arXiv:2301.13783v1 [astro-ph.CO] 31 Jan 2023

一种计算贝叶斯证据的解析方法

Juan Garc´ıa-Bellido 马德里自治大学理论物理系 C-XI，坎托布兰科，28049 马德里，西班牙

2005年4月14日

摘要

贝叶斯证据是模型选择中的一项关键工具，可用于比较参数数量不同的模型。由于计算困难，它在宇宙学模型分析中的应用一直受到限制；当前的数值算法需要借助超级计算机。本文针对具有任意相关性和顶帽先验的高斯似然，给出了贝叶斯证据的精确公式；针对具有主要非高斯特征（偏度和峰度）的似然分布，则给出了近似公式。我们将这些公式应用于含有和不含等曲率扰动成分的宇宙学模型，并与此前使用数值热力学积分得到的结果进行比较。我们发现，这些结果的精度低于热力学积分，但仍足以发挥实用价值。

1 引言

模型选择是一个统计问题，旨在确定哪种模型对观测数据的描述最为理想 [1, 2]。它不同于参数估计：在参数估计中，单一模型的选择（即确定要改变哪些参数）已经完成，目标是找出这些参数的最佳拟合值及其范围。尽管参数估计技术（通常是似然拟合）已广泛应用于宇宙学数据，但模型选择统计量迄今为止的应用仍相当有限 [3, 4, 5]。这令人遗憾，因为模型选择技术对于稳健地区分参数数量不同的模型必不可少，而宇宙学中许多最有趣的问题，都涉及是否值得引入额外参数来描述新的物理效应。

在贝叶斯推断框架下，应使用贝叶斯证据 [1, 2], 进行模型选择。贝叶斯证据衡量的是结合观测数据后模型成立的概率（即先验分布上的平均似然）。贝叶斯证据为每个模型赋予一个数值，随后可按证据大小对模型进行排序，并将这些数值的比值解释为模型的相对概率。这一过程在模型的简洁性与拟合数据的能力之间形成了理想的制衡。

贝叶斯证据的应用迄今一直受限于计算上的困难。标准技术是热力学积分 [6, 7], ，它通过马尔可夫链蒙特卡罗（MCMC）方法改变温度，从而以一种兼顾后验分布和先验分布的方式对分布进行采样。然而，在近期的研究中 [5] 我们表明，在宇宙学背景下，要获得足够准确的结果，每个模型大约需要进行 10^7 次似然评估。此类分析受限于中央处理器（CPU），因为生成用于与数据比较的预测功率谱需要耗费大量时间；这一需求使问题必须借助超级计算机解决（作为比较，参数估计通常需要进行 10^5 至 10^6 次似然评估）。

本文提出并采用一种新的解析方法，通过展开似然分布函数来计算证据。该方法预先假定已获得后验分布的协方差，例如通过一次马尔可夫链蒙特卡罗参数估计运行获得；而且在目前的形式下，

要求参数的先验分布为均匀顶帽先验。[1] 虽然该方法并不适用于一般的似然分布，但我们在近似似然形状时纳入了主要的非高斯性（偏度和峰度），期望在似然分布足够简单时取得良好结果。宇宙学示例中的似然分布通常只有一个显著峰值。

我们将该方法应用于玩具模型示例和真实的宇宙学情形。特别是，我们计算了绝热模型和等曲率模型的证据；此前，我们曾在参考文献 [5] 中使用热力学积分计算这些证据。我们发现，各方法之间的差异通常不超过 ln(证据) 中的 1，这意味着解析方法的精度略低于理想水平，但已足以有效指示模型偏好。

2 贝叶斯证据

给定数据 $\mathbf{D}$ 时，模型 $\mathcal{M}$ 的参数 θ 的后验概率分布 $\mathcal{P}(\theta,\ \mathcal{M}|\mathbf{D})$ 与似然函数 $\mathcal{L}(\mathbf{D}|\theta,\ \mathcal{M})$ 的关系由贝叶斯定理给出，其中模型参数服从一组给定的先验分布函数 $\pi(\theta,\ \mathcal{M})$：

$$\mathcal{P}(\theta, \mathcal{M}|\mathbf{D}) = \frac{\mathcal{L}(\mathbf{D}|\theta, \mathcal{M})\, \pi(\theta, \mathcal{M})}{E(\mathbf{D}|\mathcal{M})}\,, \tag{1}$$

其中，E 是贝叶斯证据，即先验分布上的平均似然，

$$E(\mathbf{D}|\mathcal{M}) = \int d\theta\ \mathcal{L}(\mathbf{D}|\theta, \mathcal{M})\, \pi(\theta, \mathcal{M})\,, \tag{2}$$

其中，θ 是一个具有 n 个分量的向量，用于表征 n 个独立参数。先验分布函数 π 包含观测数据之前关于参数的所有信息，即我们的理论偏见、对模型的物理理解以及以往实验所提供的信息。

当参数数量较多（$n \gg 1$）时，无法直接计算证据积分，必须通过数值方法或解析近似求得。在数值方法中，最常用的是热力学积分 [6, 7]，但这种方法的计算量可能极其庞大 [5]。最简单的解析近似是拉普拉斯近似，适用于分布可近似为多元高斯分布的情况。当数据的数量和质量达到最佳时，这一条件或许能够满足，但在宇宙学情境中，其适用范围可能有限。

贝叶斯证据之所以重要，是因为它能够让我们在一组互斥且完备的模型中进行比较 $\{\mathcal{M}_i\}_{i=1\ldots N}$。给定数据 $\mathbf{D}$，我们可以利用贝叶斯定理计算每个假设的后验概率：

$$\mathcal{P}(\mathcal{M}_i|\mathbf{D}) \propto E(\mathbf{D}|\mathcal{M}_i)\, \pi(\mathcal{M}_i)\,, \tag{3}$$

其中，$E(\mathbf{D}|\mathcal{M}_i)$ 是模型 $\mathcal{M}_i$ 下数据的证据，而 $\pi(\mathcal{M}_i)$ 是在我们看到数据之前第i个模型的先验概率。两个相互竞争的模型的证据之比称为贝叶斯因子 [8]

$$B_{ij} = \frac{E(\mathbf{D}|\mathcal{M}_i)}{E(\mathbf{D}|\mathcal{M}_j)}\,, \tag{4}$$

而且，如果我们假设事先对任何模型都不偏好，那么这也等于模型后验概率之比，即 先验，因此 $\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2) = \ldots = \pi(\mathcal{M}_N) = 1/N$。

贝叶斯因子公式（4）在数学上体现了奥卡姆剃刀，因为更复杂的模型往往预测能力较弱，与更简单、预测能力更强的模型相比，其平均似然较低。只有当更复杂的模型能够显著改善对数据的拟合时，它们才会受到青睐。在模型的最大似然相差悬殊的简单情形下，无需采用模型选择技术；但若拟合效果的改善

[1] 扩展到高斯先验应该是可行的，但扩展到任意先验则不可行。

微乎其微，则必须借助这些技术才能恰当地讨论问题。只要新数据带来了需要增加参数的可能性，这种情况几乎不可避免，除非新数据的效力远远超过此前的数据；宇宙学示例包括引入谱倾斜、暗能量密度变化，或本文后文探讨的微弱等曲率扰动。

在本文中，我们将通过系统地考察分布的高阶累积量，得到一个近似贝叶斯证据的解析公式。其优势在于，利用这些解析公式，只要给定分布的累积量层级（假设这些累积量已在参数估计程序中针对似然分布函数预先计算），便可以计算具有任意数量参数的给定模型的证据。

需要以足够高的精度计算证据，才能得出稳健的结论。杰弗里斯提出的标准解释尺度 [1]（并在参考文献 [5], 中进行了总结）大致上会在 ln(证据) 的差值每增加 1 时加强一次其判定。因而，证据的计算精度需要高于这一尺度；ln(证据) 的不确定度为 0.1 通常已足够，即使不确定度达到其两倍通常也可以接受。这一精度要求确保模型的相对概率不会因不确定度而发生明显变化。

我们首先需要做的是刻画具有 n 参数的模型的分布函数。令$f(\mathbf{x})$ 表示该函数，并假设它已正确归一化，

$$\int_{-\infty}^{\infty} d^n\mathbf{x}\, f(\mathbf{x}) = 1\,. \tag{5}$$

那么， p点相关函数可表示为

$$\langle x_{i_1} \dots x_{i_p} \rangle = \int_{-\infty}^{\infty} d^n\mathbf{x}\; x_{i_1} \dots x_{i_p}\, f(\mathbf{x})\,. \tag{6}$$

由该分布函数总可以构造出生成泛函 ϕ（$\mathbf{u}$），其形式为傅里叶变换

$$\phi(\mathbf{u}) = \int_{-\infty}^{\infty} d^n\mathbf{x}\, e^{i\,\mathbf{u}\cdot\mathbf{x}}\, f(\mathbf{x})\,. \tag{7}$$

该函数可展开为

$$\phi(\mathbf{u}) = \exp\left[\sum_{p=1}^{\infty} \frac{i^p}{p!}\, A_{i_1\dots i_p}\, u^{i_1} \dots u^{i_p}\right], \tag{8}$$

其中，$A_{i_1\dots i_p}$ 是完全对称的p 阶张量。例如，若将阶数限制为 4，则可写为

$$\phi(\mathbf{u}) = \exp\left[i\,\mu_i u_i - \frac{1}{2!}\,C_{ij}\,u_i u_j - \frac{i}{3!}\,B_{ijk}\,u_i u_j u_k + \frac{1}{4!}\,D_{ijkl}\,u_i u_j u_k u_l + \dots + \frac{i^n}{n!}\,A_{i_1\dots i_n}\,u_{i_1}\dots u_{i_n}\right], \tag{9}$$

其中， μ_i 是变量 x_i的均值； C_{ij} 是协方差矩阵； B_{ijk} 是与第三累积量或偏度相关的三线性矩阵； D_{ijkl} 是与第四累积量或峰度相关的四阶张量； $A_{i_1\dots i_n}$ 是与 n阶累积量相关的n 阶张量。利用n点相关函数表示它们的表达式，可由公式（7）得到，只需注意到

$$\langle x_{i_1} \dots x_{i_n} \rangle = (-i)^n\, \frac{\partial^n \phi(\mathbf{u})}{\partial u_{i_1} \dots \partial u_{i_n}}\bigg|_{\mathbf{u}=0}\,. \tag{10}$$

例如，一阶项给出

$$\langle x_i \rangle = (-i)\, \frac{\partial \phi(\mathbf{u})}{\partial u_i}\bigg|_{\mathbf{u}=0} = \mu_i\,. \tag{11}$$

二阶相关函数给出

$$\langle x_i x_j \rangle = (-i)^2 \left. \frac{\partial^2 \phi(\mathbf{u})}{\partial u_i \partial u_j} \right|_{\mathbf{u}=0} = C_{ij} + \mu_i \mu_j \,, \tag{12}$$

因此，协方差矩阵通常由.....得到

$$C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \,.$$

三阶相关函数给出

$$\langle x_i x_j x_k \rangle = (-i)^3 \left. \frac{\partial^3 \phi(\mathbf{u})}{\partial u_i \partial u_j \partial u_k} \right|_{\mathbf{u}=0} = B_{ijk} + \mu_i C_{jk} + \mu_j C_{ki} + \mu_k C_{ij} + \mu_i \mu_j \mu_k \,, \tag{13}$$

因此，偏度矩阵由.....得到

$$B_{ijk} = \langle x_i x_j x_k \rangle - \langle x_i \rangle \langle x_j x_k \rangle - \langle x_j \rangle \langle x_k x_i \rangle - \langle x_k \rangle \langle x_i x_j \rangle + 2 \langle x_i \rangle \langle x_j \rangle \langle x_k \rangle \,. \tag{14}$$

四阶相关函数给出

$$\begin{aligned} \langle x_i x_j x_k x_l \rangle = (-i)^4 \left. \frac{\partial^4 \phi(\mathbf{u})}{\partial u_i \partial u_j \partial u_k \partial u_l} \right|_{\mathbf{u}=0} &= D_{ijkl} + C_{ij} C_{kl} + C_{ik} C_{jl} + C_{il} C_{jk} \\ &+ B_{ijk} \mu_l + B_{ijl} \mu_k + B_{jkl} \mu_i + B_{ikl} \mu_j \\ &+ C_{ij} \mu_k \mu_l + C_{ik} \mu_j \mu_l + C_{il} \mu_j \mu_k \\ &+ C_{jk} \mu_i \mu_l + C_{jl} \mu_i \mu_k + C_{kl} \mu_i \mu_j \\ &+ \mu_i \mu_j \mu_k \mu_l \,, \end{aligned} \tag{15}$$

因此，峰度矩阵由.....得到

$$\begin{aligned} D_{ijkl} &= \langle x_i x_j x_k x_l \rangle - \langle x_i x_j \rangle \langle x_k x_l \rangle - \langle x_i x_k \rangle \langle x_j x_l \rangle - \langle x_i x_l \rangle \langle x_j x_k \rangle \\ &- \langle x_i x_j x_k \rangle \langle x_l \rangle - \langle x_i x_j x_l \rangle \langle x_k \rangle - \langle x_i x_k x_l \rangle \langle x_j \rangle - \langle x_j x_k x_l \rangle \langle x_i \rangle \\ &+ 2 \langle x_i x_j \rangle \langle x_k \rangle \langle x_l \rangle + 2 \langle x_i x_k \rangle \langle x_j \rangle \langle x_l \rangle + 2 \langle x_i x_l \rangle \langle x_j \rangle \langle x_k \rangle + 2 \langle x_j x_k \rangle \langle x_i \rangle \langle x_l \rangle \\ &+ 2 \langle x_j x_l \rangle \langle x_i \rangle \langle x_k \rangle + 2 \langle x_k x_l \rangle \langle x_i \rangle \langle x_j \rangle - 6 \langle x_i \rangle \langle x_j \rangle \langle x_k \rangle \langle x_l \rangle \,, \end{aligned} \tag{16}$$

高阶累积量也依此类推。

3 高斯近似

我们先计算多元高斯分布的证据，也就是说，除协方差矩阵 C_{ij} 和均值 μ_i之外，所有累积量都为零的分布。在这种情况下，生成泛函和分布分别为

$$\phi(\mathbf{u}) = \exp\Big[- i \mu_i u_i - \frac{1}{2} C_{ij} \, u_i u_j \Big] \,, \tag{17}$$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} d^n \mathbf{u} \; e^{-i \, \mathbf{u} \cdot \mathbf{x}} \; \phi(\mathbf{u}) \tag{18}$$

$$= \frac{1}{(2\pi)^{n/2} \sqrt{\det C}} \exp\Big[- \frac{1}{2} C^{-1}_{ij} (x_i - \mu_i)(x_j - \mu_j) \Big] \,, \tag{19}$$

其满足

$$\langle x_i\rangle = \mu_i\,, \qquad \langle x_i x_j\rangle = C_{ij} + \mu_i\mu_j\,, \qquad \langle x_i x_j x_k\rangle = \mu_{(i}C_{jk)} + \mu_i\mu_j\mu_k\,, \qquad \cdots \tag{20}$$

其中，括号中的下标（ijk）表示循环求和。注意，所有 n 点关联函数都可以用分布的前两个矩表示，而所有高阶累积量均为零。

3.1 中心化先验

在初步计算中，我们假设先验分布为顶帽先验，并作出一个不现实的假设（稍后将予以放宽）：顶帽先验以均值为中心:

$$\pi(x,a) \equiv \begin{cases} (2a)^{-1} & -a < x-\mu < a\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{21}$$

由于顶帽函数的傅里叶变换为

$$\int_{-\infty}^{\infty} dx\, e^{iux}\,\pi(x,a) = \frac{\sin au}{au}\,\exp[i\mu u]\,,$$

我们可以用以下任一种方式写出证据

$$E(a_1,\ldots,a_n) \;=\; \int_{-\infty}^{\infty} d^n\mathbf{x}\, f(\mathbf{x}) \prod_{i=1}^{n}\pi(x_i,a_i) \;=\; \prod_{i=1}^{n}(2a_i)^{-1}\int_{-a_1}^{a_1} dx_1\cdots\int_{-a_n}^{a_n} dx_n\, f(\tilde{\mathbf{x}}) \tag{22}$$

$$\;=\; \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty} d^n\mathbf{u}\,\phi(\mathbf{u})\prod_{i=1}^{n}\frac{\sin a_i u_i}{a_i u_i}\,. \tag{23}$$

在式（22）中，我们对平移坐标 $\tilde{x}_i \equiv x_i - \mu_i$ 积分，使得 $\langle\tilde{x}_i\rangle = 0$ and $\langle\tilde{x}_i\tilde{x}_j\rangle = C_{ij}$。从现在起，我们忽略波浪号，并假定已转换到这些坐标。请注意，先验的选择并非至关重要。我们也可以选择高斯先验，结果不会有太大差别，只是窗函数 $\sin z/z$ 会变成高斯函数。现在，让我们分别在 1 个、2 个以及 n 个变量的情况下计算式（22）中的积分。

1 个变量。假设协方差仅为 $C = \sigma^2$。那么证据为

$$E(a) = \frac{1}{2a\,\sigma\sqrt{2\pi}}\int_{-a}^{a} dx\; e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} du\,\frac{\sin au}{au}\,e^{-\frac{1}{2}\sigma^2u^2} = \frac{1}{2a}\mathrm{Erf}\Big[\frac{a}{\sigma\sqrt{2}}\Big]\,, \tag{24}$$

其中，$\mathrm{Erf}[x]$ 是误差函数；当 $x \geq 2$ 或 $a \geq 3\sigma$ 时，它会非常迅速地渐近于 1。因此，对于宽度为 $2a$ 的中心顶帽先验模型，其证据可近似为 $(2a)^{-1}$。理论先验越宽，证据就越小，这正如预期。

2 变量。假设我们有两个相关变量，x_1 和 x_2，其协方差矩阵 ix

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\,. \tag{25}$$

其中，互相关 ρ 定义为

$$\rho = \frac{\langle x_1x_2\rangle}{\sqrt{\langle x_1^2\rangle\langle x_2^2\rangle}} = \frac{\langle x_1x_2\rangle}{\sigma_1\sigma_2}\,,$$

其中 σ_1 和 σ_2 分别为相应的二次色散。在这种情况下，归一化的二维分布函数为

$$f(\mathbf{x}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\,\exp\Big[\frac{-1}{1-\rho^2}\Big(\frac{x_1^2}{2\sigma_1^2} - \frac{\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{2\sigma_2^2}\Big)\Big]\,, \tag{26}$$

它具有这样的性质：对两个变量中的一个变量进行积分（“边缘化”）后，剩余变量服从一个规范化的高斯分布，

$$\int_{-\infty}^{\infty} dx_2\; f(\mathbf{x}) = \frac{1}{\sigma_1\sqrt{2\pi}}\,e^{-\frac{x_1^2}{2\sigma_1^2}}\,. \tag{27}$$

现在，我们先对 x_2 中的先验进行积分，以计算证据式（22）：

$$\frac{1}{2a_2}\int_{-a_2}^{a_2} dx_2\; f(\mathbf{x}) = \frac{e^{-\frac{x_1^2}{2\sigma_1^2}}}{\sigma_1\sqrt{2\pi}}\cdot\frac{1}{4a_2}\left[\mathrm{Erf}\Big[\frac{a_2\sigma_1+\rho\sigma_2\,x_1}{\sigma_1\sigma_2\sqrt{2(1-\rho^2)}}\Big] + \mathrm{Erf}\Big[\frac{a_2\sigma_1-\rho\sigma_2\,x_1}{\sigma_1\sigma_2\sqrt{2(1-\rho^2)}}\Big]\right]. \tag{28}$$

第一项是对 x_2 进行边缘化时得到的结果；第二项是仍依赖于 x_1 的误差函数之和，并对边缘化进行调制。我们可以将误差函数展开至二阶，

$$\frac{1}{2}\Big(\mathrm{Erf}[a+x] + \mathrm{Erf}[a-x]\Big) = \mathrm{Erf}[a] - \frac{2a\,x^2}{\sqrt{\pi}}\,e^{-a^2} + \mathcal{O}(x^4)\,,$$

将式（28）写成关于 x_1^2 的阶数形式：

$$\frac{1}{2a_2}\int_{-a_2}^{a_2} dx_2\; f(\mathbf{x}) = \frac{e^{-\frac{x_1^2}{2\sigma_1^2}}}{\sigma_1\sqrt{2\pi}}\left[\frac{1}{2a_2}\,\mathrm{Erf}\Big[\frac{a_2}{\sigma_2\sqrt{2(1-\rho^2)}}\Big] - \frac{\rho^2\,x_1^2\,e^{-\frac{a_2^2}{2\sigma_2^2(1-\rho^2)}}}{2\sigma_1^2\sigma_2(1-\rho^2)\sqrt{2\pi(1-\rho^2)}}\right]. \tag{29}$$

现在对 x_1 的先验进行积分，最终得到证据：

$$\begin{aligned} E(a_1,a_2) &= \frac{1}{4a_1a_2}\int_{-a_1}^{a_1} dx_1\int_{-a_2}^{a_2} dx_2\; f(\mathbf{x}) \\ &= \frac{1}{4a_1a_2}\,\mathrm{Erf}\Big[\frac{a_2}{\sigma_2\sqrt{2(1-\rho^2)}}\Big]\mathrm{Erf}\Big[\frac{a_1}{\sigma_1\sqrt{2}}\Big] \qquad (30)\\ &- \frac{\rho^2\,e^{-\frac{a_2^2}{2\sigma_2^2(1-\rho^2)}}}{2\sigma_1\sigma_2(1-\rho^2)\sqrt{2\pi(1-\rho^2)}}\,\frac{\mathrm{Erf}\Big[\frac{a_1}{\sigma_1\sqrt{2}}\Big]}{2a_1} + \frac{\rho^2\,e^{-\frac{a_2^2}{2\sigma_2^2(1-\rho^2)}-\frac{a_1^2}{2\sigma_1^2}}}{4\pi\sigma_1^2\sigma_2\sqrt{1-\rho^2}}\,. \end{aligned}$$

请注意，在不存在交叉相关（ $\rho\to 0$）的极限情况下，积分会因式分解，因此我们可以写出证据的精确表达式，

$$\begin{aligned} E(a_1,a_2) &= \frac{1}{4a_1a_2}\,\frac{1}{2\pi\sigma_1\sigma_2}\int_{-a_1}^{a_1} dx_1\int_{-a_2}^{a_2} dx_2\; e^{-\frac{x_1^2}{2\sigma_1^2}-\frac{x_2^2}{2\sigma_2^2}} \qquad (31)\\ &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} du_1\int_{-\infty}^{\infty} du_2\,\frac{\sin a_1u_1}{a_1u_1}\,\frac{\sin a_2u_2}{a_2u_2}\,e^{-\frac{1}{2}\sigma_1^2u_1^2-\frac{1}{2}\sigma_2^2u_2^2} \qquad (32)\\ &= \frac{1}{4a_1a_2}\mathrm{Erf}\Big[\frac{a_1}{\sigma_1\sqrt{2}}\Big]\mathrm{Erf}\Big[\frac{a_2}{\sigma_2\sqrt{2}}\Big]\,. \qquad (33) \end{aligned}$$

然而，即使存在交叉相关，如果先验较宽（$a_i \geq 2\sigma_i$），与指数项成正比的项也可以忽略不计，因此证据可以非常近似地表示为，

$$E(a_1, a_2) = \frac{1}{4a_1 a_2} \operatorname{Erf}\Big[\frac{a_2}{\sigma_2\sqrt{2(1-\rho^2)}}\Big] \operatorname{Erf}\Big[\frac{a_1}{\sigma_1\sqrt{2}}\Big] . \tag{34}$$

此外，在这种情况下，误差函数非常近似地等于 1。

n **变量** 。假设我们有 n 个相关变量，$\mathbf{x} = (x_1, \ldots, x_n)$，其协方差矩阵为

$$C_n = \begin{pmatrix} C_{11} & C_{12} & \ldots & C_{1n} \\ C_{12} & C_{22} & \ldots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \ldots & C_{nn} \end{pmatrix} . \tag{35}$$

在这种情况下，概率分布函数可以表示为

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C_n}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\Big] , \tag{36}$$

其性质是：对最后一个变量 x_n 进行边缘化后，我们得到关于 $n-1$ 个变量的相关概率分布函数，$\mathbf{x} = (x_1, \ldots, x_{n-1})$，

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big] , \tag{37}$$

其中 C_{n-1} 协方差矩阵由式 (35) 给出，但不包括最后一列和最后一行。

现在，我们将计算该多元高斯分布的证据式 (22)，首先对最后一个变量 x_n 积分，

$$\begin{aligned} \frac{1}{2a_n}\int_{-a_n}^{a_n} dx_n\ f(\mathbf{x}) &= \frac{1}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big] \\ &\quad \times \left\{ \frac{1}{2a_n} \operatorname{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathcal{O}\Big(e^{-\frac{a_n^2 \det C_{n-1}}{2\det C_n}}\Big)\right\} . \end{aligned} \tag{38}$$

现在对下一个变量 x_{n-1} 积分，可得

$$\begin{aligned} &\frac{1}{4a_n a_{n-1}}\int_{-a_n}^{a_n} dx_n \int_{-a_{n-1}}^{a_{n-1}} dx_{n-1}\ f(\mathbf{x}) = \frac{1}{(2\pi)^{(n-2)/2}\sqrt{\det C_{n-2}}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-2}^{-1}\mathbf{x}\Big] \\ &\times \left\{ \frac{1}{4a_n a_{n-1}} \operatorname{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] \operatorname{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-2}}{\det C_{n-1}}}\right] + \mathcal{O}\Big(e^{-\frac{a_n^2 \det C_{n-1}}{2\det C_n}}\Big)\right\} . \end{aligned} \tag{39}$$

继续对先验分布进行积分，我们得到 n 维分布的证据，

$$\begin{aligned} E(a_1, \ldots, a_n) &= \frac{1}{\prod_{p=1}^n 2a_p}\int_{-a_1}^{a_1}\cdots\int_{-a_n}^{a_n} d^n\mathbf{x}\ f(\mathbf{x}) \\ &= \prod_{p=1}^n \frac{1}{2a_p} \operatorname{Erf}\left[\frac{a_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right] + \mathcal{O}\left(\exp\Big[-\sum_{p=1}^n \frac{a_p^2 \det C_{p-1}}{2\det C_p}\Big]\right) , \end{aligned} \tag{40}$$

其中，协方差矩阵 C_p 按照上述方法构造，方法是依次消去 $n-p$ 最后几行和列，直到得到
$C_0 \equiv 1$ $\sum_{p=1}^{n} a_p^2$ C_{p-1} / 2 C_p $\gg$。请注意，当 (
1
34
det) (det) 时，近似效果非常好，而这种情况经常出现。另请注意，对于特定情况$n = 2$，我们可以得到前面的结果，即方程 ( )。

当 n 变量之间的互相关趋于零时，证据（40）化简为精确结果

$$E(a_1, \ldots, a_n) = \prod_{p=1}^{n} \frac{1}{2a_p} \mathrm{Erf}\left[\frac{a_p}{\sigma_p \sqrt{2}}\right] . \tag{41}$$

请注意，当我们通过令新变量x_n的先验趋于零来消除对它的需求时，证据公式（40）正确地反映了这一极限，

$$\lim_{a_n \to 0} E(a_1, \ldots, a_n) = E(a_1, \ldots, a_{n-1}) \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\det C_{n-1}}{\det C_n}} , \tag{42}$$

因此，在该极限下，我们恢复了一个适当归一化的分布， $f(x_1, \ldots\ldots, x_n) \to f(x_1, \ldots\ldots, x_{n-1})$，而仅考察似然函数则无法给出合理的答案。

另一方面，如果我们的理论偏见无法为某个变量指定具体的先验，我们会发现，随着 a 增大，证据会按 $1/2a$ 递减。因此，贝叶斯证据似乎非常适合用于区分不同的理论先验，并会惩罚纳入过多参数的做法，这正如奥卡姆剃刀所言。

3.2 非中心先验

先验不太可能恰好以分布的均值为中心，因为先验不应预先知道数据会告诉我们什么。因此，我们需要将上述内容推广到非中心先验的情形。我们仍假设先验是顶帽分布。

我们目前仍假设该概率分布可以很好地近似为均值为 μ 的高斯分布。接下来使用平移变量 $\tilde{x}_i = x_i - \mu_i$，并将高斯分布函数写成公式（36）的形式。现在，归一化的顶帽先验相对于均值而言不再居中，

$$\pi(\tilde{x}; a, b) \equiv \begin{cases} (a+b)^{-1} & -a < \tilde{x} < b \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{43}$$

对于单变量，结果是精确的，

$$E(a; b) = \int_{-\infty}^{\infty} dx \, f(x) \, \pi(x; a, b) = \frac{1}{2a + 2b} \left(\mathrm{Erf}\left[\frac{a}{\sigma\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{b}{\sigma\sqrt{2}}\right] \right) . \tag{44}$$

其中我们对偏移变量 $\tilde{x}$ 进行积分，以下将其重新记为 x。请注意，在极限 $b \to a$ 下，我们可以得到中心先验情形的结果公式（24）。

对于两个变量，分布函数如公式（26）所示，非中心贝叶斯证据为

$$E(a_1, a_2; b_1, b_2) = \frac{1}{(a_1 + b_1)(a_2 + b_2)} \int_{-a_1}^{b_1} dx_1 \int_{-a_2}^{b_2} dx_2 \, f(x_1, x_2) \tag{45}$$

$$\begin{aligned} = \frac{1}{(2a_1 + 2b_1)(2a_2 + 2b_2)} & \left\{ \left(\mathrm{Erf}\left[\frac{a_1}{\sigma_1\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{b_1}{\sigma_1\sqrt{2}}\right] \right) \right. \\ & \times \left(\mathrm{Erf}\left[\frac{a_2}{\sigma_2\sqrt{2(1-\rho^2)}}\right] + \mathrm{Erf}\left[\frac{b_2}{\sigma_2\sqrt{2(1-\rho^2)}}\right] \right) \\ & \left. - \frac{\rho}{2\pi\sqrt{1-\rho^2}} \left(e^{-\frac{a_1^2}{2\sigma_1^2}} - e^{-\frac{b_1^2}{2\sigma_1^2}} \right) \left(e^{-\frac{a_2^2}{2\sigma_2^2(1-\rho^2)}} + e^{-\frac{b_2^2}{2\sigma_2^2(1-\rho^2)}} \right) \right\} \end{aligned} \tag{46}$$

多变量情形的证据公式（36）为

$$E(\mathbf{a},\mathbf{b}) = \int_{-\infty}^{\infty} d^n\mathbf{x}\, f(\mathbf{x}) \prod_{i=1}^{n} \pi(x_i; a_i, b_i) = \prod_{i=1}^{n}(a_i+b_i)^{-1} \int_{-a_1}^{b_1} d\tilde{x}_1 \cdots \int_{-a_n}^{b_n} d\tilde{x}_n\, f(\tilde{\mathbf{x}})\,. \tag{47}$$

现在，我们从对最后一个变量 x_n 积分开始，计算多元高斯分布公式（36）的结果，

$$\begin{aligned}
\frac{1}{a_n+b_n}\int_{-a_n}^{b_n} dx_n\; f(\mathbf{x}) \;&=\; \frac{1}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big]\, \frac{1}{(2a_n+2b_n)} \\
&\times\left\{ \mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathcal{O}\left(e^{-\frac{a_n^2\,\det C_{n-1}}{2\det C_n}} + e^{-\frac{b_n^2\,\det C_{n-1}}{2\det C_n}}\right)\right\}
\end{aligned} \tag{48}$$

现在对下一个变量 x_{n-1} 积分，得到

$$\begin{aligned}
&\frac{1}{(a_n+b_n)(a_{n-1}+b_{n-1})}\int_{-a_n}^{b_n} dx_n \int_{-a_{n-1}}^{b_{n-1}} dx_{n-1}\; f(\mathbf{x}) = \\
&\frac{1}{(2\pi)^{(n-2)/2}\sqrt{\det C_{n-2}}} \exp\Big[-\frac{1}{2}\,\mathbf{x}^T C_{n-2}^{-1}\mathbf{x}\Big]\, \frac{1}{(2a_n+2b_n)(2a_{n-1}+2b_{n-1})} && (49)\\
&\times\left\{\left(\mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right)\right. && (50)\\
&\times\left(\text{误差}\right. \\
&\text{函数}\left[\;\frac{a_{n-1}}{\sqrt{2}}\sqrt{\frac{\det C_{n-2}}{\det C_{n-1}}}\right] + \mathrm{Erf}\left[\frac{b_{n-1}}{\sqrt{2}}\sqrt{\frac{\det C_{n-2}}{\det C_{n-1}}}\right]\Big) && (51)\\
&+\;\left.\mathcal{O}\left(e^{-\frac{a_n^2\,\det C_{n-1}}{2\det C_n}} + e^{-\frac{b_n^2\,\det C_{n-1}}{2\det C_n}}\right)\times\left(e^{-\frac{a_{n-1}^2\,\det C_{n-2}}{2\det C_{n-1}}} + e^{-\frac{b_{n-1}^2\,\det C_{n-2}}{2\det C_{n-1}}}\right)\right\}.
\end{aligned}$$

继续对先验分布积分，最终得到 n 维分布的证据，

$$\begin{aligned}
E(\mathbf{a},\mathbf{b}) \;&=\; \frac{1}{\prod_{p=1}^{n}(a_p+b_p)}\int_{-a_1}^{b_1}\cdots\int_{-a_n}^{b_n} d^n\mathbf{x}\; f(\mathbf{x}) \\
&=\; \prod_{p=1}^{n}\frac{1}{(2a_p+2b_p)}\left(\mathrm{Erf}\left[\frac{a_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right] + \mathrm{Erf}\left[\frac{b_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right]\right) \\
&\qquad + \mathcal{O}\left(\prod_{p=1}^{n}\left[\exp\Big(-\frac{a_p^2\,\det C_{p-1}}{2\det C_p}\Big) + \exp\Big(-\frac{b_p^2\,\det C_{p-1}}{2\det C_p}\Big)\right]\right),
\end{aligned} \tag{52}$$

其中，协方差矩阵 C_p 按照上述方法构造，依次去掉最后 $n-p$ 行和列，直到 $C_0 \equiv 1$。请注意，只要指数很大，近似就非常准确：$\sum_{p=1}^{n}(a_p^2\det C_{p-1})/(2\det C_p) \gg 1$；这种情况很常见。还要注意，在极限 $b \to a$ 下，我们会得到中心化先验分布的证据表达式公式（40）。

现在让我们评估一个归一化到似然分布极大值的分布的证据，

$$f(\mathbf{x}) = \mathcal{L}_{\max}\exp\Big[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\Big] \tag{53}$$

在这种情况下，证据由公式（52）给出，并乘以归一化产生的因子 $\mathcal{L}_{\max} \times (2\pi)^{n/2}\sqrt{\det C_n}$。然后，我们可以忽略指数级微小的修正项，计算证据的对数：

$$\begin{aligned}\ln E &= \ln \mathcal{L}_{\max} + \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln\det C_n - \sum_{p=1}^{n}\ln(2a_p + 2b_p) \\ &\quad + \sum_{p=1}^{n}\ln\left(\mathrm{Erf}\left[\frac{a_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right] + \mathrm{Erf}\left[\frac{b_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right]\right) . \end{aligned} \tag{54}$$

无关联情形。假设我们有一个变量之间不存在相关性的多元高斯分布，即 $C_{ij} = \sigma_i^2\delta_{ij}$ 是一个对角矩阵；那么证据恰好为，

$$E(\mathbf{a},\mathbf{b}) = \frac{1}{\prod_{p=1}^{n}(a_p+b_p)}\int_{-a_1}^{b_1}\cdots\int_{-a_n}^{b_n} d^n\mathbf{x}\; f(\mathbf{x}) = \prod_{p=1}^{n}\frac{1}{2(a_p+b_p)}\left(\mathrm{Erf}\left[\frac{a_p}{\sigma_p\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{b_p}{\sigma_p\sqrt{2}}\right]\right), \tag{55}$$

其中 σ_p 是每个变量 $\tilde{x}_p$的离散度，因此证据的对数变为 es

$$\ln E = \ln \mathcal{L}_{\max} + \frac{n}{2}\ln(2\pi) + \sum_{p=1}^{n}\ln\sigma_p - \sum_{p=1}^{n}\ln(2a_p+2b_p) + \sum_{p=1}^{n}\ln\left(\mathrm{Erf}\left[\frac{a_p}{\sigma_p\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{b_p}{\sigma_p\sqrt{2}}\right]\right) \tag{56}$$

拉普拉斯近似。对证据的拉普拉斯近似假设分布是相关高斯分布，并且先验分布足够宽，使得整个分布都能轻松地落在先验范围内。在这种情况下，误差函数近似为 1，不会对证据产生贡献；由公式 (54) 可得

$$\ln E = \ln \mathcal{L}_{\max} + \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln\det C_n - \sum_{p=1}^{n}\ln\Delta\theta_p\,, \tag{57}$$

其中 $\Delta\theta_p = a_p + b_p$ 是与先验相关的参数区间。在下一节中，我们将比较不同的近似方法。

4 非高斯修正项

这种方法的优点在于，给定分布的任意一组矩，就可以系统地计算给定模型及其自身先验分布的证据。在此，我们将考察协方差矩阵之外的前两个矩，即偏度项和峰度项，见式（9）。

4.1 偏度

让我们从高斯近似的第一个修正项——三线性项 B_{ijk}——开始。为此，我们将生成泛函（9）写为

$$\phi(\mathbf{u}) = \exp\left[i\,\mu_i u_i - \frac{1}{2!}\,C_{ij}\,u_i u_j - \frac{i}{3!}\,B_{ijk}\,u_i u_j u_k\right] . \tag{58}$$

通过进行变量替换 $u_i = y_i - i\,C_{ik}^{-1}(x_k - \mu_k)$，我们可以计算傅里叶变换积分，并得到规范化的概率分布函数

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det C_n}} \exp\Big[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\Big] \\ &\quad \times \left(1 - \frac{1}{2}B_{ijk}\,C_{ij}^{-1}C_{kl}^{-1}\,x_l + \frac{1}{6}B_{ijk}\,C_{il}^{-1}C_{jm}^{-1}C_{kn}^{-1}\,x_l x_m x_n\right), \end{aligned} \tag{59}$$

其中 x_k 是偏移坐标（$x_k - \mu_k$）。该偏斜分布函数满足

$$\langle x_i\rangle = 0\,, \qquad \langle x_i x_j\rangle = C_{ij}\,, \qquad \langle x_i x_j x_k\rangle = B_{ijk}\,, \qquad \langle x_i x_j x_k x_l\rangle = 0\,, \quad \cdots \tag{60}$$

通过直接计算即可验证这一点。现在，让我们计算该偏斜模型的公式（22）中的证据。由于公式（59）括号中的附加项关于 x 都是奇函数，在居中矩形先验公式（21）所对应的这类关于中心对称的积分区间上进行积分时，它们对证据的贡献会相互抵消，因此该偏斜模型的最终证据与高斯模型公式（40）的证据并无不同。如果先验相对于均值存在偏移，例如公式（43）中的情况，那么奇函数项对证据的贡献将不会抵消。下面来计算它们的贡献。

对于单变量（$n = 1$），正确归一化的似然函数可以写成

$$f(x) = \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}}\,\left(1 - \frac{B\,x}{2\sigma^4} + \frac{B\,x^3}{6\sigma^6}\right),$$

满足 $\langle x\rangle = 0$、$\langle x^2\rangle = \sigma^2$、$\langle x^3\rangle = B$，贝叶斯积分可以精确地计算为

$$E(a,b) = \frac{1}{2a+2b}\left(\mathrm{Erf}\left[\frac{a}{\sigma\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{b}{\sigma\sqrt{2}}\right]\right) - \frac{B\sigma^{-3}}{6\sqrt{2\pi}}\left[\left(1 - \frac{a^2}{\sigma^2}\right)e^{-\frac{a^2}{2\sigma^2}} - \left(1 - \frac{b^2}{\sigma^2}\right)e^{-\frac{b^2}{2\sigma^2}}\right]\frac{1}{a+b}\,. \tag{61}$$

请注意，对于偶对称（以零为中心）的先验分布，在 $b = a$ 的情况下，证据可简化为公式（24）。

对于任意数量的变量，计算会更复杂。我们从第 n 个变量开始，为了计算积分，定义辅助函数

$$\begin{aligned} g(\lambda) &= \int_{-a_n}^{b_n} dx_n\, x_n\, \frac{\exp\Big[-\frac{\lambda}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\Big]}{(2\pi)^{n/2}\sqrt{\det C_n}} = \frac{\exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \times \\ &\quad \times \frac{1}{\lambda\sqrt{2\pi}}\left(\exp\Big[-\frac{\lambda a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\Big] - \exp\Big[-\frac{\lambda b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\Big]\right), \end{aligned} \tag{62}$$

使得，利用误差函数$'[x] = \frac{2}{\sqrt{\pi}}\,e^{-x^2}$，

$$\begin{aligned} -2g'(\lambda = 1) = \int_{-a_n}^{b_n} dx_n\, x_n\, \frac{(\mathbf{x}^T C_n^{-1}\mathbf{x})\,\exp\Big[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\Big]}{(2\pi)^{n/2}\sqrt{\det C_n}} &= \frac{\exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \times \\ \times \frac{1}{\sqrt{2\pi}}\left\{\left(2 + a_n^2\frac{\det C_{n-1}}{\det C_n}\right)\exp\Big[-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\Big] - \left(2 + b_n^2\frac{\det C_{n-1}}{\det C_n}\right)\exp\Big[-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\Big]\right\} & \end{aligned} \tag{63}$$

因此，利用式（63），对 x_n 非居中先验上的偏度校正分布函数（式（59））进行积分，可得

$$\begin{aligned} \int_{-a_n}^{b_n} dx_n\; f(\mathbf{x}) &= \frac{\exp\Big[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\Big]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}}\left\{\frac{1}{2}\left(\mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right)\right. \\ &\left. - \frac{1}{6}B_{ijn}\,C_{ij}^{-1}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\left[\left(1 - a_n^2\frac{\det C_{n-1}}{\det C_n}\right)e^{-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}} - \left(1 - b_n^2\frac{\det C_{n-1}}{\det C_n}\right)e^{-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}}\right]\right\} \end{aligned} \tag{64}$$

定义两个新函数:

$$E_i(a_i,b_i) = \frac{1}{2}\left(\mathrm{Erf}\left[\frac{a_i}{\sqrt{2}}\sqrt{\frac{\det C_{i-1}}{\det C_i}}\right]+\mathrm{Erf}\left[\frac{b_i}{\sqrt{2}}\sqrt{\frac{\det C_{i-1}}{\det C_i}}\right]\right), \tag{65}$$

$$F_i(a_i,b_i) = \frac{1}{6\sqrt{2\pi}}\sqrt{\frac{\det C_{i-1}}{\det C_i}}\left[\left(1-a_i^2\frac{\det C_{i-1}}{\det C_i}\right)e^{-\frac{a_i^2}{2}\frac{\det C_{i-1}}{\det C_i}}-\left(1-b_i^2\frac{\det C_{i-1}}{\det C_i}\right)e^{-\frac{b_i^2}{2}\frac{\det C_{i-1}}{\det C_i}}\right].$$

依次对 x_{n-1}、.....、x_1 积分，最终得到经过三阶修正的概率分布函数 $f(\mathbf{x})$ 的贝叶斯证据,

$$E(\mathbf{a},\mathbf{b}) = \prod_{p=1}^{n}\frac{E_p(a_p,b_p)}{(a_p+b_p)}\left[1-\sum_{k=1}^{n}B_{ijk}\,C_{ij}^{-1}\,\frac{F_k(a_k,b_k)}{E_k(a_k,b_k)}\right]. \tag{66}$$

除非 $B_{ijk}\,C_{ij}^{-1}$ 非常大，否则对误差函数的修正会受到指数级抑制，因此我们预计不会显著偏离高斯情形公式（40）。还需注意，如果先验是对称的，很容易看出积分中的偏度部分会消失，$F_k(a_k、\ b_k)\to 0$，取 $b_k\to a_k$ 即可明确验证这一点。

4.2 峰度

偏度之后的下一项修正是四阶矩（即峰度），由公式（9）中的 D_{ijkl} 项给出。暂且忽略三阶偏度，并写作

$$\phi(\mathbf{u}) = \exp\left[i\,\mu_i u_i-\frac{1}{2!}\,C_{ij}\,u_iu_j+\frac{1}{4!}\,D_{ijkl}\,u_iu_ju_ku_l\right]. \tag{67}$$

通过执行相同的变量替换 $u_i = y_i - i\,C_{ik}^{-1}\,(x_k-\mu_k)$，现在可以计算傅里叶变换，并得到适当归一化的概率分布函数

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det C_n}}\exp\left[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\right]\left(1+\frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}\right.\\ &\left.-\frac{1}{4}D_{ijkl}\,C_{ij}^{-1}C_{km}^{-1}C_{ln}^{-1}\,x_mx_n+\frac{1}{24}D_{ijkl}\,C_{im}^{-1}C_{jn}^{-1}C_{kp}^{-1}C_{lq}^{-1}\,x_mx_nx_px_q\right). \end{aligned} \tag{68}$$

计算这些积分后，很容易看出该分布满足

$$\langle x_ix_j\rangle = C_{ij}\,,\qquad \langle x_ix_jx_kx_l\rangle = D_{ijkl}+C_{ij}C_{kl}+C_{ik}C_{jl}+C_{il}C_{jk}\,,\qquad \dots \tag{69}$$

请注意，要使新的似然分布（68）正定，必须满足$D_{ijkl}C_{ij}^{-1}C_{kl}^{-1} < 4$；若进一步要求中心处只有一个极大值，则还必须满足 $D_{ijkl}C_{ij}^{-1}C_{kl}^{-1} < 2$。这些条件限制了证据相对于高斯分布的证据可能产生的最大偏差。

现在来计算这个峰度模型的证据，见公式（22）。公式（68）括号中的额外项都是 x的偶函数，即使对于中心先验分布也不能忽略。

对于单变量（$n=1$），正确归一化的似然函数可写为

$$f(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\left(1+\frac{D}{8\sigma^4}-\frac{D\,x^2}{4\sigma^6}+\frac{D\,x^4}{24\sigma^8}\right),$$

满足 $\langle x\rangle = 0$、$\langle x^2\rangle = \sigma^2$、$\langle x^3\rangle = 0$、$\langle x^4\rangle = D+3\sigma^4$ 等条件。贝叶斯积分可以精确地计算为

$$E(a,b) = \frac{1}{2a+2b}\left(\mathrm{Erf}\left[\frac{a}{\sigma\sqrt{2}}\right]+\mathrm{Erf}\left[\frac{b}{\sigma\sqrt{2}}\right]\right)+\frac{D\sigma^{-4}}{8\sqrt{2\pi}}\left(\frac{a}{\sigma}\left(1-\frac{a^2}{3\sigma^2}\right)e^{-\frac{a^2}{2\sigma^2}}+\frac{b}{\sigma}\left(1-\frac{b^2}{3\sigma^2}\right)e^{-\frac{b^2}{2\sigma^2}}\right)\frac{1}{a+b}. \tag{70}$$

对于任意数量的变量，计算过程同样会复杂得多。让我们从第 n 个变量开始，并且为了计算第一个积分，定义一个新的辅助函数

$$h(\lambda) = \int_{-a_n}^{b_n} dx_n \frac{\exp\left[-\frac{\lambda}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\right]}{(2\pi)^{n/2}\sqrt{\det C_n}} = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\right]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \times$$
$$\times\frac{1}{2\sqrt{\lambda}}\left(\mathrm{Erf}\left[\frac{a_n\sqrt{\lambda}}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n\sqrt{\lambda}}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right), \tag{71}$$

使得，

$$-2h'(\lambda=1) = \int_{-a_n}^{b_n} dx_n \frac{(\mathbf{x}^T C_n^{-1}\mathbf{x})\,\exp\left[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\right]}{(2\pi)^{n/2}\sqrt{\det C_n}} = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\right]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \times$$
$$\times\left\{\frac{1}{2}\left(\mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right)\right. \tag{72}$$
$$\left. - \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\left(a_n\,\exp\left[-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right] + b_n\,\exp\left[-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right]\right)\right\}.$$

$$4h''(\lambda=1) = \int_{-a_n}^{b_n} dx_n \frac{(\mathbf{x}^T C_n^{-1}\mathbf{x})^2\,\exp\left[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\right]}{(2\pi)^{n}\sqrt{\det C_n}} = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\right]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}} \times$$
$$\times\left\{\frac{3}{2}\left(\mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right)\right. \tag{73}$$
$$- \frac{3}{\sqrt{2\pi}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\left(a_n\,\exp\left[-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right] + b_n\,\exp\left[-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right]\right)$$
$$\left. - \frac{a_n^2}{\sqrt{2\pi}}\left(\frac{\det C_{n-1}}{\det C_n}\right)^{3/2}\left(a_n\,\exp\left[-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right] + b_n\,\exp\left[-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}\right]\right)\right\}.$$

因此，利用式 (72) 和 (73)，峰度校正分布函数 (68) 在 x_n 先验上的积分变为

$$\int_{-a_n}^{b_n} dx_n\; f(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^T C_{n-1}^{-1}\mathbf{x}\right]}{(2\pi)^{(n-1)/2}\sqrt{\det C_{n-1}}}\left\{\frac{1}{2}\left(\mathrm{Erf}\left[\frac{a_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right] + \mathrm{Erf}\left[\frac{b_n}{\sqrt{2}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\right]\right) + \right. \tag{74}$$
$$\left. + \frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\det C_{n-1}}{\det C_n}}\left[a_n\left(1 - \frac{a_n^2}{3}\frac{\det C_{n-1}}{\det C_n}\right)e^{-\frac{a_n^2}{2}\frac{\det C_{n-1}}{\det C_n}} + b_n\left(1 - \frac{b_n^2}{3}\frac{\det C_{n-1}}{\det C_n}\right)e^{-\frac{b_n^2}{2}\frac{\det C_{n-1}}{\det C_n}}\right]\right\}.$$

现在我们可以定义一个新函数

$$G_i(a_i,b_i) = \frac{1}{8\sqrt{2\pi}}\sqrt{\frac{\det C_{i-1}}{\det C_i}}\left[a_i\left(1 - \frac{a_i^2}{3}\frac{\det C_{i-1}}{\det C_i}\right)e^{-\frac{a_i^2}{2}\frac{\det C_{i-1}}{\det C_i}} - b_i\left(1 - \frac{b_i^2}{3}\frac{\det C_{i-1}}{\det C_i}\right)e^{-\frac{b_i^2}{2}\frac{\det C_{i-1}}{\det C_i}}\right]. \tag{75}$$

对 x_{n-1}、.....、x_1 进行迭代积分后，我们得到四阶校正概率分布函数 f（$\mathbf{x}$）的贝叶斯证据，

$$E(\mathbf{a},\mathbf{b}) = \prod_{p=1}^{n}\frac{E_p(a_p,b_p)}{(a_p+b_p)}\left[1 + D_{ijkl}\,C_{ij}^{-1}\,C_{kl}^{-1}\sum_{m=1}^{n}\frac{G_m(a_m,b_m)}{E_m(a_m,b_m)}\right]. \tag{76}$$

因此，除非 $D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}$ 非常大，否则对误差函数的校正会受到指数抑制，我们预计不会显著偏离高斯情形，即公式（40）。

为了比较模型，通常会计算证据的对数。假设给定一个以最大似然归一化、且包含直至四阶修正项的似然分布函数，

$$f(\mathbf{x}) = \mathcal{L}_{\max}\exp\left[-\frac{1}{2}\mathbf{x}^T C_n^{-1}\mathbf{x}\right]\left(1+\frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}\right)^{-1}\Bigg(1-\frac{1}{2}B_{ijk}\,C_{ij}^{-1}C_{kl}^{-1}\,x_l+\frac{1}{6}B_{ijk}\,C_{il}^{-1}C_{jm}^{-1}C_{kn}^{-1}\,x_l x_m x_n$$
$$+\,\frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}-\frac{1}{4}D_{ijkl}\,C_{ij}^{-1}C_{km}^{-1}C_{ln}^{-1}\,x_m x_n+\frac{1}{24}D_{ijkl}\,C_{im}^{-1}C_{jn}^{-1}C_{kp}^{-1}C_{lq}^{-1}\,x_m x_n x_p x_q\Bigg)\,. \tag{77}$$

请注意，该分布经过归一化，使得极大值对应于均值中心化分布，即 $\mathbf{x}\,=0$。在这种情况下，归一化分布的证据为

$$E(\mathbf{a},\mathbf{b}) = \mathcal{L}_{\max}\;(2\pi)^{n/2}\sqrt{\det C_n}\left(1+\frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}\right)^{-1}\times \tag{78}$$
$$\prod_{p=1}^{n}\frac{E_p(a_p,b_p)}{(a_p+b_p)}\left[1-\sum_{k=1}^{n}B_{ijk}\,C_{ij}^{-1}\,\frac{F_k(a_k,b_k)}{E_k(a_k,b_k)}+D_{ijkl}\,C_{ij}^{-1}\,C_{kl}^{-1}\sum_{m=1}^{n}\frac{G_m(a_m,b_m)}{E_m(a_m,b_m)}\right].$$

于是，我们可以通过下式计算证据的对数：

$$\begin{aligned}\ln E \;\;=\;\; & \ln\mathcal{L}_{\max}+\frac{n}{2}\ln(2\pi)+\frac{1}{2}\ln\det C_n-\ln\left(1+\frac{1}{8}D_{ijkl}\,C_{ij}^{-1}C_{kl}^{-1}\right)-\sum_{p=1}^{n}\ln(2a_p+2b_p)\\ & +\;\sum_{p=1}^{n}\ln\left(\mathrm{Erf}\left[\frac{a_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right]+\mathrm{Erf}\left[\frac{b_p}{\sqrt{2}}\sqrt{\frac{\det C_{p-1}}{\det C_p}}\right]\right)\\ & +\;\ln\left(1-\sum_{k=1}^{n}B_{ijk}\,C_{ij}^{-1}\,\frac{F_k(a_k,b_k)}{E_k(a_k,b_k)}+D_{ijkl}\,C_{ij}^{-1}\,C_{kl}^{-1}\sum_{m=1}^{n}\frac{G_m(a_m,b_m)}{E_m(a_m,b_m)}\right).\end{aligned} \tag{79}$$

请注意，条件 $D_{ijkl}C_{ij}^{-1}C_{kl}^{-1}\,<\,2$ 限制了峰度修正对证据的最大贡献。

无关联情形。如果似然分布中的不同变量之间不存在相关性，则贝叶斯证据的精确表达式为

$$\ln E = \ln\mathcal{L}_{\max}+\frac{n}{2}\ln(2\pi)+\sum_{p=1}^{n}\ln\sigma_p-\sum_{p=1}^{n}\ln(2a_p+2b_p)+\sum_{p=1}^{n}\ln\left(\mathrm{Erf}\left[\frac{a_p}{\sigma_p\sqrt{2}}\right]+\mathrm{Erf}\left[\frac{b_p}{\sigma_p\sqrt{2}}\right]\right) \tag{80}$$
$$-\;\ln\left(1+\frac{1}{8}D_{iijj}\,\sigma_i^{-2}\sigma_j^{-2}\right)+\;\ln\left(1-\sum_{k=1}^{n}B_{iik}\,\sigma_k^{-2}\,\frac{F_k(a_k,b_k)}{E_k(a_k,b_k)}+D_{iijj}\,\sigma_i^{-2}\sigma_j^{-2}\sum_{m=1}^{n}\frac{G_m(a_m,b_m)}{E_m(a_m,b_m)}\right),$$

其中，σ_p 是变量 x_p 的相应离散度，而函数 E_i，F_i 和 G_i 是式 (65) 和 (75) 中对应于无关联矩阵的极限函数。

5 模型比较

最后，我们将转向上述形式体系的具体应用。首先，我们会进行一些玩具模型测试，以评估其性能；然后考察一些真实的宇宙学应用，而我们此前已通过热力学积分获得了这些应用的结果 [5]。

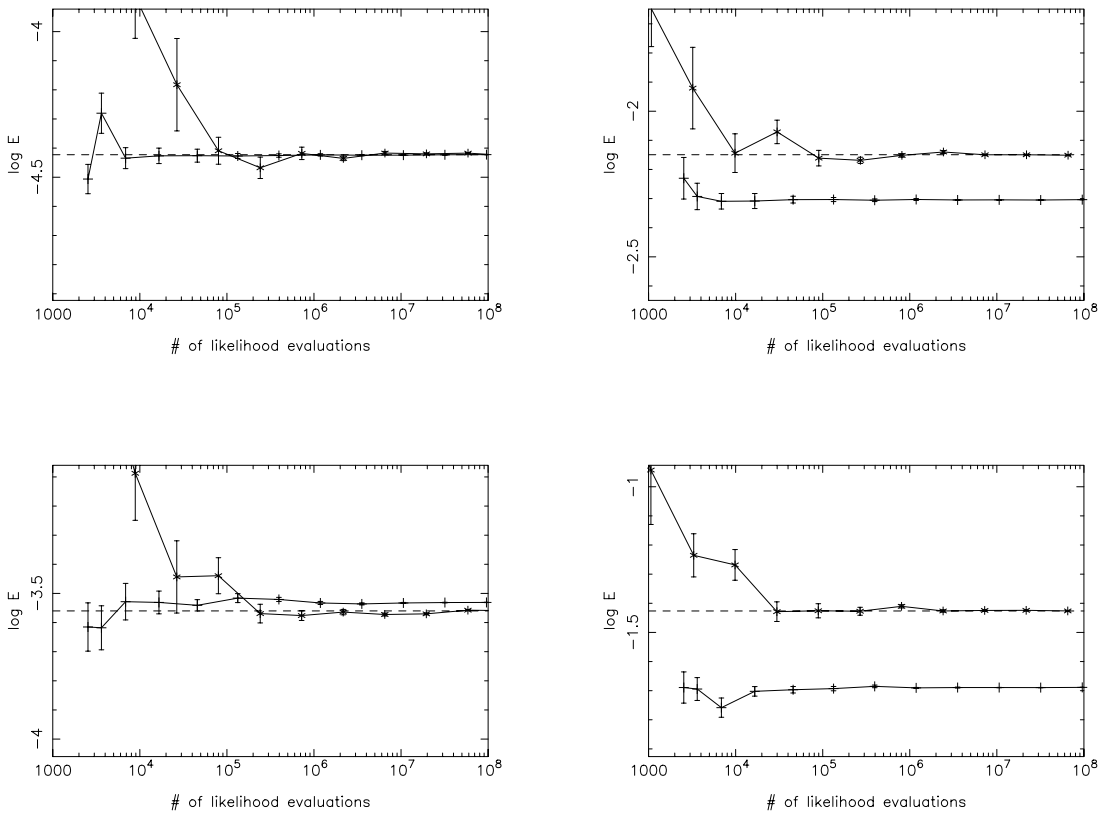


图1：该图展示了计算得到的证据随似然评估次数的变化。请注意，横轴采用对数刻度。实线表示热力学积分。点线和点划线分别表示应用和未应用非高斯修正的解析方法。水平虚线表示通过直接积分得到的数值。上面两个面板对应 $\mathcal{L}_g$，下面两个面板对应 $\mathcal{L}_{ng}$。左侧面板对应两个参数的宽平坦先验 $(-7{,}10)$，右侧面板对应两个参数的窄先验分布 $(-2, 3)$。讨论见正文。

5.1 Ababy 玩具模型比较

我们先从一个非常简单的二维玩具模型入手。本节旨在说明热力学积分的低效，并初步展示我们在此提出的方法的性能。此外，这个二维模型足够简单，可以通过蛮力直接数值积分计算证据，从而同时检验计算精度。我们采用以下两种似然形式：

$$\mathcal{L}_g(x,y) = \exp\left[-\frac{2x^2-2(y-1)^2-xy}{2}\right] \tag{81}$$

$$\mathcal{L}_{ng}(x,y) = \exp\left[-\frac{2x^2-2(y-1)^2-xy}{2}\right] + \exp\left[-\frac{2x^2-2y^2-3xy}{2}\right] \tag{82}$$

下标 g 和 ng 分别表示高斯情形和非高斯情形。

首先，我们计算 使用式 (56) 和 (80) 以及协方差，通过解析方法计算证据 e

矩阵，这些矩阵是通过使用标准 Metropolis–Hastings 算法（提议宽度固定）对似然进行采样推断得到的。使用了从几条到数百万个样本不等的马尔可夫链。我们还使用参考文献 [5] 中介绍的热力学算法计算证据。同样，我们调整算法参数，以获得精度各异的证据值。图1绘出了所得证据随似然评估次数的变化，并同时给出了通过直接数值积分推断出的正确值。似然评估次数至关重要，因为在宇宙学参数估计和模型比较中，这是限制计算速度的步骤。结果与预期一致。我们注意到，先验范围的大小似乎并非至关重要。考虑到解析方法需要了解真实协方差信息，而我们只能提供根据先验截断似然估算出的协方差矩阵，这一点令人安心。我们还注意到，在所有情况下，热力学积分最终都会收敛到正确值。不过，这需要进行非常多次似然评估；即使对于二维问题，通常也需要约一百万次。解析方法在样本数达到一万时就已受到系统误差的限制。对于高斯情形，按构造不存在系统误差；而非高斯情形则存在约为 0.1 的系统误差（以 $\ln E$ 计）。非高斯修正将误差减小了约一半，因此正确估计了纯高斯近似所带来的不确定性。在宽先验情形下，唯一具有显著量级的非高斯修正是 $\ln(1 + D_{ijkl}C_{ij}^{-1}C_{kl}^{-1}/8)$。

5.2 玩具模型比较

现在，我们继续计算表 I 所示的具有 5 个和 6 个参数的简单玩具模型的贝叶斯证据。其目的是再次将结果与通过热力学积分获得的结果进行比较，但这一次使用的模型与宇宙学中通常遇到的典型问题更加相似。

参数	Mean	先验范围	模型
x_1	0.022	$[0.0001, 0.044]$	toy5，toy6
x_2	0.12	$[0.001, 0.3]$	toy5，toy6
x_3	1.04	$[0.8, 1.4]$	toy5，toy6
x_4	0.1	$[0.01, 0.3]$	toy5,toy6
x_5	3.1	$[2.6, 3.6]$	toy5,toy6
x_6	0.98	$[0.5, 1.5]$	toy6

表1：用于解析评估玩具模型证据的参数，分别对应 5 个和 6 个参数。玩具模型的似然极大值（任意地）取为 $\mathcal{L}_{\max} = 1$。

从五参数模型开始，我们首先假设其似然分布为不相关的多元高斯分布。在这种情况下，目标是测试热力学积分法，该方法给出 $\ln E^{\rm num}_{\rm toy5} = -8.65 \pm 0.03$，而精确表达式给出 $\ln E^{\rm ana}_{\rm toy5} = -8.66$。因此，我们得出结论，热力学积分法在获得正确的模型证据方面表现相当不错。对于不相关分布，式 (57) 中的拉普拉斯近似也表现良好，$\ln E^{\rm Lap}{}_{\text{玩具5}} = -8.67$。

现在我们考虑一个具有相关协方差矩阵 C_{ij} 的似然函数，其均值和离散度与前一种情况相同，但存在显著的相关性。所需的解析公式，即公式 (54)，不再是精确的，[2] 并给出 $\ln E^{\rm ana}_{\rm toy5c} = -7.32$。作为比较，热力学积分给出 $\ln E^{\rm num}_{\rm toy5c} = -7.28 \pm 0.06$，同样在误差范围内完全一致。在这种情况下，拉普拉斯近似显著失效，$\ln E^{\rm Lap}{}_{\rm toy5c} = -6.89$，原因在于所选的相关性使后验分布与先验分布的边界发生了显著接触。

现在让我们回到无关联情形，并像表 I 中那样引入一个新参数 x_6，然后评估因这一新参数而出现的不同证据，以考察系统误差对贝叶斯证据评估的敏感性，以及它们对模型比较的影响。数值

[2] 可以旋转参数基底以消除相关性，但这样一来先验分布就不会是顶帽先验了。

结果为 $\ln E^{\rm num}_{\rm toy6} = -10.75 \pm 0.03$，而精确解析表达式给出 $\ln E^{\rm ana}_{\rm toy6} = -10.74$，两者在误差范围内完全一致。对于无关联分布，式 (57) 中的拉普拉斯近似再次表现良好：$\ln E^{\rm Lap}{}_{\text{玩具6}} = -10.74$。

当似然函数中的相关性较强，且先验范围不太大时，朴素的拉普拉斯近似（公式 (57)）的表现不如解析近似（公式 (54)）。

5.3 面积模型比较

在本小节中，我们将使用参考文献 [5]，中的结果。在该文中，我们评估了 5 参数和 6 参数绝热模型，以及三种 10 参数混合绝热与等曲率模型的证据。所用的先验范围列于表 II。后几种模型对数据的拟合略好，但需要更多参数；这正是需要采用模型选择技术来得出可靠结论的情形。在参考文献 [5] 中，我们使用热力学积分计算证据，并表明等曲率模型不如绝热模型受青睐，但显著性水平仅为轻微。[3]

从最简单的绝热模型开始，该模型采用哈里森–泽尔多维奇谱。我们将上述解析公式（公式 (54)）与 cosmoMC 程序 [10], 提供的协方差矩阵结合使用，得到了 $\ln E^{\rm ana}_{\rm ad} = -854.07$；而热力学积分得到 $\ln E^{\rm num}_{\rm ad} = -854.1 \pm 0.1$ [5]。两者吻合极佳；这是因为绝热模型的分布函数可以很好地用高斯分布近似，而且先验范围相当大，因此公式 (54) 与拉普拉斯近似所得结果 $\ln E^{\rm Lap}{}_{\text{绝热}} = -854.08$ 非常接近。

参数	Mean	先验范围	模型
$\omega_{\rm b}$	0.022	[0.018, 0.032]	AD-HZ,AD-$n_{\rm s}$,ISO
$\omega_{\rm dm}$	0.12	[0.04, 0.16]	AD-HZ，AD-$n_{\rm s}$，ISO
θ	1.04	[0.98, 1.10]	AD-HZ，AD-$n_{\rm s}$，ISO
τ	0.17	$[0, 0.5]$	AD-HZ，AD-$n_{\rm s}$，ISO
$\ln[10^{10}\mathcal{R}_{\rm rad}]$	3.1	[2.6, 4.2]	AD-HZ,AD-$n_{\rm s}$,ISO
$n_{\rm s}$	1.0	[0.8, 1.2]	AD-$n_{\rm s}$,ISO
$n_{\rm iso}$	1.5	$[0, 3]$	ISO
$\delta_{\rm cor}$	1.5	$[-0.14,\ 0.4]$	ISO
$\sqrt{\alpha}$	0	$[-1, 1]$	ISO
β	0	$[-1, 1]$	ISO

表 2: 模型中使用的参数；命名方式及其他详情请参阅参考文献 [5] 。对于 AD-HZ 模型 $n_{\rm s}$ 被固定为 1，而 $n_{\rm iso}$、$\delta_{\rm cor}$、α 和 β 被固定为 0。在 AD-$n_{\rm s}$ 模型中，$n_{\rm s}$ 也会变化。所有等曲率模型对整组参数采用相同的先验分布。

然而，对于 $n_{\rm s}$变化的绝热模型，解析方法的表现较差：解析方法和拉普拉斯方法均得到 $\ln E_{\rm AD-n_s} = -853.4$，而数值方法给出的值更小，为 -854.1，差异接近 1。

接下来讨论等曲率情形。对于冷暗物质等曲率模型，我们得到了极佳的结果：根据公式 (54)，其值为 $\ln E^{\rm ana}_{\rm cdi} = -855.08$，而热力学积分给出的值为 $\ln E^{\rm num}_{\rm cdi} = -855.1 \pm 0.1$。这一点令人惊讶，因为至少三个变量的非高斯性相对较大：$n_{\rm iso}$、β 和 $\delta_{\rm cor}$；这些变量的先验分布并未以均值为中心。不过，中微子等曲率情形的结果一致性要差得多，差异为 0.6。这表明，冷暗物质等曲率模型的比较结果如此接近，在某种程度上可能只是统计上的偶然，实际采用的方法准确性较低。

不同模型的摘要见表 3。

[3]最近，Trotta [9] 采用了不同的方法分析一类受限的等曲率模型，该模型只包含一个额外参数，并发现它极不受支持。得出不同结论的主要原因，是他为等曲率振幅选择了截然不同的先验；在该先验下，几乎所有模型都由等曲率模主导，因而与数据的吻合度很差。

模型	$\ln \mathcal{L}^{\max}$	$\ln E^{\mathrm{num}}$	$\ln E^{\mathrm{ana}}$	$\ln E^{\mathrm{Lap}}$
toy5	0	-8.65 ± 0.03	-8.66	-8.67
toy5c	0	-7.28 ± 0.06	-7.32	-6.89
toy6	0	-10.75 ± 0.03	-10.74	-10.74
toy6c	0	-9.73 ± 0.06	-9.71	-9.63
AD	-840.78	-854.1 ± 0.1	-854.1	-854.1
AD-n_{s}	-838.50	-854.1 ± 0.1	-853.4	-853.4
CDI	-838.05	-855.1 ± 0.2	-855.1	-854.5
NID	-836.60	-855.1 ± 0.2	-854.5	-854.5
NIV	-842.53	-855.1 ± 0.3	-854.9	-854.9

表 3: 不同模型（包括玩具模型和真实模型）的最大似然及证据 s.

5.4 Savage–Dickey 方法

另一种计算证据的数值方法是 Savage–Dickey 方法，该方法最早在参考文献 [11]中描述，最近又在参考文献 [9]中使用。只要两个模型是嵌套的，即其中一个模型包含在另一个模型的参数空间中，这种技术就能通过对用于参数估计的马尔可夫链进行简单快速的分析，计算两个模型的证据比。例如，AD 模型嵌套在 AD-n_{s} 模型中，而 AD 和 AD-n_{s} 模型都嵌套在 CDI、NID 和 NIV 模型中。在马尔可夫链的语境下，Savage–Dickey 方法本质上是衡量采样器在嵌套模型中停留的时间，并根据两个模型各自的体积进行加权。当外层模型具有额外参数时，该方法依赖于这样的近似：将嵌套模型视为在额外参数方向上具有可忽略窄先验分布的模型。然而，我们注意到，当存在许多额外参数时，由于与网格法参数估计在参数众多的模型中失效的原因类似，该方法必然会失效。鉴于先验体积比很大，马尔可夫链蒙特卡罗参数估计的动态范围不足以探测这两个模型。

AD 和 AD-n_{s} 模型相差一个参数。使用与解析方法相同的 AD+ns 样本（即用于提取协方差矩阵的样本），我们得到 ln(E_{AD}/E_{AD+n_s}) =0。03。精确热力学积分得到的结果 ln($E_{\mathrm{AD}}/E_{\mathrm{AD}-n_s}$) $= 0 \pm 0.1$ 与此高度吻合。AD-n_{s} 和 CDI（或 NID、NIV）模型相差四个参数。对于大多数简单的参数化选择（尤其包括等曲率和交叉相关倾斜参数），AD-n_{s} 并非一个点，而是等曲率模型参数空间内的一个超曲面（即 $\alpha = 0$ 及其他三个参数是虚设的、不受约束的参数，不会影响证据）。在这些情况下，随着额外参数的先验分布收紧至嵌套模型附近，Savage–Dickey 方法给出的证据比并不收敛，尽管其与热力学积分确定的数值相差不到一个 ln E单位。

6 讨论与结论

我们建立了一套分析形式体系，用于在似然分布任意、具有一系列非高斯修正项，且先验分布为任意居中或非居中的顶帽先验的情况下计算贝叶斯证据。在当前宇宙学背景下，这项分析可为模型比较问题提供极大帮助，因为观测数据仍无法排除基于 ΛCDM 暴胀范式的大多数标准模型扩展。

作为将针对似然分布近似为高斯分布的模型所得贝叶斯证据精确公式和近似公式付诸应用的一个例子，我们将解析预测的数值与采用热力学积分方法、耗时较长的算法计算出的数值进行了比较。解析得到的数值与数值方法得到的结果出奇地吻合。虽然可以估算解析公式中高阶修正的大小，但要

估算数值方法的系统误差却非常困难。因此，我们可以用这种解析方法检验热力学积分方法中的系统误差。目前为止，两种方法所得数值一致，因此数值方法似乎是估算证据的有效工具。不过，要正确完成计算需要投入大量精力和机器时间。因此，我们建议采用解析估算，其修正项是可控的，因为可以计算下一阶修正并证明其影响很小。

补记：我的工作完成多年后，出版了一本书 [12]，全面讨论了宇宙学中的贝叶斯方法。

参考文献

[1] H. Jeffreys，《概率论》，第3版，牛津大学出版社（1961）。

[2] D. J. C. MacKay，《信息论、推断与学习算法，剑桥大学出版社（2003）。

[3] A. Jaffe，《天体物理学杂志》**471**, 24 (1996); P. S. Drell, T. J. Loredo, and I. Wasserman I，《天体物理学杂志》**530**, 593 (2000); M.V. John and J. V. Narlikar，《物理评论D》 **65**, 043506 (2002); M. P. Hobson, S. L. Bridle, and O. Lahav，《皇家天文学会月刊》 **335**, 377 (2002); A. Slosar 等，《皇家天文学会月刊》**341**, L29 (2003); T. D. Saini, J. Weller, and S. L. Bridle，《皇家天文学会月刊》 **348**, 603 (2004); A. Niarchou, A. H. Jaffe, and L. Pogosian，《物理评论D》 **69**, 063515 (2004); P. Marshall, N. Rajguru, and A. Slosar，《物理评论D》**73**, 067302 (2006)。

[4] A. R. Liddle，《皇家天文学会月刊》**351**，L49（2004）。

[5] M. Beltran，J. Garc´ıa-Bellido，J. Lesgourgues，A. R. Liddle 和 A. Slosar，《物理评论D》**71**，063532（2005）。

[6] J. J. K. ´O'Ruanaidh 和 W. J. Fitzgerald，《数值贝叶斯方法应用于信号处理》，施普林格出版社，纽约（1996）。

[7] M. P. Hobson 和 C. McLachlan，《皇家天文学会月刊》**338**，765（2003）。[8] R. E. Kass 和 A. E. Raftery，《美国统计学会杂志》**90**，773（1995）。[9] R. Trotta，《皇家天文学会月刊》**378**，72（2007）。[10] A. Lewis 和 S.Bridle，《物理评论D》**66**，103511（2002）。[11] J. M. Dickey，Ann.Math. Stat **42**，204（1971）。

[12] M. P. Hobson、A.H. Jaffe、A. R. Liddle、P. Mukherjee 和 D. Parkinson，宇宙学中的贝叶斯方法，剑桥大学出版社（2010）。