【统计理论与方法】

MCMC 方法的发展与现代贝叶斯的复兴
——纪念贝叶斯定理发现 250 周年

刘乐平，高 磊，杨 娜
（天津财经大学 统计学系，天津 300222）

摘要：信息科学的进步如何影响统计学理论与方法的发展，是 21 世纪大数据时代统计学面临的重要问题。回顾 20 世纪 40 年代 MCMC 方法的起源与发展，基于千岛湖植物考察与 R 软件直观性解释 Metropolis – Hastings 算法，讨论 MCMC 方法的发展对现代贝叶斯统计复兴具有至关重要的作用。基于云计算随机模拟，统计学与信息科学密切结合、贝叶斯与频率统计相互交融的统计分析新范式，可能成为大数据研究的新思路。

关键词：MCMC 方法；Metropolis-Hastings 算法；千岛湖植物考察；贝叶斯定理
中图分类号：C829.29 **文献标志码**：A **文章编号**：1007－3116(2014)02－0003－09

一、引 言

1763 年 12 月 23 日，理查德·普莱斯(Richard Price)在伦敦皇家学会会议上宣读了托马斯·贝叶斯的遗世之作，从此贝叶斯统计学诞生了。在整个 19 世纪，贝叶斯统计倍受争议和冷落，主要原因是不知道如何恰当地处理先验概率。20 世纪上半叶，一个与贝叶斯统计完全不同的理论——频率统计学(Frequentist statistics)迅速发展，在统计学领域几乎占主导地位。尽管如此，贝叶斯统计思维仍在意大利的布鲁诺·菲尼蒂和英国的哈罗德·杰弗里斯等少数统计学家的努力下，得以传承[1]50－51。

现代贝叶斯运动的复兴，始于 20 世纪下半叶，领军人物是美国的吉米·萨维奇(Jimmy Savage)和英国的丹尼斯·林德利(Dennis Lindley)，由于贝叶斯后验分布在高维计算上的困难，贝叶斯统计推断非常难以实现，贝叶斯方法的应用受到了很大的限制[2]。随着计算机信息技术的发展和贝叶斯方法的改进，特别是 MCMC 方法的发展、WinBUGS 软件以及 R 语言的广泛应用，原来复杂的数值计算问题如今变得简单便捷，参数后验分布的模拟也趋于方便，现代贝叶斯分析的理论和应用又重新得到了迅速的发展[3]。

MCMC(Markov Chain Monte Carlo)方法的发展，对现代贝叶斯分析的复兴起着至关重要的作用。笔者在回顾 MCMC 方法发展历史的基础上，给出 MCMC 方法直观性的实例，并介绍 R 软件中 MCMC 的相关程序包，以此纪念贝叶斯定理 250 周年，因为 MCMC 方法和现代贝叶斯两者的发展和兴起，不仅给出了解决问题的方法，重要的是改变了考虑问题的思路。

收稿日期：2013－09－13；**修复日期**：2013－11－11
基金项目：国家自然科学基金项目《Solvency II 框架下非寿险准备金风险度量与控制研究》(71171139)、《多重风险相依情形下的最优保险问题研究》(71371138)；天津社会科学规划项目《宏观统计数据可靠性评估方法研究》(TJTJ10－651)；全国统计科研计划项目《贝叶斯统计与频率统计相融合的大数据分析新范式》(2013LY024)；天津财经大学研究生创新基金资助
作者简介：刘乐平，男，江西萍乡人，经济学博士，教授，博士生导师，研究方向：贝叶斯数据分析，精算与风险管理；
高 磊，男，山东德州人，博士生，研究方向：精算与风险管理；
杨 娜，女，河北唐山人，硕士生，研究方向：精算与风险管理。

二、MCMC 方法的起源

MCMC 方法起源于军事需要，来源于物理问题，理论基础是概率统计，研究手段是计算机技术。MCMC 方法产生于第二次世界大战期间，在研制原子弹的“曼哈顿计划”过程中，1953 年由美国物理学家 Metropolis 等人提出，经加拿大统计学教授 Hastings 与其博士生于 1970 年推广和完善。

(一)蒙特卡洛方法(Monte Carlo，MC)

蒙特卡洛方法诞生于第二次世界大战期间的美国军事研究重要基地——新墨西哥州的洛斯阿拉莫斯国家实验室。20 世纪 40 年代中后期，在研制原子弹的“曼哈顿计划”过程中，Stanislaw Ulam 在病床上为了解决一个棘手的组合计算问题(纸牌游戏“solitaire”中获胜概率的计算)，产生了蒙特卡洛方法的最初理念。约翰·冯·诺伊曼(John von Neumann)将这种思路直接应用到核问题的中子扩散研究中[4]。Nicholas Metropolis 建议将这种方法命名为“蒙特卡洛”①。

1946 年 2 月，经过三年的研发，世界上第一台计算机——ENIAC 诞生。ENIAC 的问世与蒙特卡洛方法的发展密切相关。1947 年，约翰·冯·诺依曼对蒙特卡洛方法进行了拓展，以此来解决热核和裂变问题。同年，Ulam 和约翰·冯·诺依曼发明了逆向和接受－拒绝(accept－reject)技术来模拟均匀分布。1949 年，在由兰德、国家统计局和橡树岭实验室(Rand，NBS and the Oak Ridge laboratory)支持的关于蒙特卡洛的研讨会上，Metropolis 和 Ulam 共同发表了第一篇关于蒙特卡洛方法的论文[5]。

(二) MCMC 方法的开篇之作

MCMC 算法的起源与世界上第二台计算机——MANIAC 有关。1952 年初，在 Metropolis 指导下，MANIAC 在洛斯阿拉莫斯国家实验室诞生。Nicolas Metropolis 是创建洛斯阿拉莫斯国家实验室的 15 名科学家之一(1915 年生于美国，1941 年获得芝加哥大学物理学博士学位，1943 年 4 月来到洛斯阿拉莫斯国家实验室，直到 1999 年去世，在实验室工作了 56 年)，Nicolas Metropolis 既是物理学家又是数学家，既是美国科学院的院士又是美国数学会和物理学会的会士(Fellow)。20 世纪 50 年代初，借助改进的计算机设备，Micolas Metropolis 他与 Ulam 一起合作，负责设计研发氢弹[5]。

1953 年 6 月，在化学物理期刊(Journal of Chemical Physics)上 Metropolis 等人发表了 MCMC 方法的开篇之作《通过快速计算机器计算状态方程》(Equations of state calculations by fast computing machines)，论文的主要关注点是在 R^{2N} 内计算以下积分公式(类似于贝叶斯的后验分布)②：

$$\Im=\int F(\theta)\exp\{-E(\theta)/kT\}d\theta/\int\exp\{-E(\theta)/kT\}d\theta$$

其中 θ 表示 R^2 里的 N 个粒子，粒子的能量 E 表示为：

$$E(\theta)=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}V(d_{ij})$$

其中 V 是一个势函数，d_{ij} 是 θ 中粒子 i 和 j 之间的欧氏距离。

通过温度 T、玻尔兹曼(Boltzmann)常数 k 和正规因子

$$Z(T)=\int\exp\{-E(\theta)/kT\}d\theta$$

可以将波尔兹曼分布 $\exp\{-E(\theta)/kT\}$ 参数化。

由于 θ 是 $2N$ 维向量，所以不可能用数值积分计算。对于高维计算问题，由于 $\exp\{-E(\theta)/kT\}$ 对于粒子系统随机配置的大多数实现值结果都非常小，所以标准的蒙特卡洛方法不能正确地逼近积分值 $\Im$。

为提高蒙特卡洛方法的有效性，Metropolis 等人(1953)提出了 N 个粒子随机游走的修正方法，即对 N 个粒子中的每个粒子 (X,Y) 都做一个微小的随机移动，设 ξ_1,ξ_2 是 $(-1,1)$ 上的随机数，对 X 和 Y 分别做一个随机的变换，令：

$$X'=X+\alpha\xi_1 \qquad Y'=Y+\alpha\xi_2$$

产生新粒子 (X',Y')，并且假定新粒子以概率

$$\min\{1,\exp(-\Delta E/kT)\}$$

被接受，否则将重复原粒子。新粒子被接受后，可以计算新粒子结构和原粒子结构的能量差 ΔE。要特别注意的是，Metropoli 等人文章中一次只移动一个粒子，而不是一起移动所有的粒子，因此使这种算法看起来像 Gibbs 样本法的雏形[6]。

① 之所以用赌城' 蒙特卡洛' 来命名这种随机模拟算法，Metropolis(1987)说是由于乌拉姆(Ulam)的叔叔特别沉迷于摩纳哥的大赌场——Monte Carlo Casino(http://en.wikipedia.org/wiki/Monte_Carlo_method)。

② 此文章到 2013 年 9 月 10 日被引用次数为 25 191 次。

（三）Metropolis 算法的推广

Nicholas Metropolis 是物理学家，他在核物理研究中碰到的粒子分布高纬计算问题，对统计学提出了挑战。Hastings（1970）和他指导的唯一博生生 Peskun（1973）圆满地解决了这个难题，攻克了常规蒙特卡洛方法遇到的维度瓶颈问题，将 Metropolis 算法一般化，并推广为一种统计模拟工具，形成了 Metropolis－Hastings(M－H)算法。W. Keith Hastings1930 年出生于加拿大，1962 年博士毕业于多伦多大学数学系（当时统计学在数学系中），1964 年在贝尔实验室工作两年后，在母校数学系任统计教职，他的研究兴趣主要关注概率分布随机抽样的蒙特卡洛方法①。当时多伦多大学化学系 John Valleau 教授的研究团队面临一个“粒子系统的平均势能估计难题”，即在一个确定势场（defined potential field）中，100 个粒子组成了一个粒子系统，由于每个粒子有 6 维坐标，所以粒子系统就有 600 个维度。那么，如何使用 Metropolis 方法估计高纬粒子系统的平均势能。当 Hastings 得知这个问题可以利用马尔可夫链、并通过抽取高纬分布的随机样本轻松解决时，他意识到这个方法对于统计学的重要意义，于是便全身心投入此方法及其变化的研究中，完成了 MCMC 方法史上里程碑的论文——Monte Carlo Sampling Methods Using Markov Chains and their Applications（《马尔可夫链蒙特卡洛抽样方法及其应用》），发表在 1970 年的 Biometrika 上[7]。

M－H 算法相对于 Metropolis 方法而言，看起来更像是专业的统计模拟工具，最重要的是，M－H 算法不要求“建议分布函数”必须是对称的，从而应用起来更加灵活方便。Hastings 在文章里举了三个例子：第一个目标分布是泊松分布，采用的建议分布是随机游走；第二个目标分布是正态分布，建议分布是均匀随机游走，但此均匀分布的中心不是马氏链的当前值 θ_t，而是 $-\theta_t$；第三个目标分布是多元分布，Hastings 引进了 Gibbs 抽样的策略，每次只更新其中一个变量，这样的转移矩阵同样满足平稳条件，因为每次都离开了目标不变变量。三年后，Peskun 发表了题为 Optimum Monte－Carlo Aampling Using Markov Chains（《最优马尔可夫链蒙特卡洛抽样方法》）的文章，比较了 Metropolis 和 Barker 的接受概率的形式，也证明了在离散情形下 Metropolis 是最优的选择，这里的最优性可以通过经验平均值的渐近方差来表示[8]。

三、MCMC 方法的发展

从 Peskun 之后，在统计学研究领域里关于 MCMC 方法的研究沉寂了近十年之久。随后，在模式识别、图像分析和空间统计学等领域中，出现了关于 MCMC 方法应用的文章，其中 Geman and Geman 发表了在 MCMC 方法史上具有重要突破性的文章[9]。该文基于随机松弛（Stochastic relaxation）算法，采用 Gibbs 分布对图像的贝叶斯恢复进行了研究，提出了 Gibbs 采样的概念并将其引入到统计应用领域，Robert 和 Casella（2011）将此文称为“革命的种子”，吹响了 MCMC 方法革命的号角。

1987 年，Tanner 和 Wong 在论文中采用基于多个条件分布进行模拟的方法，这种思路等价于从联合分布进行模拟，基本具备了 Gibbs 采样的雏形。这篇文章被选入了美国统计学会杂志（Journal of the American Statistical Association）的讨论论文[10]。需要指出的是，Tanner 和 Wong 的论文虽然已经基本具备了 Gelfand 和 Smith（1990）文章的雏形，但与后者相比，其影响有限。原因之一是 Tanner 和 Wong 的论文似乎只是应用到缺失数据问题，特别是论文的题目中就有“数据补全”这样的字眼；另一原因是 Tanner 和 Wong 论文的理论基础不是马尔可夫链，而是函数分析，特别是要求马尔可夫转移核一致有界而且连续，也许正因如此影响了很多数学背景不强的潜在研究人员。

一系列以 MCMC 方法和贝叶斯为主题的学术会议，展示了 Gibbs 抽样爆炸式的发展过程。1986 年夏，Adrian Smith 做了关于分层模型（Hierarchical models）的系列学术演讲；1989 年 6 月，在魁北克省舍布鲁克市（Sherbrooke，Québec）举行的贝叶斯学术会议上，Adrian Smith 第一次详细阐释了 Gibbs 抽样的本质，这种方法的广度与深度震撼了与会者。1990 年，Gelfand 和 Smith 发表论文 Sampling－based approaches to calculating marginal densities（《基于抽样的边际分布计算方法》），将这种思想阐述得更为深刻和完整，成为主流统计学界大规模使用 MCMC 方法的真正起点[11]。

20 世纪 90 年代，是 MCMC 方法发展的黄金时期，并在理论研究方面获得了很多突破：1991 年，Alan Gelfand，Prenatal Goel 和 Adrian Smith 在俄亥俄州立大学（Ohio State University）举办了 MCMC 方法会

① 关于 Hastings 的资料较少，较详细的网络资料见 http://probability.ca/hastings/。

议，相关讨论成果都已成为 MCMC 领域非常有影响力的论文。1992 年的 5 月，皇家统计学会（Royal Statistical Society）召开了关于“Gibbs 抽样与其他 MCMC 方法”的会议，有四篇论文得到了众多学者的重视，并发表在 JRSSB 1993 年的第一期上。

在理论研究获得飞速发展的同时，MCMC 的应用研究也取得了可喜的成果，MCMC 方法之所以发展迅速，一个重要的原因就是它的简便实用。以 Gelfand 和 Smith 提出的非常简单的随机效应模型为例：

$$Y_{ij}=\theta_i+\varepsilon_{ij}\quad(i=1,\cdots,K;j=1,\cdots,J)$$

其中 $\theta_i\sim N(\mu,\sigma_\theta^2)$，$\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$，独立于 θ_i。

在频率学派看来，对方差分量进行估计比较困难；对于传统贝叶斯学派来说，这也是一个噩梦，因为此时的积分难以求得；对以 MCMC 为特征的现代贝叶斯分析来说，在给定参数 $\mu,\sigma_\theta^2,\sigma_\varepsilon^2$ 先验信息的情况下，问题可以很方便地通过 Gibbs 抽样解决。基于该随机效应模型，学者们利用 Gibbs 抽样迅速得到了贝叶斯线性混合模型的参数估计，该随机效应模型的其他应用还包括空间统计、变点分析、回归中的变量选择和基因组学等。

到了 20 世纪 90 年代中期，MCMC 理论渐趋成熟，理论突破开始变缓，但仍然取得了一些成果。许多动态系统可以用状态空间模型（state－space model）来描述，而序贯蒙特卡洛模拟（Sequential Monte Carlo，SMC）是分析状态空间模型的重要工具。Roberts 和 Rosenthal（2007）建立了 SMC 与 MCMC 的联系，提出了适应性 MCMC[12]。

Green（1995）提出的逆跳的马尔可夫链蒙特卡洛模拟（Reversible Jump Markov Chain Monte Carlo，RJMCMC），可被视为 MCMC 方法第二次革命的开端，所构建的马氏链不仅可以在一个模型的参数空间内进行转移，还可以在不同模型（维度可以不同）之间跳跃，从而为贝叶斯模型选择提供了强大工具[13]。Richardson 和 Green（1997）将 RJMCMC 应用到混合阶估计中；Brooks，Giudici 和 Roberts 提出了提高 RJMCMC 转移效率的方法；Marin 和 Robert 将 RJMCMC 应用到变量选择中。由于 MCMC 方法得到的并非独立样本，具有自相关性，因此不能直接用于参数估计，尤其是不能用于直接估计参数的标准误差。为了克服自相关性，Hobert 等人提出在原始马氏链转移链基础上间隔抽取构成新的样本序列，从而克服自相关性，可以得到参数估计以及估计误差[14]。

四、MCMC 方法直观性实例

MCMC 方法实质上就是利用马尔可夫链进行蒙特卡洛模拟。它可以分解成两个 MC：前一个 MC 是马尔可夫链（Markov Chain），后一个 MC 是蒙特卡洛方法（Monte Carlo）。在此，按照蒙特卡洛方法、马尔可夫链和 Metropol－Hastings 算法的次序，介绍 MCMC 方法的直观性实例。

（一）蒙特卡洛方法（Monte Carlo）：圆周率 π 的估计

蒙特卡洛方法与数值计算中的其他确定性算法不同，是一种以概率统计理论为基础的非确定性随机模拟算法，也称统计模拟方法。它使用随机数（或伪随机数）解决很多不确定性的计算问题，将所求解的问题同一定的概率分布相联系，并用电子计算机实现统计模拟或抽样，从而获得问题的近似解。一般来说，蒙特卡洛方法可以粗略地分成两类：

一类是所求解的问题本身具有内在的随机性，借助计算机的运算能力可以直接模拟这种随机的过程。例如在核物理研究中，分析中子在反应堆中的传输过程。中子与原子核作用受到量子力学规律的制约，人们只能知道它们相互作用发生的概率，却无法准确获得中子与原子核作用时的位置以及裂变产生的新中子的行进速率和方向。科学家依据其概率进行随机抽样得到裂变位置、速度和方向，进而在模拟大量中子的行为后经过统计就能获得中子传输的范围，以此作为反应堆设计的依据。

另一类是所求解的问题可以转化为某种随机分布的数字特征值，比如随机事件出现的概率，或者随机变量的期望值。通过随机抽样的方法，以随机事件出现的频率估计其概率，或者以样本的数字特征估算随机变量的数字特征，并将其作为问题的解。这种方法多用于求解复杂的多维积分问题（http://zh.wikipedia.org/）。

例如，假设要计算一个不规则图形的面积，图形的不规则程度和分析性计算（比如积分）的复杂程度成正比，与利用定积分方法计算不规则图形面积的方法（任意细分后求和再取极限得到精确值）不同，蒙特卡洛方法的简单直观思路为：假想有一袋小玻璃珠子，把玻璃珠子均匀地朝这个图形上撒，再数此图形中有多少颗玻璃珠子，而玻璃珠子的数目就近似代表图形的面积。当玻璃珠子越小、撒得越多的时候，结果就越精确。借助计算机程序可以生成大量均匀分布坐标点（代替玻璃珠子），然后统计出图

形内的点数，通过它们占总点数的比例和坐标点生成范围的面积，就可以求出图形面积。

利用蒙特卡洛方法，可以非常简便地近似计算圆周率 π：在一个边长为单位 1 的正方形内，画出一个半径为 1 的 1/4 圆，见图 1。

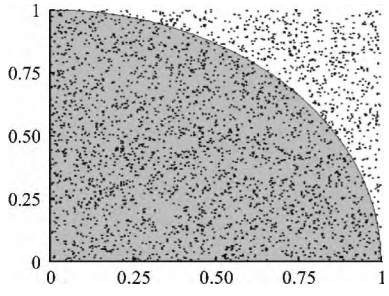

图 1　利用蒙特卡洛方法计算圆周率图

向正方形内随机投点，设总投点数为 N；统计圆内的点数，用 n 表示，则 n 与总点数 N 的比值就近似等于 1/4 圆的面积($\pi/4$)和正方形面积(1)之比，即 $\pi/4$，所以 $\pi\approx 4n/N$。

通过计算机来模拟上述随机投点过程：首先，每次随机生成两个服从均匀分布的 0 到 1 之间的随机数，设为 x 和 y，分别以 x 和 y 为横纵坐标，就得到图中的一个随机投点；其次，再判断这个点是否在圆内，而固定 x 只需判别 y 是否小于或等于 $\sqrt{1-x^2}$ 即可；最后，生成 N 个随机点，加总圆内点的总数 n，计算 $4n/N$ 即可。当随机点 N 取值越大时，其结果越接近于圆周率(当投点达到 3 万次时，用蒙特卡洛方法可以得到 $\pi\approx 3.143\ 6$)。

(二) 马尔可夫链 (Markov Chain)：股市的表现

理解了蒙特卡洛方法的基本思路，再看马尔可夫链及其平稳分布的相关理论。马尔可夫链(以下简称马氏链)的定义如下：

$$P(X_{t+1}=x\mid X_t,X_{t-1},\cdots)=P(X_{t+1}\mid X_t)$$

简言之，即马氏链的下一个状态只依赖于当前状态，而与已经过去的状态没有关系。它是介于独立和相关之间的一种理想形式，例如在研究时间序列时，为了便于分析，希望昨天、今天和明天表示的三种随机状态(或变量)完全独立，但实际上可能它们完全相关，变量之间相互独立太理想，完全相关又太复杂，马氏链对此进行了简化，假设明天只与今天相关，而与昨天无关(独立)①。换言之，即在马氏链过程中，在给定当前知识或信息的情况下，只有当前的状态用来预测将来，过去(即当前以前的历史状态)对于预测将来(即当前以后的未来状态)是无关的。

下面以来自 wikipedia 的例子来说明马氏链及其平稳性：假设股票市场在一周内的表现有三种可能状态：熊市—状态 1、牛市—状态 2、平衡市—状态 3。股市行情的转化见图 2。

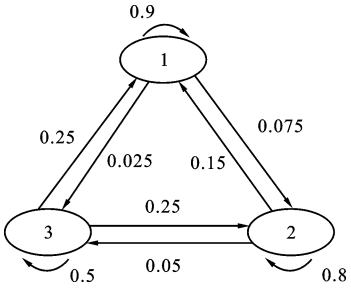

图 2　马氏链三种状态转移示意图

由图 2 可见，如果当前股市为牛市，则下周为熊市的概率为 7.5%、平衡市的概率为 2.5%、依然为牛市的概率为 90%，其他行情转化与之类似，则该马氏链的一步转移概率矩阵为：

$$P=\begin{bmatrix}0.9 & 0.075 & 0.025\\ 0.15 & 0.8 & 0.05\\ 0.25 & 0.25 & 0.5\end{bmatrix}$$

考虑三种初始情况：当前行情为牛市、当前行情为熊市、当前行情为平衡市，它们对应的初始概率分布为：$\pi_0=(1,0,0)$、$\pi_0=(0,1,0)$、$\pi_0=(0,0,1)$。分析在这三种情况下股市行情在未来 30 个星期之内的转移变化情况，见图 3。

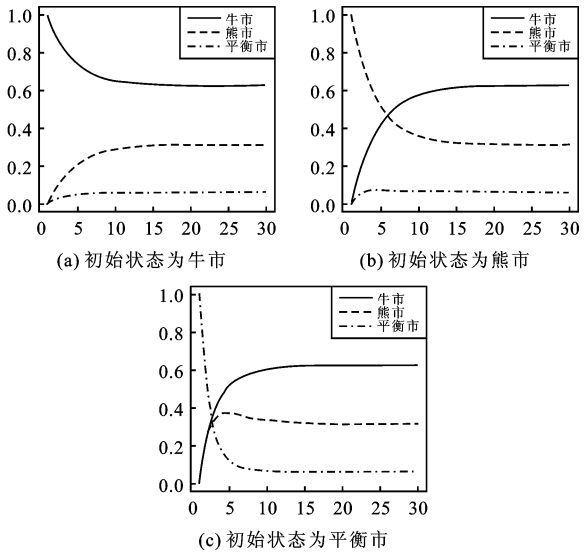

图 3　马氏链的平稳分布图

从图 3 发现，从第 20 周开始，三种情况下分布

① 马氏链直观的解释是荷花池中青蛙的跳跃过程；有趣的解释是，芭蕾舞团在招考少儿芭蕾舞女演员时，预计她未来的体型，只需看她的母亲而不需要看她的姥姥。

都开始收敛。最为奇特的是，不论初始状态是三种行情中的哪一种，最终都会收敛到平稳概率分布 $\pi=(0.625,0.312,0.063)$。也就是说，收敛行为与初始状态无关，收敛行为主要是由概率转移矩阵 P 决定的。

所有的 MCMC 方法均以马氏链定理为理论基础。根据马氏链定理，满足一定条件的马氏链的转移序列会形成一个具有平稳分布的样本。一个自然的想法是，如果想获得某个分布（称为目标分布）的样本，是否可以设计一个马氏链（其平稳分布为目标分布），然后采集这个马氏链的转移序列来作为目标分布的随机样本呢？答案是肯定的。

至此可以看到，用 MCMC 方法进行随机模拟的关键是设计一条马尔科夫链，使其平稳分布与目标分布一致，而纷繁众多的 MCMC 方法之间的区别就在于从不同的角度出发设计马氏链。

（三）Metropol－Hastings 算法："千岛湖植物考察"

为了更加生动形象地说明 Metropol－Hastings 算法，用"千岛湖植物考察"为例进行说明①。

某旅游风景点"千岛湖"由大大小小上千个岛组成，千岛湖风景区植物种类非常丰富。植物学家要去各小岛进行植物考察研究，考察的时间安排是依据各岛上植物种类数来确定的，植物种类数多的岛屿考察时间长，植物种类数少的岛屿考察时间短。但在考察之前，植物学家并不清楚各个岛的植物种类数，甚至不知道一共有多少个岛，所以不能事先确定留在每一个岛上的天数，也无法得知在各岛考察时间占总考察时间的比率——考察时间分布。那么，如何帮助植物学家安排在各岛的行程路线和考察时间？为了用图形进行直观介绍，将岛屿简化为 9 个，并假设岛屿一字排开（如图 4 中 a），植物学家来到某一岛屿后，不仅可以了解该岛上的植物种类数，还可以了解相邻岛的植物种类数。

精通 MCMC 的统计学者向植物学家建议用 M－H 算法的思想，通过"两步随机试探法"来决定考察路线和停留时间。植物学家到达某岛屿后，由于岛屿是一字排开，所以面临三种选择（三个状态）：一是继续停留在岛屿上（保持原状态）；二是到临近左面岛屿（转移到前状态）；三是到临近右面岛屿（转移到后状态）。这三种状态可以转化为两种随机选择：一是去或留，二是左或右。但 M－H 算法突破常理，用"逆向思维"来确定选择过程，先确定左或右，再决定去或留（究其原因可能是先基于左或右岛的情况，再与当前岛屿比较，最后决定是走还是留）。"两步随机试探法"的具体实施过程如下：

第一步，随机确定方向—向左还是向右：植物学家在某岛屿上考察时（假设在第 5 个岛屿），就"可能"计划前往邻近岛屿（第 4 个岛或第 6 个岛），到底是向左还是向右呢？植物学家通过随机掷一枚均匀硬币来决定，如果硬币为正面，就计划向左去第 4 个岛；如果硬币为反面，则计划向右去第 6 个岛。

第二步，基于第一步的信息，部分随机确定意愿—出发还是停留：假设第一步硬币为正面，根据第一步规则，植物学家应计划"可能"向左去第 4 个岛，但"可能"去，也"可能"不去。如何确定呢？统计学者给出的策略是，从当前岛（第 5 个岛）的管理者那里可以得到目标岛（左面第 4 个岛）的植物种类数，比较目标岛和当前岛的植物种类数量，再决定是否出发。

如果目标岛的植物种类数比所处当前岛的植物种类数多，那么植物学家就一定前往目标岛进行考察研究；如果目标岛的植物种类数比所处当前岛的少，那么植物学家又需要"另外一步随机试探法"来确定意愿—依概率出发或是停留。

什么是依概率出发呢？举例说明：假设目标岛有 150 种植物，而所处当前岛有 200 种植物，则前往目标岛的概率就为 0.75（150/200），继续留在当前岛的概率为 0.25（1－0.75）。那么植物学家到底是出发还是停留呢？似乎还没有确切答案，此时的随机判断方法是：在地上画一条 1 米长的线段，然后随机向该线段内投一个石子，如果石子落在 0 和 0.75 米之间，则出发去目标岛考察研究；如果石子落在 0.75 和 1 之间，则继续留在当前岛。

以上植物学家的随机试探方法本质上就是马氏链的转移矩阵，那么该植物学家设计的马式链的平稳分布（各个岛屿的考察研究的时间分布）是否与其目标分布（即各个岛屿的植物种类数分布）相同呢？用 R 软件来模拟结果。

① 本例受 John K. Kruschke(2011)Doing Bayesian Data Analysis：A Tutorial with R and BUGS 书中" 总统候选人巡岛演讲拉选票" 章节启发。

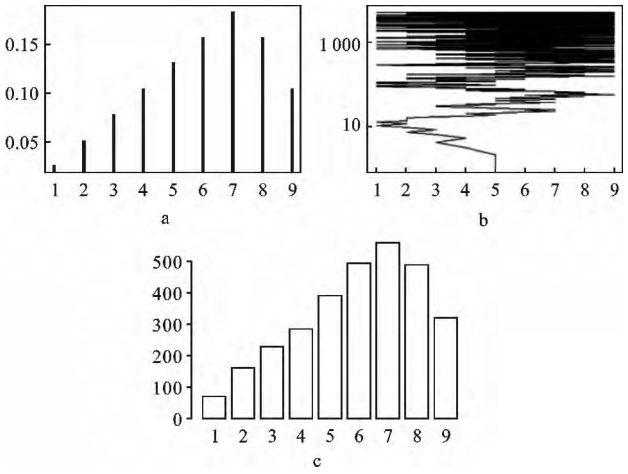

图 4　a 目标分布　b 行动轨迹　　c 频数分布

图 4 中 a 显示出各个岛的相对植物种类数，这也是植物学家在各个岛屿逗留时间的目标分布。图 4 中 b 显示出植物学家按照前述“两步随机试探法”进行了 5 000 次转移，转移轨迹如图 4b 所示。注意图 4 中 b 的移动轨迹：第一天，植物学家在最中间的岛即第 5 个岛上；第 2 天它仍然停留在第 5 个岛屿上；第 3 天转移到了第 4 个岛屿上；第 4 天转移到了第 3 个岛屿上；第 5 天又回到了第 4 个岛屿上……图 4 中的 c 显示在各个岛屿植物考察研究的频数直方图，可以看出各岛屿被考察的相对频数，与图 4 中 a 显示出的各个岛的相对植物种类数相似，两图走势是一样的，这也说明植物学家设计的马式链的平稳分布与目标分布相同。

上述过程反映了 M－H 算法抽取样本的基本原理。“千岛湖植物考察”的例子只是 M－H 算法的一种特殊情况：目标分布是一维离散分布，建议分布也是离散的，分别以 0.5 的概率建议向左或向右移动。M－H 算法更一般的是研究高维连续型分布，建议分布也更复杂，但实质与此例是相同的。以下给出 M－H 算法的一般步骤：

1. 任意设定马氏链的起始点 $X_0=x_0$。

2. 构造建议分布 $q(x'|x)$。

3. 链在时刻 t 处于状态 x_t，即 $X_t=x_t$，$t=0,1,2,\cdots$，循环以下步骤直至马氏链达到平稳状态：

(1)从建议分布 $q(x'|x_t)$ 中产生一个潜在的转移 $x_t\to x'$。

(2)从 $U(0,1)$ 中产生随机数 u。

(3)$X_{t+1}=\begin{cases}x' & u<\alpha(x_t,x')\\ x_t & u\geqslant\alpha(x_t,x')\end{cases}$

其中　$\alpha(x_t,x')=\min\left(1,\dfrac{q(x'|x_t)\pi(x')}{q(x_t|x')\pi(x_t)}\right)$，也就是说若 $u<\alpha(x_t,x')$，则接受建议状态，否则保持原状态。

$\alpha(x_t,x')=\min\left(1,\dfrac{q(x'|x_t)\pi(x')}{q(x_t|x')\pi(x_t)}\right)$，此等式可以保证马尔可夫链的平稳分布是目标分布 $\pi(x)$。Metropolis 算法将建议分布限制在对称分布：$q(x_t|x')=q(x'|x_t)$，因此 $\alpha(x_t,x')=\min\left(1,\dfrac{\pi(x')}{\pi(x_t)}\right)$，这也是 Metropolis 算法与 M－H 算法的区别之处。

不断循环以上过程，为什么就得到了目标分布的随机样本了呢？尤其是从一个与 $\pi(x)$ 毫不相关的建议分布 $q(x'|x)$ 产生随机数，然后经过一个简单的判断过程，最终竟然得到了服从目标分布的随机数！事实上，建议分布 $q(x'|x)$ 和接受概率 $\alpha(x_t,x')$ 共同扮演了马氏链的转移核（类似于离散马式链的转移矩阵）的作用，根据马氏链定理，可证明其会收敛于目标分布 $\pi(x)$（见图 5）。

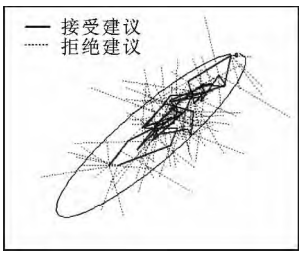

图 5　Metropolis－Hastings 采样轨迹图

图 5 所示为对一个二元正态分布进行 MCMC 随机模拟的路径（实线），其中实线表示的是接受建议分布产生的转移建议，细虚线表示建议分布产生的转移建议被拒绝。由此可见，建议分布 $q(x'|x)$ 产生的转移建议就像脱缰的“野马”，四处乱窜，而接受概率就像一条随机出现的“无形缰绳”，时时牵拽着这头野马，让它在目标分布所形成的“水草丰美”的区域内漫步。

五、MCMC 方法、R 软件与现代贝叶斯分析

贝叶斯定理简捷直观，贝叶斯推断清晰自然，由于应用领域广泛，受到了众多学者的青睐。除了先验分布的选择之外，贝叶斯分析在发展过程中遭遇最大的瓶颈就是后验分布的计算，特别是面对复杂高维的问题时，难以用解析的方法求得明确的后验分布。

计算技术的进步和 MCMC 方法的发展，重新唤醒了人们对贝叶斯分析这个古典统计思想的认识。Metropolis 等学者另辟蹊径，没有沿用先通过

复杂的数学分析方法积分得到后验分布，再求其期望值、中位数和置信区间这些统计特征值的传统方法，而是充分利用现代计算技术，基于马尔可夫理论，使用蒙特卡洛模拟方法，回避后验分布表达式复杂的计算，创造性地使用 MCMC 方法，直接对后验分布的独立随机样本进行模拟，再通过分析模拟样本来获得相关统计特征值信息。MCMC 方法的发展，解开了贝叶斯分析计算的“紧箍咒”，现代贝叶斯分析重获春天，再次生机勃发。

在近年颇为流行的统计分析软件 R 中，基于 MCMC 采样的现代贝叶斯分析越来越方便，许多可以实现 MCMC 的 R 程序包如雨后春笋般出现。在 R 主站上搜索关于 MCMC 的程序包，目前共有 103 个，而相对于传统的贝叶斯计算软件包（如 BUGS、WinBUGS、OpenBUGS 等），这些程序包功能更加强大与灵活。R 中 MCMC 相关程序包大致可以分为 5 类：

1. 用于一般统计模型拟合。比较核心的包有：mcmc 包，可以自定义后验分布函数，并从中进行 M－H 采样；MCMCpack 包，包含了许多经典统计模型的贝叶斯分析工具；bayesSurv 包，提供了进行贝叶斯生存回归分析的函数；DPpackage 包，所包含的函数可以完成贝叶斯非参数、半参数分析；bayesm 包，可以用于微观经济研究中的各种贝叶斯分析，如线性回归模型、多项 logit 模型、多项 probit 模型、密度估计等。

2. 用于特殊统计模型的贝叶斯分析。例如 MCMCglmm 包，专门求解贝叶斯广义线性混合模型；lmm 包，拟合贝叶斯线性混合模型；BayesTree 包，完成贝叶斯回归树分析。

3. 将 R 与其他采样器连接起来的程序包。R 的优势在于方便灵活的统计分析，进行超大规模的模拟运算可能逊色于专业的采样器（如 WinBUGS，BUGS）。R 提供了与各种专业采样器连接起来的接口，并封装在程序包中，可自由加载使用。在 R 中定义贝叶斯分析模型，然后利用接口传递给采样器并在采样器中进行 MCMC 模拟，模拟结束后，再利用接口将结果传递到 R 中，在 R 中对模拟结果进行分析。这类包有：R2WinBUGS（与 WinBUGS 的接口）；BRUGS（与 OpenBUGS 的接口）；rjags、R2jags 与 runjags（与 JAGS 的接口）。

4. 处理 MCMC 样本的包。coda 包是核心的处理 MCMC 样本的包，可以用于收敛诊断与输出分析，几乎所有关于 MCMC 的包都会依赖于 coda 包，足见此包的重要性。

5. 辅助学习贝叶斯分析课程的包。这是 R 最人性化的一面，不仅提供了分析问题的工具，还是学习贝叶斯分析方法的良师益友。在有关贝叶斯分析的参考书目里，经典的当属《Bayesian Data Analysis》（《贝叶斯数据分析》Gelman，2003），为了方便学习，该书对应的包 BayesDA 可以从 CRAN 加载使用，该包涵盖了书中所涉及的数据集和函数，对于贝叶斯爱好者来说，这无异于一学习利器。类似的包还有 LearnBayes（《Bayesian Computation with R》《基于 R 的贝叶斯计算书中的对应包》Jim Albert，2009.），此包由浅入深，其中编写的 rwmetrop() 和 rgibbs() 两个函数对于理解 M－H 算法和 Gibbs 采样很有帮助。

六、结论与启示

MCMC 方法产生于物理和原子弹工作的研究中，在计算物理学（如粒子输运计算、量子热力学计算、空气动力学计算）等领域应用广泛，之后基于 MCMC 的现代贝叶斯分析在生物医学、金融工程学和宏观经济学等方面的应用，也得到了蓬勃的发展[15－16]。

2013 年被称为“大数据元年”，随着信息技术的疾速发展，自然科学中基因测序高维数据和社会科学中互联网海量数据对统计分析理论和方法提出了严峻的挑战。统计学向何处去？“离数学越来越远，与信息科学越来越近”的“统计与信息”相结合的趋势似乎不可阻挡；基于云计算的随机模拟算法、贝叶斯统计与频率统计相融合的大数据分析新范式也初见倪端。

MCMC 方法的发展和现代贝叶斯分析的复兴改变了我们解决问题的思路，它意味着一种思维范式的转换。现代贝叶斯拓宽了分析的视野，即从样本空间的“有限范围”放眼到样本外“无边区域”；MCMC 方法改变了计算的重点，即从“精确解析”到“算法模拟”。MCMC 带领我们进入一个全新的统计世界，在那里“小样本”变成了“大数据”，“准确”也被高精度的“近似”所替代。

60 年前，当 Ulam 和约翰·冯·诺依曼发明蒙特卡洛方法时，他们无法想象目前在处理基因组学和气候学的巨型数据集时使用 MCMC 方法；250 年前，当理查德·普莱斯宣读贝叶斯论文时，他也无法预测 18 世纪的贝叶斯定理会成为 21 世纪 Google 计算的新力量（用 Google 去搜索“18 世纪的贝叶斯

定理成为 Google 计算的新力量”)；当今的统计学界是否也会有一种方法将被 60 年后的学者采用？是否也将有一个定理会被 250 年后的同仁们所纪念？这个问题似乎可以基于 MCMC 方法，用贝叶斯定理去分析。

参考文献：

[1] 陈希孺. 数理统计学简史[M]. 长沙：湖南教育出版社，2002，50－51.

[2] 刘乐平，袁卫. 现代贝叶斯分析与现代统计推断[J]. 经济理论与经济管理，2004(6).

[3] Andrieu C，De Freitas N，Doucet A，et al. An Introduction to MCMC for Machine Learning[J]. Machine learning，2003(1).

[4] Robert C. Casella G. A Short History of Markov Chain Monte Carlo：Subjective Recollections from Incomplete Data[J]. Statistical Science，2011(1).

[5] Metropolis N，Ulam S. The Monte Carlo Method[J]. Journal of the American Statistical Association，1949(4).

[6] Metropolis N，Rosenbluth A W，Rosenbluth M N，et al. Equations of State Calculations by Fast Computing Machines [J]. Journal of Chemical Physics，1953(2).

[7] Hastings W K. Monte Carlo Sampling Methods Usng Markov Cins and their Applications[J]. Biometrika，1970(1).

[8] Peskun P H. Optimum Monte－Carlo Sampling Using Markov Chains[J]. Biometrika，1973(3).

[9] Geman S，Geman D. Stochastic Relaxation，Gibbs Distributions and the Bayesian Restoration of Images[J]. IEEE，Transactions on Pattern Analysis and Machine Intelligance，1984(6).

[10] Tanner M，Wong W. The Calculation of Posterior Distributions by Data Augmentation[J]. Journal of the American Statistical Association，1987(2).

[11] Gelfand A，Smith A. Sampling Based Approaches to Calculating Marginal Densities[J]. Journal of the American Statistical Association，1990(4).

[12] Roberts G O，Rosenthal J S. Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms[J]. Journal of applied probability，2007(4).

[13] Green P. Reversible Jump MCMC Computation and Bayesian Model Determination[J]. Biometrika，1995(4).

[14] Hobert J P，Jones G L，Presnell B，et al. On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo [J]. Biometrika，2002(4).

[15] 丁东洋，周丽莉. 基于贝叶斯方法的信用评级模型构建与违约概率估计[J]. 统计与信息论坛，2010(9).

[16] 周丽莉，丁东洋. 基于 MCMC 模拟的贝叶斯分层信用风险评估模型[J]. 统计与信息论坛，2011(12).

Development of MCMC Methods and Revival of Modern Bayesian
Celebrating 250 Years of Bayes's Theorem

LIU Le-ping，GAO Lei，YANG Na

(Department of Statistics，Tianjin University of Finance and Economics，Tianjin 300222，China)

Abstract：It is important issues of statistics in the era of big data that how advances in information science influence the development of statistical theory and methods. The paper briefly describes the origins and development of MCMC methods，and then through MCMC methods intuitive examples－－Qiandao Lake Plant Visits，discuss the development of MCMC methods to modern Bayesian，in order to commemorate the 250th anniversary of Bayes' theorem discovery. New ideas in stochastic simulation based on cloud computing，statistics combined with information science，Bayesian statistics mingled with frequency will change our big data research methods.

Key words：MCMC；Metropolis－Hastings；Qiandao Lake Plant Visits；Bayes's theorem

(责任编辑：郭诗梦)