

doi: 10.3969/j.issn.1000-8349.2013.03.05

伽玛暴数据处理中的贝叶斯方法

卜庆翠, 陈 黎, 李兆升, 王德华

(北京师范大学 天文系, 北京 100875)

摘要: 贝叶斯推断是建立在贝叶斯定理上的一种参数估计方法。根据贝叶斯定理, 当根据经验, 对待估计的参量 θ 的分布密度 $p(\theta)$ (称为“验前分布”) 有所了解时, 在给定观测数据 D 的情况下, 可以计算出待估参数 θ 的“验后分布”—— $p(\theta|D)$ 。 $p(\theta|D)$ 反映了观测结果对 $p(\theta)$ 的修正。所有贝叶斯统计推断都是以验后分布为基础的。贝叶斯估计法是数据分析中的有力工具, 其在伽玛暴 (GRBs) 数据分析窗口展现了多方面的应用, 例如分析光变结构, 确定参数分布, 检验是否存在某种谱线特征, 比较和选取模型, 等等。

关 键 词: 贝叶斯估计; GRBs; 统计学

中图分类号: P172.3 **文献标识码:** A

1 引 言

从数据中挖掘信息需要丰富的统计学知识。但天文研究中的统计手段大多局限在一些经典的数据处理方法上, 即所谓 Fisher 模式或频率统计 (frequentist) 的方法^[1,2]。天文学家们最常用的工具是: 时变分析的傅里叶变换, 最小二乘法, χ^2 拟合优度检验, 非参数的柯尔莫哥洛夫检验和斯米尔诺夫双样本检验 (K-S 检验), 以及针对多变量的主成分分析法^[3]。但是由于受“高斯综合征”^①(Gaussian syndrome) 的影响, 使得在用“ \sqrt{n} 准则”来估计测量误差时, 只停留在一个比较低层次的水准上^[4]。天文学家偏爱频率统计最主要的原因就是它的便利性。当遇到一种新情况时, 频率统计可以简便地使用最大似然函数法, 从经验上来看, 人们认为出现大误差的几率较小, 并且较容易得到最佳近似值^[5]。虽然不少人认为频率统计比贝叶斯估计法更具有优越性, 但是物理学家们从直觉上认为贝叶斯估计比频率统计更有优势。

近 10 年来, 越来越多的天文学家意识到贝叶斯估计的潜力。一系列关于“现代天文统

收稿日期: 2013-01-15; 修回日期: 2013-05-02

资助项目: 国家自然科学基金 (NSFC 11173024, NSFC 10778716); 973 项目 (2009CB824800); 中央高校基本科研业务费专项资金

^①源于对数据整合规则不加批评地使用。比如习惯性地利用误差平方的倒数来计算权重, 或习惯性地把用拟合曲线所得出的理论数值与原始数据之间差值的平方和等同于 χ^2 来计算。

计中的挑战”的会议在宾夕法尼亚州立大学召开^[6]。James Berger, William Jefferys, Thomas Loredo 和 Alanna Connors 等人都为贝叶斯估计在天文学中的应用作了开拓性的贡献^[7-11]。近年来,贝叶斯估计被越来越多的天文学家用来研究伽玛暴 (GRBs)、引力波、宇宙学常数、类星体,以及超新星和星系分类等。根据 NASA 天体物理数据查询系统 (ADS) 提供的数据做出的统计显示,在过去 10 年发表的论文中,其中标题和摘要中都提及“Bayesian”的文章数目由 121 篇增加至 1369 篇 (见图 1)。

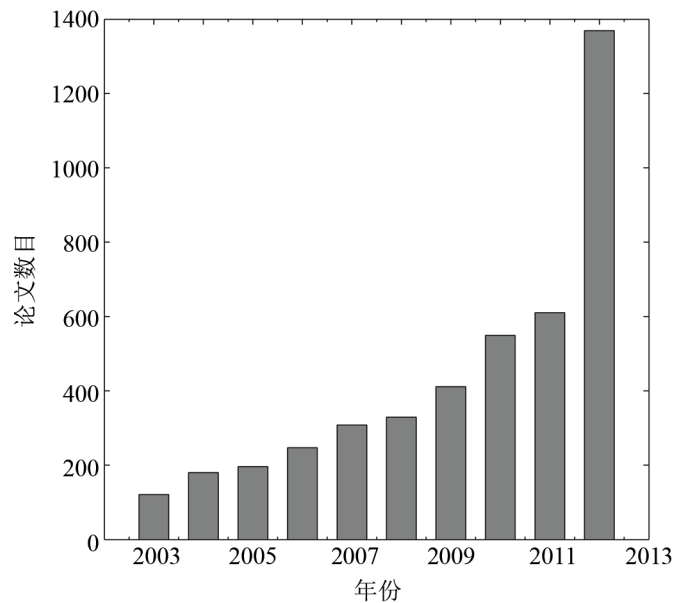


图 1 关于贝叶斯理论的论文数目近 10 年的逐年统计结果

本文的目的不是详细讲解贝叶斯理论,而是介绍用贝叶斯估计处理 GRBs 数据时得到的一些例子,展示如何利用贝叶斯估计来处理数据。

2 频率统计的基本概念

频率统计建立在随机变量 X 的样本服从独立同分布的假设基础上。一个随机变量 X , 其分布函数包含未知成分, 可记为:

$$F(x, \theta) = P(X \leq x) \quad , \quad (1)$$

这里 θ 用来确定分布函数 F , 称为参数。

统计学的任务就是通过对随机变量 X 的抽样 X_1, X_2, \dots, X_n 来估计 θ 。频率统计中, 在对参数 θ 进行估计的时候, 一般有两种典型的方法: (1) 在给定样本容量 n 的基础上, 最优化风险函数; (2) 样本容量 $n \rightarrow \infty$ 时, 最优化渐近近似函数^[12]。频率学派视概率为事件大量

重复实验频率的稳定值, 如果随机变量 X 的概率分布密度形式为 $p(x, \theta_1, \dots, \theta_m)$, 那么参数 $\theta_1, \dots, \theta_m$ 是客观存在的常数, 可以通过对随机变量 X 的大量抽样来估计。常用的参数估计方法主要有 3 种: 点估计、最小方差无偏估计以及区间估计。最大似然估计是统计学奠基人之一费希尔 (Fisher) 最先提出来的一种广泛使用的参数点估计方法, 其直观想法为: 如果试验的结果得到样本观测值 x_1, x_2, \dots, x_n , 则我们应当选取使这组样本观测值出现的可能性最大的 θ 值作为 θ 的估计值, 也就是选取似然函数

$$L = \prod_{i=1}^n p(x_i, \theta) \quad (2)$$

的最大值点 $\hat{\theta}$ 为参数 θ 的估计值。

对于同一个未知参数, 不同的方法得到的估计量可能不同, 于是产生了一系列评价参数好坏的标准: 无偏性 ($E\hat{\theta} = \theta$)、均方误差 (mean squared error)、有效性。最小方差无偏估计就是建立在无偏估计基础上的一种参数估计方法, 目的是在知道一个无偏估计的基础上, 构造一个新的无偏估计, 使得其方差比原来的方差小并在一切无偏估计类中能达到最小。最后, 为了回答估计值的可靠度与精度问题, 引入了区间估计的方法。对于给定的置信度, 可以根据样本观测值来确定未知参数 θ 的置信区间 (confidence interval)。而当我们需要根据样本对随机变量分布性质的某个断言进行检验时, 则引入了假设检验的概念。在频率统计中, 假设检验的方法一般采用适用于大样本的 χ^2 拟合优度检验、适用于小样本的 K-S 检验、P 值检验、似然比检验以及 F 检验等。

需要强调的是, 频率统计学认为概率具有“客观性”, 统计推断的信息完全来自样本, 不存在“主观臆断的”先验信息, 因而要求大量重复试验之后的频率才能获得可靠的估计。在置信区间的解释上, 常常会被误解。例如, 误认为参数 θ 的置信度为 90% 的置信区间就是该区间包含 θ 的概率为 90%。事实上, 置信度为 90% 的置信区间的含义是: 如果有 100 次抽样, 大约有 90 次得出的置信区间可能覆盖 θ 。

此外, 在假设检验中, 经典统计学对原假设 (模型 1) H_0 和备选假设 (模型 2) H_1 的判定标准是: 若 H_0 为真时, 小概率事件发生, 则拒绝原假设 H_0 。由于参数被认为是常数, 所以不存在两个模型概率的大小。这和下文的贝叶斯模型选取有明显不同。

3 贝叶斯理论的基本概念^[12,13]

3.1 贝叶斯定理

贝叶斯公式描述了随机事件的条件分布密度 $p(D|\theta)$, $p(\theta|D)$ 和分布密度 $p(\theta)$ 之间关系:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \quad (3)$$

其中 $p(\theta)$ 是参数 θ 的验前分布 (在不引起误读的情况下, 后文中“分布密度”简写为“分

布”，用小写字母 p 表示，而概率则由 P 表示)；它给出的是先于观测数据 D 之前的经验分布。 $p(\theta|D)$ 是结合验前分布 $p(\theta)$ 和观测数据 D 得出的 θ 的验后分布； $p(D|\theta)$ 是观测数据 D 的分布模型，即似然函数。验前分布和验后分布是贝叶斯估计专属的说法。形式上看，式 (3) 描述了如何利用数据 D 中包含的信息来更新验前信息。一个“信息型验前分布”可以给出一个变量详细确切的信息。当验前信息未知时，一般采用所谓“无信息型验前分布”，这种验前分布一般形状平缓，对最后的分析结果不会有太大影响。最简单的无信息型验前分布即均匀分布。例如^[14]，假设观测某源得到一组带有背景污染的数据和一组纯背景的观测数据 D 。源曝光时间记为 τ_S ，纯背景曝光时间记为 τ_B 。由于计数 D 服从泊松分布，设 λ_B 和 λ_S 分别表示源的背景和源光子所服从的泊松分布的指数，则似然函数可表示为：

$$p(D|\lambda_B, \lambda_S) = \frac{e^{-(\lambda_B + \lambda_S)} (\lambda_B + \lambda_S)^D}{D!}, \quad (D = 0, 1, 2, \dots), \quad (4)$$

通常把 λ_S 和 λ_B 的验前分布取为 Γ 分布，或者，把 λ_B 看做已知，认为 λ_S 的验前分布具有形式 $p(\lambda_S) \propto \frac{1}{(\lambda_S + \lambda_B)^m}$ ^[15]，其中 $\lambda_S \geq 0$ ， $\lambda_B \geq 0$ ， $0 \leq m \leq 1$ ，不管验前分布取哪种， λ_S 的边缘分布总可以由积分

$$p(\lambda_S|D) = \int_0^\infty p(\lambda_S, \lambda_B|D) d\lambda_B \quad (5)$$

计算出来。如果直接减去背景可能会得出负的计数，但是贝叶斯估计不存在这种麻烦。

3.2 贝叶斯推断

贝叶斯推断建立在验后分布 $p(\theta|D)$ 上^[12]，例如利用验后分布计算贝叶斯估计值 $\hat{\theta}_B$ 和贝叶斯方差 $\sigma^2(\theta_B)$ ：

$$\begin{cases} \hat{\theta}_B = E(\theta) = \int \theta p(\theta|D) d\theta \\ \sigma^2(\theta_B) = E(\theta - E(\theta))^2 = \int |\theta - E(\theta)|^2 p(\theta|D) d\theta \end{cases}, \quad (6)$$

$\hat{\theta}_B$ 是满足与 θ 偏差最小的估计。如果参数中包含一部分计算时不关心的参数 θ' ，就可以把它们积分掉，通过计算验后分布 $p(\theta, \theta'|D)$ 的边缘分布：

$$p(\theta|D) \propto \int p(\theta, \theta'|D) d\theta', \quad (7)$$

进而计算模型中指定参数的贝叶斯可信区间 (credibility interval)^[12]。可信度为 ξ 的可信区间 $[\theta_1, \theta_2]$ 满足：

$$\xi = \frac{\int_{\theta_1}^{\theta_2} p(\theta|D) d\theta}{\int_{\text{all } \theta} p(\theta|D) d\theta}, \quad (8)$$

常取 $\xi = 1\sigma = 0.683$ 。贝叶斯可信区间和频率统计中的置信区间 (confidence interval)^[12] 有类似的作用，但意义不同。在贝叶斯统计中，参数 θ 被看作一个随机变量，它服从某种概率分布，可信区间指明 θ 在其中的概率为 ξ ；而频率统计中， θ 被认为是一个固定值，置信区间是随机的。置信度为 ξ 的置信区间应该解释为：这个区间覆盖 θ 真值的概率为 ξ 。可信区间和置

信区间都不是唯一确定的, 可信区间依赖于验前分布, 而置信区间仅依赖观测数据本身; 此外, 两者对多余参数的处理方式也完全不同, 前者通过积分求边缘分布, 后者则基于似然比检验和显著性检验给出所关心参数的置信区间。对于低维空间来说, 我们可以用分析法或者数值方法来计算贝叶斯置信度和置信区间; 但对于高维空间来说, 计算验后平均数、方差、置信区间以及其他数据的高难度积分需要采用蒙特卡罗方法来处理。

3.3 贝叶斯模型的选择

贝叶斯模型的选择是基于贝叶斯因子的一种模型筛选法。通常须在已知数据 D 的情况下, 从两个模型 M_1 和 M_2 中选择一个。贝叶斯因子 B_{21} 定义为:

$$B_{21} = \frac{p(D|M_2)}{p(D|M_1)}, \quad (9)$$

其中 $p(D|M_j)$ 是个边缘密度, 通过对参数 θ 的积分求得:

$$p(D|M_j) = \int p(\theta|M_j)p(D|\theta, M_j)d\theta, \quad (10)$$

$p(D|M_j)$ 也称为第 j 个模型的“证据”(evidence)。所以 B_{21} 可以写作:

$$B_{21} = \frac{\int p(\theta_2|M_2)p(D|\theta_2, M_2)d\theta_2}{\int p(\theta_1|M_1)p(D|\theta_1, M_1)d\theta_1}, \quad (11)$$

$B_{21} > 1$ 表示数据更支持模型 M_2 。Jeffreys 给出过 B_{21} 的一个解释, 见表 1。

表 1 关于 B_{21} 的解释^[16]

序号	B_{21}	证据力度
1	< 1 : 1	负支持 (支持 M_1)
2	1 : 1 至 3 : 1	M_1 几乎不值一提
3	3 : 1 至 10 : 1	支持 M_2
4	10 : 1 至 30 : 1	强支持 M_2
5	30 : 1 至 100 : 1	极强地支持 M_2
6	> 100 : 1	确定模型就是 M_2

对于一维参数, 可以用被积函数在最大似然值 $\hat{\theta}$ 处的值 $p(\hat{\theta}|M)p(D|\hat{\theta}, M)$ 乘以验后分布 $p(\theta|D, M)$ 的宽度 $\sigma(\hat{\theta}|D)$ 来近似“证据”:

$$p(D|M) \approx p(\hat{\theta}|M)p(D|\hat{\theta}, M)\sigma(\hat{\theta}|D). \quad (12)$$

如果参数为多维情况, 可以用下面的公式来估计 $p(D|M)$ ^[17]:

$$p(D|M) = p(\hat{\theta}_1, \dots, \hat{\theta}_m|M)(2\pi)^{m/2} \sqrt{\det V} L_{\max}, \quad (13)$$

其中 m 是 θ 的维数, L 是似然函数, $L_{\max} = p(D|\hat{\theta}, M)_{\hat{\theta}=(\hat{\theta}_1, \dots, \hat{\theta}_m)}$, $(\hat{\theta}_1, \dots, \hat{\theta}_m)$ 是参数的极大似然估计。 $V = \left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right)^{-1} \Big|_{\theta=\hat{\theta}}$ 是协方差矩阵, 贝叶斯因子显示出观测数据对验后比的贡献。其中,

贝叶斯验后比 O_{21} 定义为:

$$O_{21} = \frac{p(M_2|D)}{p(M_1|D)} = \frac{p(M_2)p(D|M_2)}{p(M_1)p(D|M_1)} = B_{21} \frac{p(M_2)}{p(M_1)} . \quad (14)$$

验后比和贝叶斯因子只差一个因子, 即验前比 $\rho = p(M_2)/p(M_1)$, 它可以根据美学原理或经验确定: 如简单的模型验前概率更高; 也常取均匀的验前分布, 这时 ρ 是个常数; 在对两个模型没有偏好的情况下, 可以认为 $\rho = 1$, 于是 $O_{21}=B_{21}$ 。一般来说, 增加模型参数的个数就会使拟合的结果更好, 因而使相应模型的“证据”增加。但是, 一个最好的模型一定不是最复杂的模型。物理崇尚“奥卡姆剃刀原理”(Occam's Razor), 也就是“简单有效原理”。所以, 引进“Occam 因子”来平衡拟合优度和模型复杂度之间的问题^[18]:

$$O_{21} = B_{21} \frac{\text{Occam}(M_2)}{\text{Occam}(M_1)} , \quad (15)$$

在一维情形, 可以取 $\text{Occam}(M) = p(D|\hat{\theta}, M)\sigma(\hat{\theta}|D)$; 类似地, 在多维情况下 $\text{Occam}(M) = (2\pi)^{m/2} \sqrt{\det \bar{V}} L_{\max}$, 其中各量说明见 (12 - 14) 式。

3.4 贝叶斯信息准则 (Bayesian Information Criterion, BIC)

当数据的分布形式是所谓“指数族”时 (通常遇到的正态分布、指数分布、 Γ 分布, 泊松分布, 二项分布等许多分布都属于这类分布), 可以导出一个渐近的结果——BIC^[19]:

$$\text{BIC} = -2\ln L_{\max} + m \ln(N) , \quad (16)$$

其中, L_{\max} 意义如 (13) 式, m 是待估参数 θ 的维数, N 为样本容量。(16) 式右端的第一项反映了数据对给定模型拟合的误差, 第二项反映了模型的复杂程度。因此, BIC 越小, 得到的模型越好。在正态分布的假设下, BIC 简化为:

$$\text{BIC} = N \ln \left(\frac{\text{RSS}}{N} \right) + m \ln(N) , \quad (17)$$

其中 RSS 是模型的残差平方和。

4 用贝叶斯统计法分析 GRBs 数据

贝叶斯理论常被用来研究 GRBs 的多种特性。

4.1 贝叶斯信息准则的应用

在天文统计中, 往往利用 BIC 作为模型选择评判的一个统计标准, 它要求观测数据的分布属于指数族。对于两个待选的模型 (θ_1, M_1) 和 (θ_2, M_2) , 定义 $\text{BIC}_{21} = \text{BIC}_2 - \text{BIC}_1$, 表示模型 M_2 相对于模型 M_1 的优越性, 由公式 (16) 可得:

$$\text{BIC}_{21} = -2(\ln p(D|\theta_2, M_2) - \ln p(D|\theta_1, M_1)) + (m_2 - m_1) \ln(N), \quad (18)$$

其中, m_1, m_2 分别是参数 θ_1 和 θ_2 中维数, 当样本容量 $N \rightarrow \infty$ 时^[17], BIC_{21} 和公式 (9) 中的 B_{21} 有如下关系:

$$\frac{\text{BIC}_{21} + 2 \ln B_{21}}{2 \ln B_{21}} \rightarrow 0. \quad (19)$$

因此, 在样本容量 $N \rightarrow \infty$ 时, $\text{BIC}_{21} \approx -2 \ln B_{21}$ 。BIC 更吸引人的地方就是即使验前分布未知, 也可以按照 (16) 式来计算^[17]。类似 Jeffreys 对 B_{21} 的解释 (表 1), Kass 为 BIC_{21} 的值提供了一个判断依据, 见表 2。

表 2 关于 BIC_{21} 的解释^[17]

序号	BIC_{21}	反对模型 2 的证据
1	0 : 2	模型 2 几乎不值一提
2	2 : 6	支持模型 2
3	6 : 10	强烈支持模型 2
4	> 10	极强支持模型 2

基于 BIC 算法以及其判断准则, Swenson 等人^[20]提出了一种寻找 GRBs 耀发 (flare) 的新算法, 应用于在紫外/光学波段探测 GRBs 光变曲线中的耀发现象。他们在 GRBs 光变曲线上利用断点函数 (breakpoint function)^①来寻找曲线上的潜在断点, 进而确定是否存在耀发。所谓断点, 就是在 GRBs 曲线上无法用幂律函数拟合的那些点, 这些点很可能是耀发成分。Swenson 等人分析了 Swift 观测到的紫外/光学望远镜 (UVOT) 余辉目录上的光变曲线样本, 把光变曲线上潜在的耀发开始时间 T_{start} , 结束时间 T_{stop} 以及峰值时间 T_{peak} 当做要寻找的一些特殊断点, 即待定参数。假设存在一个耀发成分, 在耀发之前能够完好拟合底层光变曲线的最后一个数据指定为 T_{start} , 出现耀发后, 底层光变曲线最佳拟合得到的第一个数据点指定为该耀发的 T_{stop} 。如果光变曲线有 n 个数据点, 则取由 $n-1$ 个断点分开的线段构成的折线拟合会有最小的残差平方和 ($\text{RSS} = 0$)。但是断点越多, 意味着参数越多, 对应于 (16) 式中 BIC 的第二项也就会越大。这样的拟合并不能把耀发部分凸显出来。一个好的断点函数的选取, 应该具有尽可能少的断点, 凸显耀发的成分。Swenson 等人使用 $\text{BIC}_2 - \text{BIC}_1 > 6$ 的判据, 以确定是否以更多耀发成分的光变曲线模型 M_2 来取代较少耀发成分的光变曲线模型 M_1 , 从而最终确定光变曲线中耀发的数量。他们利用这个方法, 计算了来自于 Swift 紫外/光学望远镜 (UVOT) GRB 余辉目录上的 201 个光变曲线, 并在其中 69 个光变曲线中发现了 119 个耀发。

Liu 等人^[21]利用 GRBs 的观测数据来约束宇宙结构学参数和暗能量模型。他们首先使用最新发布的 Union 2 SNe 数据来约束宇宙结构学参数: 即减速 (q_0)、加速 (j_0) 和突变参数 (s_0), 然后在 $z \leq 1.4$ 的条件下校准了 GRBs 的 5 个光度关系式。Liu 假设 GRB 的光度与红移无关, 获得了 66 个高红移 GRBs 的距离模数。最后, 他们把包含微波背景辐射和重子声学振

^①R 软件包 strucchange 里自带的一个函数。

荡 (Baryon Acoustic Oscillation) 的观测数据和 116 个已知红移的 GRBs 来约束一些暗能量模型。利用贝叶斯信息准则 BIC 判断后, 他们发现 Λ CDM 宇宙学模型是最优模型。

4.2 贝叶斯块分析 (The Bayesian block analysis, BB)

BB 是一个时域算法, 目的是为了找出光变曲线比较显著的转折点, 用来分析不同类型的 GRB 光变数据 (TTE, BIN, TTS)^[22, 23]。BB 算法的大体思路是: 假设所考虑的光变时间区间是 T , 模型 M_1 表示在整个 T 上, 计数率为常数; 模型 M_2 表示 T 分为两个子区间 $T_1 + T_2 = T$ 分别有两个不同的计数率。然后根据两个模型的验后比来决定是否把区间进一步细分。BB 算法输入的是原始的光子计数数据, 而输出则是最佳变换点。在光子到达率不存在显著的统计差异情况下, 这些变换点将时间序列分为最有可能的时间段; 每一个时间段称为一个 “Block”, “Block” 内的计数率为常数, 还要保证此 Block 的总计数与原光变总计数一致。这样, 就把光变曲线分为一系列 “Block” 的拼接 (见图 2)。

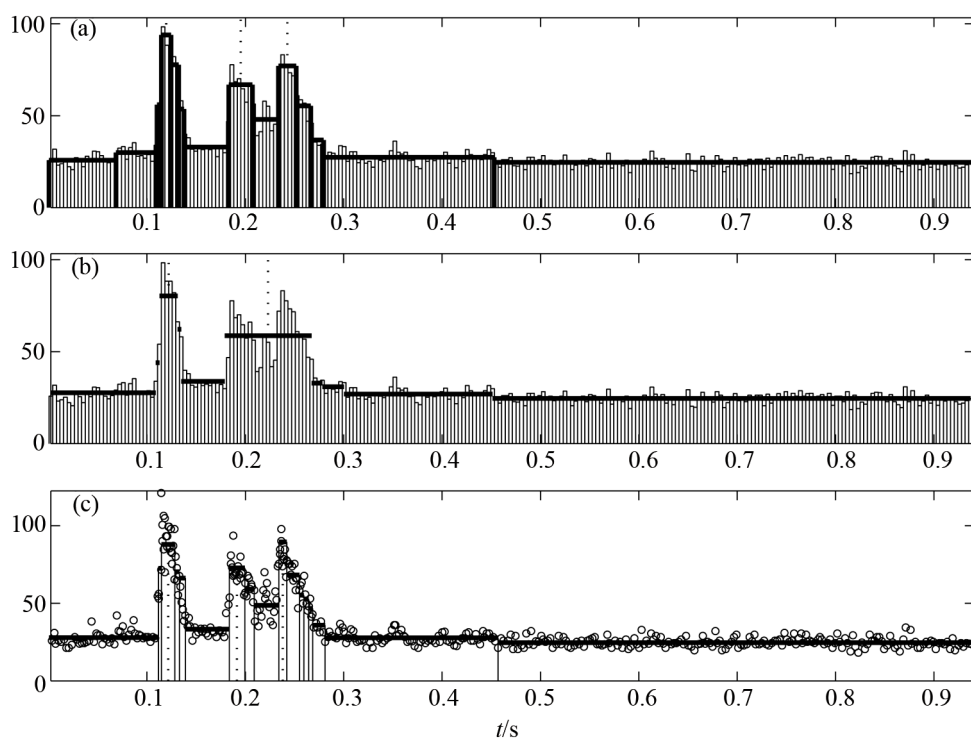


图 2 对 TTE 数据进行 BB 处理的结果, 我们看到 BB 把光变分成了若干个 Block, 在每个 Block 中, 计数率为常数 (来自文献 [22] 图 2)。

通过 BB 算法可以研究 GRBs 的持续时间, 并可以通过分析不同能量段的光变曲线来探测其脉冲的瞬时结构, 揭示其脉冲形状以及表征其强度变化^[24]。Lin 等人^[25]利用 BB 法来研究短暴的延伸辐射。他们在样本中发现了 3 个短暴: GRB 050724, GRB 061006 和 GRB 070714 很有可能存在延展辐射。Qin 等人^[26]利用 BB 法来研究 T_{90} 的双峰分布。通过计算 GBM 探测器探测到的 315 个 T_{90} Fermi 伽玛暴, 并和 CGRO/BATSE、BeppoSAX/GRBM、HETE-

2/FREGATE 以及 Swift/BAT 中的样本进行比较, 提出 T_{90} 的双峰分布可能源自于仪器的选择效应。利用 BB 算法, Scargle 等人^[27] 发现了最短的短暴 (100 μs), 它表现为叠加在主暴 GRB 920229 上的一个尖峰信号。他们改进了 1998 年提出的 BB 方法^[28]: 针对 Swift 光变曲线中由光子计数和高斯分布误差组成的数据, 结合几种不同的数据类型, 重新推导了 BB 拟合函数。利用改进的 BB 法, Scargle^[29] 又提出了一系列基于 BB 法的时间序列分析方法, 通过研究时变特性来阐述射电和 γ 射线波段活动星系核的物理辐射过程。Norris 等人^[30] 利用 BB 算法分析 Swift/BAT 中的短 GRBs 样本以及它们的 X 射线余辉, 来说明具有延展辐射和不具延展辐射的短暴可能存在不同的物理起源。

4.3 贝叶斯对峰值能量 E_{peak} 的估计

GRB 瞬时辐射谱的 E_{peak} 对决定 GRB 的总能量有着至关重要的作用, 并且它还在几个光度指示器中起关键性的作用。Bultner 等人^[31] 利用贝叶斯拟合方法估计 E_{peak} , 其中贝叶斯验前分布函数是由 BATSE 样本确定的。他们在验前分布函数中选择了 3 个参数: $\theta = (E_{\text{peak}}, \alpha, \beta)$, 其中 E_{peak} 是 νF_{ν} 特征能谱峰值光子流量, α 和 β 分别表示低能和高能幂指数。 E_{peak} 的验后边缘概率分布可以通过对 $p(\theta|D)$ 模型归一化以及对参数 α 和 β 积分求得。通过这个方法, 他们得到由 BAT 探测到的 218 个暴 (其中有 77 个观测到了红移) 的持续时间和谱参数。Eric 等人^[32] 对由 RHESSI 和 Swift/BAT 联合观测到的 26 个 GRBs 样本进行了光谱拟合。他们测量了观测样本的 E_{peak} 值, 并同时用贝叶斯方法与 Swift 幂律指数估计法来估计 E_{peak} 值, 比较两种结果得出, 贝叶斯方法与观测结果显示出更好的一致性。Li 等人^[33] 计算得到 4 个 GRBs 的 $\lg E_{\text{peak}}$ 和与之成协的 SNe 的峰值热星等 M_{peak} 之间存在很强的相关性 (样本相关系数 $r = -0.997$)。但 Chris^[34] 在假设 GRB 的 E_{peak} 数据服从 $\rho = 0$ (即边缘分布相互独立的) 二元高斯分布的基础上, 对于 $N = 4$ 的小样本, 得出 $|r|$ 的分布是 $[0, 1]$ 中的均匀分布, 这意味着 $|r| < 0.05$ 和 $|r| > 0.95$ 有相同的概率。因此, 为了进一步检验 Li 等人的结论, Chris 根据相关系数 $\rho \neq 0$ 的二元高斯分布中, 样本相关系数 r 所满足的分布^[35]:

$$p(r) = \frac{(N-2)\Gamma(N-1)}{\sqrt{2\pi}(N-1/2)} (1-\rho^2)^{(N-1)/2} (1-\rho r)^{-(2N-3)/2} \times (1-r^2)^{(N-4)/2} H\left[\frac{1}{2}, \frac{1}{2}; N - \frac{1}{2}; \frac{1}{2}(1+r\rho)\right], \quad (20)$$

其中 Γ 为伽玛函数, H 为超几何函数。取 ρ 的验前分布为均匀分布, 得出了 $N = 4$, $r = -0.997$ 时 ρ 的验后分布。验后分布中置信度为 0.99 和 0.95 的置信区间均包含 $\rho = 0$, 因此, 不支持 Li 的结论。进一步, Chris 利用 MCMC (Markov Chain Monte Carlo) 法来估计 E_{peak} 和与之成协的超新星的峰值热星等 M_{peak} 之间相关系数的贝叶斯验后分布, 得到两者相关系数的平均值在 $-0.33 \sim 0$ 之间, 结果暗示两者只存在很弱的相关性。Shahmoradi 等人^[36] 提出用硬度比 (HR) 来估计 GRBs 的 E_{peak} 值的方法, 它不受 GRBs 类型的限制。他们把 HR 定义为 BATSE 上 LAD 探测器 3, 4 通道的流量除以其 1, 2 通道的流量, 首先对 BATSE 里面的 249 个亮暴进行计算, 发现 HR 和 E_{peak} 之间存在强线性相关, 然后扩展到 BATSE 目录里面的全部 2130 个 GRBs 样本, 利用贝叶斯定理和 MCMC 方法, 对数据截断以及样本缺失两种情况下的观测数据进行了多变量分析; 推导出 HR- E_{peak} 关系的 90% 置信区间, 并用得到的估计值创建了一

个对应这 2130 个 GRBs 的 E_{peak} 目录; 此目录可能对长短暴、硬度-亮度关系以及利用 GRBs 研究宇宙学等方面起着重要作用。Freeman 等人^[37] 分析了 Ginga 卫星上的 GRB 870303, 该暴在低能段的两个不同时间段的谱 (分别记为 S_1 和 S_2) 上, 均显示出有吸收线成分的迹象; 用贝叶斯验后比 O_{21} 来检验连续谱加线吸收的模型 (M_2), 和单纯的连续谱模型 (M_1), 他们发现对数据 S_1 、 S_2 和联合数据 $S_1 + S_2$ 的 O_{21} 分别为 114 : 1, 7 : 1 和 40300 : 1, 结果支持连续谱加线吸收的模型, 而且联合数据对该模型的支持要强得多。

针对 GRBs 光谱的特征吸收线只被 Ginga 探测到而未被 BATSE 探测到这种情况, Band 等人^[38] 利用贝叶斯验后比来估计 Ginga 和 BATSE 探测器之间的一致性。此时贝叶斯验后比设为 H_0/H_1 , 其中零假设 H_0 是: Ginga 与 BATSE 探测结果一致; 备选假设 H_1 是: 这两颗卫星探测结果不一致 (例如, H_1 代表由于探测器故障, BATSE 无法探测到吸收线)。分析结果显示在给定探测器相关参数的前提下 BATSE 和 Ginga 的探测结果应是一致的。

4.4 和宇宙学有关的研究

近年来, GRBs 在可能把哈勃图延展到更高的红移段方面的研究引来不少关注。Donaghy 等人^[39] 提出一种基于贝叶斯推论的方法, 即对于那些已估计出内禀光度和红移的 GRBs 来说, 本身的发生率是其红移 z 的函数。首先, 他们用由 Reichart 等人^[40] 提出来的类造父变星光度估计法, 估计了从 BATSE 4B 目录里选出的 907 个长暴样本的内禀光度和红移; 然后, 他们假设 GRBs 在红移 z 处的发生率, 为 Rowan-Robinson 针对恒星形成所提出的现象学模型, 并且假设 GRBs 各向同性光度分布服从截断幂律模型; 最后利用贝叶斯近似, 计算出了此模型在给定数据下的似然函数, 并把其转换为关于参数的验后分布。Dainotti 等人^[41] 致力于寻找一个通用的关系式来描述所观测到 GRBs 的特性, 借用该公式可以直接用 GRBs 的可观测量来估计它的内禀光度。为了使 GRBs 标准化并把它们作为一种可能的距离指示器, 必须找到光度和可观测量之间的关系。一旦这种关系找到了, 就能够利用可观测量和估计到的光度来推导在红移 z 的 GRBs 光度距离 $D_L(z)$, 并构建 GRBs 哈勃图。关于这类方法的尝试主要以 $L - \tau_{\text{lag}}$ ^[42], $L - V$ ^[40], $E_{\text{iso}} - E_{\text{peak}}$ ^[43], $L - E_{\text{peak}}$ ^[44], 和 $E_{\gamma} - E_{\text{peak}}$ ^[45-47] 这 5 类关系式为主。

其中一个和贝叶斯方法有关的关系式是时间-光度关系式, 它是由 Dainotti 等人^[41] 在 GRBs 的 X 射线余辉中发现的。根据 Willingale 等人^[48] 的工作, Swift 上所有 GRBs 样本的 X 射线余辉光变曲线都可用相同的函数形式 $f(t, \alpha_a, t_a, T_a)$ 来拟合 (见图 3)。

因此, Dainotti 等人利用这组数据来寻找在截断时间 T_a 处的 X 射线光度 L_x 和时间 T_a 的可能关系式, 并利用贝叶斯统计来确定拟合带来的不确定度, 他们发现 $\lg[L_x(T_a)]$ 和 $\lg[T_a/(1+z)]$ 之间的线性关系存在固有弥散 $\delta = 0.33$ 。

为了获得一个独立于宇宙学之外的标准烛光, 就需要对光度指示器进行校准。Liang^[49] 根据贝叶斯理论, 用红移 z 在基准红移 ($z_c = 1$ 或 $z_c = 2$) 附近极小区域的 GRBs, 对他们所提出的光度指示器

$$E_{\text{iso}} = c(\Omega) E_p^{a_1} t_b^{a_2} \quad (21)$$

进行了校准。式中 E_{iso} 是 GRBs 各向同性能量, E_p 是谱的峰值能量, t_b 是 GRBs 余辉光变的截断时间, $c = c(\Omega)$ 是系数, Ω 是一组宇宙学参数, $A = \{a_1, a_2\}$ 是待定的参数。挑选 GRBs 的

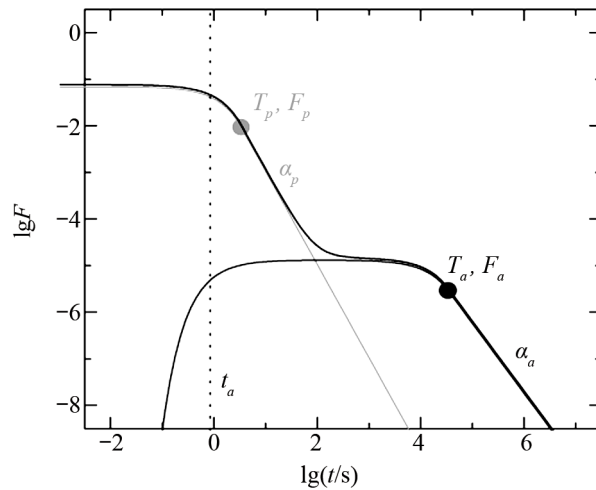


图 3 上方的曲线显示了瞬时辐射 (prompt emission) 的光变, 下方的曲线显示了余辉的光变
 两条曲线在 T_a 处重合。其中, 余辉光变曲线由初始状态上升在 t_a 时刻之后保持稳定, 然后在 T_a 时刻开始呈幂律形式衰减 (来自参考文献 [48] 中的图 1)。

样本时, 要求既可以作为光度指示器, 且红移 z 分布在足够窄的范围内 ($z \in z_c \pm \Delta z$)。用多元回归法可推导出参数 A , 记之为 $A(\Omega, z)$, 校准法可以概括为下面两个步骤。

(1) 利用下式校准 A :

$$A(z) = \frac{\int_{\Omega} A(\bar{\Omega}, z)p(\bar{\Omega}, z)d\bar{\Omega}}{\int_{\Omega} p(\bar{\Omega}, z)d\bar{\Omega}}, \quad (22)$$

其中 $p(\Omega, z) \propto \exp(-\chi^2_{\min}/2)$, χ^2_{\min} 是拟合后的 χ^2 。

(2) 利用给定的 GRBs 样本, 在合理的范围内, 对宇宙学参数进行边缘化积分, 从而校正 c , 最终达到校正 E_{iso} 的目的。Liang 的光度指示器在 $\Delta z \approx 0.3$ 内, 此时参数 A 基本不依赖宇宙学常数。

Schaefer 等人^[50] 列出了一个由 69 个含可观测量的 GRBs 的目录, 这个目录录入了 5 个当时已存在的双参数关系式。利用每一个关系式可以估计出距离模数 μ_z , 并取权重, 最终得出一个 GRBs 哈勃图。Cardone 等人^[51] 在 Schaefer 工作的基础上对哈勃图进行了更新。首先, 他们增加了一个 Dainotti^[41] 提出的 L_X-T_a 关系式, 扩大了 GRBs 样本数目, 并且减小了 μ_z 的不确定性; 然后, 利用贝叶斯拟合的方法, 采用 Wilkinson Microwave Anisotropy Probe5(WMAP5) 的数据, 利用标准的 Λ CDM 宇宙学模型对这 6 个关系式进行了校准, 进而更新了 GRBs 哈勃图; 最后, 他们提出不依赖任何宇宙学模型来对 GRBs 的经验公式进行校准, 给出了一个新的 GRBs 哈勃图。

Λ CDM 宇宙学模型预测宇宙从早期到红移 $z \approx 0.9$ 之间是减速膨胀的, 之后变为加速膨胀; 而基于共形引力的宇宙学模型 (cosmological model based on conformal gravity) 则预测宇宙从一开始就是加速膨胀的。为了区分这两种不同的宇宙学模型, Diaferio 等人^[52] 利用

GRBs 来进行计算。由于在红移 $z \approx 0.9$ 处被认证的 I_a 型超新星数目太少, 所以建议用 GRBs 来探测宇宙膨胀史。他们利用贝叶斯估计法推算宇宙学参数以及描述 GRBs 数据所需的参数, 把 GRB 样本和 SNe 样本结合起来, 发现贝叶斯估计法可以有效地避免“循环性问题”, 这是所有其他计算方法都被困扰的难题。他们的结果表明, 已知的 SNe 和 GRBs 样本都可以很好地被 Λ CDM 模型和共形引力模型容纳, 并且不排斥连续的加速膨胀条件。但是 GRBs 作为宇宙学探针, 目前还不是十分有效, 因为他们还不足以区分上述两种宇宙膨胀史。Freitas^[53] 同样利用 GRBs 的宇宙学探测器作用, 来研究广义的 Chaplygin 气体模型 (GCGM)。为了避免出现循环问题, 他们采用了一组和宇宙学无关的数据, 利用贝叶斯统计来对模型参数施加限制条件。

5 总结和展望

贝叶斯方法在 GRBs 数据分析中的应用主要表现在两个方面: 模型选择和参数估计。利用 BB 算法可以对 GRBs 光变曲线进行分析, 揭示其脉冲形状及强度变化; 可以研究短暴的延伸辐射; 为 T_{90} 的双峰分布可能源自于仪器的选择效应提供依据; 发现了最短的短暴; 说明具有延展辐射和不具延展辐射的短暴可能存在不同的物理起源。

利用贝叶斯拟合, 可以对 GRBs 峰值能量 E_{peak} 进行估计, 进而研究 GRBs 的总能量。用 E_{peak} 估计值创建的目录可能对长短暴、硬度-亮度关系以及利用 GRBs 研究宇宙学等方面研究起着重要作用。同时, 贝叶斯拟合在校正参数和模型筛选方面也取得了令人信服的结果。GRBs 可能把哈勃图延展到更高的红移段, 该方向引起人们高度重视。为了使 GRBs 标准化并把它们作为一种可能的距离指示器, 天文学家们做了大量的工作, 来寻找光度与可观测量之间的关系。利用贝叶斯拟合的方法更新了 GRBs 哈勃图, 还提出了不依赖任何宇宙学模型来对 GRBs 的经验公式进行校准, 给出了一个新的 GRBs 哈勃图。同时, 贝叶斯方法也被用于宇宙学模型方面的研究, 例如: 区分两种主流的宇宙膨胀史, 约束宇宙结构学参数和暗能量模型。

上面只是用了很有限的例子, 来展示贝叶斯估计的用途。贝叶斯估计在很大程度上需要根据特定的情况使用, 因此它需要更多的技巧和经验。随着天文学的发展, 发现频率统计存在一些不理想的特征, 在一些特定情况下贝叶斯估计反而会有优势。

在一些重要的状况下, 用频率统计的方法可能无法计算出正确的 P 值, 且经典统计置信区间中的枢轴量的构造是相当复杂的, 而贝叶斯方法巧妙地避开了这一困境。尽管缺少所需的某些特定条件, 但是利用似然比 (likelihood ratio test, LRT) 检验和 F 检验, 来探测某一个源或谱线模型的发射线已经变成了一种时尚。Protassov 等人^[54] 提出, 在参数空间边缘对假设进行检验时, 不应使用 LRT 检验的 χ^2 分布和 F 检验的 F 分布; 因为在未知相关参数的分布情况下, 统计检验是无效的, 既不能对某个观测值进行校准, 也不能计算 P 值。例如, 在探测谱的某种发射线或对其吸收特性进行检测时, 由于相关参数分布是未知并且不可预测的, 所以无法利用 LRT 和频率统计。使用 LRT 检验和 F 检验必须满足两个条件: 首先, 待比较的

两个模型应该是分级的巢式模型。也就是说, 相对于原假设模型——零模型 (Null Model), 复杂模型仅仅是多了一个或者多个附加参数。其次, 附加参数的空值 Null Value 不能处在可能的参数值的边界上。比如, 在对发射线进行检验时, 因为谱线流量必须是非负数, 并且其空值 Null Value 是 0, 而 0 恰好是非负数的边界, 因此不满足 LRT 和 F 检验的条件。Protassov 对 GRB 970508 余辉的 X 射线谱, 采用了不含 Fe K α 发射线的模型和含有 Fe K α 发射线的两种模型, 并采用了两种不同的模型拟合方法: XSPEC 中的 χ^2 拟合和贝叶斯拟合; 最后他们计算出 P 值来对 Fe K α 发射线的证据进行估计, 指出贝叶斯方法是一种简单、有力并直观的方法。

近 20 年来, 贝叶斯估计在天文统计中的地位发生了戏剧性的变化。在 1991 年召开的第一届现代天文学中的统计学挑战会议上, 贝叶斯方法不仅是非常新的方法, 并且广受争议。而现在, 不仅消除了对贝叶斯方法的争议, 而且还被普遍使用。Loredo^[55] 指出人们对贝叶斯方法的三个主要误区: (1) 错误地把变化性等同于不确定性, 让数据分析学家们误以为频率统计比其实际计算要简单得多。(2) 错误地认为贝叶斯方法的高维积分更加困难。其实, 对于频率统计法和贝叶斯方法都存在多维积分计算困难的难题, 在高维计算上, 两者都趋向于使用蒙特卡洛方法来进行计算。(3) 错误地认为贝叶斯就是验前分布。验前分布固然重要, 但是似然函数的调节作用更为重要。特别是, 对于天文观测的实际情况, 各类扁平验前分布对于结果的影响很小。最为重要的是, 贝叶斯方法可以对参数空间进行测量。

贝叶斯估计已经被越来越多的天文学者认可, 而且已经被广泛应用于各个与计算相关的行业, 它将在未来的天文计算方面展示它迷人的魅力。

参考文献:

- [1] Efron B. The American Statistician, 1986, 40(1): 1
- [2] Fenton N E, Neil M. Risk Assessment and Decision Analysis with Bayesian Network. London: CRC Press, 2012: 116
- [3] Feigelson E D, Babu G J. astro-ph/0401404, 2004
- [4] D'Agostini G. Proceedings of the 18th International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis. von der Linden W, Dose V, Fischer R, eds. USA: Springer, 1999: 157
- [5] Cousins R D. Am.J.Phys, 1995, 63(5): 398
- [6] Babu G J, Feigelson E D, eds. Statistical Challenges in Modern Astronomy IV: proceedings of a conference held at Pennsylvania State University. USA: Astronomical Society of the Pacific, 2007: 189
- [7] Jefferys W, Basso B. American Astronomical Society, 1997, 28: 1099
- [8] Berger J O. Statistical Challenges in Modern Astronomy II. Babu G J, Feigelson E D, eds. New York: Springer, 1997: 15
- [9] Loredo T J, Wasserman I M. ApJ, 1998, 502: 75
- [10] Connors A. Statistical Challenges in Modern Astronomy II. Babu G J, Feigelson E D, eds. New York: Springer, 1997: 39
- [11] Loredo T J. astro-ph/0409386, 2004
- [12] Samaniego F J, ed. A Comparison of the Bayesian and Frequentist Approaches to Estimation. New York: Springer, 2010: 15
- [13] Wikipedia. Bayes theorem. http://en.wikipedia.org/wiki/Bayes_theorem, 2013
- [14] van Dyk D A, Connors A, et al. ApJ, 2001, 548: 224
- [15] Zhu Y S. Chinese Physics C. 2008, 32(5): 363
- [16] Jeffreys H, ed. Mathematical Proceedings of the Cambridge Philosophical Society. UK: Cambridge University, 1935, 31(2): 203

- [17] Kass R E, Raftery A E. *J. Am. Stat. Assoc.*, 1995, 90: 773
- [18] Nemenman I, Bialek W. *Phys. Rev. E*, 2002, 65: 6137
- [19] Schwarz G. *Ann. Stat.*, 1978, 6: 461
- [20] Swenson C A, Roming P W, De Pasquale M, et al. arXiv:1303.0286, 2013
- [21] Liu J W, Wang F Y. *Chinese Astronomy and Astrophysics*, 2012, 36: 155
- [22] Scargle J D. *ApJ*, 1998, 504: 405
- [23] Scargle J D, Norris J P, Jackson B, et al. arXiv:1207.5578, 2012
- [24] Veres P, Bagoly Z, Kelemen J, et al. *AIPC*. 2008, 1065: 35
- [25] Lin L, Liang E W, Zhang B B, et al. *AIPC*. 2008, 1065: 39
- [26] Qin Y, Liang E W, Yi S X, et al. arXiv:1205.1188, 2012
- [27] Scargle J D, Norris J P, Bonnell J T. *astro-ph/9712016*, 1997
- [28] Scargle J, Norris J, Jackson B, et al. *Studies in Astronomical Time Series Analysis. VI. Optimal Segmentation: Blocks, Triggers, and Histograms*. <http://trotsky.arc.nasa.gov/jeffrey/global.pdf>, 2006
- [29] Scargle J D. arXiv:1006.4643, 2010
- [30] Norris J P, Gehrels N, Scargle J D. *astro-ph/1101.1648*, 2011
- [31] Butler N R, Kocevski D, Bloom J S, et al. *ApJ*, 2007, 671: 656
- [32] Eric C, Mark E, Steven E, et al. *AIPC*. 2008, 1000: 154
- [33] Li L X. *MNRAS*, 2006, 372: 1357
- [34] Chris K. *MNRAS*, 2009, 393:1370
- [35] Muirhead R J, ed. *Aspects of Multivariate Statistical Theory*. New York: John Wiley, 2009: 161
- [36] Shahmoradi A, Nemiro R J. *MNRAS*, 2010, 407(4): 2075
- [37] Freeman P E, Graziani C, Lamb D Q, et al. *ApJ*, 1999, 524: 753
- [38] Band D L, Ford L A, Matteson J L, et al. *ApJ*, 1994, 434: 560
- [39] Donaghy T, Lamb D Q, Reichant D E, et al. *AIPC*. 2003, 662: 450
- [40] Reichart D, Lamb D Q, Fenimore E, et al. *ApJ*, 2001, 552: 57
- [41] Dainotti M G, Capozziello S. *MNRAS*, 2008, 391: 79
- [42] Norris J P, Marani G F, Bonnell J T. *ApJ*, 2000, 534: 248
- [43] Amati L, Frontera F, Tavani M. et al. *A&A*, 2002, 390: 81
- [44] Schaefer B E, Gerardy C L, Hoffich P. *ApJ*, 2003, 588: 387
- [45] Ghirlanda G, Ghisellini G, Lazzati D, et al. *ApJ*, 2004, 616: 331
- [46] Dai Z G, Liang E W, Xu D, et al. *ApJ*, 2004, 612: 101
- [47] Ghirlanda G, Ghisellini G, Firmani C, et al. *A&A*, 2006, 452: 839
- [48] Willingale R, O' Brien P T, Osborne J P, et al. *ApJ*, 2007, 662: 1093
- [49] Liang E W, Zhang B. *MNRAS*, 2006, 369: 37
- [50] Schaefer B E, Collazzi A C. *astro-ph/0701548*, 2007
- [51] Cardone V F, Capozziello S, Dainotti M G. *MNRAS*, 2009, 400: 775
- [52] Diaferio A, Ostorero L, Cardone V. arXiv:1103.5501, 2011
- [53] Freitas R C, Goncalves S V B, Velten H E. arXiv:1004.5585, 2011
- [54] Protassov R, van Dyk D A, Connors A. *ApJ*, 2001, 571: 545
- [55] Loredano T J. arXiv:1208.3036, 2012

Bayesian Approaches to Gamma Ray Bursts Data Analysis

BU Qing-cui, CHEN Li, LI Zhao-sheng, WANG De-hua

(Department of Astronomy, Beijing Normal University, Beijing 100875, China)

Abstract: Bayesian method is based on Bayes' Theorem in which one deems it sensible to consider a probability distribution function (pdf) for the unknown parameter θ , $p(\theta)$, called the "prior" pdf for θ . $p(\theta)$ reflects our knowledge before observation. We let $p(\theta|D)$ be the "posterior" pdf of θ (given the data D), which reflects our modified beliefs after incorporating the results of the observation. All Bayesian inferences are based on the posterior distribution. Bayesian approach is a powerful tool for the gamma ray bursts (GRBs) data analysis, such as analyzing structure in photon counting data, combining different lightcurves, determining various parameters' distributions, testing if there is spectral line, and so on.

During the last decade, more and more astronomers realized the potential of the Bayesian approach. Our aim here is not to provide a detailed explanation of Bayesian theory but rather to show what we have learned from GRBs data analysis by using Bayesian approach. As a comparison, we first give a brief introduction to frequentist approach and Bayesian approach. Then, we present specific examples to study various characteristics of GRBs by using Bayesian method. We emphasize the following aspects: using the BIC (Bayesian information criterion) to select model, using BB (Bayesian block) analysis to find optimal change-points in GRBs light curves, using Bayesian fit method to estimate GRBs peak energy, using this method to extend the Hubble diagram to a very high redshift. Bayesian approach has two mainly usages in GRBs data analysis: model selection and parameter estimation. To some degrees, Bayesian approach requires a great deal of thought about the given situation to apply sensibly, therefore, it seemed to need more techniques and experiences.

Key words: gamma ray bursts; Bayesian method; statistical