

多个粗差定位的抗掩盖型 Bayes 方法

归庆明^{1,2} 李新娜¹

(1 信息工程大学理学院,郑州市科学大道 62 号,450001)
(2 信息工程大学测绘学院,郑州市陇海中路 66 号,450052)

摘 要:在综合利用先验信息与观测信息的基础上,提出了多个粗差探测的 Bayes 方法。为了有效地防止掩盖和湮没现象的发生,在分析掩盖和湮没现象发生原因的基础上,从识别向量的样本相关系数阵的特征结构出发,提出了多个粗差定位的抗掩盖型 Bayes 方法,并设计了相应的算法——自适应 MCMC 抽样算法。

关键词:Bayes 方法;掩盖;自适应 MCMC 抽样算法
中图法分类号:P207

文献[1]针对文献[2,3]提出的粗差探测 Bayes 方法涉及一系列事件后验概率的繁杂计算,且没有给出判别粗差的后验概率值的阈值以及粗差估值的方法的缺陷,基于识别变量的后验概率提出了一种新的粗差探测 Bayes 方法,并设计了一种 MCMC(Markov chain monte carlo)抽样方法以计算识别变量的后验概率值。笔者通过大量试验证实,对某些观测数据,若直接采用该方法进行粗差定位可能产生掩盖和湮没现象,从而导致粗差探测失败。进一步分析发现,MCMC 抽样对初始条件比较敏感。当初始条件把含有粗差的观测值视为正常观测值时,MCMC 抽样可能会收敛到错误的后验分布,从而产生掩盖和湮没现象,导致粗差定位失败。同时,当含有粗差的观测值之间发生掩盖时,它们对应的识别变量之间是强相关的,这种强相关性将大大降低 MCMC 抽样的收敛速度^[4],同样也会导致粗差定位失败。

1 识别变量的相关性分析

考虑测量平差的 Gauss-Markov 模型:

$$\begin{cases} \boldsymbol{L} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\Delta} \\ \boldsymbol{\Delta} \sim N_n(0, \sigma_0^2 \boldsymbol{P}^{-1}) \end{cases} \quad (1)$$

假定:

1) 不含粗差的观测值 \boldsymbol{L}_i 对应的观测误差 $\boldsymbol{\Delta}_i$

服从正态分布 $N(0, \tau^{-1} p_i^{-1})$, 含粗差的观测值 \boldsymbol{L}_i 对应的观测误差 $\boldsymbol{\Delta}_i$ 服从方差膨胀正态分布 $N(0, k_i^2 \tau^{-1} p_i^{-1})$ ($k_i \neq 1$ 为一给定的常数);

2) 每个观测值含有粗差的先验概率相等,且都为 α , 则 $\boldsymbol{\Delta}_i \sim (1 - \alpha) N(0, \tau^{-1} p_i^{-1}) + \alpha N(0, k_i^2 \tau^{-1} p_i^{-1})$, $i = 1, \dots, n$, 即 $\boldsymbol{\Delta}_i$ 服从污染正态分布;

3) (\boldsymbol{X}, τ) 服从正态-Gamma 先验分布,其具体形式为 $p(\boldsymbol{X}, \tau) = p(\boldsymbol{X} | \tau) p(\tau)$, 其中 $\{\boldsymbol{X} | \tau\} \sim N_t(\boldsymbol{X}_0, \tau^{-1} \boldsymbol{\Sigma}_0^{-1})$, $\tau \sim \Gamma(\alpha_0, \alpha_1)$, \boldsymbol{X}_0 为给定的 $t \times 1$ 向量, $\boldsymbol{\Sigma}_0$ 为给定的 $t \times t$ 正定矩阵, α_0 和 α_1 为给定的常数。

对应每个观测值 \boldsymbol{L}_i 引入一个识别变量:

$$\delta_i = \begin{cases} 1 & \boldsymbol{L}_i \text{ 服从 } N(0, k^2 \tau^{-1} p_i^{-1}) \\ 0 & \boldsymbol{L}_i \text{ 服从 } N(0, \tau^{-1} p_i^{-1}) \end{cases}, i = 1, \dots, n$$

记识别向量 $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ 。按照文献[1], 当 $q_i \doteq P\{\delta_i = 1 | \boldsymbol{L}\} > 0.5$ 时, 认为观测值 \boldsymbol{L}_i 含有粗差, 反之, 认为观测值 \boldsymbol{L}_i 是正常观测值。

以下分 3 种情况分析和讨论识别变量之间的相关系数, 以获取识别变量之间的相关性信息。

① 当第 i 个含有粗差的观测值被第 j 个含有粗差的观测值掩盖时, 有 $P\{\delta_i = 1 | \delta_j = 1, \boldsymbol{L}\} \approx 1$, 而 $q_i \approx 0, q_j \approx 1$ 。如果对应的识别变量之间相互独立, 则应有 $q_i = P\{\delta_i = 1 | \boldsymbol{L}\} = P\{\delta_i = 1 | \delta_j = 1$,

表 2 基于 MCMC 抽样的识别变量的后验概率

Tab. 2 Posterior Probability of Classification Variable
Based on MCMC Sampling Algorithm

观测值序号	q_i	观测值序号	q_i	观测值序号	q_i
1	0.037 3	7	0.035 8	13	1.000 0
2	0.046 9	8	0.035 7	14	1.000 0
3	0.036 0	9	0.042 7	15	0.067 8
4	0.036 4	10	0.039 4	16	0.049 7
5	0.035 9	11	0.036 9	17	0.037 8
6	0.035 7	12	0.038 3	18	1.000 0
19	0.049 2				

表 3 基于自适应 MCMC 抽样的识别变量的后验概率

Tab. 3 Posterior Probability of Classification Variable
Based on adaptable MCMC Sampling Algorithm

观测值序号	q_i	观测值序号	q_i	观测值序号	q_i
1	0.036 5	7	0.035 7	13	0.060 9
2	0.040 6	8	0.035 7	14	1.000 0
3	0.035 9	9	0.038 9	15	1.000 0
4	0.036 0	10	0.037 4	16	0.041 8
5	0.035 8	11	0.036 3	17	0.036 7
6	0.035 7	12	0.036 9	18	0.036 2
19	0.041 5				

5 结 语

1) 在分析掩盖和湮没发生原因的基础上,对识别向量的样本相关系数矩阵进行特征结构分析,探寻掩盖和湮没现象发生时观测值对应的识别变量之间的相关性信息,进而重新选择 MCMC 抽样的初始子集,改进了基于识别变量后验概率的粗差定位方法。

2) 识别变量的引入,不但克服了以往探测方法的模糊性及探测标准选择的困难,更有利于分析混入多个粗差后观测值之间的相关性信息。同时,仅从一个受污染的观测值向量不可能获得观测向量对应的识别向量的相关系数矩阵或其估计,而先验信息的利用,又能成功地获得了识别向量的样本相关系数矩阵,进而采取有效的措施揭示出隐含其中的相关性信息,这也正是利用先验信息进行粗差探测的一个显著的优势。

参 考 文 献

[1] 李新娜,归庆明,许阿裴. 基于识别变量的粗差探测的 Bayes 方法[J]. 测绘学报, 2008,37(3):355-360

[2] 归庆明,宫铁松,李国重,等. 粗差探测的 Bayes 方法[J]. 测绘学报, 2006,35(4):303-307

[3] Gui Qingmin, Gong Yisong, Li Guochong, et al. Bayesian Approach for Detection of Gross Errors Based on Posterior Probability[J]. J Geod, 2007,81(6):651-659

[4] Smith A F M, Roberts G O. Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods[J]. Journal of the Royal Statistical Society(Ser. B), 1993,55:3-24

[5] 周江文,黄幼才,杨元喜,等. 抗差最小二乘法[M]. 武汉:华中理工大学出版社,1997

第一作者简介:归庆明,教授,博士生导师,现主要从事测量误差理论与数据处理的研究。
E-mail:guiqm@public.zz.ha.cn

Bayesian Unmasking Method for Positioning Multiple Blunder

GUI Qingming^{1,2} LI Xinna¹

(1 Institute of Science, Information Engineering University,62 Kexue Road, Zhengzhou 450001,China)
(2 Institute of Surveying and Mapping, Information Engineering University, 66 Middle Longhai Road, Zhengzhou 450052, China)

Abstract: Combining prior information with observing information, Bayesian methods for blunder detection are imposed. Especially a lot of effective measures are used to overcome the masking and swamping. When multiple blunder influence each other, the Bayesian method for blunder positioning based on the posterior probabilities of classification variables sometimes gives birth to masking and swamping which leads to the failure of positioning blunder. Hence, on the basis of seeking the reason of masking and swamping, and analyzing the eigenstructure of sampling correlation matrix of classification variables, the Bayesian unmasking method for positioning multiple blunder is introduced. The corresponding algorithm-adaptable MCMC sampling algorithm is implemented.

Key words: Bayesian method; masking; algorithm-adaptable MCMC sampling algorithm

About the first author: GUI Qingming, professor, Ph.D supervisor. He is engaged in the theoretic research on data processing.
E-mail: guiqm@public.zz.ha.cn