arXiv:2301.05257v1 [math.ST] 12 Jan 2023

用柯西分布阐释推断模型

ChuanhaiLiu* 和 Ryan Martin†

2023年1月16日

摘要

统计推断作为一种将经验转化为知识的正式科学方法，实践起来一向困难重重。尽管在当代，频率学派和贝叶斯方法已被公认为两大主流学派，但在过去一百年间，人们始终致力于开发更为可靠的方法，既从哲学层面探讨推断的科学意义，也从数学层面追求精确性和有效性。这些方法包括费希尔枢轴论证、Dempster-Shafe信念函数理论、广义枢轴推断、置信分布，以及近来提出的推断框架——推断模型。由于对柯西分布进行精确且有效的推断一向极具挑战性，本文以此为例，阐释统计推断领域不同学派的思想。研究表明，推断模型的标准方法能够对柯西分布的位置参数和尺度参数进行精确、有效且无先验的概率推断，而其他现有方法都存在各种困难。

1 引言

尽管统计推断已广泛应用于科学研究，但在统计学基础的发展中，它仍是一个难以捉摸的问题。一个很好的切入点，是两大学派——贝叶斯学派和频率学派——之间长达250年的争论。本文从将经验转化为知识的角度，简要讨论这两大主流学派。与此相应，本文也简要评述费希尔枢轴论证、Dempster-Shafe信念函数理论、广义枢轴推断和置信分布。

从将经验转化为知识的角度来看，Martin和Liu（2015b）指出，一个健全的推断框架应当根据已知信息，对我们感兴趣的未知量作出断言，并给出有效且高效的不确定性评估。一个令人满意的特例是，对所有未知量的先验分布均已知时进行贝叶斯推断，因为其推断结果具有有效性、高效性和概率性。然而，这并不一定意味着在这种情况下使用后验分布进行推断的科学家就是贝叶斯学派。贝叶斯统计学家常把这样一类人称为贝叶斯推断者：他们为所有未知量指定先验分布，这些分布或基于其主观意见（也称个人概率），或旨在产生经过频率校准的推断结果。为了区分这两类贝叶斯学派人士，后一类或许可以称为广义贝叶斯学派人士；这一术语曾用于基于贝叶斯决策理论的点推断研究。因此，从

*普渡大学统计学系
†北卡罗来纳州立大学统计学系

Martin 和 Liu（2015b）的观点来看，这类贝叶斯方法最值得质疑之处在于其推断结果的有效性或意义，这是延续数百年的贝叶斯学派与频率学派之争中的一个关键问题。

可以说，为了在科学研究中获得有效的不确定性评估，而在一定程度上牺牲理想概率推断所具有的情境特异性，这一需求使频率学派方法在20世纪广受欢迎。置信区间和p值是实践中最常用的两种频率学派概念，但它们缺乏理想的、情境特异性的概率解释，这至少在哲学层面令人不安。近期发展的推断模型（IM）框架在一定程度上有助于解决这一问题。更确切地说，Martin 和 Liu（2014）证明，在较弱的条件下，置信区间就是可信区间，p值就是可信度；而可信度这一概念具有情境特异性和概率性。从这个意义上说，IM将经过完善的频率学派作为其特例纳入其中。

柯西概率分布在物理学、数学以及光谱学等多个相关学科中发挥着重要作用。它也是学生t分布族中两个极端情形之一，另一个是人们熟悉的高斯分布。由于柯西分布的精确推断众所周知地难以实现，我们以此为例，阐明统计推断中不同的思想流派。研究表明，IM方法能够根据样本，对柯西分布的位置参数和尺度参数进行精确且有效的推断，而其他思想流派则面临各种困难。

第2节回顾柯西分布及其估计。第3节简要回顾IM方法。文中通过介绍如何利用单个观测值，在尺度已知的情况下估计柯西位置参数的解，来说明IM方法的推断程序。第4节讨论尺度已知时对位置参数的推断，以及同时对位置参数和尺度参数进行推断。第5节介绍仅针对位置参数和仅针对尺度参数的边际推断。本例中的推断问题为评估不同学派的观点提供了一个基准示例。因此，文中也讨论了针对这一问题的贝叶斯方法和信任推断方法。最后，我们在第6节中作简要总结。文中指出，尽管讨论仅限于柯西分布的推断，但具体的IM方法可以推广到一般的位置参数和尺度参数推断，也就是群变换。

2 柯西分布

标准柯西分布 $C(0,\ 1)$的概率密度函数为

$$f(x;0,1)=\frac{1}{\pi}\frac{1}{1+x^2},\qquad (x\in\mathbb{R})$$

或者等价地，通过累积分布函数

$$F(x;0,1)=\frac{1}{\pi}\arctan(x)+\frac{1}{2},\qquad (x\in\mathbb{R})$$

一般柯西分布 $C(\mu,\ \sigma)$ 可通过引入位置参数 $\mu\in\mathbb{R}$ 和尺度参数 $\sigma\in\mathbb{R}_+=\{\sigma$ ，从 $C(0,\ 1)$ 轻松得到： $\sigma\in\mathbb{R},\ \sigma>0\}$。因此， $C(\mu,\ \sigma)$ 的概率密度函数为

$$f(x;\mu,\sigma)=\frac{1}{\pi\sigma}\frac{1}{1+\frac{(x-\mu)^2}{\sigma^2}},\qquad (x\in\mathbb{R}) \tag{2.1}$$

其中，$\theta \equiv (\mu,\ \sigma)' \in \Theta \equiv \mathbb{R} \otimes \mathbb{R}_+$.。本文考虑的问题是：从容量为 n 的样本 X_1，...，X_n 中，对位置参数 μ 进行推断，其中尺度 σ 已知或未知。

正如Zhang（2010）所指出的，估计$C(\mu,\sigma)$的参数似乎极其困难。对于 $n > 3$，Copas（1975）和Gabrielsen（1982）证明，μ 和 σ 的联合似然函数是单峰的。就牛顿-拉夫森法和EM算法等计算方法而言，这一点似乎颇具吸引力；然而，固定 σ 时μ 的似然函数会随机地呈多峰，其局部极大值（不包括全局极大值）的数量渐近服从均值为 $1/\pi$的泊松分布，这一点令人困扰；参见Barnett（1966）、Reeds（1985）以及Bai和Fu（1987）。

众所周知，柯西分布具有稳定性，即$\frac{1}{n}\sum_{i=1}^{n} X_i \sim C(\mu,\sigma)$。这意味着，样本均值作为 μ 的估计量效率低下。文献中已研究了截尾均值法以及更一般的L估计量等替代方法；参见Rothenberg等（1964）、Bloch（1966）、Cane（1974）和Zhang（2010）。

另一个有趣的 μ 估计量是在 σ 固定时定义的，称为 Pitman 估计量（Freue，2007；Pitman，1939）。该估计量是使用平坦先验计算得到的 μ 贝叶斯后验均值，即

$$\hat{\mu}_P = \frac{\int_{-\infty}^{\infty} u \prod_{i=1}^{n} f(x_i; u, \sigma) du}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i; u, \sigma) du} \tag{2.2}$$

在给定 $n > 1$的条件下。我们将在第4节证明，使用平坦先验得到的 μ 贝叶斯后验进行推断，在某种意义上等同于 IM 方法所得的推断结果。将 Lindley（1958）的方法应用于条件似然函数，即可确立这一结果。频率学派条件推断方法与我们称为条件IM的方法相关。

McCullagh（1992）研究了同时包含位置参数和尺度参数的柯西模型上的条件推断。为了对位置参数 μ进行推断，McCullagh（1992）采用了费希尔枢轴推断方法，并证明该方法针对两种不同的辅助统计量会产生不同的结果。这一发现又为费希尔枢轴推断法存在困难这一问题增添了有趣的例证。关于这一问题的更详细讨论见第 5.3 节。

3 推断模型

3.1 基本框架

为了实现无先验概率推断，推断模型（IM）的核心思想是根据观测事件，预测未观测但可预测的随机量，从而对不确定性进行评估。若随机变量的分布完全已知，则称其为可预测的当且仅当。令 X 表示其抽样空间 $\mathbb{X}$中的观测数据，令 θ 表示参数空间 Θ中的未知参数。具体而言，对于假设的模型，我们引入一个分布已知的辅助未观测随机量 $U \in \mathbb{U}$，并用一个方程组表示该模型，将 X与 θ 和 U联系起来：

$$a(X, \theta, U) = 0 \tag{3.1}$$

IM 推断通过使用所谓的有效随机集（见第 3.2 节）来预测 U 的未观测实现，记为 $U^\star$；它通过式（3.1）与观测数据 X 相关联。更具体地，令 $\mathcal{S}$ 表示预测随机集。对于任何感兴趣的断言 $\theta \in A \subset \Theta$，支持 A 的证据定义为所有可能的 θ 值都满足

(3.1)

$$\Theta_{X,S} = \{\theta : a(X, \theta, U) = 0 \text{ for some } U \in \mathcal{S}\} \tag{3.2}$$

落入 A，即

$$\mathrm{Bel}\left(A|X\right)=\Pr\left(\Theta_{X,S}\subseteq A\right). \tag{3.3}$$

函数 Bel $(.|X)$ 称为信念函数。关于信念函数这一概念的近期讨论，参见 Martin 等人（2010）及其中的参考文献。

与通常的概率不同，概率具有可加性，即，$\Pr(A)+\Pr(A^c)=1$，而信念函数具有次可加性，即，

$$\mathrm{Bel}\left(A|X\right)+\mathrm{Bel}\left(A^c|X\right)\leq 1$$

对所有 $A\subseteq\Theta$ 均成立。信念函数具有次可加性，是因为存在这样一个复杂情况：$\Theta_{X,\ S}$可能同时与 A 和 A^c 重叠。这种重叠产生了一种新的情况，即既不属于 A，也不属于 A^c。因此，通过引入另一个函数——似然函数，便可完整刻画 A 的不确定性评估。该函数定义为

$$\mathrm{Pl}\left(A|X\right)=\Pr\left(\Theta_{X,S}\cap A\neq\emptyset\right)\qquad(A\subseteq\Theta). \tag{3.4}$$

由此可得

$$\mathrm{Pl}\left(A|X\right)=1-\mathrm{Bel}\left(A^c|X\right)\qquad(A\subseteq\Theta). \tag{3.5}$$

例 3.1. 考虑根据单个观测值 X 对柯西分布$(\mu,1)$进行推断。通过以下关联表示抽样模型

$$X=\mu+Z\qquad(Z\sim\mathrm{Cauchy}\left(0,1\right)).$$

假设我们关注所有单点断言 $\{\mu\}$ 的同时推断。取对称随机集

$$\mathcal{S}=[-|C|,|C|]\qquad(C\sim\mathrm{Cauchy}\left(0,1\right)).$$

它会生成以下推断结果：

1. 信念函数$(\{\mu\}|X)=0$；以及

2. $\mathrm{Pl}\left(\{\mu\}|X\right)=\Pr\left(\{X-\mu\}\in\mathcal{S}\right)=\Pr\left(|X-\mu|\leq|C|\right)=2\Pr\left(C\leq-|X-\mu|\right),$

对于所有 $\mu\in(-\infty,\infty)$。

由 $X=0$ 给出的似然曲线如图 1 所示。对于 μ的点估计，可以采用 μ的最可能值。最可能值取决于底层随机集的构造。然而，从这个例子可以看出，所有不确定性都由可预测随机变量 Z 的分布刻画；在统计学文献中，它被称为枢轴量。对关于 μ 的断言使用似然，而不是给 μ 赋予通常的概率，是为了避免通过关联 $\mu=X-Z$，由 Z 的概率推导 μ的概率时遇到的困难。关于这一问题的更多讨论见第 5.3 节。

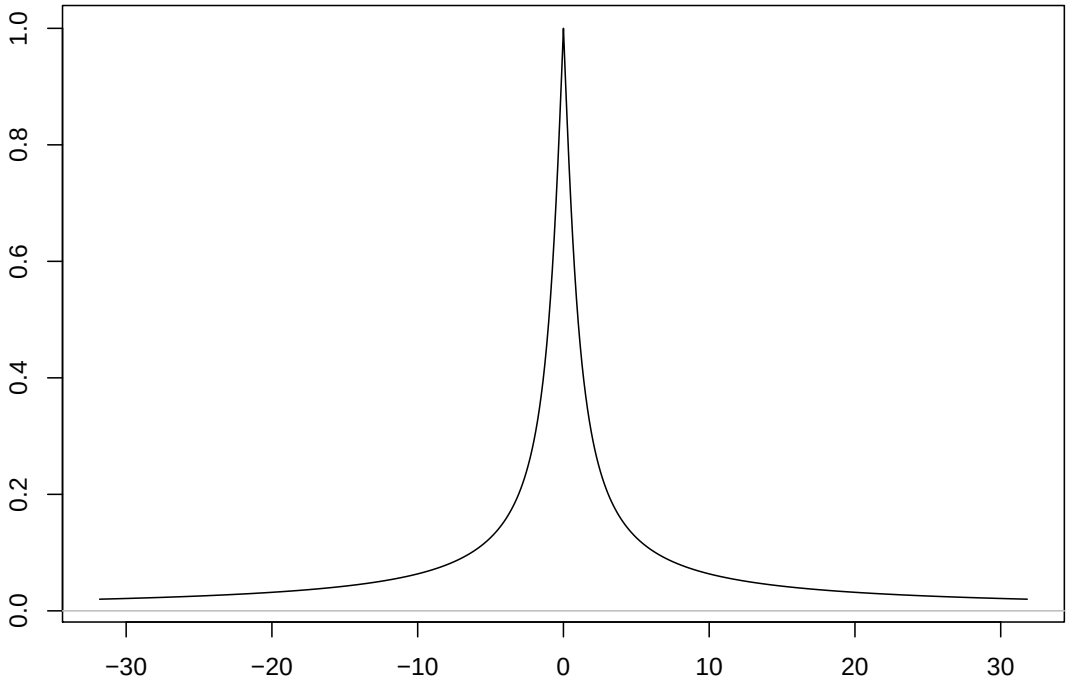

图 1：例 3.1 中的似然函数 Pl($\{\mu\}|X=0$)。

3.2 随机集

尽管 IM 的理念源于 Fisher 的费希尔枢轴论证，旨在实现无先验概率推断（Martin 和 Liu, 2013；Martin等人，2010；Zhang和Liu，2011），但 IM 通过处理辅助随机变量 U来对不确定性进行推理。由于 U 定义明确，概率运算不会出错。费希尔枢轴论证则并非如此。例如，假设对于给定的 X， U 与 θ 之间存在一一映射。写作

$$\theta = \theta_X(U). \tag{3.6}$$

遗憾的是，映射 (3.6) 无法实现有效的概率质量传播，因为映射会随着 (U， θ) 对的变化而改变。这解释了为什么 IM 使用随机集来预测 $U^\star$。使用随机集实际上是在进行概率计算/积分。IM背后的这一思路有助于确立所需的频率校准性质，在 IM 框架中称为有效性。

定义3.1 (有效性)。 若随机集 $\mathcal{S} \subset \mathbb{U}$ 与 $U^\star$ 独立，则称其在预测$U^\star$时有效，当且仅当 $\mathcal{S}$覆盖 $U^\star$ 的概率在分布上不小于标准均匀随机变量，也就是说，

$$Pr(Pr(U^\star \in \mathcal{S}|U^\star) \leq u) \geq u \tag{3.7}$$

对所有 $u \in [0,1]$。

Martin 和 Liu（2013）指出，使用有效随机集可以保证关于 θ 的推断是有效的，即在 A 为真的情况下，在重复实验中，作为反对 A 的证据度量，信念函数 Bel($A^c|X$) 的上界由标准均匀随机变量给出。

4 信息组合

4.1 在已知 σ的情况下对 μ 进行推断

考虑从样本量为 n的样本 $X_1, ..., X_n$中，对位置参数未知 μ 、尺度已知 σ 的柯西分布进行推断。不失一般性，我们取单位尺度 $\sigma = 1$。一个简单的关联为

$$X_i = \mu + U_i \qquad (i = 1, ..., n; U_i \overset{iid}{\sim} \mathrm{C}(0, 1)) \tag{4.1}$$

为了对 μ进行有效推断，仍可在 n维空间 $\mathbb{R}^n$ 中使用预测随机集来预测未观测的 U_i'，但这种做法效率不高。为了通过预测进行有效推断，我们将（4.1）改写为

$$\begin{aligned} X_1 &= \mu + U_1 & (4.2)\\ W_i \equiv U_i - U_1 &= X_i - X_1 \qquad (i = 2, ..., n) & (4.3) \end{aligned}$$

显然，只需预测 U_1 就可以对 μ 进行推断。此外，（4.3）中的 n维量 $W \equiv （W_2, ..., W_n）'$所包含的信息可用于更有效地预测 U_1。这种实现高效预测的思路催生了所谓的条件IM。Martin和Liu（2015a）建立了一般的CIM理论，并阐明了其与贝叶斯推断和Fisher 条件推断的关系。例如，Martin和Liu（2015a）证明，CIM 程序能够产生有效推断。

U_1 的条件分布可通过常规代数运算直接求得，结果如下:

$$f_{U_1|W}(u|W = w) \propto \frac{1}{1 + u^2} \prod_{i=2}^{n} \frac{1}{1 + (u + w_i)^2} \qquad (u \in \mathbb{R}) \tag{4.4}$$

IM 信念和似然　值可使用标准数值方法相应地计算

。由 (4.2) 和 (4.3) 给出的 CIM 构造似乎并非唯一。关键在于，$n - 1$的一个 $(n - 1)$ 维量 $U \equiv (U_1, ... U_n)'$ 是完全观测到的。要对 μ进行推断，我们只需根据观测信息预测一个一维量。以下结果讨论了从下列 CIM 指数类中选择不同构造时的唯一性问题： $a \equiv (a_1, ... a_n)' \in \mathbb{R}^n$，其中 $\sum_{i=1}^{n} a_i = 1$ 和 $a_1 \neq 0$:

$$\begin{aligned} \sum_{i=1}^{n} a_i X_i &= \mu + T & (4.5)\\ W_i \equiv U_i - T &= X_i - \sum_{j=1}^{n} a_j X_j \qquad (i = 2, ..., n) & (4.6) \end{aligned}$$

其中 $T \equiv \sum_{i=1}^{n} a_i U_i$。

定理 1。 对于 T 上的任意预测随机集 $\mathcal{T}$，在以 $W = w$ 为条件时，该集合

$$\{\mu : \sum_{i=1}^{n} a_i X_i = \mu + t \ for \ some \ t \in \mathcal{T}\} \tag{4.7}$$

$_i$对所有 a 都相同，其中 $a \equiv (a_1, ..., a_n)' \in \mathbb{R}^n$，且 $\sum_{i=1}^{n} a_i = 1$ 和 $a_1 \neq 0$。

证明。 由 (4.6) 可知，我们只需证明：对于所有 $\mu_0 \in \mathbb{R}$，给定 $W = w$ 时 $T \geq \sum_{i=1}^{n} a_i X_i - \mu_0$ 的条件概率与 a 无关。由于从 $(U_1, ..., U_n)'$ 到 (T, W) 的变换的雅可比行列式为 $|a_1|$，因此有

$$\begin{aligned} f_{T|W=w}(t) &\propto \frac{1}{1 + \frac{(t - \sum_{k=2}^{n} a_k (t + X_k - \sum_{j=1}^{n} a_j X_j))^2}{a_1^2}} \prod_{k=2}^{n} \frac{1}{1 + (t + X_k - \sum_{j=1}^{n} a_j X_j)^2} \\ &= \prod_{k=1}^{n} \frac{1}{1 + (t + X_k - \sum_{j=1}^{n} a_j X_j)^2} \end{aligned}$$

对于 $t \in \mathbb{R}$。由此可得

$$\Pr\left(T \geq \sum_{i=1}^{n} a_i X_i - \mu_0\right) \propto = \frac{\int_{-\infty}^{\mu_0} \prod_{k=2}^{n} \frac{1}{1+(X_k - u)^2} du}{\int_{-\infty}^{\infty} \prod_{k=2}^{n} \frac{1}{1+(X_k - u)^2} du} \tag{4.8}$$

对于所有 $t_0 \in \mathbb{R}$，证明完毕。

需要指出的是，虽然可以考虑更一般的一类条件推断模型（CIM），用 $(U_1 - T, ..., U_n - T)$ 的任意线性变换来表示 $(n-1)$ 维量，但很容易看出，结论仍然相同，只是相应的论证需要使用更繁琐的符号。

□

注 4.1。 从定理 *1* 的证明中的结果（*4.8*）可以看出，在已知 σ 时，对位置参数 μ 进行有效推断，本质上与采用关于 μ 的平坦先验所得的贝叶斯后验相同。对于来自如下形式联合分布的任意样本 $X_1, ..., X_n$，这一结论均成立： $F(X_1 - \mu, ..., X_n - \mu)$。这一结果可通过将 ***Lindley***（*1958*）应用于给定 X_1 时 μ 的似然来证明，不过这里采用的是给定 W 时 X_1 的条件模型。

4.2 关于 μ 和 σ的联合推断

对于关于 μ和 σ的联合推断，来自 C(μ, σ) 的大小为 n 的样本（≥ 2）， X_1，……， X_n的关联可写为

$$X_i = \mu + \sigma U_i \qquad (i = 1, ..., n; U_i \overset{iid}{\sim} \mathrm{C}(0, 1)) \tag{4.9}$$

沿用 Fisher（1934）和 McCullagh（1992）的做法，我们使用次序统计量（$X_{(1)}$，……， $X_{(n)}$）′，并将用于合并信息的关联写为

$$\begin{aligned} X_{(1)} &= \mu + \sigma T & (4.10) \\ X_{(2)} - X_{(1)} &= \sigma S & (4.11) \\ \frac{X_{(i)} - X_{(1)}}{X_{(2)} - X_{(1)}} &= W_i, \qquad (i = 3, ..., n) & (4.12) \end{aligned}$$

其中，$(U_{(1)}, ..., U_{(n)})'$表示对 $U_1, ..., U_n$进行排序，其中 $U_{(1)} \leq ... \leq U_{(n)}$, $T = U_{(1)}$, $S = U_{(2)} - U_{(1)}$, 且 $W_i = \frac{U_{(i)} - U_{(1)}}{U_{(2)} - U_{(1)}}$适用于 $i = 3, ..., n.$。因此，可以根据给定观测信息 $(W_3, ..., W_n)$ 的预测分布，通过预测 (T, S)，对 (μ, σ) 进行推断。

定理 2。 给定 $W=(W_3,...,W_n)'=(w_3,...,w_n)$ 时，(T,S) 的条件密度为

$$f_{T,S|W=w}(t,s)\propto s^{n-2}\frac{1}{1+t^2}\frac{1}{1+(t+s)^2}\prod_{i=3}^{n}\frac{1}{1+(t+w_is)^2},\qquad (t\in\mathbb{R},s\in\mathbb{R}_+)\tag{4.13}$$

证明。 众所周知，$(U_{(1)},...,U_{(n)})'$ 的密度函数为

$$f_{U_{(1)},...,U_{(n)}}(u_1,...,u_n)=n!\prod_{i=1}^{n}\frac{1}{\pi}\frac{1}{1+u_i^2},\qquad (u_1\le u_2\le...\le u_n).$$

该变换的雅可比行列式

$$\begin{aligned}U_{(1)}&=T\\U_{(2)}&=T+S\\U_{(i)}&=T+SW_i,\qquad (i=3,...,n)\end{aligned}$$

可轻易求得为 s^{n-2}。因此，给定 $W=(W_3,...,W_n)'=(w_3,...,w_n)$ 时，(T,S) 的条件密度为

$$f_{T,S|W=w}(t,s)\propto s^{n-2}\frac{1}{1+t^2}\frac{1}{1+(t+s)^2}\prod_{i=3}^{n}\frac{1}{1+(t+w_is)^2},\qquad (t\in\mathbb{R},s\in\mathbb{R}_+)$$

□

5 边际推断

本节讨论双参数柯西模型 $C(\mu,\sigma)$ 中仅关于 σ 的推断，以及仅关于 μ 的推断。这两个问题都属于所谓的边际推断，即针对未知参数中低维量的推断。Martin 和 Liu（2015c）讨论了 IM框架中边际推断的一般方法。事实证明，在贝叶斯可信区间和 IM 可信度区间方面，对未知 μ 的 σ 进行 IM 推断，以及对未知 μ 的 σ 进行推断，与采用 Pitman 先验的贝叶斯推断一致

$$\frac{1}{\sigma}d\mu d\sigma,\qquad (\mu\in\mathbb{R},\sigma\in\mathbb{R}_+)\tag{5.1}$$

以贝叶斯可信区间和 IM 可信度区间的形式。

McCullagh（1992）指出，基于两种不同辅助统计量的费希尔枢轴推断并不唯一，而双参数柯西模型正是一个例子。第 5.3 节从边际IM的角度讨论相关问题。

5.1 对未知 μ的 σ 推断

对 σ 进行有效边际推断的基本思路是，通过预测一维辅助随机变量而非二维辅助随机变量来获得有效推断。其有效性的直观解释如下。对于任意用于未观测的 (T,S)、并在 (4.10) 和 (4.11) 中定义的有效预测随机集$\mathcal{S}_{T,\ S}$，投影预测随机集

$$\mathcal{S}_S=\{s:(t,s)\in\mathcal{S}_{T,S}\text{ for some }t\}$$

为仅针对 S 提供了有效预测随机集，从而可通过 (4.10) 对 σ 进行有效边际推断。这意味着，可以通过构造有效的二维预测随机集来实现对 σ 的有效推断，这些预测随机集在 S的空间上的投影具有较小的随机性。

对于双参数柯西分布模型，可以通过根据（4.13）或其边际分布预测（4.11）中未观测到的 S ，对 σ 进行有效的边际推断

$$f_{S|W=w}(s) \propto s^{n-2} \int_{-\infty}^{\infty} \frac{1}{1+t^2} \frac{1}{1+(t+s)^2} \prod_{i=3}^{n} \frac{1}{1+(t+w_i s)^2} dt, \qquad (s \in \mathbb{R}_+)$$

令 $\mathcal{S}$ 为针对未观测到的 S的有效预测随机集。 $\mathcal{S}$ 在 σ 空间上的传播随机集为

$$\{(X_{(2)} - X_{(1)})/s : s \in \mathcal{S}\}.$$

对于单侧有效预测随机集

$$\mathcal{S}_S = \{s : s \geq S\}, \qquad (S \sim f_{S|W=w}(s))$$

传　播随机集在 σ 空间上的特征由其上界无界性决定

$$G \equiv \frac{X_{(2)} - X_{(1)}}{S}.$$

不难证明， G 的密度具有如下形式

$$f_{G|W=w}(g) \propto \frac{1}{g^{n+1}} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \frac{1}{1+\frac{(X_{(i)}-u)^2}{g^2}} du, \qquad (s \in \mathbb{R}_+) \tag{5.2}$$

这建立了 IM 与贝叶斯方法在双参数柯西模型 $C(\mu, \sigma)$ 中对 σ 进行边际推断之间的联系。

5.2 对 μ 的推断（ σ 未知）

与对 σ 的边际推断类似，通过利用边际关联，可对 μ 进行有效推断

$$\frac{X_{(1)} - \mu}{X_{(2)} - X_{(1)}} = \frac{T}{S} \tag{5.3}$$

具体而言，是基于观测到的 W 对 $M \equiv T/S$ 进行预测。 M 的预测分布密度为

$$f_{M|W=w}(m) \propto s^{n-2+1} \int_{0}^{\infty} \frac{1}{1+m^2 s^2} \frac{1}{1+(ms+s)^2} \prod_{i=3}^{n} \frac{1}{1+(ms+\frac{X_{(i)}-X_{(1)}}{X_{(2)}-X_{(1)}} s)^2} ds, \qquad (m \in \mathbb{R})$$

上述预测分布与基于 Pitman 先验 (5.1) 的贝叶斯推断有内在联系，而非基于杰弗里斯先验。为此，注意，对于 M 上的任意预测随机集 S，通过 (5.3) 中的映射，其在 μ 空间中的传播随机集为

$$\mu = X_{(1)} - (X_{(2)} - X_{(1)})M$$

对于使用单侧预测随机集处理 M 的情形， μ 上的随机集也为单侧。其非平凡端点的密度为

$$f_{Z|W=w}(z) \propto \int_{0}^{\infty} \frac{1}{s^n} \prod_{i=1}^{n} \frac{1}{1+\frac{(z-X_{(i)})^2}{s^2}} \frac{1}{s} ds, \qquad (z \in \mathbb{R})$$

这一结果对应于所声称的在双参数柯西分布模型中对 μ的IM边际推断的贝叶斯解释。

5.3 IM与费希尔枢轴推断

推断模型的研究最初受到枢轴推断以及 Dempster-Shafer 信念函数理论的启发。正如 Liu和Martin（2015）所阐述的，虽然 IM 提供无需先验的精确推断，但枢轴推断并非无需先验。更多讨论可见 Liu和Martin（2015）以及 Martin和Liu（2015b）。这里，我们以 McCullagh（1992）在双参数模型下发现的条件推断和枢轴推断中的非唯一性问题为例，从 IM框架的视角说明枢轴论证的困难。

基于文献中柯西模型 $C(\mu,\ \sigma)$ 的良好性质，包括 Menon（1962，1966）、Pitman和Williams（1967）以及Williams（1969）的研究，McCullagh（1992）利用复数引入了参数 $(\mu,\ \sigma)$ 的一种颇具吸引力的表示形式

$$\theta = \mu + i\sigma$$

并将 $C(\mu,\ \sigma)$ 写作 $C(\theta)$。此处将用到的 McCullagh（1992）结果可概括为以下定理。

定理3。若 $Y \sim C(\theta)$，则

$$Y^* = \frac{aY+b}{cY+d} \sim C\left(\frac{a\theta+b}{c\theta+d}\right).$$

取 $c=1$ 和 $a=b=d=0$，可知

$$R_i = \frac{1}{X_i}, \qquad (i=1,...,n)$$

构成来自 $C(1/\theta)$ 或 $C\left(\frac{\mu}{\mu^2+\sigma^2}, \frac{\sigma}{\mu^2+\sigma^2}\right)$ 的一个大小为 n 的样本。第4节和第5节的讨论表明，关于

$$\mu^* \equiv \frac{\mu}{\mu^2+\sigma^2} \qquad \text{and} \qquad \sigma^* \equiv \frac{\sigma}{\mu^2+\sigma^2}$$

使用条件关联

$$\begin{aligned} R_{(1)} &= \mu^* + \sigma^* U_{(1)} \\ R_{(2)} - R_{(1)} &= \sigma^*(U_{(2)} - U_{(1)}) \\ \frac{R_{(i)} - R_{(1)}}{R_{(2)} - R_{(1)}} &= \frac{U_{(i)} - U_{(1)}}{U_{(2)} - U_{(1)}}, \qquad (i=3,...,n) \end{aligned}$$

在这种情况下，该结果相应的贝叶斯解释是对 $(\mu^*,\ \sigma^*)$ 使用 Pitman 先验：

$$\frac{1}{\sigma^*} d\mu^* d\sigma^* = \frac{1}{\sigma^*} d\mu^* d\sigma^*, \qquad (\mu^* \in \mathbb{R}, \sigma^* \in \mathbb{R}_+)$$

枢轴推断的要求颇高，因为它实际上将这个先验视为 $(\mu^*,\ \sigma^*)$ 上的隐式先验分布，并给人一种最终推断不依赖先验的印象。因此，在针对 $C(\mu,\ \sigma)$ 的条件推断和枢轴推断这两种构造中使用不同的先验分布，即，

$$\frac{1}{\sigma} d\mu d\sigma$$

and

$$\frac{1}{\sigma^*}d\mu^* d\sigma^* = \frac{\mu^2+\sigma^2}{\sigma}d\mu d\sigma,$$

会导致相互冲突的概率推断。

IM不存在枢轴论证所面临的困难。当所关注的问题是关于 $\mu^* = \mu/(\sum^+ \sigma^2)$或 $\sigma^* = \sigma/(\mu^2+\sigma^2)$的推断时，定理3中的变换仅用于IM框架中。然而，这个例子很好地说明了Liu和Martin（2015）所论述的困难。

6 结论性评述

毫无疑问，统计推断在科学研究中非常重要。为推断建立坚实的基础，已被证明是最困难的问题之一。本文通过柯西模型阐释了IM框架。由于篇幅有限，本文重点基于直觉解释柯西模型下的基本IM、条件IM和边际IM；对于该模型，所有以往思想流派似乎都认为推断十分困难。柯西模型相关的大量研究成果充分证明了这一点，其中包括Pitman（1939）、Menon（1962，1966）、Rothenberg等（1964）、Bloch（1966）、Pitman和Williams（1967）、Williams（1969）、Cane（1974）、Copas（1975）、Gabrielsen（1982）、Reeds（1985）、Bai和Fu（1987）、McCullagh（1992）、Freue（2007）、Zhang（2010）、Martin和Liu（2015a）以及其中的参考文献。广受欢迎的矩法和最大似然法是标准的频率学派方法。这些方法并不精确且效率低下。McCullagh（1992）发现的非唯一性问题揭示了枢轴论证的困难。它还表明，在进行贝叶斯推断时必须谨慎，尤其是在希望获得频率校准推断的情况下。

尽管我们的讨论仅限于对柯西分布的推断，但具体的IM方法可以扩展到对一般位置参数和尺度参数的推断，即群变换。相关主题值得单独开展深入研究。

关于IM的最新进展可参见Martin和Liu（2015b），其中还包括一些新的发展，例如Liu和Martin（2020）以及Cella和Martin（2022），这里只举几例。尽管IM仍有必要进一步发展，但我们相信，以观测到的随机变量为条件来预测未观测随机变量这一基本思想，将继续成为发展精确且高效的概率推断的基础。

参考文献

Bai, Z. 和 J. Fu（1987）。关于柯西分布位置参数的最大似然估计量。《加拿大统计学杂志》15(2)，137–146。Barnett, V.（1966）。似然方程有多个根时对最大似然估计量的评估。Biometrika 53(1/2)，151–165。Bloch, D.（1966）。关于柯西分布位置参数估计的一则注记。《美国统计学会会刊》61(315)，852–855。Cane, G. J.（1974）。基于样本分位数对柯西分布参数进行线性估计。《美国统计学会会刊》69(345)，243–245。Cella, L. 和 R. Martin（2022）。有效性、相容似然度量与保形预测。《国际近似推理杂志》141，110–130。

Copas, J. (1975)。关于柯西分布似然的单峰性。Biometrika 62 (3)，701–704。Fisher, R. A.（1934）。数学似然的两个新性质。伦敦皇家学会会刊：A辑，数学与物理特征论文集 144(852), 285–307。Freue, G. V. C.（2007）。柯西位置参数的 Pitman 估计量。统计规划与推断杂志 137(6), 1900–1913。Gabrielsen, G.（1982）。关于柯西分布似然的单峰性：一些评论。Biometrika 69(3), 677–678。Lindley, D. V.（1958）。信 fiducial 分布与贝叶斯定理。皇家统计学会杂志：B辑（方法论）20, 102–107。Liu, C. 和 R. Martin（2015）。无先验后验概率推断的框架。Wiley跨学科评论：计算统计学 7(1), 77–85。Liu, C. 和 R. Martin（2020）。推断模型与可能性测度。arXiv 预印本 arXiv:2008.06874。Martin, R. 和 C. Liu（2013）。推断模型：无先验后验概率推断的框架。《美国统计学会会刊》108(501), 301–313。Martin, R. 和 C. Liu（2014）。关于将 p值解释为可信度的说明。《中华统计学志》，1703–1716。Martin, R. 和 C. Liu（2015a）。条件推断模型：结合信息进行无先验概率推断。皇家统计学会杂志：B辑（统计方法论）77(1), 195–217。Martin, R. 和 C. Liu（2015b）。推断模型：不确定性推理，第145卷。CRC出版社。Martin, R. 和 C. Liu（2015c）。边际推断模型：针对关注参数的无先验概率推断。《美国统计学会会刊》110(512), 1621–1631。Martin, R.、J. Zhang 和 C. Liu（2010）。邓普斯特–谢弗理论与基于弱信念的统计推断。统计科学，72–87。McCullagh, P.（1992）。条件推断与柯西模型。Biometrika 79(2), 247–259。Menon, M.（1962）。柯西分布的刻画。《数学统计年鉴》33(4), 1267–1271。Menon, M.（1966）。柯西分布的另一项特征性质。《数学统计年鉴》37(1), 289–294。Pitman, E.（1939）。关于位置参数和尺度参数的假设检验。Biometrika 31(1/2), 200–215。Pitman, E. 和 E. Williams（1967）。柯西随机变量的柯西分布函数。《数学统计年鉴》，916–918。

Reeds, J. A.（1985）。柯西位置似然方程根的渐近数量。《统计年鉴》，775–784。Rothenberg, T. J.、F. M. Fisher 和 C. B. Tilanus（1964）。关于柯西样本估计的一则说明。《美国统计学会会刊》59(306), 460–463。Williams, E.（1969）。柯西随机变量的柯西分布函数及柯西分布的刻画。《数学统计年鉴》，1083–1085。Zhang, J.（2010）。柯西分布位置参数的高效 L估计量。《计算统计学》25(1), 97–105。张 J. 和 C. Liu（2011）。基于弱信念的邓普斯特-谢弗推理。《中华统计学志》，475–494。