ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.16, No.14, May . 2020

E-mail: xsjl@dnzs.net.cn
http://www.dnzs.net.cn
Tel:+86-551-65690963 65690964

基于MCMC方法的顾客满意度的Bayes估计

赵琪

(山东英才学院 基础部,山东 济南 250104)

摘要:该文提出运用MCMC方法,解决顾客满意度的Bayes模型中的数值计算问题。简要介绍了MCMC方法。运用顾客满意度的Bayes模型进行了实例分析,并且运用OpenBUGS软件给出了在先验分布为Dirichlet分布情况下的Bayes模型的数据仿真。

关键词:顾客满意度;Bayes模型;MCMC方法;OpenBUGS软件

中图分类号:O212.8 文献标识码:A

文章编号:1009-3044(2020)14-0133-02

开放科学(资源服务)标识码(OSID):

DOI:10.14004/j.cnki.ckt.2020.1526

The Bayes Estimation of Customer Satisfaction Based on MCMC Method

ZHAO QI

(Department of Basic Courses, Shandong Yingcai University, Jinan 250104,China)

Abstract: In this paper, using MCMC method, numerical computing problems of the Bayes model of customer satisfaction is solved. Firstly, we introduced MCMC method. Afterwards, We has carried on the instance analysis using the customer satisfaction of the Bayes model, and the Bayes model of simulation data is given under the prior distribution for the Dirichlet distribution using OpenBUGS software.

Key words: customer satisfaction; Bayes model; MCMC method; OpenBUGS software

自20世纪70年代,顾客满意度(Customer Satisfaction)的研究在欧美等西方国家开始兴起。20世纪末,我国也开始进行国家满意度指数模型的设计工作。目前,对顾客满意度进行测量和研究已成为企业的一项重要工作内容,顾客满意度是衡量企业整体质量的重要指标,另外它还广泛应用与国家、行业等宏观、中观分析。已有许多高校和市场研究公司参与到顾客满意度模型的研究中来。其中,苗敬毅[1]提出了顾客满意度的Bayes模型。本文运用马氏链蒙特卡洛方法(MCMC),解决顾客满意度的Bayes模型中的数值计算问题。

1 顾客满意度的Bayes模型

1.1 Bayes统计的基本思想

贝叶斯统计学是统计学的重要流派,具有广泛的应用与影响力,它与经典统计学的主要区别是使用先验信息(经验或历史资料)。假设θ是我们感兴趣的一个未知量(θ所有可能值的集合称为参数空间,记为Θ),按贝叶斯的观点,我们将它看做随机变量,可以用一个概率分布去描述它,这个分布称为先验分布[2]。贝叶斯公式结合了θ的先验信息和样本信息(x)得到了样本给定下θ的条件分布称为θ的后验分布,记为$\pi(\theta|x)$,一切决策和推断都是基于后验分布做出,下面给出后验分布的确定方法。

设θ的先验分布为$\pi(\theta)$,样本X依赖于θ的条件密度函数为$p(x|\theta)$。θ与X的联合密度函数为:

$$h(x,\theta)=p(x|\theta)\pi(\theta) \tag{1}$$

而X具有边际密度:

$$M(x)=\int_{\Theta}h(x,\theta)d\theta=\int_{\Theta}p(x|\theta)\pi(\theta)d\theta \tag{2}$$

因此:

$$\pi(\theta|x)=\frac{h(x,\theta)}{M(x)}=\frac{p(x|\theta)\pi(\theta)}{\int_{\Theta}p(x|\theta)\pi(\theta)d\theta} \tag{3}$$

后验分布是在样本信息给定的前提下集中了样本信息与先验分布中有关θ的一切信息,因此后验分布$\pi(\theta|x)$要比先验分布$\pi(\theta)$更贴近实际情况,具有更高的可信度。作为θ的估计一般采用后验分布$\pi(\theta|x)$的某个位置特征量,例如后验分布的分位数或数学期望。本文采用$\theta=E[\pi(\theta|x)]$。

1.2 顾客满意度的Bayes估计

顾客满意度的Bayes估计实际上是一种多级评分的Bayes估计。多级评分的Bayes估计文献[3]中已有详细的证明,在这里我们引用其主要结论。假设在某种指标体系下,每个指标的可能测评分数为$0,1,\ldots,K$。假设指标体系由n个指标构成,测评分数为0分的指标有x_0个,……,评K分的有x_K个。对每一测评

收稿日期:2020-01-15

者来说，其测评分数记为：$x=(x_0,x_1,...,x_K)'$，则 $\sum_{i=0}^{K}x_i=n$。设某一测评者的测评分数为 $T=(T_0,T_1,...,T_K)'$。记 $\theta_0=\frac{T_0}{n},...,\theta_K=\frac{T_K}{n},\theta=(\theta_0,...,\theta_K)'$，则 θ 可作为测评分数 $0,1,...,K$ 的概率估计。当 θ 已知时，分数 x 的分布是多项分布：

$$p(x|\theta)=n!\prod_{i=0}^{K}\left|\frac{\theta_i^{x_i}}{x_i!}\right| \tag{4}$$

其中 $x_i=0,1,...,n(i=0,1,...,K),\sum_{i=0}^{K}x_i=n,\sum_{i=0}^{K}\theta_i=1$；当对 m 个回答者进行调查时，记样本为：$X=(x^{(1)},...,x^{(m)}),x^{(j)}=(x_{0j},...,x_{Kj})',j=1,...,m$。$x^{(j)}$ 表示总体的第 j 个回答者的评分，因此总有：$\sum_{i=0}^{K}x_{ij}=n,j=1,...,m$。假设 θ 的先验分布为 $\pi(\theta)$，则当 θ 确定时，X 的联合分布函数为：

$$f(x|\theta)=(n!)^m\prod_{i=0}^{K}\left|\frac{\theta_i^{\sum_{j=1}^{m}x_{ij}}}{\prod_{j=1}^{m}x_{ij}!}\right| \tag{5}$$

我们将研究 θ 的 Bayes 估计。

2 马氏链蒙特卡洛方法（MCMC）

贝叶斯统计中的许多问题都可归结为关于后验分布的积分计算问题，但是当后验分布很复杂时，这样的计算通常都难以进行。因此，我们有必要探讨一些新的计算方法。MCMC方法就是解决此类问题的一种非常有效的办法。通过建立一个平稳分布为 $\pi(x)$ 的 Markov 链来得到 $\pi(x)$ 的样本，然后就可以利用这些样本信息作各种统计推断，这就是MCMC方法的基本思想。换言之，MCMC方法实际上是一个Monte Carlo综合程序，它的随机样本的产生与一条马氏链相关联。Gibbs[4]抽样是最简单却是最常用的一种MCMC方法，它是一种基于条件分布的迭代抽样方法。

设给定一个 m 维联合分布 $\pi(x_1,...,x_m)$，构造如下的转移核

$$P_{x,y}\triangleq P(x,y)=\prod_{k=1}^{m}\pi(y_k|y_1,...,y_{k-1},x_{k+1},...,x_m) \tag{6}$$

其中 $x=(x_1,...,x_m),y=(y_1,...,y_m),x_i\in D,y_i\in D$（$D$ 为 m 维空间的一个区域），而 $\pi(y_k|y_1,...,y_{k-1},x_{k+1},...,x_m)$ 是在除第 k 个分量外，将第1至第 $k-1$ 个分量固定为 $y_1,...,y_{k-1}$，并将第 $k+1$ 至第 m 个分量固定为 $x_{k+1},...,x_m$ 的条件下，第 k 个分量在 y_k 处的条件分布。

在Gibbs抽样中，由 $X_n(\omega)$ 的样本得到 $X_{n+1}(\omega)$ 的样本的具体步骤如下：

1)从分布 $\pi(y_1|x_2,...,x_m),y_1\in D_1$ 中抽取样本 y_1（注意 $x_2,...,x_m$ 来自 $X_n(\omega)$）；

2)从分布 $\pi(y_2|y_1,x_3,...,x_m),y_2\in D_2$ 中抽取样本 y_2，依次下去，从分布 $\pi(y_k|y_1,...,y_{k-1},x_{k+1},...,x_m,y_k\in D_k$ 中抽取样本 $y_k(k=1,...,m-1)$；

3)最后从分布 $\pi(y_m|y_1,...,y_{m-1}),y_m\in D_m$ 中抽取样本 y_m，则 $(y_1,...,y_m)$ 就是 $X_{n+1}(\omega)$ 的一个样本。

现任取一个初值 $X_0(\omega)=y^{(0)}$，按上述方法得到 $X_1(\omega)$ 的一个样本 $y^{(1)}$，对 n 归纳地得到 $X_2(\omega),...,X_n(\omega)$ 的样本 $y^{(2)},...,y^{(n)}$，当 n 充分大时，马氏链 $\{X_n(\omega)\}$ 的分布近似于 $\pi(x_1,...,x_m)$，就可以认为 $y^{(n)}$ 是近似服从 $\pi(x_1,...,x_m)$ 的一个样本。

MCMC运算的软件和应用程序已经有很多被开发出来，例如OpenBUGS[5][6]软件，对于许多常用的模型和分布进行Gibbs抽样时使用OpenBUGS都是非常方便的。因为在使用OpenBUGS时，只要设置好变量的先验分布并对所研究的数学模型进行一般性的描述，就可以很容易实现对模型的贝叶斯分析，不需要知道参数的先验密度或似然函数的精确表达式，也不需要复杂的编程。

3 数据仿真

我们以文献[1]中的实例为例，说明MCMC方法的运用过程。

此例为一企业人力资源关于公司员工满意度调查。指标体系采用国际通用型量表，涉及指标33项，样本数据79例。评分规则是：1-肯定是；2-偶尔是；3-不确定；4-偶尔不是；5-肯定不是。调查统计数据表格见文献[1]，在此不再赘述。

假设 θ 的先验分布为Dirichlet分布，OpenBUGS程序如下：

```
model
{for(i  in 1:79){
x[i,1:5]~dmulti(theta[],33)
}
theta[1:5]~ddirich(prior[]);
}
```

我们在模型运行的过程中，为确保参数的收敛性，首先进行1000次预迭代。然后丢弃最初的预迭代数据，再进行10000次迭代。下面是得到的参数 θ 的部分信息。

表1 OpenBUGS运行结果

参数	均值	标准差	MC误差	2.5%	中位数	97.5%	开始	样本
theta1	0.08496	0.005438	5.019E-5	0.07454	0.08484	0.09583	1001	10000
theta2	0.1143	0.006172	6.405E-5	0.1027	0.1142	0.1267	1001	10000
theta3	0.171	0.007322	7.127E-5	0.1568	0.1709	0.1855	1001	10000
theta4	0.2484	0.008456	8.678E-5	0.232	0.2484	0.2652	1001	10000
theta5	0.3813	0.009522	9.143E-5	0.3626	0.3813	0.3998	1001	10000

在表中可以看出，员工的满意度 θ 的Bayes估计为：$\theta=(\theta_1,...,\theta_5)'=(0.08496,0.1143,0.171,0.2484,0.3813)$，$\theta$ 是评分为1,...,5分的概率的估计，即员工对公司总体持非常满意态度的约占38%，持比较满意态度约占25%，没有特殊意见的约占17%，持不满意态度的约占11%，持非常不满意态度的约占9%。与文献[1]的结果一致。除了得到参数的均值以外，从表中还可以得到参数的、95%置信区间、后验分布的标准差和中位数等信息。从OpenBUGS软件还可以得到参数的后验分布的核密度估计、动态轨迹图、迭代历史和收敛性统计诊断图等一系列信息。另外，除了直接编写代码程序以外，在OpenBUGS中还可以使用有向图模型方式（Doodle模型）。在此不一一赘述。

（下转第137页）

依据测试数据，绘制二维编码压缩比折线，如图5。测试数据最高压缩比可以达到130倍以上，最低17.0倍左右。当图像的黑白信息相对比较集中时，压缩比例会比较大，当黑白信息相对分散时，压缩比例会比较小，压缩比例动态范围大，在较大数据量的传真通信中，MR编码具有更大的数据压缩优势。

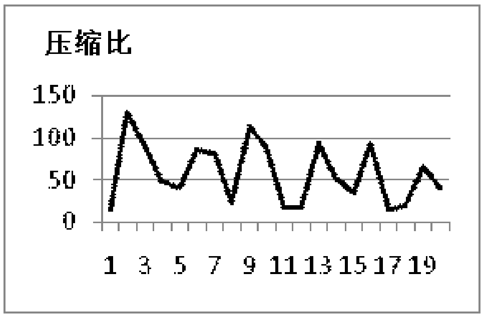

图5　二维编码压缩比折线图

4 结束语

用C#实现了MR编码，并验证了MR编码能够对图像进行二维空间的数据压缩，在较大数据量的传真通信中，MR编码具有更大的数据压缩优势。

参考文献：

[1] 田园园, 董金明. 传真图像的霍夫曼二维编码的FPGA实现[C]//第一届全国信号处理学术会议暨中国高科技产业化研究会信号处理分会筹备工作委员会第三次工作会议论文集. 北京, 2007: 245-247, 314.

[2] 李薇, 胡伟文, 沈静. 霍夫曼编码和游程编码在图像编码中的应用[J]. 舰船电子工程, 2010, 30(7): 67-69.

[3] 曾党泉. 一种改进的MH编码算法—MHZ算法[J]. 计算机与数字工程, 2014, 42(4): 578-581, 586.

【通联编辑：谢媛媛】

（上接第132页）

参考文献：

[1] 郭凤鸣,李兵,何怡刚.基于RFID技术的方向感知方法研究[J].华东师范大学学报(自然科学版),2016(2):73-80.

[2] 马旭平,马金舟,何诚,等.超市快速智能结算系统的实现[J].电脑知识与技术,2016,12(14):151-154,181.

[3] 易昌惠.条码识别技术的发展[J].企业标准化,2004(2):62-64.

[4] 张为,邹艳碧.新型超市结算支付系统的设计[J].广州大学学报(自然科学版),2006,5(2):35-39.

[5] 文军.RFID技术在未来超市中的应用[J].金卡工程,2006(6):32-33.

【通联编辑：王力】

（上接第134页）

4 结束语

MCMC方法的诞生使贝叶斯统计学如虎添翼，获得突飞猛进的发展，长期以来被认为是难以处理的模型，如今借助MCMC方法已经可以定量研究。这种方法的应用戏剧性地重塑了统计学家的工作方式，构建了以随机模拟技术为手段的统计研究模式。我们把MCMC方法运用于顾客满意度模型，显示了此方法的广泛应用性。本文的研究为这种方法应用于更多的贝叶斯模型提供了启示。

参考文献：

[1] 苗敬毅.顾客满意度的Bayes估计[J].统计与信息论坛,2004,19(2):35-39.

[2] 茆诗松,王静龙,濮晓龙.概率论与数理统计教程[M].北京:高等教育出版社,1998.

[3] 吴大伟.多级评分及其Bayes估计[J].数理统计与管理,2003,22(2):37-43.

[4] Walsh B. Markov Chain Monte Carlo and Gibbs Sampling[J]. Lecture Notes for EEB 581,2004(26).

[5] 张继巍,高文龙. Open BUGS 软件介绍及应用[J].中国卫生统计,2017,34(1):170-172.

[6] 杨维维,刘沛,巢健茜,等.贝叶斯统计分析的有力工具——OpenBUGS软件[J].中国卫生统计,2016,33(3):510-513.

【通联编辑：光文玲】