贝叶斯范式下 Azzalini 偏态正态分布与几何偏态正态分布的比较

Narayan Srinivasan*

2023年1月5日

arXiv:2301.01750v1 [stat.ME] 4 Jan 2023

摘要

正态分布的偏态推广因其在多个领域的广泛应用，一直是统计学界极为关注的研究主题。Azzalini偏态正态分布凭借直观易懂的特点，成为最受欢迎的偏态正态分布之一。然而，由于其分布本身的性质，它存在严重的推断问题。有趣的是，研究表明，贝叶斯方法能够缓解这些问题。近期，另一种偏态正态分布——几何偏态正态分布——被提出作为对偏态数据建模的替代方案；它在结构上不同于 Azzalini偏态正态分布。尽管人们对偏态正态分布颇感兴趣，但比较不同偏态分布性能的文章数量有限，尤其是在贝叶斯背景下。为弥补这一空缺，本文尝试在贝叶斯范式下，对这两种偏态正态分布在不同偏度数据建模中的表现进行比较。采用混合吉布斯-梅特ropolis算法获得几何偏态正态分布参数的后验估计，并计算其对数据的后验预测拟合。同样，对于 Azzalini偏态正态分布，本文采用吉布斯抽样算法推导其后验预测拟合。为比较模型性能，使用后验预测分布与原始数据之间的柯尔莫哥洛夫-斯米尔诺夫距离，衡量两个模型的拟合优度。本文使用这两个模型分析了一系列偏度范围广泛的真实数据集和模拟数据集，并讨论了两种模型各自的优缺点。最后，本文提出一种速度快得多的变分贝叶斯方法，用于近似几何偏态正态模型的后验分布，并讨论其性能。

关键词：偏态正态分布，贝叶斯推断，马尔可夫链蒙特卡罗，变分贝叶斯

1 引言

数据中显著的偏度会导致使用标准统计程序进行推断时出现错误。例如，均值的标准置信区间可能不准确，即实际覆盖率将不同于名义覆盖率。因此，对偏态数据进行建模的需求促成了

*印度科学教育与研究学院加尔各答分校项目学生，电子邮件：narayanseshadri98@gmail.com

许多偏态分布的发展。在应用的广泛推动下，过去几年里，人们对偏态分布的兴趣大幅增长。事实上，大多数真实数据集都表现出偏离对称性的特征，而非完全对称（[25]）。

同样，偏态分布也被用作构造对称性检验及相关推断程序的替代选择([21])。通过向对称分布添加一个控制偏度的参数，可以得到若干偏态分布。 [6] 提出的偏态正态分布是偏态分布中的一个常见例子。其他一些值得关注的例子包括偏态 t 分布([17])、偏态逻辑斯蒂分布([23])和偏态柯西分布([5])。最后，也有人提出了一种通用方法，可从对称分布生成相应的偏态分布。([26])

Azzalini 偏态正态分布（简称 ASN）的概率密度函数如下：

$$f(x) = \frac{2}{\sigma}\phi(\frac{x-\mu}{\sigma})\Phi(\lambda\frac{x-\mu}{\sigma}) \qquad x \in \boldsymbol{R}$$

其中， ϕ 和 Φ 分别表示标准正态概率密度函数和标准正态累积分布函数（CDF）。需要注意的是，正态分布是这类分布中的一个特例($\lambda = 0$)。它是单峰密度函数，偏度既可以为正($\lambda > 0$)，也可以为负($\lambda < 0$)。偏度参数 λ还控制着分布的形状，使其概率密度函数具有灵活性。此外，ASN模型可自然地推广到多元情形。([1])不出所料，该模型已用于分析来自多个领域的非对称数据集。([2], [16], [14])

尽管 ASN 分布具有许多有趣的性质，但正如 [15] 的综述论文所列举的那样，它在推断方面存在若干问题。首先，矩估计量可能不取实数值。其次，若 $\mu = 0$ 和 $\sigma = 0$，且所有观测值均为正（负），则似然函数关于 λ 单调递增（递减）。因此， λ （减号）的最大似然估计（MLE）将为无穷大。第三，即使样本中同时存在负值和正值，MLE 也不稳定。第四，参数 λ 存在可识别性问题：不同的 λ 值可能对应非常相似的偏态正态分布密度。其次，无法根据 MLE 的抽样分布解析地获得标准误和置信区间。最后，若假设三个参数均未知，则无论观测样本如何， λ 的似然函数在 $\lambda = 0$ 处都有一个驻点。最后，当 λ 趋近于零时，费舍尔信息矩阵会变为奇异矩阵。

贝叶斯方法能够令人满意地解决偏度参数的点估计和假设检验问题。 [22] 的研究为贝叶斯推断引入了 Jeffreys 参考先验，并进一步表明，在位置参数和尺度参数给定且已知的情况下，后验分布的支持集无界且后验分布适当。 [8] 通过两种不同的非信息先验凸显了贝叶斯方法的优势，第一种是 Jeffreys 先验，第二种是均匀先验。在同一篇论文中，他们还实现了一种用于贝叶斯分析的吉布斯抽样算法，并将其应用于多个数据集，包括 frontier 数据集。该数据集可在 Azzalini 的网页上获取，且其经典最大似然估计为无穷大。

ASN 分布的先验 elicitation 问题尤其具有挑战性，因为 λ 不仅控制对称性，还会影响离散程度、众数和尾部行为。 [15] 全面介绍了几种先验 elicitation 方案。针对本文的目标，采用了 [11] 提出的使用信息先验的方法。

近期，另一种基于正态分布密度与几何分布密度卷积的偏态正态分布

被称为几何偏态正态（简称 GSN）分布，由 [19]提出。单变量 GSN 分布也是一个三参数分布，是正态分布的推广。在偏度方面，GSN 分布也具有很大的灵活性；更重要的是，它不存在 ASN 的推断问题。 [19] 已实现并证明，简单的期望最大化算法能够令人满意地收敛。GSN 的多元推广也已由 [20]提出。

令人惊讶的是，尽管人们对偏态分布的兴趣日益浓厚，但只有少数论文探讨了不同偏态分布（[7], [12] 和 [18]）性能的比较。尤其是，ASN 分布的性能尚未与近期提出的 GSN 分布在多种数据集上进行比较。本文旨在贝叶斯范式下，对 ASN 和 GSN 在建模具有广泛偏度的数据集时进行全面比较，以期助益于偏态数据集的统计建模。

首先，为了实现有益且公正的比较，本文提出一种贝叶斯推断方法，并为几何偏态正态（GSN）的参数设定信息先验。接下来，本文介绍一种基于 Metropolis-Hastings 算法的马尔可夫链蒙特卡罗方法，用于计算后验分布。此外，本文还提出一种变分贝叶斯技术，用于对后验分布进行近似贝叶斯分析。最后，本文对 ASN 分布采用带有信息先验的贝叶斯推断，并将其后验预测拟合与基于马尔可夫链蒙特卡罗的 GSN 分布贝叶斯方案进行比较，所用数据集具有不同的偏度。本文结构如下：第 2 节介绍 GSN 分布并讨论其性质。第 3 节介绍贝叶斯方案及先验假设。第 4 节使用模拟数据集和真实数据集比较 ASN 和 GSN 分布。第 5 节介绍使用变分推断对 GSN 分布进行贝叶斯推断的另一种近似方案。最后，第 6 节总结全文，并根据所得结果展望未来的研究方向。

2 几何偏态正态分布（GSN)

本节介绍 [19] 提出的几何偏态正态（GSN）分布。它是正态分布的偏态版本，因为正态分布是该分布族中的一个特例。GSN 具有若干值得注意的性质。例如，GSN 分布的密度可以是单峰或多峰的；其矩生成函数存在显式形式；GSN 分布也可以是具有较厚尾部的对称分布。GSN 分布由三个参数 μ、 σ^2 和 p 描述。具体而言，下面给出 GSN 的定义。

设 N 和 X_i 为相互独立的随机变量， $i = 1$，2，.....，其分布分别为 $N \sim$ GE(p) 和 $X_i \sim N(\mu, \sigma^2)$，其中 GE 表示几何分布。定义

$$X = \sum_{i=1}^{N} X_i.$$

X 称为服从几何偏态正态分布 称为服从几何偏态正态分布，记作 $X \sim GSN(\mu, \sigma, p)$.

当 p=1 时，该分布退化为正态密度。联合概率密度函数

对于 X 和 N，其中 $p \neq 1$，可表示为

$$f_{X,N}(x,n) = \phi((x-n\mu)/\sqrt{n}\sigma)p(1-p)^{n-1}/\sqrt{n}\sigma,$$

其中， $\phi(\cdot)$ 是标准正态分布的密度函数。根据 X 和 N 的联合密度，可以得到 X 的边缘密度：

$$f_X(x) = \sum_{k=1}^{\infty} \phi((x-k\mu)/\sqrt{k}\sigma)p(1-p)^{k-1}/\sqrt{k}\sigma. \tag{1}$$

我们将在后续研究 N 给定 $X = x$ 时的条件密度。具体而言， N 和 N^{-1} 的条件期望推导如下（见 [19]）：

$$E(N \mid X = x) = \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}e^{-\frac{1}{\sigma^2 n}(x-n\mu)^2}.\sqrt{n}}{\sum_{k=1}^{\infty}(1-p)^{k-1}e^{-\frac{1}{\sigma^2 k}(x-k\mu)^2}/\sqrt{k}} \tag{2}$$

and

$$E(N^{-1} \mid X = x) = \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}e^{-\frac{1}{\sigma^2 n}(x-n\mu)^2}/n^{3/2}}{\sum_{k=1}^{\infty}(1-p)^{k-1}e^{-\frac{1}{\sigma^2 k}(x-k\mu)^2}/\sqrt{k}}. \tag{3}$$

3 几何偏态正态的贝叶斯推断

进行贝叶斯分析时，我们首先需要为感兴趣的参数指定先验分布。下一小节将详细介绍这些先验分布。

3.1 先验假设

我们考虑以下关于 μ、 σ 和 p 的信息共轭先验：

$$\begin{aligned}(\mu,\sigma^2) &\sim N-\Gamma(v_0,n_0,\alpha,\beta), \text{ and}\\ p &\sim \text{Beta}(a,b),\end{aligned}$$

其中， $N-\Gamma$ 表示正态-逆伽马分布。超参数的选择将在第 4.1 节和第 4.2 节中讨论，因为这些选择取决于数据。

3.2 后验分析与贝叶斯框架

设 X_1、 X_2...、 X_m 为来自密度如式 (1) 所示的 GSN 的独立同分布随机变量。则似然为

$$L(\mu,\sigma,p) = \prod_{i=1}^{m}\sum_{k=1}^{\infty}\phi(x_i - k\mu/\sqrt{k}\sigma)p(1-p)^{k-1}/\sqrt{k}\sigma. \tag{4}$$

根据对 p 和 (μ、σ^2) 所作的先验假设（见第 3.1 小节），结合式 (4) 给出的似然，可得完整的后验分布 π(μ、σ、p)，如下：

$$\pi(\mu,\sigma,p|\mathbf{x}) \propto \prod_{i=1}^{m}\sum_{k=1}^{\infty}\phi(x_i - k\mu/\sqrt{k}\sigma)p^a(1-p)^{k+b-2}\frac{1}{\sqrt{k}}\left(\frac{1}{\sigma^2}\right)^{\alpha+2}e^{-\frac{2\beta+n_0(\mu-v_0)^2}{2\sigma^2}},$$

其中 $\mathbf{x} = \{x_1, x_2, \ldots, x_m\}$ 是$\mathbf{X} = \{X_1, X_2, \ldots, X_m\}$的实现值。然而，由于后验密度涉及求和，因此不可处理；相应地，参数的完整条件密度也不可处理。为此，我们考虑如下 GSN 随机变量的层次表示，并为每个 X_i 对应一个潜变量 N_i，对所有 $i = 1, 2, \ldots, m,$ 均如此，从而能够针对所选的共轭先验直接计算条件后验分布。层次表示如下：

$$N_i \mid p \sim GE(p). \tag{5}$$

$$X_i \mid N_i\,, \mu\,, \sigma\,, p \sim N(N_i\mu, N_i\sigma^2) \text{ for } i = 1, 2, \ldots, m. \tag{6}$$

利用这一表示，可以得到参数的完整条件密度，从而便于通过吉布斯抽样从联合后验分布中进行模拟。

令 $\mathbf{N}$ 表示随机向量 $\{N_1, N_2, \ldots, N_m\}$，并将 N_i 的实现值记为 n_i，其中 $i = 1, 2, \ldots, m$。然后，利用我们选择的先验分布，可得到 p 以及 (μ, σ^2) 的完整条件密度：

$$p \mid \mu, \sigma, X, \mathbf{N} \sim \text{Beta}(m + a, \sum_{i=1}^{m} n_i - m + b) \tag{7}$$

$$(\mu, \sigma^2) \mid p, X, \mathbf{N} \sim N - \Gamma(\mu^*, n^*, \alpha^*, \beta^*), \tag{8}$$

其中

$$\mu^* = \frac{n_0 v_0 + \sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i + n_0}, n^* = \sum_{i=1}^{m} n_i + n_0, \alpha^* = \alpha + \frac{n}{2}, \text{ and}$$

$$\beta^* = 2\beta + n_0(u^* - v_0)^2 + \sum_{i=1}^{m}\frac{1}{n_i}(x_i - n_i\mu^*)^2.$$

完整条件密度的推导见附录 8.1。

由于 $\mathbf{N}$ 未被观测到，我们需要增加一个步骤，以构建完整的抽样算法，从而获得模型中所有未知量的联合后验分布。对于每个 $i = 1$，2、.....、m，对应于 X_i，存在一个潜变量 N_i，其后验分布仅取决于 X_i以及模型的其他参数。对于 N_i（$i = 1$，2、.....、m），其后验分布可在忽略归一化常数的情况下推导如下：

$$p(N_i \mid \mu, \sigma, p, X) \propto \frac{e^{-\frac{(X_i - N_i\mu)^2}{2N_i\sigma^2}}}{\sqrt{N_i}}(1-p)^{N_i - 1}. \tag{9}$$

3.3 Gibbs-梅特罗波利斯-黑斯廷斯混合算法

本节将介绍两种混合吉布斯-梅特罗波利斯-黑斯廷斯算法，用于从上一小节所述的后验分布中进行模拟。首先，我们使用带有几何分布提议的梅特罗波利斯-黑斯廷斯算法，更新每个 N_i（$i = 1$，2、.....、m）。[24]采用类似方法实现了基于多元 GSN 的贝叶斯聚类。其次，我们还提出了一种随机游走算法，用于更新每个 N_i。

3.3.1 带几何分布提议的 Gibbs-梅特罗波利斯-黑斯廷斯算法

在迭代 t 时

1. 抽取样本 $N^{(t)}$（取决于 $\sigma^{2^{(t-1)}}$ $\mu^{(t-1)}$ 和 $p^{(t-1)}$）

   - 从均匀分布(0,1)中抽取样本 p_{prop}
   - 对于每个 $i = 1,\ 2, \ldots, m-$ 从几何分布(p_{prop})中抽取提议样本 N_i^*- 使用方程 (9) 计算 N_i 的接受概率，得到

$$r = \frac{p(N_i^* \mid X_i, \mu^{(t-1)}, \sigma^{2^{(t-1)}}, p^{(t-1)}) q(N_i^{(t-1)})}{p(N_i^{(t-1)} \mid X_i, \mu^{(t-1)}, \sigma^{2^{(t-1)}}, p^{(t-1)}) q(N_i^*)},$$

   其中 q 为几何分布(p_{prop})。$-$以 min(1, r) 的概率接受 N_i^* 作为新的样本 N_i^t。

2. 根据方程 (7) 和 (8) 给出的相应条件后验分布，抽取 $\sigma^{2^{(t)}}$（取决于 $N_{(t)i}$）、$\mu^{(t)}$（取决于 $N_{(t)i}$ 和 $\sigma^{2^{(t)}}$）以及 $p^{(t)}$（取决于 $N_i^{(t)}$）。

3.3.2 Gibbs-随机游走算法

在第 t 次迭代时

1. 抽取 $N^{(t)}$（取决于 $\sigma^{2^{(t-1)}}$、$\mu^{(t-1)}$ 和 $p^{(t-1)}$）

   - 对于每个 i=1,2,...,m$-$抽取 ϵ，使得 P($\epsilon = 1$) = 1/2 且 P(ϵ = -1) = 1/2$-$模拟 N_i^*，如下：$N_i^* = N_i + \epsilon$（约束条件为 $N_i^* \geq 1$）。也就是说，这是一个在 1 处具有反射边界的随机游走。$-$ 根据方程 (9) 计算 N_i^* 的接受概率，如下

$$r = \frac{p(N_i^* \mid X_i, \mu^{(t-1)}, \sigma^{2^{(t-1)}}, p^{(t-1)})}{p(N_i^{(t-1)} \mid X_i, \mu^{(t-1)}, \sigma^{2^{(t-1)}}, p^{(t-1)})}$$

   $-$以概率 min(1，r) 接受 N_i^* 作为新的样本 N_i^t

2. 从方程（7）和（8）给出的各自条件后验分布中抽取$\sigma^{2^{(t)}}$（取决于 $N^{(t)}$）、$\mu^{(t)}$（取决于 $N^{(t)}$ 和$\sigma^{2^{(t)}}$）以及 $p^{(t)}$（取决于 $N^{(t)}$）的样本。

4 比较研究

4.1 模拟研究

为了比较 ASN 和 GSN 在拟合偏态数据方面的表现，我们从极小偏度到大偏度模拟了不同范围的偏度。为了公平比较 ASN 和 GSN，其中大多数数据集由 ASN 模拟生成。极小偏度、小偏度和大偏度的数据由 ASN 生成。另有一个中等偏度的数据集由 GSN 模拟生成，用于说明。生成 ASN 样本时，我们使用了 R 包 'sn'。GSN 样本则使用方程（5）和（6）生成，具体如下：

- 从几何分布 Geometric(p) 中生成 N
- 从正态分布 Normal($N\mu$,$N\sigma$) 中生成 X。因此，X 服从 GSN(μ, σ,p)

我们获得 GSN 和 ASN 的后验预测拟合，并使用拟合结果与原始数据之间的 Kolmogorov-Smirnov 距离（KSD）比较拟合效果。我们还报告参数的最大后验（MAP）估计值。我们使用 R 函数 'map estimate' 获得 MAP。

我们还进行了一项最终模拟研究，使用来自不同于 ASN 和 GSN 的偏斜分布的数据。具体而言，我们从对数正态分布生成模拟数据，并在贝叶斯范式下使用 ASN 和 GSN 分布拟合这些数据。在这种情况下，我们也报告 KSD，以比较 ASN 和 GSN 的拟合优度。

在所有这些模拟研究中，均从相应分布中抽取 100 个样本。

为了对 GSN 的参数进行推断，我们采用混合 Metropolis-Gibbs 算法，并通过随机游走步骤更新潜变量。我们注意到，使用几何提议分布（[24]）而非简单随机游走会产生相似的结果。

我们采用 [11] 的方法获得 ASN 的后验预测拟合。附录中简要介绍了吉布斯抽样算法

两个模型的先验 elicitation 均旨在降低先验分布的影响。通过适当设置超参数，我们获得了近乎非信息先验。GSN 模型中参数 p 的先验设为 Beta(1,1)，对应非信息均匀分布。通过增大方差（设置超参数 $n_0 = 0.001$），为均值 elicitation 了弥散正态先验。同样，对于 ASN，也为均值和偏度参数 λ elicitation 了弥散正态先验。

4.1.1 极小偏度

本小节中，数据由 ASN ($\mu = 0, \sigma = 1, \lambda = -0.5$) 生成，因此偏度很小。事实上，生成的样本的皮尔逊偏度系数为 -0.038。表1给出了两个模型的后验预测拟合结果。由于数据由 ASN 生成，因此我们可以将参数的 MAP 估计与真实值进行比较。表1显示，ASN 分布的参数 MAP 估计值与真实值非常接近。还可以看到，参数的可信区间长度很小（为简洁起见，未报告相关图形）。图1展示了 ASN 和 GSN 的后验预测密度与观测数据密度的对比

。图中显示，GSN 的拟合效果优于 ASN，这也得到了表1最后一列所列 KSD统计量的证实。

模型	均值(μ)	标准差（σ）	λ(ASN) 或 p(GSN)	KSD
ASN	0.01	0.89	-0.5	0.119
GSN	-0.41	0.97	0.93	0.049

表1：极小偏度数据：报告了 ASN 和 GSN 的参数 MAP值，以及原始数据与 ASN 和 GSN 后验预测拟合之间的 KSD。GSN 的 KSD 略优。

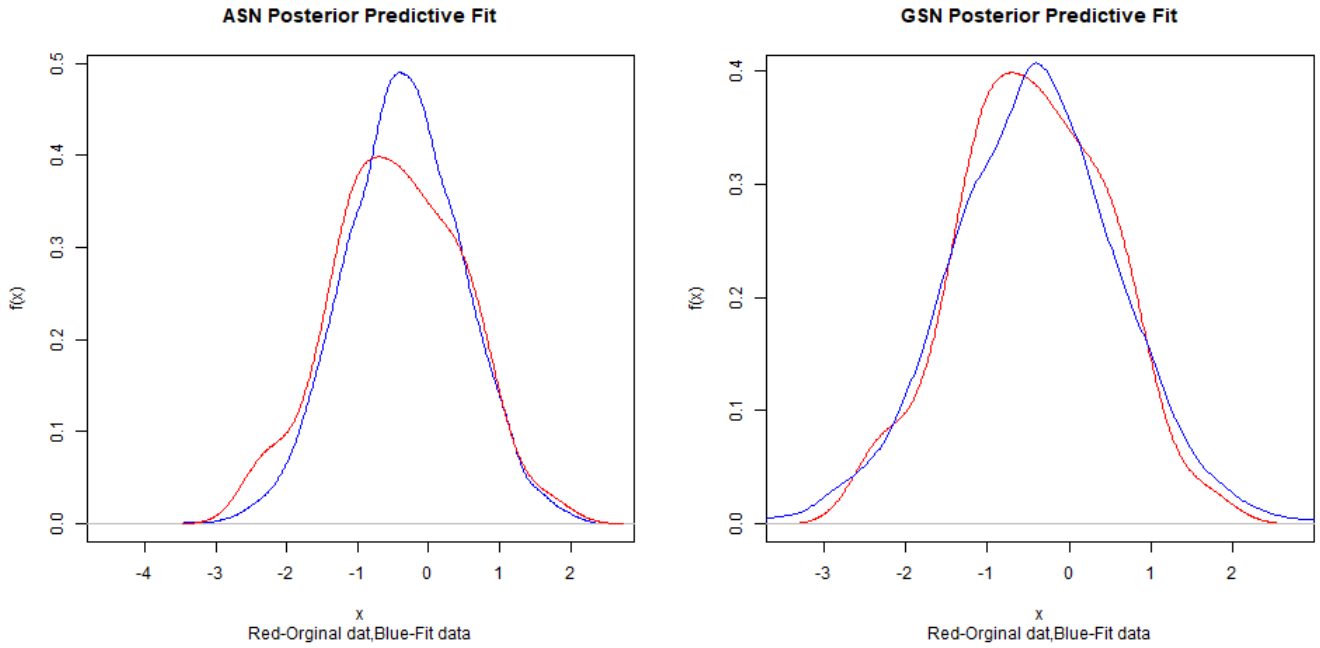

图1：极小偏度数据：红色曲线为 ASN((μ=0,σ=1, λ=-0.5) 的 100 个样本的核密度，蓝色曲线则分别为左侧 ASN 模型和右侧 GSN 模型的后验预测拟合密度。两种情况下均采用了带信息先验的贝叶斯推断方案。

4.1.2 小偏度

如第 4.1 节开头所述，模拟数据由适用于小偏度的 ASN 分布生成。ASN 的参数选为 $\mu = 0$, $\sigma = 1, \lambda = -1$。计算得出的皮尔逊偏度系数为 -0.370，符合预期，表明偏度为负。请注意，偏度的幅度大于此前的模拟数据。

在此，我们也分别得到了 ASN 和 GSN 中相关参数的 MAP。由于模拟数据来自 ASN，因此可以将 ASN 参数的 MAP 与之进行比较。表2给出了这些参数的 MAP。此外，从 ASN 的 μ 和 σ 的后验密度可以看出，可信区间的长度较小。尽管可以看出，ASN 参数的 MAP 对真实参数值的估计相当准确，且

相应可信区间的长度较小，但 GSN 的 KSD 略小于 ASN。

模型	均值(μ)	标准差 (σ)	λ(ASN) 或 p(GSN)	KSD
ASN	0	0.98	-1.22	0.084
GSN	0.58	0.89	0.86	0.059

表2：小偏度数据：报告了 ASN 和 GSN 参数的 MAP值。此外，还给出了原始数据与 ASN 和 GSN 的后验预测拟合之间的 KSD。

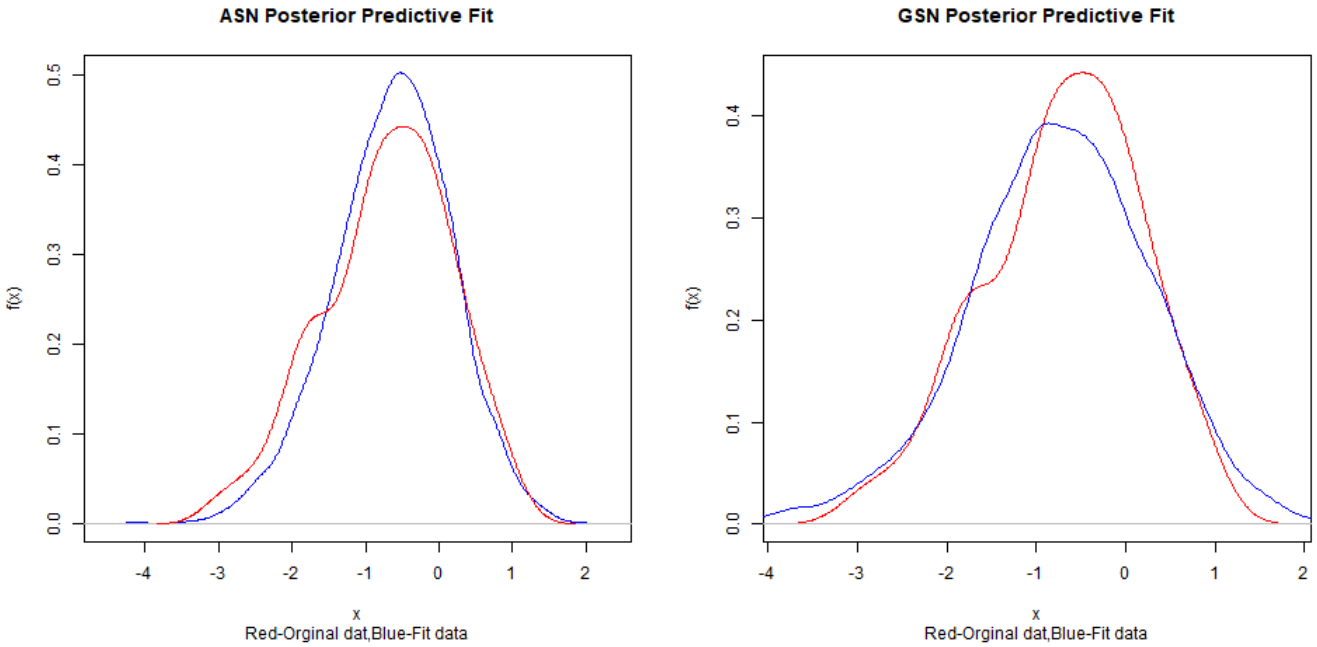

图2：小偏度数据：红色曲线为 ASN 的 100 个样本的核密度（(μ= 0,σ=1,λ=-1)，蓝色曲线为 ASN 和 GSN 模型的后验预测拟合密度，分别显示在左侧和右侧。两种情况下均采用了带信息先验的贝叶斯推断方案。

4.1.3 中等偏度

为了生成具有中等偏度的样本，我们借助 GSN。从 GSN 中抽取了 100 个数据点，模型参数为 $\mu = 1$、$\sigma = 1$ 和 $p = 0.8$。模拟数据的皮尔逊偏度系数为 0.748，其绝对值大于前两个模拟数据集。

在本次模拟研究中，可以将 GSN 的参数最大后验估计与其真实参数值进行比较。显然（见表3），GSN 模型的最大后验估计能够很好地估计模拟数据的真实参数。虽然此处未报告，但参数的可信区间长度也相当合理。重要的是，GSN 模型对应的 KSD 显著低于

ASN 模型对应的 KSD。这或许并不令人意外，因为数据是由 GSN 本身模拟生成的。不过，值得注意的是，即使数据集由 ASN 生成，GSN 对应的 KSD 也与 ASN 相当（实际上还低于 ASN）。图3展示了 ASN 和 GSN 对原始数据的拟合情况。显然，GSN 模型对真实数据的近似效果更好，尤其是在右尾附近。在两个尾部，ASN 都低估了真实密度。

模型	均值(μ)	标准差（σ）	λ(ASN) 或 p(GSN)	KSD
ASN	0	1.01	2.88	0.223
GSN	0.94	1.06	0.83	0.09

表3：中度偏斜数据：报告了 ASN 和 GSN 参数的 MAP 值。此外，还给出了 ASN 和 GSN 的原始数据与后验预测拟合之间的 KSD。

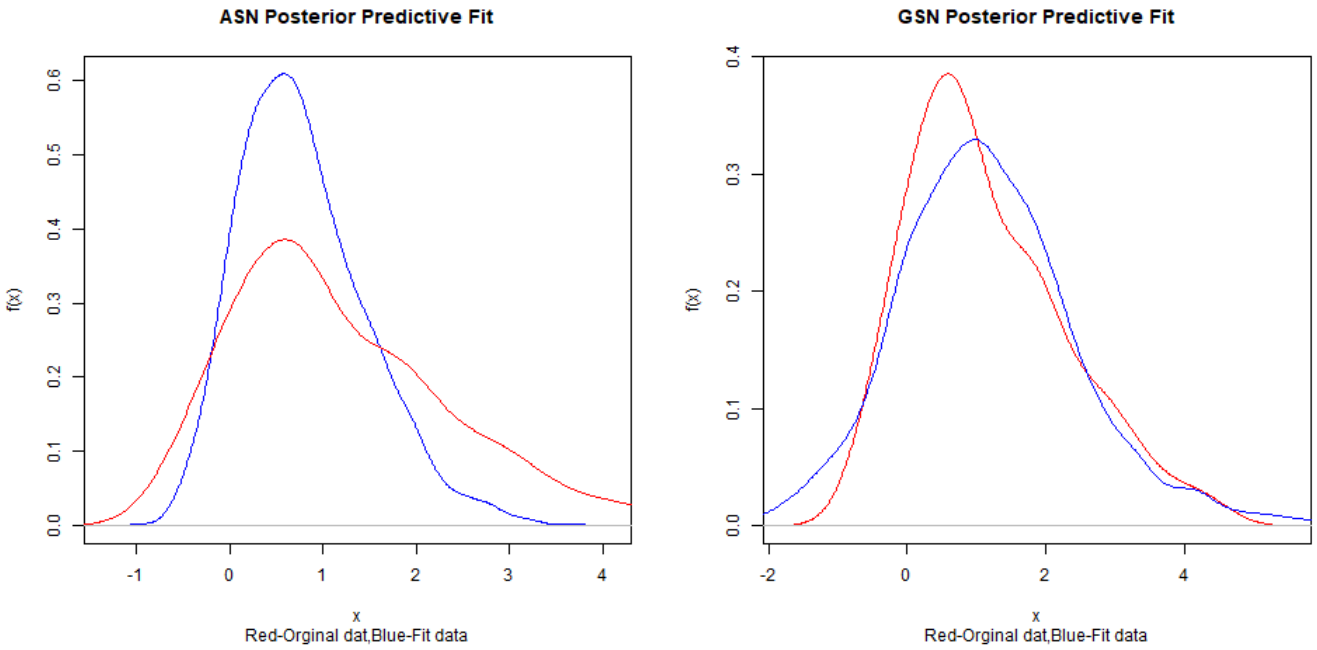

图3：中度偏斜数据：红色表示 GSN((μ=1,σ=1,λ=0.8) 的 100 个样本的核密度，蓝色表示 ASN 和 GSN 模型后验预测拟合的密度，分别对应左图和右图。两种情况下均采用了带信息先验的贝叶斯推断方案。

4.1.4 大偏度

模拟了具有极高 λ 值的 ASN 重度正偏态数据。我们从 ASN 中生成了 100 个随机观测值，其中 μ= 0、σ=1、 λ=100。模拟数据的皮尔逊偏度系数（0.956）表明，与此前模拟的数据集相比，其正偏度较大。ASN 和 GSN 的参数最大后验估计如表4所示。表4中还报告了核 Stein 差异值。

需要注意以下几点：当偏度高达 0.956 时，ASN 的 MAP 表现不佳。事实上，对于μ 和 σ，95%可信区间均未能涵盖真实值。此外，ASN 的核 Stein 差异值远大于 GSN。图4也证实了这一发现。与 GSN 相比，ASN 的拟合密度表现较差。显然，对于该数据集，GSN 的表现优于 ASN。

模型	均值(μ)	标准差 (σ)	λ(ASN) 或 p(GSN)	KSD
ASN	0.48	0.64	0.94	0.207
GSN	0.36	0.26	0.44	0.07

表4：大偏度数据：报告了 ASN 和 GSN 参数的 MAP 值。此外，还给出了原始数据与 ASN 和 GSN 后验预测拟合之间的 KSD。

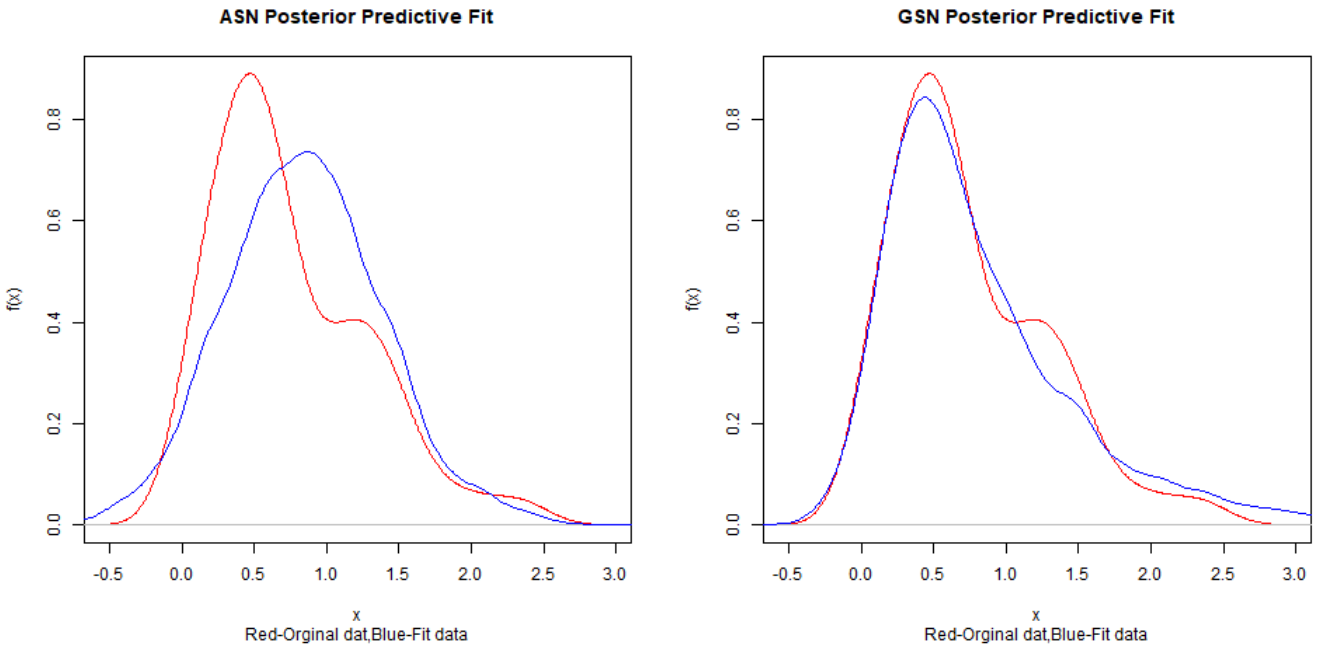

图4：大偏度数据：红色表示 ASN（(μ= 0,σ=1,λ=100) 的 100 个样本的核密度，蓝色表示左侧和右侧分别为 ASN 和 GSN 模型的后验预测拟合密度。两种情况下均采用了带信息先验的贝叶斯推断方案。

4.1.5 对数正态

作为最后一项模拟研究，我们从参数为 μ=0 和 $\sigma^2 = 0.6$ 的对数正态密度中模拟了 100 个随机观测值。我们考虑了生成样本中的负值，从而得到 -1.641 的负皮尔逊偏度系数。因此，在此处考虑的所有数据集中，这个数据集的偏度（绝对值）最大。对于这种情况，我们给出了 ASN 和 GSN 的真实密度与拟合密度图（图5）。与 ASN 的拟合密度相比，GSN 模型的拟合密度对数据的描述更好。GSN 与

后验预测和原始样本之间的 KSD 为 0.063，而 ASN 的这一数值为 0.278。即使对于这个高度负偏的数据集，GSN 的表现也显著优于 ASN。

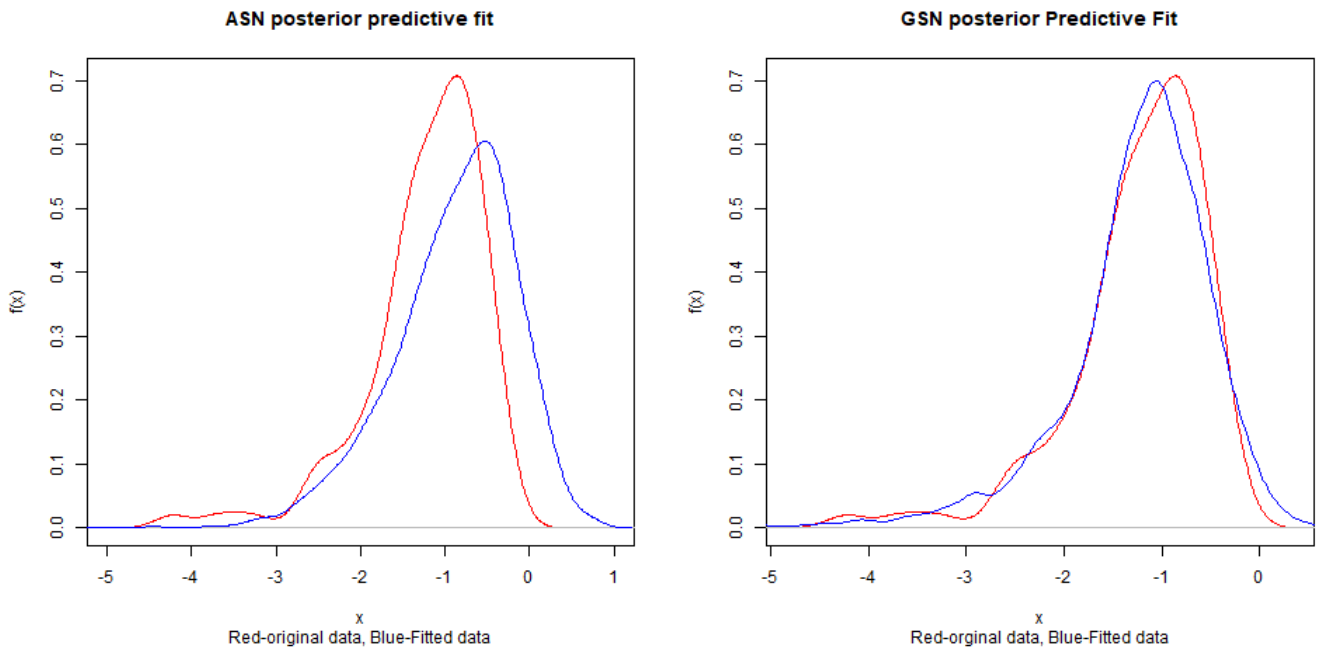

图5：对数正态数据：红色曲线为从对数正态密度中抽取的100个样本的核密度。该样本的偏度系数为较大的负值（-1.641）。蓝色曲线分别表示左侧 ASN 和右侧 GSN 模型的后验预测拟合密度。两种情况下均采用了基于信息先验的贝叶斯推断方案。

4.1.6 模拟研究的发现

从结果中我们注意到以下几点：

1. 首先，对于偏离正态性程度较小的数据集，即偏度系数较小的数据集，ASN 和 GSN 的拟合效果相当。2. GSN 能更好地捕捉尾部较重的数据。3. 在存在中等偏度的情况下，我们再次看到 GSN 的表现更好。4. 当数据集的偏度较大时，GSN 模型明显优于 ASN。这一点并不令人意外，因为 ASN 分布所允许的偏度系数最大仅为 0.995，且仅在偏度参数 $\lambda = \pm\infty$

4.2 真实数据分析

本节使用两个真实数据集比较 ASN 和 GSN 的表现。

4.2.1 Frontier 数据集

首先，我们分析 Frontier 数据集，以比较两种偏态正态分布。Adelchi Azzalini 的网页上提供了“Frontier 数据集” `http://azzalini.stat.unipd.it/index-en.html`。该数据集的偏度参数经典最大似然估计为 ∞，因此得到半正态密度；然而，很容易看出，取有限值的偏度参数能带来好得多的拟合。 [8]使用该数据集来强调贝叶斯方案的优势。对于这个数据集，我们再次报告 ASN 和 GSN 的后验预测拟合（图 6）与原始样本之间的 KSD，分别为 0.096 和 0.089。这再次证实了 [8], 的结论：贝叶斯方法能够用于经典最大似然估计失效的数据，而且 GSN 模型也可作为 ASN 的替代方案。

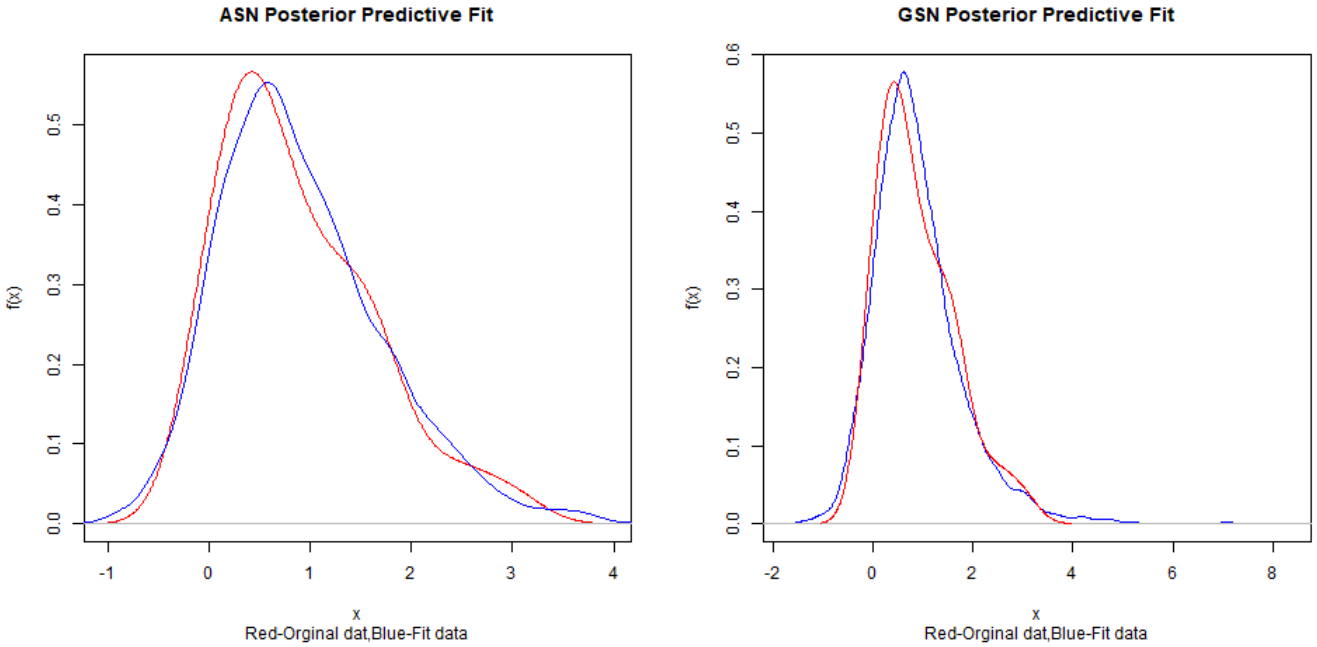

图 6：红色曲线表示来自 Azzalini 网页的 Frontier 数据集的核密度。蓝色曲线分别表示左侧 ASN 模型和右侧 GSN 模型的后验预测拟合密度。两种情况下均采用了信息先验下的贝叶斯推断方案。

4.2.2 豚鼠数据

最后，我们考察一个真实数据集，其中记录了注射不同剂量结核杆菌的豚鼠的生存时间。该数据集取自 [19], ，作者使用它来说明，在经典最大似然估计推断中，采用 GSN 比采用 ASN 更具优势。该数据集的皮尔逊偏度系数相当大，为 1.796。因此，从图 7 可以明显看出，GSN 模型的拟合效果优于 ASN 模型。此外，ASN 和 GSN 的后验预测拟合对应的核 Stein 差异值分别为 0.313 和 0.12。值得一提的是，我们想指出，在论文 [19], 中，作者计算了基于最大似然估计值拟合的分布与数据原始密度之间的核 Stein 差异值，结果为 0.13。因此，我们可以得出结论：我们的 GSN 贝叶斯方案略优于经典最大似然估计

推断。不过，对于这个数据集，无论在经典方案还是贝叶斯方案下，GSN 模型的表现都优于 ASN 模型。

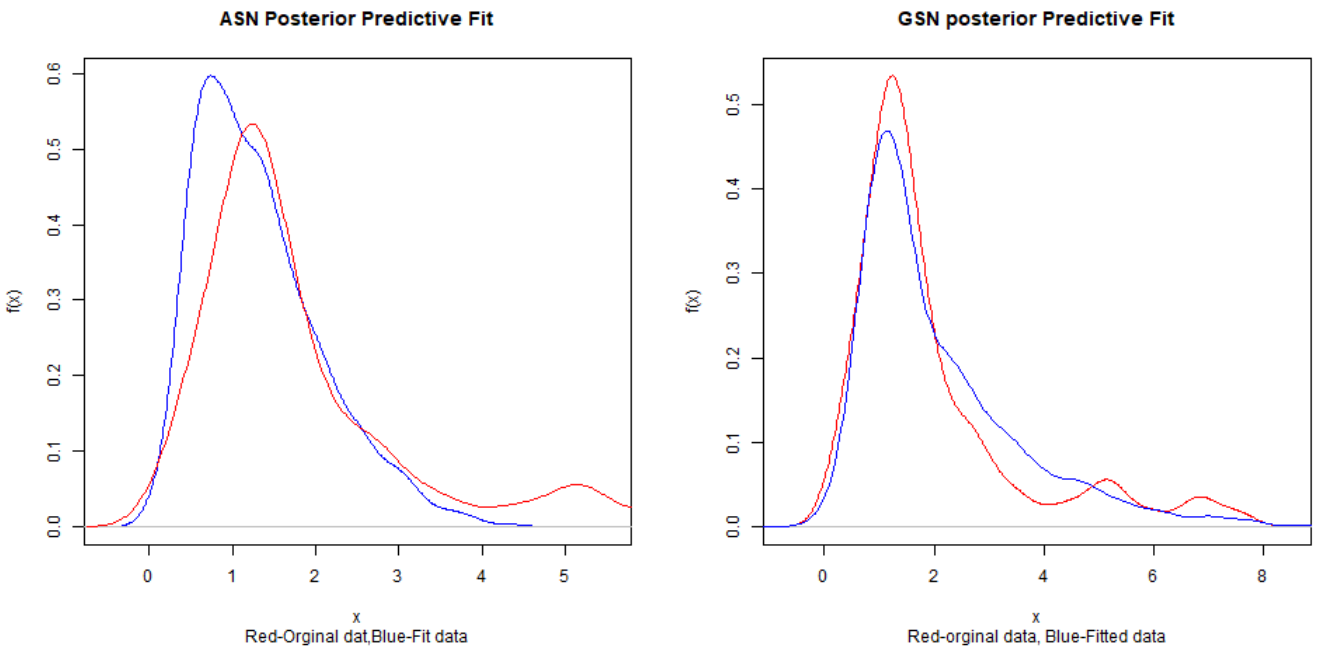

图7：红色曲线表示 [19]中的豚鼠数据集的核密度。该样本的偏度系数较大，为 1.796。蓝色曲线分别表示左侧 ASN 模型和右侧 GSN 模型的后验预测拟合密度。两种情况下均采用了带有信息先验的贝叶斯推断方案。

5 几何偏斜正态的变分推断

5.1 变分推断（VI）简介

变分推断（VI）已被广泛用于近似贝叶斯模型的后验密度。与马尔可夫链蒙特卡罗相比，变分推断通常速度快得多，也易于扩展到大型数据集。它已应用于大规模文档分析、计算神经科学和计算机视觉等问题。然而，与马尔可夫链蒙特卡罗不同，变分推断无法保证渐近精确收敛到后验分布。我们介绍[10]所述的变分推断基础知识。

令 $\boldsymbol{x} = \{x_1, x_2, ...x_n\}$ 为一组观测变量， $\boldsymbol{z} = \{z_1, z_2, ...z_m\}$ 为一组潜变量，其联合密度为 $p(\boldsymbol{z}, \boldsymbol{x})$。我们假设，所有感兴趣的未知量都表示为潜在随机变量。这包括可能决定数据的参数，以及与每个数据点相关的潜变量。推断问题的目标是计算给定观测值时潜变量的条件密度 $p(\boldsymbol{z} \mid \boldsymbol{x})$，可将其写为

$$p(\boldsymbol{z} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{z}, \boldsymbol{x})}{p(\boldsymbol{x})}.$$

分母是边缘密度或模型证据，前提是 s

$$p(\boldsymbol{x}) = \int p(\boldsymbol{z}, \boldsymbol{x}) dz.$$

对于许多模型，该证据积分无法以闭式形式求得，或在计算上难以实现。这些模型的推断很困难，因为我们需要证据来根据联合分布计算条件分布。

在变分推断中，我们指定一个关于潜变量的密度族 $\mathcal{Q}$ 。 $\mathcal{Q}$ 中的每个元素，例如 $q(\boldsymbol{z})$，都是用于近似精确条件分布的候选密度。我们的目标是找到最佳候选项，也就是与精确条件分布的 KL 散度最小的候选项。现在，推断问题可简化为求解以下优化问题：

$$q^*(\boldsymbol{z}) = \arg\min_{q(\boldsymbol{z}) \epsilon \mathcal{Q}} KL(q(\boldsymbol{z}) \mid\mid p(\boldsymbol{z} \mid \boldsymbol{x})),$$

其中，KL散度表示为：

$$KL(q(\boldsymbol{z}) \mid\mid p(\boldsymbol{z} \mid \boldsymbol{x})) = E[\log q(\boldsymbol{z})] - E[\log p(\boldsymbol{z} \mid \boldsymbol{x})].$$

请注意，所有期望 值均是相对于 $q(\boldsymbol{z})$ 计算的。展开条件分布后，我们得到 t

$$KL(q(\boldsymbol{z}) \mid\mid p(\boldsymbol{z} \mid \boldsymbol{x})) = E[\log q(\boldsymbol{z})] - E[\log p(\boldsymbol{z}, \boldsymbol{x})] + \log p(\boldsymbol{x}), \tag{10}$$

它取决于 log p(x)，因此难以计算。为了解决这个问题，我们不直接计算 KL 散度，而是最大化一个替代目标；该目标在加上一个常数后，等价于最小化 KL 散度。这个替代目标函数称为证据下界（ELBO），其形式如下：

$$\text{ELBO}(q) = E[\log p(\boldsymbol{z}, \boldsymbol{x})] - E[\log q(\boldsymbol{z})]. \tag{11}$$

之所以称为证据下界，是因为对于任意 $q(\boldsymbol{z})$，ELBO 都给出了证据的对数下界，即 $\log p(x) \geq$ ELBO(q)。为说明这一点，注意公式（10）和（11）给出了证据的以下表达式，

$$\log p(x) = KL(q(\boldsymbol{z}) \mid\mid p(\boldsymbol{z} \mid \boldsymbol{x})) + ELBO(q),$$

因此，对于任意 $q \in \mathcal{Q}$，log($p(\boldsymbol{x})$) ≥ELBO(q)。

5.2 均值场近似

如果潜变量可以分组成相互独立的组，那么证据下界的计算就可以简化。为此，我们采用平均场近似（参见 [9] 的章节.....）。具体而言，对于 GSN，平均场近似的思想会非常方便。设 $\boldsymbol{z}$ 的元素被划分为互不相交的组 z_i，其中 $i = 1, \ 2 \ldots m$。平均场近似将分布族 $\mathcal{Q}$ 限定为满足以下因式分解形式的分布 $q(z)$：

$$q(\boldsymbol{z}) = \prod_{i=1}^{m} q_i(z_i).$$

在所有可因式分解的分布 $q(\boldsymbol{z})$ 中，我们寻求使证据下界(q)最大的分布。因此，我们希望对所有分布 $q_i(z_i)$ 关于 ELBOL(q) 进行优化。我们用于求解这一优化问题的算法是坐标上升变分推断（CAVI）（详情参见 [9] ）。CAVI 在固定其他因子的同时，迭代优化平均场近似密度中的每个因子。它逐步提升证据下界，直至达到局部最优解。

根据 [9], 最优解， $q_j^*(z_j)$ 是相对于$q_i(z_i)$ 的 log $p(\boldsymbol{x},\ \boldsymbol{z})$ 期望，对所有 $i \neq j$ 均成立，即

$$\log q_j^*(z_j) = E_{i\neq j}[\log p(\boldsymbol{x}, \boldsymbol{z})] + \text{const.}$$

我们初始化所有因子 $q_i(z_i)$ 、$i = 1$、 2、.....、 m，然后依次遍历这些因子，并使用其他所有因子的当前估计值，根据上述方程依次将每个因子替换为修订后的估计值。GSN 的详细算法见第 5.4 节。

5.3 GSN 的变分推断

从 GSN 随机变量 X的概率密度函数中，我们观察到，GSN 分布可以看作正态随机变量的几何混合，即

$$f_{X|(\mu,\sigma,p)}(x) = \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{\sqrt{k}\sigma}\phi(x - k\mu/\sqrt{k}\sigma).$$

已有研究表明，对于无法使用吉布斯抽样的混合模型，VI 的表现可能优于更通用的马尔可夫链蒙特卡罗技术（[10]）。这促使我们使用 CAVI算法探索适用于 GSN 的变分贝叶斯技术。

与马尔可夫链蒙特卡罗算法类似，我们对 μ、 σ^2、 p 假设如下信息先验：

$$p \sim \text{Beta}(a, b),\ \text{and}$$
$$(\mu, \sigma^2) \sim N - \Gamma(v_0, n_0, \alpha, \beta),$$

其中 $N - \Gamma$ 表示正态-逆伽马分布。回顾一下 GSN 随机变量 X_1、 $...X_n$ 的模型：

$$N_i \mid p \sim Geometric(p)$$
$$X_i \mid N_i\,, \mu\,, \sigma\,, p \sim N(N_i\mu, N_i\sigma^2) \qquad i = 1, 2, ...m$$

根据变分推断的术语，令潜变量向量为 $\boldsymbol{z}$ =($\boldsymbol{N}$、 μ、 σ^2、 p)，其中包含 $\boldsymbol{N}$ = (N_1、 N_2、 $...N_m$)，包括参数和未观测变量。假设变分分布 q 分解为：

$$q(\boldsymbol{z}) = q(\boldsymbol{N})q(\mu, \sigma, p).$$

我们得到潜变量 $\boldsymbol{N}$ 的如下最优变分分布（完整推导请见附录）：

$$q^*(\boldsymbol{N}) = \prod_{i=1}^{m} \frac{1}{\boldsymbol{C}_i\sqrt{n_i}} e^{An_i + B\frac{x_i^2}{n_i}}, \tag{12}$$

其中

$$\boldsymbol{C}_i = \sum_{i=1}^{\infty} \frac{1}{\sqrt{n_i}} e^{An_i + B\frac{x_i^2}{n_i}}, \text{ with } A = \psi(b^*) - \psi(a^* + b^*) - \frac{1}{2n^*} - \frac{(\mu^*)^2\alpha^*}{2\beta^*}, \text{ and } B = -\frac{\alpha^*}{2\beta^*}.$$

得益于共轭先验的选择，我们得到模型参数（p， μ， σ）的最优变分分布的以下闭式解:

$$q^*(p) \sim \text{Beta}(a^*, b^*), \tag{13}$$

其中 $a^* = n + a$，以及 $b^* = E_N(\sum_{i=1}^m n_i) - n + b.$

$$q^*(\mu, \sigma^2) \sim N - \Gamma(\mu^*, n^*, \alpha^*, \beta^*), \tag{14}$$

其中

$$\mu^* = \frac{\sum_{i=1}^m x_i + n_0 v_0}{n_0 + E_N\left(\sum_{i=1}^m n_i\right)}, \; n^* = n_0 + E_N\left(\sum_{i=1}^m n_i\right), \; \alpha^* = \alpha + 1 + \frac{n}{2}, \text{ and}$$
$$\beta^* = 2\beta + n_0(\mu^* - v_0)^2 + E_N\left(\sum_{i=1}^m \frac{1}{n_i}(x_i - n_i\mu^*)^2\right).$$

此外，我们注意到，关于 N 的期望可利用式（2）和（3）中给出的期望计算如下:

$$E_N\left(\sum_{i=1}^m n_i\right) = \sum_{i=1}^m E_{N_i}(n_i) \tag{15}$$

$$E_N\left(\sum_{i=1}^m \frac{1}{n_i}(x_i - n_i\mu^*)^2\right) = \sum_{i=1}^m \left(x_i^2 E_{N_i}\left(\frac{1}{n_i}\right) + \mu^{*2} E_{N_i}(n_i) - 2x_i\mu^*\right) \tag{16}$$

5.4 算法与模拟

下面介绍用于 GSN 的 CAVI算法，以变分分布近似后验分布:

- 初始化 $N_{(0)\,i}$， $\mu_{(0)}$ $\sigma^{2}{}_{(0)}$ $p_{(0)}$ 并计算 $ELBO_{(0)}(q^*(N_{(0)\,i}$， $\mu_{(0)}$ $\sigma^{2}{}_{(0)}$ $p_{(0)})$
- 对于迭代 t，当 $ELBO_{(t-1)} - ELBO_{(t-2)} > \epsilon$

  1. 使用 $\mu^{(t-1)}$， $\sigma^{2\,(t-1)}$， $p^{(t-1)}$，根据方程 (12) 计算变分分布 $q^*(N^{(t)})$，并使用方程 (2) 和 (3) 计算所需的期望。

  2. 根据方程 (13) 和 (14) 计算 $\mu^{(t)}$， $\sigma^{2\,(t)}$， $p^{(t)}$ 的变分分布，并重新计算步骤 1 中计算的期望。

  3. 根据方程（11）计算 $ELBO^{(t)}(q^*(N, \mu, \sigma, p))$（证据下界的完整表达式见附录）。

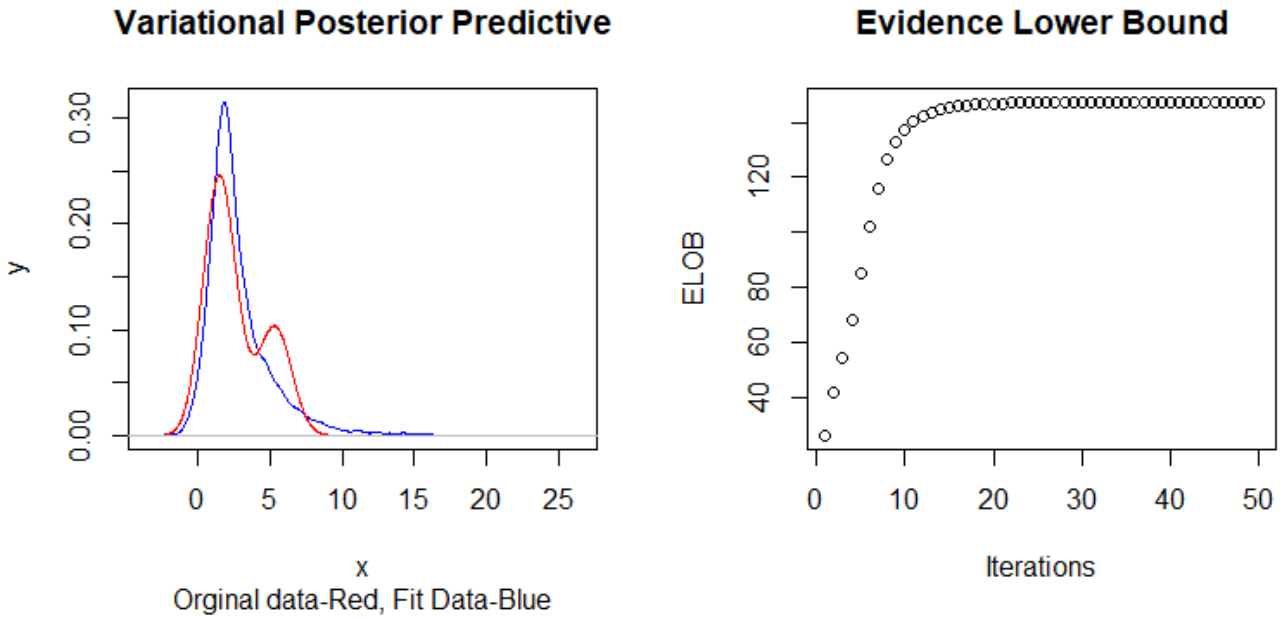

图 8：左：将变分后验预测拟合（蓝色）与来自 $GSN(\mu = 2, \sigma = 1, p = 0.6)$ 的原始 100 个样本（红色）进行对比。变分后验分布通过 CAVI算法 得到。右：GSN 模型的证据下界图。证据下界随着每次迭代而增加，并迅速收敛，表明 CAVI算法 优化了证据下界。

数据集	GSN- 变分推断	GSN-MCMC	ASN
极小偏度	0.125	0.049	0.119
小偏度	0.129	0.059	0.084
中等偏度	0.131	0.09	0.223
大偏度	0.130	0.07	0.207
对数正态	0.148	0.063	0.278
Frontier数据	0.164	0.089	0.096
豚鼠数据	0.203	0.12	0.313

表 5：比较研究中使用的数据集的变分后验预测与原始数据的柯尔莫哥洛夫-斯米尔诺夫距离（KSD），以及 ASN 和 GSN 模型对应的 KSD 结果（第 4.1 节）

- 返回 $q^*(\mu, \sigma, p)$

为了展示 CAVI 算法的收敛性，我们从$GSN(\mu = 2, \sigma = 1, p = 0.6)$生成了一个随机数据集，并运行了该算法。图 8 展示了变分后验预测以及证据下界的收敛情况。

为了比较变分推断（VI）与马尔可夫链蒙特卡罗技术在 GSN 和 ASN 上的表现，我们在第 4.1 节和第 4.2 节所考察的模拟数据集和真实数据集上实施了变分方法。核 Stein 差异值见表 5。对于极小

偏度数据，采用马尔可夫链蒙特卡罗技术的 ASN 表现优于采用变分方法的 GSN。然而，一旦偏度增至中等或较大，采用变分贝叶斯（这是一种近似方法）的 GSN 表现便优于 ASN。此外，在真实数据集上也观察到了类似结果。不过，在所有这些研究中，基于马尔可夫链蒙特卡罗的 GSN 表现优于另外两种方法。与马尔可夫链蒙特卡罗技术相比，采用变分方法的 GSN 表现较差并不令人意外，因为前者是一种近似方法。不过，需要牢记的是，与马尔可夫链蒙特卡罗方法相比，变分方法显著缩短了计算时间。

6 结论

我们介绍了如何处理数据中的偏度问题，以及两种偏态正态分布：广为人知的阿扎利尼偏态正态分布([6])和近期提出的几何偏态正态([19])。我们指出了贝叶斯推断用于阿扎利尼偏态正态分布的优势，并进一步介绍了使用混合 Metropolis-Hastings 算法、结合信息先验对 GSN 进行贝叶斯推断的方法，同时提出了一种新颖的变分贝叶斯方法。

根据模拟研究和真实数据分析的结果，我们发现，对于小到中等偏度，ASN 和 GSN 的拟合效果相当。然而，对于具有大偏度和厚尾的数据，GSN 的拟合效果更好。我们还想指出，GSN 的贝叶斯推断实现和信息先验的选取更加容易。此外，结果表明，一个简单的变分算法在 GSN 情形下能够快速收敛。

在偏离正态性较小的情况下，ASN 具有优势，因为偏度可以通过单个参数控制；而对于 GSN，均值(μ)和概率(p)都会决定偏度。然而，正如引言中所述，ASN 的偏度参数(λ)的先验 elicitation 可能较为困难，因为 λ 不仅控制对称性，也决定分布的形状。此外，信息贝叶斯方法的实现可能更具挑战性，因为其条件后验分布属于统一偏态正态分布族([11])。

最后，我们提到 GSN（[20]）和 ASN（[1]）都有多元扩展，贝叶斯推断也可以进一步扩展到这两种分布，并进行类似的比较。多元 GSN 或 ASN 分布的混合分布可用于聚类问题（[24]）。GSN 可应用于贝叶斯回归，以比较 GSN 与 ASN 在处理偏态正态变量时的表现。变分推断可推动开发更快速且易于扩展的高维贝叶斯回归方法（[13]）。

7 致谢

我要感谢我的导师 Satyaki Mazumder 博士，他在偏态正态分布这一主题上给予了我真知灼见，并在整个项目过程中持续支持我，直至本文的撰写与编辑完成。

8 附录

8.1 条件后验分布的推导

再次令 $\mathbf{X} = \{X_1, X_2, \ldots, X_m\}$ 为 GSN 随机变量，令 $\mathbf{N} = \{N_1, N_2, \ldots, N_m\}$为潜变量。考虑 p 的以下先验分布：

$$p \sim \text{Beta}(\alpha, \beta)$$

根据方程 (5)， p 的条件后验分布可在相差一个常数的意义下求得，如下：

$$\begin{aligned} p(p \mid \mu, \sigma, \mathbf{X}, \mathbf{N}) &\propto p(\mathbf{X} \mid \mathbf{N}, \mu, \sigma, p) p(\mathbf{N} \mid p) p(p) \\ &\propto \prod_{i=1}^{m} p(1-p)^{n_i - 1} p^{a-1} (1-p)^{b-1} \\ &\propto p^{m} (1-p)^{\sum_{i=1}^{m} n_i - m} p^{a-1} (1-p)^{b-1} \\ &\propto p^{m+a-1} (1-p)^{\sum_{i=1}^{m} n_i - m + b - 1} \end{aligned}$$

因此，我们得到

$$p(p \mid \mu, \sigma, X, N) \sim Beta(m + a, \sum_{i=1}^{m} n_i - m + b).$$

接下来，对于均值和方差参数（μ， σ^2），考虑以下先验分布：

$$\mu \mid \sigma^2 \sim N(v_0, \frac{\sigma^2}{n_0})$$

$$\sigma^2 \sim Inv - Gamma(\alpha, \beta)$$

根据方程 (6)， $p(\mathbf{X} \mid \mu$， σ^2， p， $\mathbf{N})$，的似然为：

$$\begin{aligned} p(\mathbf{X} \mid \mu, \sigma^2, p, \mathbf{N}) &\propto \prod_{i=1}^{m} e^{-\frac{(x_i - n_i \mu)^2}{2 n_i \sigma^2}} \\ &\propto e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i \mu)^2} \end{aligned}$$

再次根据方程（6），我们得到 (μ, σ^2) 的条件联合后验分布：

$$\begin{aligned} p(\mu, \sigma^2 \mid p, \mathbf{X}, \mathbf{N}) &\propto p(X \mid \mu, \sigma, p, \mathbf{X}, \mathbf{N}) p(\mu, \sigma^2) \\ &\propto \frac{1}{\sigma^m} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i \mu)^2} \left(\frac{1}{\sigma^2}\right)^{\alpha + 1 + 1/2} e^{-\frac{1}{2\sigma^2} \left(2\beta + n_0 (u - v_0)^2\right)} \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{m}{2} + \alpha + 1 + \frac{1}{2}} e^{-\frac{1}{2\sigma^2} \left(2\beta + n_0 (u - v_0)^2 + \sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i \mu)^2\right)} \end{aligned}$$

对于 μ的条件后验分布，首先注意到， μ 的先验分布为：

$$p(\mu \mid \sigma^2) \propto e^{-\frac{n_0 (\mu - v_0)^2}{2\sigma^2}}.$$

因此我们有:

$$\begin{aligned}
p(\mu \mid \sigma^2, p, \mathbf{X}, \mathbf{N}) &\propto p(\mathbf{X} \mid \mu, \sigma^2, p, \mathbf{N}) p(\mu \mid \sigma^2) \\
&\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\frac{1}{n_i}(x_i - n_i\mu)^2} . e^{-\frac{n_0(\mu - v_0)^2}{2\sigma^2}} \\
&\propto e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}\frac{1}{n_i}(x_i^2 + n_i^2\mu^2 - 2x_i n_i \mu) + n_0(\mu - v_0)^2\right)} \\
&\propto e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}\frac{x_i^2}{n_i} + \mu^2\sum_{i=1}^{m} n_i - 2\mu\sum_{i=1}^{m} x_i + n_0\mu^2 + n_0 v_0^2 - 2n_0 v_0 \mu\right)} \\
&\propto e^{-\frac{1}{2\sigma^2}\left((\sum_{i=1}^{m} n_i + n_0)\mu^2 - 2\mu(\sum_{i=1}^{m} x_i + n_0 v_0) + \sum_{i=1}^{m}\frac{x_i^2}{n_i} + n_0 v_0^2\right)} \\
&\propto e^{-\frac{(\sum_{i=1}^{m} n_i + n_0)}{2\sigma^2}\left(\mu - \frac{n_0 v_0 + \sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i + n_0}\right)^2}
\end{aligned}$$

在最后一步中, 我们忽略了与 μ 无关的项, 并配成了完全平方。因此,

$$\mu \sim N(\mu^*, \frac{\sigma^2}{n^*}),$$

其中

$$\mu^* = \frac{n_0 v_0 + \sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i + n_0}$$

$$n^* = \sum_{i=1}^{m} n_i + n_0$$

σ^2 的条件后验分布

由于我们有,

$$p(\mu \mid \sigma^2, p, X, N) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{n^*(\mu - \mu^*)^2}{2\sigma^2}}$$

将 σ 和 μ 的联合后验分布除以 μ 的全条件后验分布, 得到的 σ 的条件后验分布不依赖于 μ, 因此我们得到,

$$\begin{aligned}
p(\sigma^2 \mid \mu, p, X, N) &\propto \frac{p(\mu, \sigma^2 \mid p, X, N)}{p(\mu \mid \sigma^2, p, X, N)} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1+\frac{1}{2}-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}\left(2\beta + n_0(u - v_0)^2 + \sum_{i=1}^{m}\frac{1}{n_i}(x_i - n_i\mu)^2 + n^*(\mu - \mu^*)^2\right)}
\end{aligned}$$

σ 的条件后验分布与 μ无关, 我们通过固定 $\mu = \mu^*$ 来计算后验分布, 得到:

$$p(\sigma^2 \mid \mu, p, X, N) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1} e^{-\frac{1}{2\sigma^2}\left(2\beta + n_0(u^* - v_0)^2 + \sum_{i=1}^{m}\frac{1}{n_i}(x_i - n_i\mu^*)^2\right)}$$

最后我们得到

$$\sigma^2 \sim Inv - Gamma(\alpha^*, \beta^*)$$

$$\alpha^* = \alpha + \frac{n}{2}$$

$$\beta^* = 2\beta + n_0(u^* - v_0)^2 + \sum_{i=1}^{m}\frac{1}{n_i}(x_i - n_i\mu^*)^2$$

N_i的条件后验分布

在这种情况下，条件分布没有闭式形式，不过我们仍然可以得到以下比例关系，

$$\begin{aligned} p(p \mid \mu, \sigma, \mathbf{X}, \mathbf{N}) &\propto p(\mathbf{X} \mid \mathbf{N}, \mu, \sigma, p) p(\mathbf{N} \mid p) \\ &\propto \prod_{i=1}^{m} \frac{1}{\sqrt{n_i}} e^{-\frac{(x_i - n_i\mu)^2}{2n_i\sigma^2}} (1-p)^{n_i - 1} \end{aligned}$$

这给出了每个独立 N_i的分布

8.2 变分更新方程的推导

沿用相同的记号，令**X** 和**N** 分别表示 GSN 随机变量和潜变量。现在，令变分分布为 $q(N, p, \mu, \sigma^2)$，并假设潜变量和参数的分布按如下方式因子分解：

$$q(N, p.\mu, \sigma^2) = q(N) q(p, \mu, \sigma^2)$$

潜变量的变分分布

$$\begin{aligned} \log q^*(N) &= E_{p,\mu,\sigma^2}(\log p(X \mid N, \mu, \sigma^2, p)) + E_{p,\mu,\sigma^2}(\log \pi(N \mid p)) + \text{const} \\ &= E_{\mu,\sigma^2}\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i\mu)^2 - \sum_{i=1}^{m} \log \sqrt{n_i} - \frac{n}{2} \log \sigma^2 \right) + \\ &\quad E_p\left(\left(\sum_{i=1}^{m} n_i - n \right) \log p(1-p) \right) + \text{const} \\ &= -\sum_{i=1}^{m} \log \sqrt{n_i} + E_p(\log(1-p)) \left(\sum_{i=1}^{m} n_i \right) - \\ &\quad E_{\mu,\sigma^2}\left(\frac{1}{2\sigma^2} \sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i\mu)^2 \right) + \text{const} \end{aligned} \tag{17}$$

关于参数的期望可通过使用下一节推导的参数变分后验分布求得

参数的变分分布

p, μ, σ^2 的变分分布为，

$$\begin{aligned} \log q^*(p, \mu, \sigma^2) = E_N(\log p(\mathbf{X} \mid \mathbf{N}, \mu, \sigma^2, p)) + E_N(\log p(N \mid p)) + E_N(\log \pi(p)) + \\ E_N(\log p(\mu, \sigma^2)) \end{aligned} \tag{18}$$

我们可以进一步将 $q(p, \mu, \sigma^2)$ 分解为 $q(p)q(\mu, \sigma^2)$，并将 $q(\mu, \sigma^2)$ 分解为 $q(\mu \mid \sigma^2)q(\sigma^2)$。根据对 p 的先验假设和式 (18)，我们有

$$\begin{aligned}\log q^*(p) &= E_N(\log \pi(N \mid p)) + E_N(\log \pi(p)) + \text{const} \\ &= n \log p + \left(E_N \left(\sum_{i=1}^{m} n_i \right) - n \right) \log(1-p) + (a-1)\log(p) + \\ &\quad (b-1)\log(1-p) + \text{const} \\ &= (n + a - 1)\log p + (E_N(\sum_{i=1}^{m} n_i) - n + b - 1)\log(1-p) + \text{const},\end{aligned}$$

这进而意味着

$$q^*(p) = \text{Beta}\left(n + a, E_N \left(\sum_{i=1}^{m} n_i \right) - n + b \right)$$

此外，对于 μ， σ^2 的联合密度，我们有

$$\log q^*(\mu, \sigma^2) = \log q^*(\mu \mid \sigma^2) + \log q^*(\sigma^2)$$

与推导条件后验分布类似，我们再次利用先验假设的共轭性和式 (18)，得到:

$$q^*(\mu \mid \sigma^2) = N(\mu^*, \frac{\sigma^2}{n^*}),$$

其中

$$\mu^* = \frac{\sum_{i=1}^{m} x_i + n_0 v_0}{n_0 + E_N\left(\sum_{i=1}^{m} n_i \right)}$$

$$n^* = n_0 + E_N(\sum_{i=1}^{m} n_i).$$

对于 σ^2 ，我们使用以下关系

$$\log q^*(\sigma^2) = \log q^*(\mu, \sigma^2) - \log q^*(\mu \mid \sigma^2)$$

化简上述表达式后，得到

$$q^*(\sigma^2) = Inv - Gamma(\alpha^*, \beta^*),$$

其中

$$\alpha^* = \alpha + 1 + \frac{n}{2}, \text{ and}$$

$$\beta^* = 2\beta + n_0(\mu^* - v_0)^2 + E_N \left(\sum_{i=1}^{m} \frac{1}{n_i} (x_i - n_i \mu^*)^2 \right).$$

GSN 分布的证据下界（ELBO）

$$ELBO(q) = \sum_N \int\int\int q(\mathbf{N}, p, \mu, \sigma^2) log\{\frac{p(\mathbf{X}, \mathbf{N}, p, \mu, \sigma^2)}{q(\mathbf{N}, p, \mu, \sigma^2)}\} dp d\mu d\sigma^2$$

我们得到最终表达式为:

$$\begin{aligned} ELBO(q) &= E\left[log\, p(\mathbf{X}, \mathbf{N}, p, \mu, \sigma^2)\right] - E\left[log\, q(\mathbf{N}, p, \mu, \sigma^2)\right] \\ &= E\left[p(\mathbf{X} \mid \mathbf{N}, \mu, \sigma^2)\ \right] + E\left[log\, p(\mathbf{N} \mid p)\right] + E\left[log\, p(p)\right] + E\left[log\, p(\mu, \sigma^2)\right] \\ &\quad - E\left[log\, q(\mathbf{N})\right] - E\left[log\, q(p)\right] - E\left[log\, q(\mu, \sigma^2)\right] \end{aligned}$$

所有期望均是针对相应参数的变分分布 (q) 计算的。

8.3 ASN 的信息性贝叶斯推断

我们简要介绍用于比较研究的吉布斯抽样算法。关于采用信息先验对 ASN 进行贝叶斯推断的完整细节，请参见 [11]。沿用 [11], 的相同记号，ASN 的概率密度函数表示为

$$f(x) = \frac{2}{\omega}\phi(\frac{x-\xi}{\omega})\Phi(\alpha\frac{x-\xi}{\omega}) \qquad x \in \boldsymbol{R}$$

我们用 ξ 表示均值，用 ω^2 表示方差，用 α 表示偏度参数（此前分别记作 μ、σ^2 和 λ）

沿用 [8], 和 [4], 的方法，基于 [6], 提出的 ASN 分布的随机表示，我们引入独立标准正态潜变量 η_1、.....、η_n。给定这些潜变量，我们可以将第 i 个观测值视为服从正态分布，其均值为 $\delta(y_i - \xi)$，方差为 $w^2(1-\delta^2)$。这种解释使得在以下先验选择下，参数 (ξ、ω^2) 的条件后验分布具有共轭性:

$$(\xi, \omega^2) \sim N - \Gamma(\xi_0, \kappa\omega^2, a, b)$$

我们假设偏度参数服从 ASN 先验分布， α

$$\alpha \sim ASN(\alpha_0, \psi_0, \lambda_0)$$

[11] 已表明，条件后验分布，$\pi(\alpha \mid y)$，属于统一偏态正态分布类，即 SUN。([3])。

考虑独立同分布的 ASN 样本 $y = (y_1, \ldots, y_n)$，我们得到条件后验分布 $\pi(\alpha \mid y)$

$$\alpha \mid y \sim SUN_{1,n+1}(\alpha_0, \gamma_2, \psi_0, 1, \Delta_2, \Gamma_2), \tag{19}$$

其中 $\Delta_2 = [\delta_i]_{i=1,\ldots,n+1}$ 且 $\delta_i = \psi_0 z_i(\psi_0^2 z_i^2 + 1)^{-1/2}$ 以及 $z = (y^T, \lambda_0\psi_0^{-1})^T$, $\gamma_2 = (\Delta_{2;1:n}\alpha_0\psi_0^{-1}, 0)$, $\Gamma_2 = \mathcal{I} - D(\Delta_2) + \Delta_2\Delta_2^T$.

关于先验 elicitation，我们再次参阅 [11]

吉布斯抽样算法

1. 根据其全条件后验分布更新 η_i

$$\eta_i \sim TN_0(\delta(y_i - \xi), w^2(1-\delta^2))$$

其中 δ 为 $\alpha/\sqrt{\alpha^2+1}$，且 TN_τ（μ，σ^2）是均值为 μ、方差为 σ^2、下限截断的正态分布 τ

2. 样本

$$\xi \sim N(\hat{\mu}, \hat{\kappa}\omega^2)$$

$$\omega^2 \mid \xi \sim Inv - Gmma(a + (n+1)/2, b + \hat{b})$$

其中

$$\widehat{\mu} = \frac{\kappa \sum_{i=1}^{n}(y_i - \delta\eta_i) + (1-\delta^2)\xi_0}{n\kappa + (1-\delta^2)}, \hat{\kappa} = \frac{\kappa(1-\delta^2)}{n\kappa + (1-\delta^2)}$$

$$\hat{b} = \frac{1}{2(1-\delta^2)}\left(\delta^2 \sum_{i=1}^{n}\eta_i^2 2\delta \sum_{i=1}^{n}\eta_i(y_i - \xi) + \delta \sum_{i=1}^{n}(y_i - \xi) + \frac{1-\delta^2}{\kappa}(\xi - \xi_0)^2\right).$$

3. 从 α 中抽取样本

$$\alpha \sim \pi(\alpha \mid y^*)$$

其中 $y_i^* = (y_i - \xi)/\omega$ 对于 $i = 1 \ldots n$ 和 $\pi(\alpha \mid y)$，由方程 (19) 给出。

参考文献

[1] AAzzalini, A. 和 Valle, A. D. (1996). 多元偏态正态分布。*Biometrika*，83(4):715–726。

[2] Adcock, C.、Eling, M. 和 Loperfido, N. (2015). 金融与精算学中的偏态分布：综述。《*The European Journal of Finance*》，21(13-14):1253–1281。

[3] Arellano-Valle, R. B. 和 Azzalini, A. (2006). 关于偏态正态分布族的统一。《*Scandinavian Journal of Statistics*》，33(3):561–574。

[4] Arellano-Valle, R. B.、Genton, M. G. 和 Loschi, R. H. (2009). 多元偏态正态分布的形状混合。《*Journal of Multivariate Analysis*》，100(1):91–101。

[5] Arnold, B. C. 和 Beaver, R. J. (2000). 偏斜柯西分布。《统计与概率快报》，49(3):285–290。

[6] Azzalini, A. (1985). 一类包含正态分布的分布。《*Scandinavian Journal of Statistics*》，12(2):171–178。

[7] Azzalini, A., Browne, R. P., Genton, M. G. 和 McNicholas, P. D.（2014）。比较两种偏态分布的表述，并特别关注基于模型的聚类。*arXiv* 预印本 *arXiv:1402.5431*。

[8] Bayes, C. 和 Branco, E.（2007）。标量偏态正态分布偏度参数的贝叶斯推断。巴西概率与统计学杂志，21：141–163。[9] Bishop, C. M.（2006）。模式识别与机器学习（信息科学与统计）。Springer-Verlag，柏林，海德堡。[10] Blei, D. M.、Kucukelbir, A. 和 McAuliffe, J. D.（2017）。变分推断：统计学家综述。美国统计学会会刊，112(518)：859–877。[11] Canale, A. 和 Scarpa, B.（2013）。偏态正态分布的信息性贝叶斯推断。[12] Eling, M.（2012）。将保险索赔拟合到偏态分布：偏态正态分布和偏态 Student 分布是好的模型吗？保险：数学与经济学，51(2)：239–248。[13] Fasano, A.、Durante, D. 和 Zanella, G.（2019）。用于高维二元回归模型的可扩展且准确的变分贝叶斯方法。[14] Figueiredo, F. 和 Gomes, M.（2013）。统计过程控制中的偏态正态分布。*Revstat* - 统计学期刊，11：83–104。[15] Ghaderinezhad, F.、Ley, C. 和 Loperfido, N.（2020）。斜对称分布的贝叶斯推断。对称性，12：491。[16] Hossain, A. 和 Beyene, J.（2015）。应用偏态正态分布检测微小 RNA 数据中的差异表达。应用统计学杂志，42(3)：477–491。[17] Jones, M. C. 和 Faddy, M. J.（2003）。t 分布的偏态扩展及其应用。英国皇家统计学会杂志。*B* 辑（统计方法学），65(1)：159–174。[18]Kazemi, R. 和 Noorizadeh, M.（2015）。偏态逻辑斯蒂分布与偏态正态分布的比较。*MATEMATIKA*：马来西亚工业与应用数学杂志，31(1)：15–24。[19] Kundu, D.（2014）。几何偏态正态分布。$Sankhy\bar{a}$：印度统计学杂志，*B* 辑（*2008-*），76(2)：167–189。[20] Kundu, D.（2017）。多元几何偏态正态分布。*Statistics*，51(6)：1377–1397。[21] Ley, C.（2013）。偏态分布。美国癌症协会。[22] Liseo, B. 和 Loperfido, N.（2006）。关于标量偏态正态分布参考先验的说明。统计规划与推断杂志，136：373–389。[23] Nadarajah, S.（2009）。偏态逻辑斯蒂分布。*AStA* 统计分析进展，93(2)：187–203。[24] Redivo, E.、Nguyen, H. D. 和 Gupta, M.（2020）。使用几何偏态正态分布对偏态和多峰数据进行贝叶斯聚类。计算统计学与数据分析，152：107040。

[25] Simon, H. A.（1955年）。关于一类偏斜分布函数。*Biometrika*，42(3/4):425–440。[26] Wang, J.、Boyer, J. 和 Genton, M. G.（2004年）。多元分布的偏斜对称表示。*Statistica Sinica*，14(4):1259–1270。