满足： $1 - \theta'_1 x - \theta'_2 x^2 - \dots - \theta'_q x^q = 0$ 的根在圆外

$\theta = \theta'$

假设 $x = \{x_1, x_2, \dots, x_n\}$ 为 MA(q)模型的样本数据。令 $\hat{\rho}(x) = (\hat{\rho}_1(x), \hat{\rho}_2(x), \dots, \hat{\rho}_q(x))$ 为前 q 阶样本自相关系数， $S(x) = (\hat{\gamma}_1(x), \hat{\gamma}_2(x), \dots, \hat{\gamma}_q(x))$ 为前 q 阶样本自协方差函数。文献[8]讨论了在白噪声序列的方差已知情况下 MA(2)模型 ABC 的参数估计，采用的统计量为 $S(x)$ 。本文首先给出在白噪声序列的方差 σ^2 已知情况下 MA(q)模型参数 ABC 估计。

算法 4

输入：样本数据 $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ ；模型阶数 q ；伽马分布的参数 α, β ； θ 样本容量 N ；容忍度 ε ； R^d 距离 $d(\cdot, \cdot)$ ；白噪声标准差 σ^2 。

输出： θ 的 ABC 估计 $\hat{\theta}$ 。

算法步骤：

1. 计算样本 x^* 的自相关系数 $\hat{\rho}(x^*) = (\hat{\rho}_1^*(x), \hat{\rho}_2^*(x), \dots, \hat{\rho}_q^*(x))$ ；

2. for $i = 1$ to N do

repeat

利用**算法 3** 随机生成 $\theta = (\theta_i : i = 1, 2, \dots, q)$;

随机生成系数为 θ , 白噪声方差 σ^2 的 MA(q)模型的时间序列 $x = (x_1, x_2, \dots, x_n)$, 计算样本自相关系数 $\hat{\rho}(x) = (\hat{\rho}_i(x), i = 1, 2, \dots, q)$;

until $d(S(x), S(x^*)) \leq \varepsilon$

$\theta_i = \theta^i$

end for

3. 计算模型系数 ABC 估计 $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i$ 。

下面我们讨论在白噪声序列的方差未知情况下 MA(q)模型参数的 ABC 估计。由**命题 2**可知, 两个可逆的 MA(q)模型系数 θ 相同当且仅当它们的前 q 阶自相关系统相同。这说明 q 过程的模型系数 θ 可以由前面 q 阶确定, 与白噪声的方差无关。因此在估计参数 θ 时, 可以假设 MA(q)的白噪声方差已知(不妨假定 $\sigma = 1$), 这样利用**算法 4** 给出 θ 的 ABC 估计 $\hat{\theta}$; 然后再用样本方差作为统计量, 给出 σ 的 ABC 估计, 具体算法如下:

算法 5

输入: 样本数据 $x^* = (x_1^*, x_2^*, \dots, x_n^*)$; 模型阶数 q ; 伽马分布的参数 α, β ; θ 的样本容量 N ; 容忍度 $\varepsilon_1, \varepsilon_2$; R^d 的距离 $d(\cdot, \cdot)$ 。

输出: θ, σ 的 ABC 估计 $\hat{\theta}, \hat{\sigma}$ 。

算法步骤:

1. 计算样本 x^* 的样本标准差 σ^* 和自相关系数 $\hat{\rho}(x^*) = (\hat{\rho}_1^*(x), \hat{\rho}_2^*(x), \dots, \hat{\rho}_q^*(x))$;

2. for $i = 1$ to N do

repeat

利用**算法 3** 随机生成一个 $\theta = (\theta_i : i = 1, 2, \dots, q)$;

随机生成系数 θ , 白噪声标准差 $\sigma = 1$ 的 MA(q)模型时间序列 $x = (x_1, x_2, \dots, x_n)$, 计算样本自相关系数 $\hat{\rho}(x) = (\hat{\rho}_i(x), i = 1, 2, \dots, q)$;

until $d(\hat{\rho}(x), \hat{\rho}(x^*)) \leq \varepsilon_1$

$\theta_i = \theta^i$

end for

3. 计算模型系数的 ABC 估计 $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i$;

4. for $i = 1$ to N do

repeat

生成随机数 $\tau \sim \Gamma(\alpha, \beta)$, 令 $\sigma = \frac{1}{\tau}$;

随机生成系数为 $\hat{\theta}$, 白噪声标准差 σ 的 MA(q)模型时间序列 $x = \{x_1, x_2, \dots, x_n\}$, 计算 $x = \{x_1, x_2, \dots, x_3\}$ 的样本标准差 σ ;

until $|\sigma^* - \sigma| \leq \varepsilon_2$

$\sigma_i = \sigma$

end for

5. 计算白噪声标准差的 ABC 估计 $\hat{\sigma} = \frac{1}{N} \sum_{i=1}^N \sigma_i$;

5. 数值模拟

为了比较自协方差函数和自相关函数作为统计量的 ABC 参数估计效果, 我们假定模型为

$$x_k = u_k - 0.6u_{k-1} - 0.2u_{k-2}, k = 0, \pm 1, \pm 2, \dots$$

其中 $\{u_k\}$ 为高斯白噪声, 且标准差为 1。

首先根基模型随机生成时间序列样本 $x^* = (x_1^*, x_2^*, \dots, x_{10000}^*)$, 然后根据给废除分别由自协方差函数和自相关函数作为参量统计量模拟的后验样本。为了便于比较, 我们没有采用在**算法 4** 中给定容忍度条件下, 选取 θ 的样本, 而是先生成容量 100000 的 θ 样本, 然后按照距离由小到大进行排序, 选取距离最小的前 1000 个作为 θ 作为后验样本。**图 1** 和**图 2** 分别为自协方差函数和自相关函数作为统计量近似后验分布的 $\theta = (\theta_1, \theta_2)$ 的样本散点图。与自协方差的函数相比, 由自相关函数作为统计量得到样本点更加集中在 $\theta = (0.6, 0.2)$ 附近。**图 3** 和**图 4** 分别表示由自协方差函数和自相关函数对后验分布抽样的 θ_1 与 θ_2 的直方图。**表 1** 给出了自协方差函数和自相关函数作为参数统计量 ABC 估计的 θ 均值和方差。显然由自相关函数作为统计量得到的估计精度大大优于自协方差统计量。

下面我们给出在白噪声未知的情况下 ABC 估计。我们仍选用上面的 MA(2)模型, 这里假设白噪声的先验分布为 $\sigma = \frac{1}{\tau}, \tau \sim \Gamma(2, 5)$ 。首先根据模型随机生成时间序列样本 $x^* = (x_1^*, x_2^*, \dots, x_{10000}^*)$ 。然后选取容忍度 $\varepsilon = 0.01$, 利用**算法 5** 模拟 $\theta = (\theta_1, \theta_2)$ 的容量为 1000 后验样本, 最后给出参数 θ 的估计。在此基础上选用选取容忍度 $\varepsilon = 0.01$, 模拟 σ 容量为 1000 后验样本, 然后给出 σ 的 ABC 估计值。为了 ABC 估计的效果, 我们和极大似然估计进行比较, 结果见**表 2**。

从**表 2** 可以看出, 与极大似然估计相比较, ABC 估计精确更高。

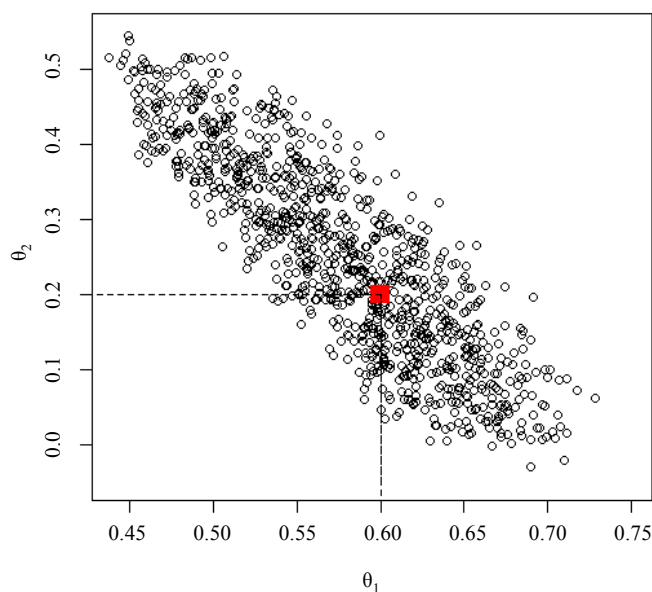


Figure 1. The posterior sample scatters with autocovariance function as statistic

图 1. 自协方差函数为统计量的后验样本散点图

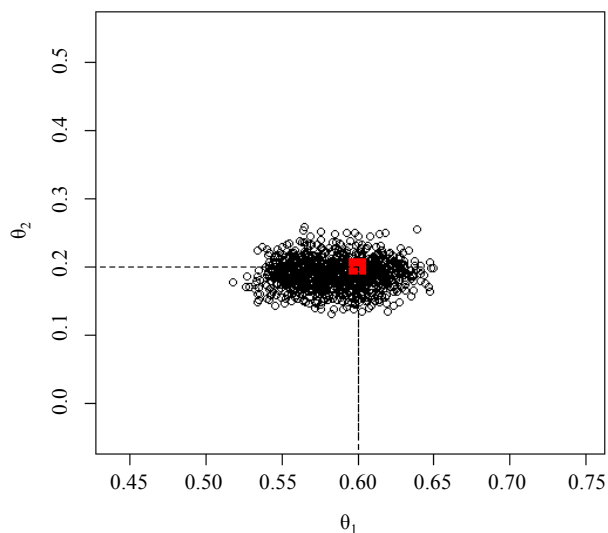


Figure 2. The posterior sample scatters with autocorrelation functions as statistics

图 2. 自相关函数为统计量的后验样本散点

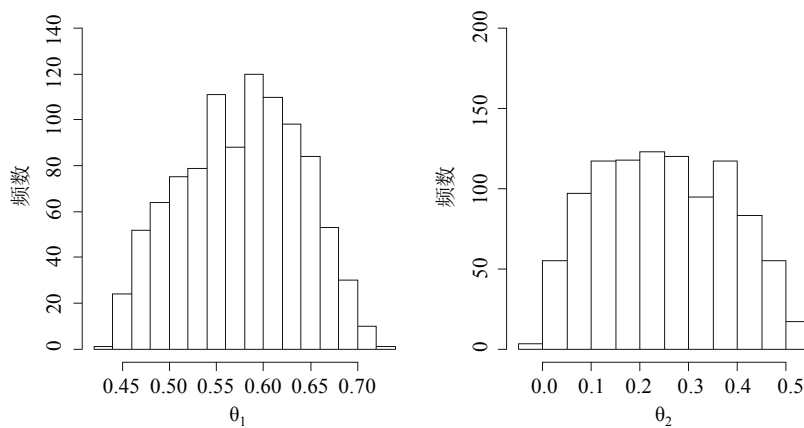


Figure 3. Histogram of sampling posterior distribution from autocovariance function

图 3. 由自协方差函数对后验分布抽样的直方图

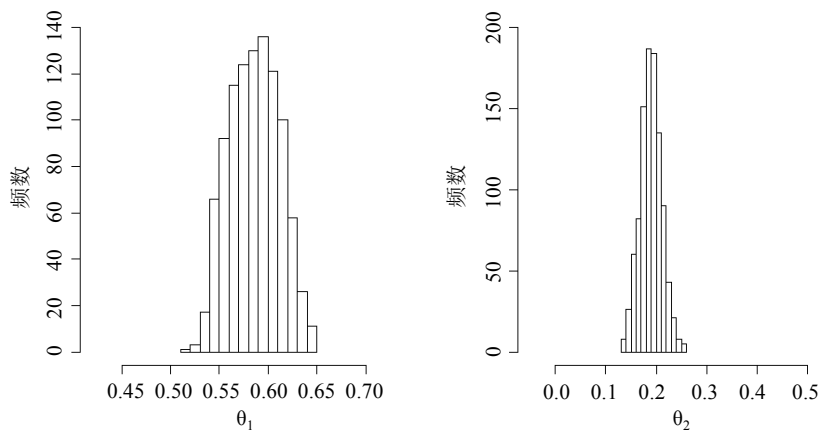


Figure 4. Histogram of sampling posterior distribution from autocorrelation function

图 4. 由自相关函数对后验分布抽样的直方图

Table 1. Comparison of ABC estimation results with autocovariance function and autocorrelation function as statistics
表 1. 自协方差函数和自相关函数为统计量的 ABC 估计结果对比

	自协方差函数为统计量的 ABC 估计		自相关函数为统计量的 ABC 估计	
	均值	方差	均值	方差
θ_1	0.577071	0.06291456	0.5859509	0.02553763
θ_2	0.2494374	0.1328754	0.1936627	0.02143973

Table 2. Comparison of the maximum likelihood estimation and ABC estimation with autocorrelation function as statistic
表 2. 极大似然估计和自相关函数为统计量的 ABC 估计对比

	极大似然估计		ABC 估计	
	估计值	相对误差	估计值	相对误差
θ_1	0.6265	4.42%	0.5859509	2.34%
θ_2	0.1899	5.05%	0.1936627	3.17%
σ	0.9957	0.43%	0.9984724	0.15%

6. 实例分析

疾病预防控制中心自 2004 年 1 月起每月定期发布全国法定传染病疫情概况, 其中乙肝是目前各类病毒性肝炎中危害最严重的, 也是发病数最多的一个型别, 并且不像其他类型的肝炎那样呈现出明显的季节性或趋势性变化, 因此对于乙肝发病数的建模比其他类型的肝炎更复杂。我们选取 2004 年 1 月至 2018 年 6 月期间的月乙肝发病人数数据进行建模, 然后通过 2018 年 7~9 月数据来说明模型的拟合效果, 最后利用模型对 2018 年 10~12 月乙肝发病数做出预测。

自相关函数是识别一个 MA 模型阶的有用工具, 图 5 和图 6 分别给出了乙肝发病数的时间序列图和样本自相关函数图。由图 6 可见, 该序列的 1 阶和 2 阶自相关系数显著, 因此选用 MA(2)对时间序列数据进行建模。

乙肝发病人数时序图

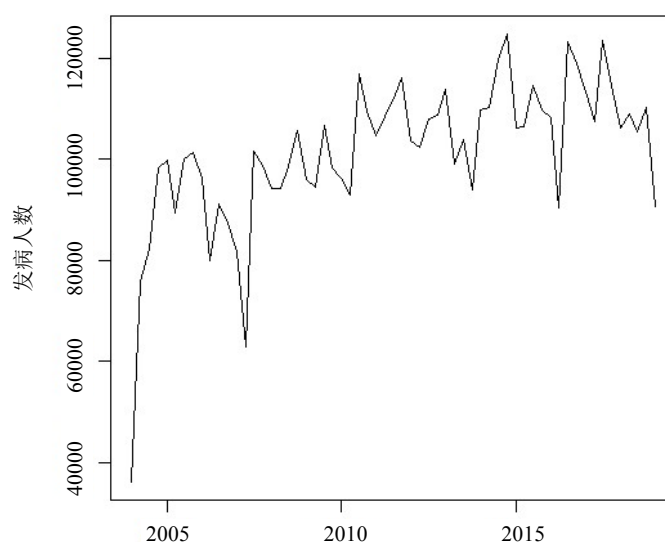


Figure 5. Time series chart of the number of patients with Hepatitis B
图 5. 乙肝发病人数的时间序列

样本自相关函数图

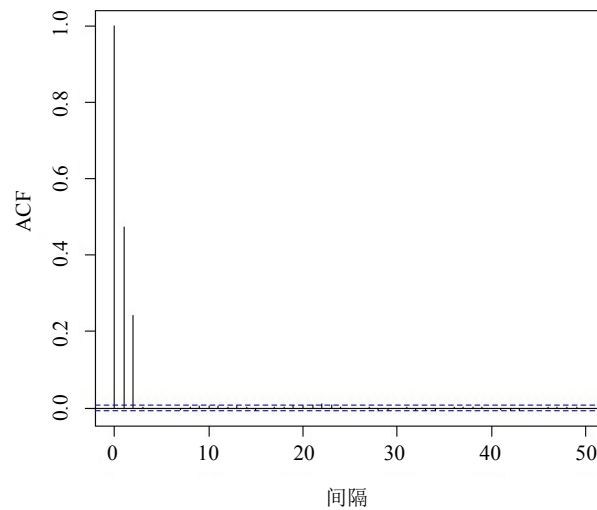


Figure 6. Autocorrelation function chart of the number of patients with Hepatitis B

图 6. 乙肝发病人数样本自相关函数

MA(2)模型如下:

$$x_k = c + u_k - \theta_1 u_{k-1} - \theta_2 u_{k-2},$$

其中 $(u_k)_{k \in \mathbb{Z}}$ 为高斯白噪声, 即 $u_i \sim N(0, \sigma^2)$ 。

显然 c 的估计为样本均值 $\hat{c} = 99708$ 。通过**算法 5** 对参数进行 ABC 估计, 得出 $\theta_1, \theta_2, \sigma$ 的估计值分别为:

$$\hat{\theta}_1 = 0.5143, \hat{\theta}_2 = 0.8455, \hat{\sigma} = 2286.843$$

因此该 MA(2)模型为:

$$y_k = 99708 + u_k - 0.5143u_{k-1} - 0.8455u_{k-2},$$

其中 $u_k \sim N(0, 2286.843^2)$ 。

表 3 分别给出 2018 年 7, 8, 9 三个月我国乙肝发病人数的实际值, 预测值与相对误差, 以及未来三个月我国乙肝发病人数的 95%置信区间和预测值。

Table 3. Prediction of the number of patients with Hepatitis B in China by ABC algorithm

表 3. ABC 算法对我国乙肝发病人数预测结果

时间	实际值	预测值	相对误差	时间	95%置信区间	预测值
2018 年 7 月	103,809	102,179	1.570%	2018 年 10 月	(92,716, 101,680)	98,415
2018 年 8 月	105,068	105,943	0.833%	2018 年 11 月	(93,740, 103,820)	101,657
2018 年 9 月	95,312	98,143	2.970%	2018 年 12 月	(93,402, 106,014)	99,823

基金项目

浙江省自然科学基金(LQ18A010007)。

参考文献

- [1] Marjoram, P., Molitor, J., Plagnol, V., *et al.* (2003) Markov Chain Monte Carlo without Likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **26**, 15324-15328.
- [2] Robert, C. and Casella, G. (2004) Monte Carlo Statistical Methods. 2nd Edition, Springer, New York. <https://doi.org/10.1007/978-1-4757-4145-2>
- [3] Voss, J. (2014) An Introduction to Statistical Computing, a Simulation-Based Approach. John Wiley Sons, Ltd, Chichester, 15-18, 181-192.
- [4] Rubin, D.B. (1984) Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, **12**, 1151-1172. <https://doi.org/10.1214/aos/1176346785>
- [5] Tavaré, S., Balding, D.J., Griffith, R.C. and Donnelly, P. (1997) Inferring Coalescence Times from DNA Sequence Data. *Genetics*, **145**, 505-518.
- [6] Pritchard, J., Seielstad, M., Perez-Lezaun, A. and Feldman, M. (1999) Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites. *Molecular Biology and Evolution*, **16**, 1791-1798. <https://doi.org/10.1093/oxfordjournals.molbev.a026091>
- [7] Beaumont, M.A., Zhang, W. and Balding, D.J. (2002) Approximate Bayesian Computation in Population Genetics. *Genetics*, **162**, 2025-2035.
- [8] Marin, J.-M., Pudlo, P., Robert, C.P., *et al.* (2012) Approximate Bayesian Computational Methods. *Statistics and Computing*, **22**, 1167-1180. <https://doi.org/10.1007/s11222-011-9288-2>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: aam@hanspub.org