一种用于 Feldman-Cousins置信区间的重要性抽样方法

Lukas Berns*
Tohoku University, Japan
（日期：2023年3月21日）

在各种高能物理领域中，例如中微子振荡实验，典型渐近置信区间构造所依据的若干假设并不成立，因此必须采用计算成本高昂的方法（如 Feldman-Cousins 方法），以获得具有适当统计覆盖率的置信区间。按照构造方式，要计算高置信水平下的区间，需要拟合数百万甚至数十亿个伪实验，同时却将大部分计算成本浪费在低置信水平下过于精确的区间上。本文介绍了一种简单的重要性抽样方法：将针对所有待检验参数值生成的伪实验复用于同一个混合分布。这样可以显著降低临界值估计的误差，尤其是在高置信水平下；同时，还能在生成伪实验所对应的参数值之间，对这些临界值进行正确插值。本文通过中微子振荡分析中的一个简单示例，以数值方式展示了理论计算所得的性能，并讨论了该方法与统计力学及 p-值计算中类似技术之间的关系。

arXiv:2303.11290v1 [hep-ex] 20 Mar 2023

I. 引言

任何实验的重要组成部分，都是通过统计分析从测量结果中提取模型参数（例如物理常数）的相关信息。由于测量不可避免地包含统计涨落，人们通常以置信区间（高维情况下则为置信区域）的形式报告这些约束。这些区间是在参数空间中根据观测数据计算得出的，其构造方式确保：对于参数的任何真实值，至少有预先设定比例的可能实验结果会产生一个覆盖该真实参数值的区间。可能实验结果中这一预先设定的比例称为置信水平（CL）。

本文余下部分采用以下记号： x 表示实验结果，它可以是单次实验中多个观测值组成的向量。 θ 表示模型参数，其中可以包含一个或多个维度的连续自由度，也可以包含离散自由度。 $p(x \mid \theta)$表示给定某些模型参数时实验结果的概率分布函数。对于给定的实验结果，将 $p(x \mid \theta)$视为 θ 的函数，称为似然函数，记作 $L(\theta \mid x) := p(x \mid \theta)$。使似然达到最大值的参数值记作 $\theta(x) := \arg\max_\theta L(\theta \mid x)$，某个参数值处的对数似然与最大似然处的对数似然之差记作 $\Delta\chi^2(\theta \mid x) := -2\log L(\theta \mid x)/L(\theta(x) \mid x)$。置信水平记作 $1-\alpha$。

在许多情况下，可以应用 Wilks 的一个有用定理 [1] ，从而大大简化此类置信区间的构造。该定理指出，在渐近极限下，在真实参数值处计算的 $\Delta\chi^2(\theta \mid x)$服从自由度为 k 的卡方分布，其中 k 是参数空间 θ的维数，而参数空间必须是连续的。该定理成立的前提是满足适当条件，从而保证能够在真实参数值附近找到似然的最大似然值，并且似然可以作泰勒二次展开。根据这一渐近分布，可以通过选取所有满足 $\Delta\chi^2(\theta \mid x) \leq \Delta\chi^2_c$的$\theta$ 值来构造置信区间，其中临界值 $\Delta\chi^2_c$ 可由卡方分布的分位数函数轻松计算得到。

由于需要满足相应的假设，基于威尔克斯定理的置信区间并不适用于观测值数量较少的情况，也不适用于因物理边界（例如 $\theta \geq 0$）、离散自由度，或无法通过二次展开刻画的周期性而不适用的参数空间。以中微子振荡实验为例，它们存在上述所有问题，稍后我们将给出一个例子。在这种情况下，必须针对选定的参数值实际生成伪实验系综，以研究适用于构造置信区间的统计量的分布。

一种常用方法是 Feldman-Cousins（FC）方法 [2], ：对于每个假定真实值为 θ_t 而生成的伪实验 x' ，计算真实参数值处的 $\Delta\chi^2(\theta_t \mid x')$ 值，以获得其分布。

* lukasb@epx.phys.tohoku.ac.jp

然后，通过该分布的经验 $1-\alpha$ 百分位数得到临界值 $\Delta\chi^2_c$。由于 $\Delta\chi^2(\theta_t \mid x')$ 的分布通常会随真实参数值而变化，因此临界值现在是计算它们时所对应真实值的函数，我们将其记为 $\Delta\chi^2_c(\theta_t)$。最后，通过选择 $\Delta\chi^2(\theta \mid x) \leq \Delta\chi^2_c(\theta)$，为实际观测到的数据 x 构造置信区间。实际上，只能在选定的参数值处计算 $\Delta\chi^2_c(\theta)$，因此需要进行插值（例如线性插值），以计算置信区间。

Feldman-Cousins 方法在计算高置信水平区间时效率很低，因为根据定义，只有少部分伪实验会参与分位数计算。例如，在粒子物理学中，“发现”的阈值通常设为 $\alpha = 5.7 \times 10^{-7}$（“$5\sigma$”阈值），此时，根据定义，只有每 170 万次伪实验中才有一次的 $\Delta\chi^2$（$\theta_t \mid x'$）值会高于临界值。因此，为了得到所需的临界值，往往需要拟合数百万次伪实验，同时却将大部分计算时间“浪费”在较低置信水平下精度过高的临界值上。出于这些原因，实际中通常只计算到 2σ（$\alpha = 4.6 \times 10^{-2}$）或 3σ 置信水平（$\alpha = 2.7 \times 10^{-3}$）的 FC 置信区间。

本文表明，引入一种能更频繁地产生高置信水平伪实验的替代抽样分布实际上极其简单：只需将参数取若干值时生成的伪实验重用于构造混合分布。通过适当的重新加权，可使高置信水平下临界值的误差呈指数级降低。该方法还提供了一种方法，可在真实参数值的子集之间正确插值临界值，从而无需采用常见的朴素插值方法。

本文安排如下。首先，我们回顾传统 FC 方法。接着，我们定义新方法，并通过讨论一种理想的重要性抽样分布来推导该方法。我们计算重要性抽样权重的上界，并据此计算与传统 FC 方法相比，估计临界值的误差降低幅度。随后，我们讨论临界值插值的能力，以及误差和其他诊断量的计算。接着，我们采用一个源自中微子振荡分析的玩具示例，对比这两种计算方法，并将改进效果与上一节中的理论上界进行核验。最后，我们讨论该方法与统计力学中类似技术及 p 值计算的关系、与贝叶斯边缘似然的关系，以及存在干扰参数时该方法的适用范围。

II. 传统 Feldman-Cousins 方法中的临界值

为引入符号，我们简要回顾传统 Feldman-Cousins 方法中临界值的计算。首先，我们在参数空间中选取 S 个点，记为 θ_s，其中 s 从 1 到S。现在，在每个 θ_s处，我们通过从 $p(x \mid \theta_s)$中采样，生成一个由 n_{exp} 个伪实验组成的系综 $\{x\}_s$。虽然假定所有伪实验都位于同一空间中，但表示系综的花括号上的 s 下标会标明生成这些实验所用的分布。对于每个伪实验 $x \in \{x\}_s$，现在计算 $\Delta\chi^2(\theta_s \mid x)$，并通过任何合适的估计量求出 $1-\alpha$ 分位数$\Delta\chi^2_{c,\ s}$。例如，可以直接对 $\Delta\chi^2(\theta_s \mid x)$的值排序，并将第 $\lfloor \alpha \times n_{\text{exp}} \rfloor$ 大值作为$\Delta\chi^2_{c,\ s}$，此时我们有

$$\sum_{x \in \{x\}_s} I\big(\Delta\chi^2(\theta_s \mid x) \geq \Delta\chi^2_{c,s}\big) = \lfloor \alpha \times n_{\text{exp}} \rfloor. \tag{1}$$

此处，$I(\cdot)$是指示函数：若括号中的逻辑语句为真，则返回 1，否则返回 0。$\lfloor \cdot \rfloor$表示向下取整函数。

最后，通过某种插值方案得到临界值函数 $\Delta\chi^2_c(\theta)$。例如，可以设定$\Delta\chi^2_c(\theta_s) := \Delta\chi^2_{c,\ s}$ 并对其间任意 θ 值进行线性插值。为减小插值误差，通常必须在迭代方案中手动或自动 [3]调整抽样参数值 $\{\theta\}_S$ 的选取。

临界值的渐近方差与二项误差 $\alpha(1-\alpha)/n_{\text{exp}}$成正比，因此，高置信水平（$\alpha \ll 1$）通常意味着需要 $n_{\text{exp}} \gg 1/\alpha$ 才能获得可靠的临界值。由于整个过程要针对参数空间中的所有 S点重复进行，生成（并拟合）的伪实验总数为 $S \times n_{\text{exp}} \gg S/\alpha$。

III. 混合分布 Feldman-Cousins 方法

A. 定义

我们将这种新方法称为“混合分布 Feldman-Cousins”方法。它与传统方法的主要区别在于：在计算所有目标各自的临界值时，会重复使用生成的伪实验

参数空间点 θ_t ，并赋予额外的权重

$$w(x \mid \theta_t) := \frac{p(x \mid \theta_t)}{\frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)} = \frac{1}{\frac{1}{S}\sum_{s=1}^{S} \exp[-\frac{1}{2}\{\Delta\chi^2(\theta_s \mid x) - \Delta\chi^2(\theta_t \mid x)\}]}. \tag{2}$$

分母中的值是 $x \in \{x\}_{\rm mix} := \bigcup_{s=1}^{S}\{x\}_s$的抽样概率分布，即所有 θ_s 值对应的 $p(x \mid \theta_s)$的混合分布。由于权重基于抽样概率，而抽样概率就是似然函数，因此可以使用与计算每个伪实验的 $\Delta\chi^2(\theta \mid x)$相同的过程来计算权重。由于取两个 $\Delta\chi^2$ 值之差，最小 χ^2 在$\hat{\theta}(x)$处的贡献，以及任何与 θ无关的偏移量（例如泊松似然中的 $n!$因子）都会在分母中抵消，因此无需精确获知这些量。传统方法只需计算生成伪实验时所对应 θ_s 值的 $\Delta\chi^2(\theta_s \mid x)$；而在这里，我们需要计算所有 $\theta_{s'}$ （包括 $s' \neq s$）和 θ_t对应的值。

现在，我们可以将临界值 $\Delta\chi^2_{c,t}$ 定义为 $x \sim \{x\}_{\rm mix}$ 对应的 $\Delta\chi^2(\theta_t \mid x)$ 的 w 加权 $1-\alpha$ 分位数，例如

$$\sum_{x\in\{x\}_{\rm mix}} w(x \mid \theta_t) I\big(\Delta\chi^2(\theta_t \mid x) \geq \Delta\chi^2_{c,t}\big) \lesssim \alpha \times S n_{\rm exp}, \tag{3}$$

其中，$\lesssim$ 表示我们取满足该不等式的最小 $\Delta\chi^2_{c,t}$ 。

B. 推导

为了在较大的 $\Delta\chi^2$ 值处获得更多伪实验，从而得到更精确的高置信水平临界值，我们采用重要性抽样方法：不直接从目标分布 $p(x \mid \theta_t)$ 中抽样，而是从另一个分布中抽样，并根据概率分布之比对抽取的玩具实验重新加权，以计算目标分布下的相关量（临界值）。因此，问题变为：生成所需伪实验的理想抽样分布是什么？

需要注意的是，找到一个尽可能接近目标分布的抽样分布非常重要，同时还要使高 $\Delta\chi^2$ 伪实验更有可能生成。特别是，如果每个实验 x 由 m 次测量组成，那么这些实验是一个 m 维空间中的点，并且我们可以拉伸或压缩抽样分布的维度有 m 个。与其考虑如何估计分位数，不如考虑如何使用直方图估计 $Y(x) := \Delta\chi^2(\theta_t \mid x)$ 下的概率密度 $p(Y(x) \mid \theta_t)$。使用重新加权时，除了落入某个箱中的伪实验数量所对应的二项误差 $n_{\rm exp}p(1-p)$ 外，权重方差还会带来额外贡献：给定估计量

$$\hat{P}_b := \frac{1}{n_{\rm exp}} \sum_{i=1}^{n_{\rm exp}} w(x_i) I(y_b \leq Y(x_i) < y_{b+1}) \tag{4}$$

并使用 $I(\cdot)^2 = I(\cdot)$，得到

$$\mathbb{E}[\hat{P}_b] = \pi_b \mathbb{E}_b[w] \tag{5}$$

$$\mathrm{Var}[\hat{P}_b] = \frac{1}{n_{\rm exp}} \left(\pi_b(1-\pi_b)\mathbb{E}_b[w]^2 + \pi_b \mathrm{Var}_b[w]\right) \tag{6}$$

$$\pi_b := \mathbb{E}[I(y_b \leq Y(x) < y_{b+1})] \tag{7}$$

$$\mathbb{E}_b[w^k] := \frac{1}{\pi_b}\mathbb{E}[w(x)^k I(y_b \leq Y(x) < y_{b+1})] \tag{8}$$

$$\mathrm{Var}_b[w] := \mathbb{E}_b[w^2] - \mathbb{E}_b[w]^2 \tag{9}$$

其中，式（6）中的第一项是由于落入该分箱的伪实验数量所产生的常规二项误差，第二项则是由于落入该分箱的伪实验之间的权重方差所产生的附加项。

因此，我们希望增加落入高-$y(x)$ 区间的伪实验数量（$n_{\rm exp}\pi_b$），以降低二项误差，同时尽可能减小区间内的权重方差。这意味着，在理想的零方差情况下，权重应仅通过 $y(x)$ 取决于 x 。或者，等价地说，由于权重是目标分布与抽样分布之比，我们希望使用的抽样分布与目标分布之间仅相差一个 $\Delta\chi^2(\theta_t \mid x)$ 的函数因子。

关键在于理解高 $\Delta\chi^2(\theta_t \mid x)$ 值的含义。似然 $L(\theta \mid x)$ 是从 θ 中抽样得到给定伪实验 x 的概率。高 $\Delta\chi^2(\theta_t \mid x) = -2\log L(\theta_t \mid x)/L(\hat{\theta}(x) \mid x)$ 表示存在一个 $\hat{\theta}(x)$ 值，在该值处，抽样得到给定伪实验的可能性高于在“目标” θ_t 值处。因此，通过在 $\theta \neq \theta_t$ 处生成伪实验，我们可以更高效地获得 $\Delta\chi^2(\theta_t \mid x)$ 较高的伪实验。

天真地选择仅使用在某个 θ' $(\neq \theta_t)$ 处生成的伪实验，并按采样概率比值 $p(x \mid \theta_t)/p(x \mid \theta')$ 加权，效果反而会比之前更差。这是因为 $\theta(x)$ 取决于伪实验 x，因此对于某些伪实验，从 $p(x \mid \theta_t)$ 而不是从$p(x \mid \theta')$ 采样 x 可能更合适，从而导致权重方差呈指数级增长（且通常无界）。

解决方案很简单：通过在包含 的集合 $\{\theta\}_S :=$ $\{\theta_1, \theta_2, \cdots, \theta_S\}$ θ_t $p_{\text{sample}}\ x\ \geq$ 上使用混合分布 $p_{\text{sample}}(x)$ $\frac{1}{S} p\ x \mid \theta_t$ $= \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$，我们可以保证权重具有上界，因为 ( ) ( )，因此

$$w(x \mid \theta_t) \leq S. \tag{10}$$

C. 良好网格的伪实验权重界限

如果我们选择足够密集且范围足够宽广的网格（即良好网格），使得我们可以假设在 $\{\theta\}_S$ 上有一个良好的最小值 $\hat{\theta}_S(x)$，其含义是

$$\Delta\chi^2(\hat{\theta}_S(x) \mid x) \leq \begin{cases} \epsilon & \text{if } \Delta\chi^2(\theta_t \mid x) \leq \Delta\chi^2_{\max} \\ \Delta\chi^2(\theta_t \mid x) & \text{otherwise} \end{cases} \tag{11}$$

$$\Delta\chi^2(\hat{\theta}_S(x) \mid x) := \text{最小值}_s\ \Delta\chi^2(\theta_s \mid x) \tag{12}$$

对于所有 x，我们可以给出比式 (10) 严格得多的权重界限。这里，$\epsilon \lesssim 1$ 会随着 $\{\theta\}_S$ 间距变密而变小，而 $\Delta\chi^2_{\max}$ 会随着 $\{\theta\}_S$ 覆盖范围变宽而变大。由于这一附加条件也能处理 $\theta_t \notin \{\theta\}_S$ 的情况，我们定义一个符号 C：若 $\theta_t \in \{\theta\}_S$，则其值为 1；否则为 0。注意，要在 $C = 0$ 条件下保证式 (11)，通常需要 $\{\theta\}_S$ 中的参数值充分良好地包围 θ_t。例如，对于一维连续 θ 参数，需要 $\min_s \theta_s \leq \theta_t \leq \max_s \theta_s$。

首先，我们关注具有 $\Delta\chi^2(\theta_t \mid x) \leq \Delta\chi^2_{\max}$ 的伪实验，这是我们主要感兴趣的对象，并注意到

$$\frac{p(x \mid \hat{\theta}_S(x))}{p(x \mid \theta_t)} = \exp\left[\tfrac{1}{2}\{\Delta\chi^2(\theta_t \mid x) - \Delta\chi^2(\hat{\theta}_S(x) \mid x)\}\right] \geq \exp\left[\tfrac{1}{2}\Delta\chi^2(\theta_t \mid x) - \tfrac{\epsilon}{2}\right]. \tag{13}$$

概率比之和现在有如下下界

$$\sum_{s=1}^{S} \frac{p(x \mid \theta_s)}{p(x \mid \theta_t)} \geq C \times \frac{p(x \mid \theta_t)}{p(x \mid \theta_t)} + \frac{p(x \mid \hat{\theta}_S(x))}{p(x \mid \theta_t)} \geq C + \exp\left[\tfrac{1}{2}\Delta\chi^2(\theta_t \mid x) - \tfrac{\epsilon}{2}\right] \tag{14}$$

因为 $\{\theta\}_S$ 同时包含 $\hat{\theta}_S(x)$ 和（若 $C = 1$）θ_t。这意味着，对于任何具有 $\epsilon < \Delta\chi^2(\theta_t \mid x) \leq \Delta\chi^2_{\max}$ 的伪实验，从 $\frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$ 中抽取 Sn_{exp} 个样本时，抽到它的可能性高于从目标分布 $p(x \mid \theta_t)$ 中抽取 n_{exp} 个样本时。概率比之和还具有如下上界

$$\sum_{s=1}^{S} \frac{p(x \mid \theta_s)}{p(x \mid \theta_t)} = C + \sum_{s(\neq t)} \frac{L(\theta_s \mid x)}{L(\theta_t \mid x)} \leq C + (S - C)\exp\left[\tfrac{1}{2}\Delta\chi^2(\theta_t \mid x)\right] \tag{15}$$

因为 $L(\theta_s \mid x) \leq L(\hat{\theta}(x) \mid x)$。这意味着权重有如下界

$$\frac{S}{C + (S - C)\exp\left[\frac{1}{2}\Delta\chi^2(\theta_t \mid x)\right]} \leq w(x \mid \theta_t) \leq \frac{S}{C + \exp\left[\frac{1}{2}\Delta\chi^2(\theta_t \mid x) - \frac{\epsilon}{2}\right]}. \tag{16}$$

我们可以看到，界限仅通过 $\Delta\chi^2(\theta_t \mid x)$依赖于伪实验；此外还注意到，当 $\Delta\chi^2(\theta_t \mid x)$足够大时，上界 $w_{\max}$ 与下界 $w_{\min}$ 之比趋于

$$\frac{w_{\max}}{w_{\min}} \to (S - C)e^{\epsilon/2}, \tag{17}$$

这表明，只要网格点数量 S 不是特别大，权重的相对方差就很小。对于 $\Delta\chi^2(\theta_t \mid x)$超过阈值 $\Delta\chi^2_{\max}$的伪实验，我们有

$$\frac{p(x \mid \hat{\theta}_S(x))}{p(x \mid \theta_t)} = \exp\left[\tfrac{1}{2}\{\Delta\chi^2(\theta_t \mid x) - \Delta\chi^2(\hat{\theta}_S(x) \mid x)\}\right] \geq 1 \tag{18}$$

根据式（11），因此也得到了权重的一个上界

$$w(x \mid \theta_t) \leq \frac{S \times p(x \mid \theta_t)}{C \times p(x \mid \theta_t) + p(x \mid \hat{\theta}_S(x))} \leq \frac{S}{C+1}. \tag{19}$$

D. 使用良好网格时临界值估计量的性能

由于分位数（临界值）只是累积分布函数（CDF）的反函数，我们可以通过累积分布函数估计方差的降低来估计分位数估计变化的相对降低。对于可观测量 $y \sim f(y)$，二者之间的关系为 $\mathrm{Var}[\hat{y}(P)] = f(y)^{-2}\mathrm{Var}[\hat{P}(y)]$，其中 $\hat{y}(P)$ 是 $\hat{}$ 分位数函数估计量，$P(y)$ 是累积分布函数估计量，而 $f(y)$ 是概率分布函数。

按照 式（3），并使用简写符号 $Y(x) := \Delta\chi^2(\theta_t \mid x)$，我们的累积分布函数估计量 is

$$\hat{P}(y \mid \theta_t) = \frac{1}{Sn_{\exp}} \sum_{x \in \{x\}_{\mathrm{mix}}} w(x \mid \theta_t) I(Y(x) \geq y) \tag{20}$$

其中 $x \sim \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$。这是目标累积分布函数 $P(y \mid \theta_t)$ 的无偏估计量

$$\begin{aligned}
\mathbb{E}\big[\hat{P}(y \mid \theta_t)\big] &= \mathbb{E}\big[w(x \mid \theta_t) I(Y(x) \geq y)\big] && (21)\\
&= \mathbb{E}\big[I(Y(x) \geq y) \;\big|\; \theta_t\big] && (22)\\
&= P(y \mid \theta_t). && (23)
\end{aligned}$$

其中，$\mathbb{E}[\cdot]$ 表示对 $x \sim \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$取期望，$\mathbb{E}[\dot{-} \mid \theta_t]$ 表示对$x \sim p(x \mid \theta_t)$取期望。现在定义

$$y_{\max} := \max\{\Delta\chi^2_{\max}, y\}, \tag{24}$$

单次伪实验的方差为

$$\begin{aligned}
&\text{方差}\big[w(x \mid \theta_t) I(Y(x) \geq y)\big] && (25)\\
&= \mathbb{E}\big[w(x \mid \theta_t)^2 I(Y(x) \geq y)^2\big] - \mathbb{E}\big[w(x \mid \theta_t) I(Y(x) \geq y)\big]^2 && (26)\\
&= \mathbb{E}\big[w(x \mid \theta_t) I(Y(x) \geq y) \mid \theta_t\big] - P(y \mid \theta_t)^2 && (27)\\
&\leq \mathbb{E}\left[\frac{S \times I(y \leq Y(x) \leq y_{\max})}{C + \exp\left[\frac{1}{2}\big(Y(x) - \epsilon\big)\right]} \;\middle|\; \theta_t\right] + \frac{S}{C+1}\mathbb{E}\big[I(Y(x) \geq y_{\max}) \;\big|\; \theta_t\big] - P(y \mid \theta_t)^2 && (28)\\
&\leq \frac{S}{C + \exp\left[\frac{1}{2}(y - \epsilon)\right]}\mathbb{E}\left[I(y \leq Y(x) \leq y_{\max}) \mid \theta_t\right] + \frac{S}{C+1}P(y_{\max} \mid \theta_t) - P(y \mid \theta_t)^2 && (29)\\
&= S \times \left[\frac{P(y \mid \theta_t) - P(y_{\max} \mid \theta_t)}{C + \exp\left[\frac{1}{2}(y - \epsilon)\right]} + \frac{P(y_{\max} \mid \theta_t)}{C+1}\right] - P(y \mid \theta_t)^2 && (30)
\end{aligned}$$

从第二行到第三行时，我们使用了 $I(\cdot)^2 = I(\cdot)$；到第四行时，我们使用了式 (16) 和式 (19) 中的上界；到第五行时，我们使用了指示函数自变量中的 $Y(x) \geq y$。因此，累积分布函数估计量的方差为

$$\begin{aligned}
\mathrm{Var}\big[\hat{P}(y \mid \theta_t)\big] &= \frac{1}{Sn_{\exp}}\mathrm{Var}\big[w(x \mid \theta_t) I(\Delta\chi^2(\theta_t \mid x) \geq y)\big] && (31)\\
&\leq \frac{1}{n_{\exp}}\left(\frac{P(y \mid \theta_t) - P(y_{\max} \mid \theta_t)}{C + \exp\left[\frac{1}{2}(y - \epsilon)\right]} + \frac{P(y_{\max} \mid \theta_t)}{C+1} - \frac{P(y \mid \theta_t)^2}{S}\right) && (32)
\end{aligned}$$

我们注意到，前两项中的 S 因子被约去，这是因为在对每个 θ_t 值进行累积分布函数估计时，可以重复使用在所有 S 值下生成的伪实验。

作为参考，传统 FC 方法中累积分布函数估计量的方差（在以下公式中记为“conv”）由二项误差给出

$$\mathrm{Var}[\hat{P}_{\mathrm{conv}}(y) \mid \theta_t] = \frac{1}{n_{\mathrm{exp}}} \left(P(y \mid \theta_t) - P(y \mid \theta_t)^2 \right), \tag{33}$$

因此，混合-FC 方法中估计的临界值 $\hat{}y(P \mid \theta_t)$ 的方差小了以下倍数

$$\gamma := \frac{\mathrm{Var}[\hat{y}(P \mid \theta_t)]}{\mathrm{Var}[\hat{y}_{\mathrm{conv}}(P \mid \theta_t) \mid \theta_t]} \tag{34}$$

$$= \frac{\mathrm{Var}\left[\hat{P}(y \mid \theta_t)\right]}{\mathrm{Var}\left[\hat{P}_{\mathrm{conv}}(y \mid \theta_t) \mid \theta_t\right]} \tag{35}$$

$$\leq \frac{A(y) + B(y)P(y_{\max} \mid \theta_t)/P - \frac{1}{S}P}{1 - P} \tag{36}$$

$$A(y) := \frac{1}{C + \exp\left[\frac{1}{2}(y - \epsilon)\right]} \tag{37}$$

$$B(y) := \frac{1}{C+1} - A(y). \tag{38}$$

其中，y 是真实的 P 分位数，满足 $P(y \mid \theta_t) = P$。上界的典型函数形状如图1a所示。首先考虑 $P \ll P(y_{\max} \mid \theta_t)$ 的情况。对于人们通常关注的较小 $P \leq 1/2$ 值，如果 $\theta_t \in \{\theta\}_S (C = 1)$，则混合模型方法对所有 $y \geq \epsilon$ 都能比传统方法（即 $\gamma \leq 1$）得到更精确的临界值；如果 $\theta_t \notin \{\theta\}_S (C = 0)$，则对所有 $y \geq \epsilon + 2\log 2$ 都成立。随着 y 增大，相对方差先线性下降；当 $y \gtrsim 2 + \epsilon$ 时，它开始像 $\gamma \lesssim \exp(-y/2)$ 一样呈指数下降。随着 y 进一步增大并趋近于 $\Delta\chi^2_{\max}$，且 $P \ll P(y_{\max} \mid \theta_t)$ 不再成立时，$B(y)P(y_{\max} \mid \theta_t)/P$ 项将占主导地位，并在 $y \geq \Delta\chi^2_{\max}$ 时趋于 $P(y_{\max} \mid \theta_t)/P = 1$。因此，对于 $y \geq \Delta\chi^2_{\max}$，改进幅度会趋于平缓，达到 $\gamma \leq \frac{1}{C+1}$，但仍至少不逊于传统 FC 方法。通过选择合适的参数点 $\{\theta\}_S$，从而选取合适的 $\Delta\chi^2_{\max}$，便可计算出所需精度的临界值。由于方差的指数级下降抵消了累积分布函数对检验统计量通常呈指数级的依赖关系（卡方分布的情形下为 $\exp(-y/2)$），因此在较宽的检验统计量取值范围内，估计的累积分布函数的相对误差近似保持平坦（图1b）。这比传统 FC 方法高效得多：随着伪实验次数增加，低置信水平会变得过于精确，而高置信水平仍会受到较大误差的影响。

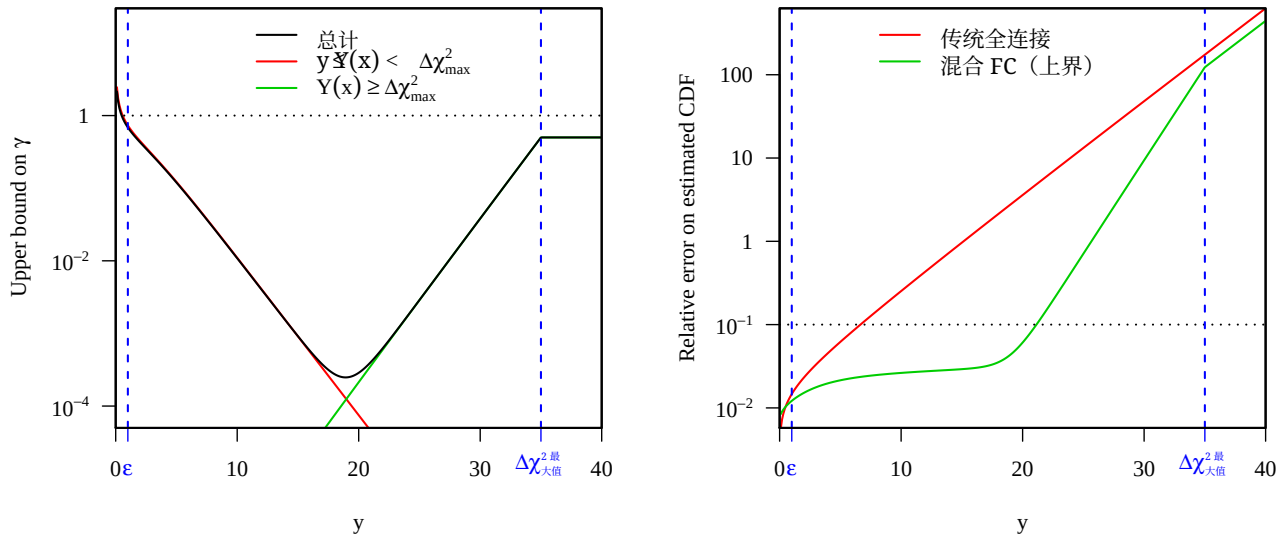
E. 插值

传统 FC 方法只能在生成伪实验时所用的参数值 θ_s 处计算临界值，而混合 FC 方法只需确保目标参数值 θ_t 足够接近采样点 $\{\theta\}_S$，且被这些采样点包围，从而满足条件式（11）。考虑到在典型设置中，需要生成的模拟实验与传统 FC 方法所用的相同，这意味着混合 FC 方法不仅降低了采样点 $\{\theta\}_S$处临界值的不确定性，还能以相近的性能对这些采样点之间的临界值进行插值。

F. 诊断和误差估计

由于混合 FC 方法利用了 $\Delta\chi^2$ 统计量与伪实验采样概率之间的关系，因此 $\Delta\chi^2$ 的计算必须与生成伪实验所用的过程相匹配。例如，不能在使用近似方法（如用于 $\Delta\chi^2$的 Pearson χ^2）的同时，从泊松随机数生成器中采样。一个简单的诊断方法是计算 $\{x\}_{\mathrm{mix}}$中所有伪实验的平均权重，并检查该值在统计涨落范围内是否等于 1。由于所有目标参数值 θ_t（权重是这些值的函数）都将使用相同的伪实验，因此不同 θ_t 值对应的这些平均权重的统计涨落会相互关联。

为了估计所计算临界值的误差，我们建议使用非参数自助法 [4] 或刀切法 [5] 等重采样方法，而不是二项误差等简单方法，从而不仅能够捕捉



(a) γ的上界

(b) 估计的累积分布函数的相对误差

图1. (a) 根据公式（36），估计的临界值方差比值上界的函数形状示例。红线表示来自具有 $y \le Y(x) < \Delta\chi^2_{\max}$ （含 $A(y)$ 的第一项）的伪实验的误差贡献；该贡献使总不确定性呈指数下降，直到来自具有 $y \ge \Delta\chi^2_{\max}$ （含 $B(y)$ 的第二项，以绿线表示）的伪实验的贡献在置信水平极高的临界值处占据主导。(b) 假设每个采样值 θ_s 对应 $n_{\text{exp}} = 10,000$ 个伪实验时，计算得到的累积分布函数估计量 $\hat{P}(y \mid \theta_t)$ 的相对误差。虚线表示 10% 的误差参考阈值。两幅图均基于以下假设构建： $\epsilon = 1$, $S = 10$, $C = 1$, $\Delta\chi^2_{\max} = 35$ Y x 1 ，并假设真实的 ( )-累积分布函数服从卡方分布，且自由度为 。在 (a) 中，绿线的指数增长因子取决于所假设的累积分布函数；而红线的衰减因子由公式（36）给出。

落在某个 $\Delta\chi^2$ 值范围内的伪实验数量的统计涨落，还能捕捉其权重的统计涨落。

IV. 单周期参数示例

我们考虑一个简单示例，该示例采用分箱泊松模型，其灵感来自在长基线中微子振荡实验中搜索 CP 破坏，具体而言是 T2K实验 [6]。该模型具有一个称为“CP 破坏相位”的单一角参数 $\delta_{\text{CP}} \in [-\pi, \pi]$ ，它受以下泊松分布观测值 $n_b \sim$Poisson(λ_b) 的约束： $B = 10$ ，其预测事件率为

$$\lambda_b(\delta_{\text{CP}}) := 10 \times (1 - \phi_b^2) \times \left(1 - \frac{1}{4}\sin(\delta_{\text{CP}} + \phi_b)\right) \tag{39}$$

$$\phi_b := \frac{b - 5.5}{10} \tag{40}$$

对于索引为 $b = 1$、2、 $\cdots$、10 的每个区间（图 2）。该模型的主要特点是，我们主要通过约 100 个观测值的总体归一化（$\sum_b n_b$）对 $\sin\delta_{\text{CP}}$ 敏感，而通过观测值随 b 变化的“形状”（旨在表示中微子能量逐渐增加的区间）对 $\cos\delta_{\text{CP}}$ 分量的敏感度较低。偏离威尔克斯定理的原因包括：$\sin\delta_{\text{CP}}$ 在 ±1 处存在物理边界（导致 $\sin\delta_{\text{CP}} = \pm1$ 附近的临界值降低）； $\cos\ \delta_{\text{CP}}$ 的符号实际上构成一个离散自由度（导致某些 $\sin\delta_{\text{CP}} \neq \pm1$ 取值处的临界值升高）；以及观测值具有泊松分布特性。在实际实验中，还会有其他与 δ_{CP} 简并的连续和离散物理参数，以及作为干扰参数处理的各种系统不确定性。不过，为了简洁清晰，我们将重点讨论

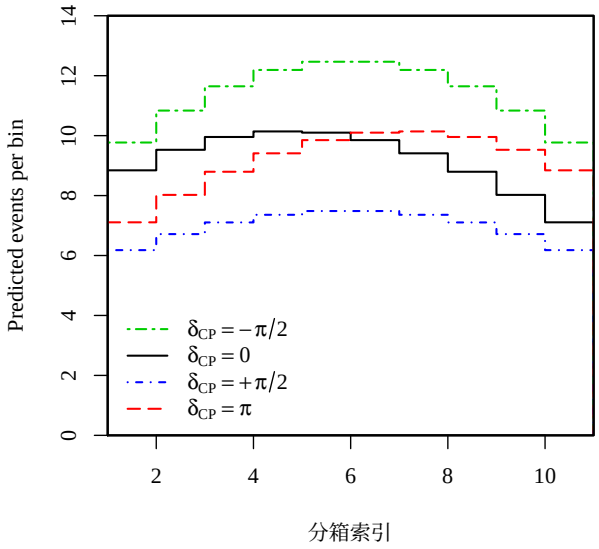


图 2. 示例中每个分箱索引 b 对应的预期事件数 $\lambda_b(\delta_{\mathrm{CP}})$。

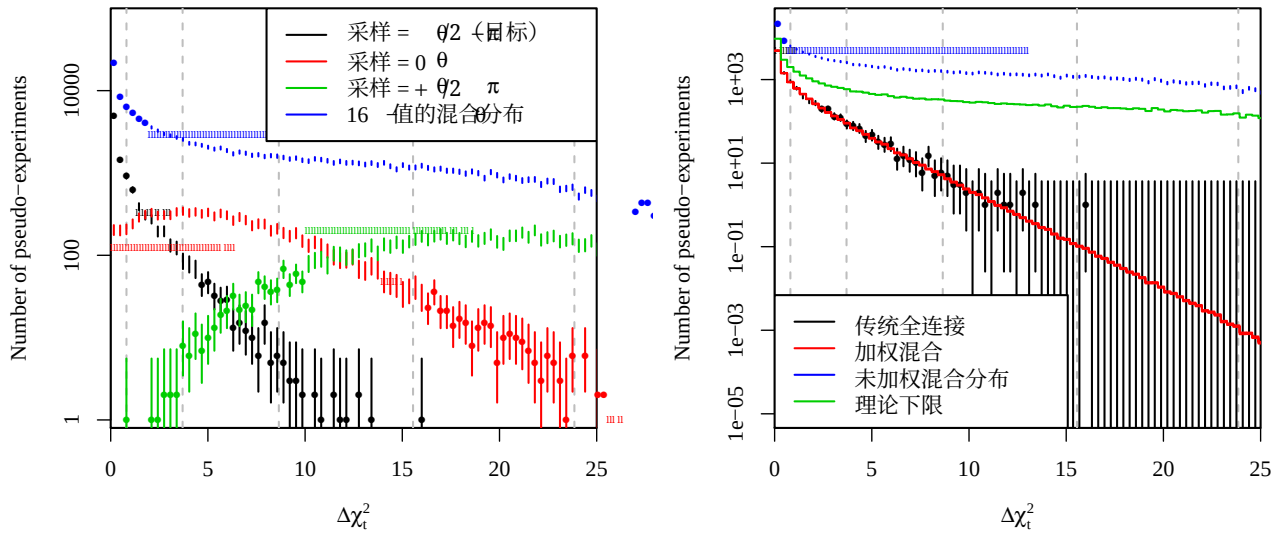
仅考虑 δ_{CP}；为与前文保持一致，以下将其称为 $\theta = (\delta_{\mathrm{CP}})$，并将观测值称为 $x = (n_1、\ n_2、\ \cdots、\ n_{10})$。

我们在参数范围 $[-\pi, \pi]$内均匀分布的 θ 的每个 $S = 16$ 个取值处，生成 $n_{\mathrm{exp}} = 10{,}000$ 个伪实验。我们首先关注 $\theta_t = -\pi/2$ 的目标值。图 3a 显示了从不同 θ_s 取值抽样得到的伪实验 x 中 $\Delta\chi^2(\theta_t \mid x)$ 的分布。在传统 FC 方法中，仅使用在$\theta_s = \theta_t$ 处生成的伪实验，这些伪实验对应于黑色直方图；对于较大的 $\Delta\chi^2(\theta_t \mid x)$，该直方图迅速下降。在混合-FC 方法中，我们还利用了在其他所有 θ_s 取值处生成的伪实验，其中 $\theta_s = 0$ 和另一极端的 $\theta_s = \pi/2$ 分别由红色和绿色直方图表示。显然，从偏移后的 θ_s 取值抽样得到的伪实验中，较大的 $\Delta\chi^2(\theta_t \mid x)$ 值所占比例显著更高。同时，可以看出仅使用在 $\theta_s = \pi/2$ 处生成的伪实验会产生一个问题：对于 $\Delta\chi^2(\theta_t \mid x)$ 较小的区域，$\theta_s = \pi/2$ 的抽样概率非常低，因此需要赋予很大的权重。然而，蓝色直方图所示的、由所有 16 θ 取值处生成的伪实验构成的混合分布，能够为所有 $\Delta\chi^2(\theta_t \mid x)$ 值生成更多伪实验；与黑色目标直方图相比，其斜率的差异表明，较大的 $\Delta\chi^2(\theta_t \mid x)$ 值对应的伪实验数量呈指数增长。图 3b 更清楚地显示了这一点，其中混合分布使用指定的权重进行了重新加权。可以看到，该分布与传统 FC 方法模拟得到的目标分布吻合良好，而且混合-FC 方法中未加权伪实验的总数超过了理论下界。

现在我们检查混合-FC 方法的一些诊断结果。图 4a 显示了重要性抽样权重$w(x \mid \theta_t)$的分布；结果发现，这些权重主要是 $\Delta\chi^2(\theta_t \mid x)$的函数，只有较小的额外方差。权重被发现完全处于式 (16) 给出的理论界内；绘制这些理论界时，通过查看图 4b 中的 $\Delta\chi^2(\hat{\theta}_S(x) \mid x)$分布，假定了一个$\epsilon = 0.3$ 值。权重之和与 1 相符（图5）。

接下来，我们考察临界值。图 6 展示了使用标准 FC 方法（黑色误差棒）和混合-FC 方法得到的临界值随（真实/目标）参数值θ 变化的情况。尽管使用了同一组伪实验，混合-FC 方法得到的临界值不确定性显著更小，尤其是在较高置信水平下；此外，该方法还能揭示θ的 16 个采样值之间函数形状的细节。

对于 1σ 临界值（图 6b），我们看到，尽管采样值的间距相对较小，但如 $\theta = \pm\pi/2$ 值旁边红色与灰色误差带未重叠所示，插值误差仍大于传统方法中二项误差带的宽度。由于这些二项误差带无法体现插值误差，其较小的宽度可能造成误导，因此混合 FC 方法的插值功能非常有用。对于 2σ （图 . 6c）和 3σ 临界值（图 . 6d）我们看到两种方法之间具有良好的一致性，同时也



(a) $\Delta\chi^2_t$ 在不同的 θ_s

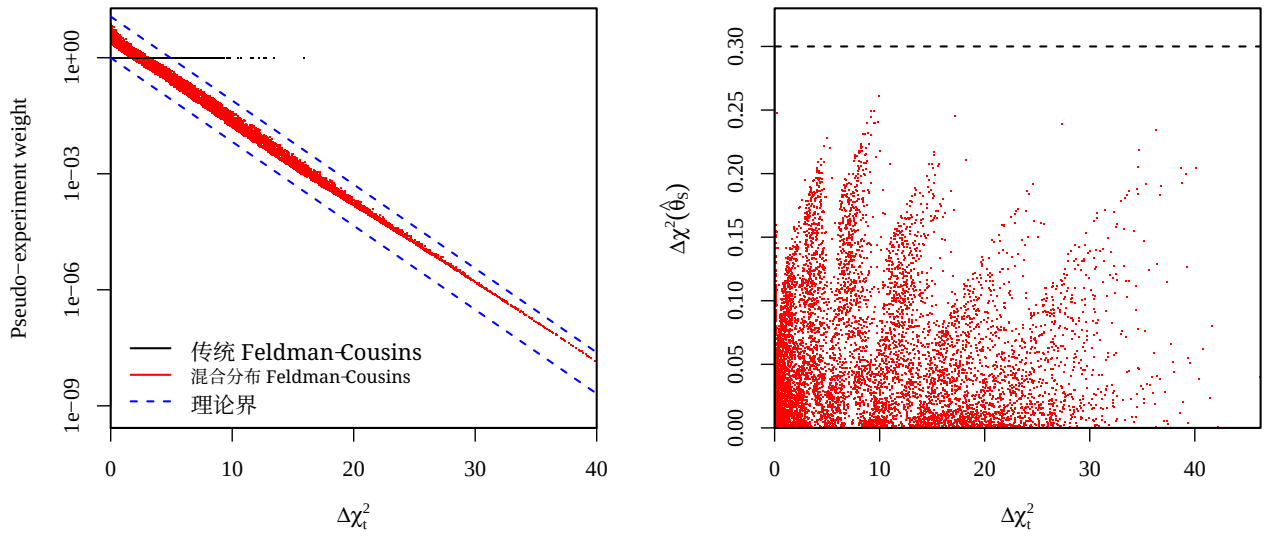
(b) 在 θ_t 处估计的 $\Delta\chi^2_t$ 分布

图 3. 目标为 $\theta_t = -\pi/2$ 的 $\Delta\chi^2_t := \Delta\chi^2(\theta_t \mid x)$ 分布：(a) 不同抽样参数值 θ_s 下的分布；(b) 使用标准 FC 方法和混合 FC 方法得到的 θ_t 处估计分布对比。在两幅图中，误差棒表示 1σ 二项置信区间。竖直虚线表示使用混合 FC 方法得到的 1，2，3，4，5σ 置信水平临界值。(a) 为清晰起见，省略了计数为零的分箱的误差棒。(b) 红色“加权混合”直方图还绘制了方框，表示每个分箱中伪实验数量及其权重方差所带来的误差；但这些误差小于线宽，因此不可见。混合分布中伪实验总数的“理论下限”，是将式（14）中的下界乘以红色“加权混合”直方图得到的，其中假设 $C = 1$ 和 $\epsilon = 0.3$。

注意到混合-FC 计算中的误差显著更小。对于 3σ 置信水平（图 6d），我们看到传统方法中的误差已经大到使临界值的某些特征难以辨认，例如 $\theta = 0, \pi$ 处的凸起，以及由泊松统计导致的临界值在 $\sin\delta_{\mathrm{CP}}$ 符号翻转时呈现的不对称性。

对于 4σ 及更高的置信水平，传统 FC 方法除了下限外，无法确定临界值。另一方面，混合 FC 方法仍能给出临界值，其相对误差大小与较低置信水平下的临界值相近。

估计得到的相对误差绘于图7，其形状与理论论证所预期的典型形状一致（图1b）。为了根据公式（36）中的 γ 绘制上界，我们基于图4b作出保守假设，即仅假设 $\Delta\chi^2(\hat{\theta}_S(x) \mid x) \leq \epsilon$，直至 $\Delta\chi^2(\theta_t \mid x) \leq 32$。在此例中，对于 $\Delta\chi^2_t < 16$，通过自助法估计的实际混合 FC 误差比 γ 给出的理论上限约小一半。这可以解释为，每个伪实验的采样有多个采样值 θ_s 参与，而理论上限的假设则是只有 $\hat{\theta}_S(x)$ 会参与。另一方面，对于 $\Delta\chi^2_t > 16 = \Delta\chi^2_{\max}/2$，理论上限开始显著增大，而通过自助法估计的实际误差仅缓慢增长。这可以解释为，我们对 $\Delta\chi^2_{\max} = 32$ 的选择过于保守：在当前示例中，所选抽样网格 $\{\theta\}_S$ 似乎在显著更高的 $\Delta\chi^2$ 值下仍然有效。这部分是因为参数 $\theta = (\delta_{\mathrm{CP}})$ 的参数空间 $\delta_{\mathrm{CP}} \in [-\pi, \pi]$有界，情况较为理想。



(a) 权重分布

(b) $\Delta\chi^2$的分布($\hat{\theta}_S \mid x$)

图4. (a) 权重 $w(x \mid \theta_t)$ 和 (b) 每次伪实验中最小 $\Delta\chi^2$ 值随 $\Delta\chi^2_t := \Delta\chi^2(\theta_t \mid x)$ 变化的分布。(a) 蓝色虚线表示假设 $\epsilon = 0.3$ 和 $C = 0$ 时的理论极限。(b) 虚线表示 $\epsilon = 0.3$。

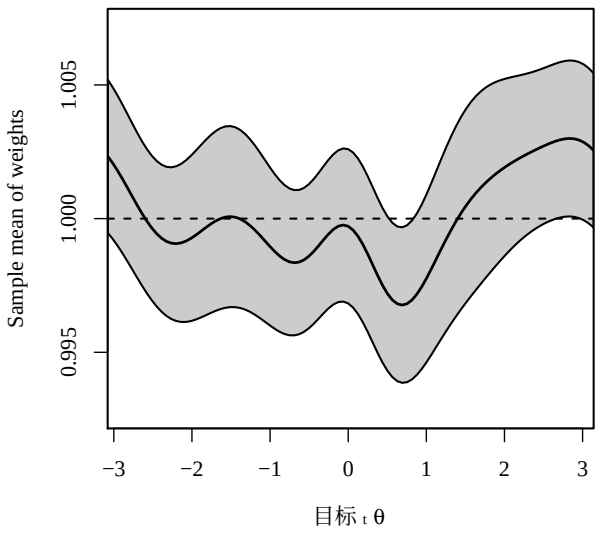
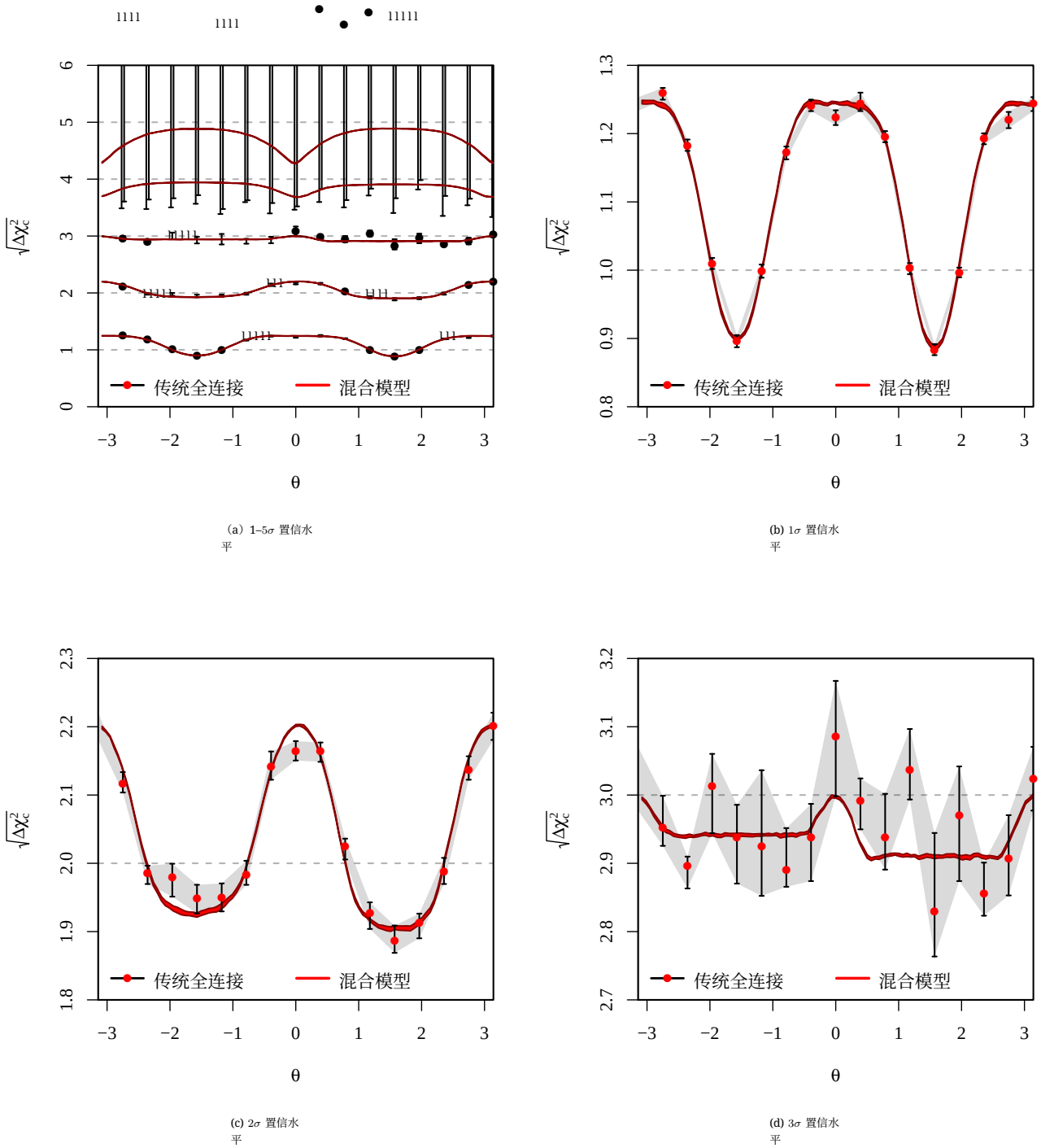


图5. 混合-FC 伪实验权重 $w(x \mid \theta_t)$ 的样本均值随目标 θ_t 值的变化。误差带表示均值的 1σ 标准误，不同目标 θ_t 值之间存在相关性。



(a) 1–5σ 置信水平

(b) 1σ 置信水平

(c) 2σ 置信水平

(d) 3σ 置信水平

图 6. 1，2，3，4，5σ 置信水平临界值（以 $\sqrt{\Delta\chi^2_c}$ 给出），由同一组伪实验分别使用标准 FC（黑色误差棒）和混合-FC 方法（红色误差带）得到。两种情况下，误差棒/误差带均表示使用二项分布误差/自助法误差分别计算标准/混合 FC 方法所得临界值的 1σ 误差。虚线表示根据威尔克斯定理得到的临界值；该定理在此处并不适用，但仍绘出以供参考。(a) 为避免重叠，4 和 5σ 对应的黑色误差棒略微错开。(b,c,d) 灰色误差带表示标准 FC 方法中误差棒端点的线性插值。

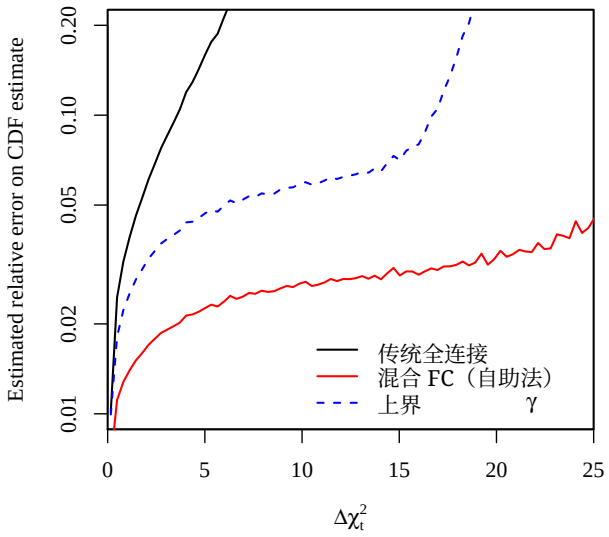


图 7. 目标为 $\theta_t = -\pi/2$ 时，累积分布函数估计量 $\hat{P}(y \mid \theta_t)$ 的相对误差估计。对于标准 FC，图中显示了二项分布的标准误（黑色实线）；计算这些误差时使用了混合-FC 方法得到的更精确的累积分布函数估计。对于混合 FC 方法，自助法误差估计（红色实线）远低于公式（36）的理论上限；该上限是在假设 $\epsilon = 0.3$ 和 $\Delta\chi^2_{\max} = 32$ 的条件下计算得到的（蓝色虚线）。

V. 讨论

A. 与统计力学技术的关系

所提出的方法在理念上与统计力学中的“多重直方图重加权”（“多直方图”）方法相似 [7]：该方法针对不同参数值模拟统计系综，并通过重新加权将其组合到所需的参数值。在多直方图方法中，组合各系综时还会赋予额外的系综权重，并调整这些权重以尽可能减小待估变量的总体误差。所提出的混合-FC 方法也可以采用类似的系综权重，而且这些额外权重还可以依赖于目标 $\Delta\chi^2_t := \Delta\chi^2(\theta_t \mid x)$值，从而尽可能降低临界值估计量的方差。

然而，与多直方图方法的一个区别是，我们不使用马尔可夫链蒙特卡洛技术对伪实验进行采样，因此每个参数值 θ 处伪实验的抽样分布 x （包括归一化常数）都是精确已知的。因此，混合-FC 方法不需要像多直方图方法那样，在最后通过迭代过程自洽地确定这些归一化常数（自由能）。

B. 与边缘分布的关系

由若干参数值 $\{\theta\}_S$ 构成的混合抽样分布，可以视为先验 $\pi(\theta) = \frac{1}{S}\sum_{s=1}^{S} \delta(\theta - \theta_s)$ 下的边缘概率分布，其中 $\delta(\cdot)$ 是狄拉克 δ 函数。正如上一段所述，额外的系综权重对应于另一种先验 $\pi(\theta \mid \Delta\chi^2_t) = \sum_{s=1}^{S} r_s(\Delta\chi^2_t)\ \delta(\theta - \theta_s)$；其中 $r_s(\Delta\chi^2_t)$ 可以在满足条件 $\sum_s r_s(\Delta\chi^2_t) = 1$（对所有 $\Delta\chi^2_t$ 均成立）的前提下进行优化，以减少误差。我们甚至可以将讨论推广到连续先验分布 $\pi(\theta \mid \Delta\chi^2_t)$；为了保留关于效率降低的论证，我们需要将公式 11 中的单点条件扩展为对 $\pi(\theta \mid \Delta\chi^2_t)$ 上一个有限大小区域的条件。

与传统 FC 方法不同，传统 FC 方法需要在每个目标参数值处进行大量伪实验；而在混合 FC 方法中，每次采样时减少伪实验数量，

同时增加所考虑的采样点数量 S 可能更可取。若固定 $Sn_{\rm exp}$，则通过降低权重的方差（其上界为 $\exp(\epsilon/2)$），可减小临界值的方差。

根据这一与边缘分布的关系，现在我们来考虑 $\Delta\chi^2(x \mid \theta_t) = -2\log L(x \mid \hat{\theta}_t)/L(x \mid \theta(x))$ 的计算，并将其近似为 $-2\log L(x \mid \theta_t)/L_m(x)$；在分母中，剖面化操作被对 θ 进行边缘化所取代，并使用了关于 θ的某个先验。因此，我们得到一个简单的似然比检验，比较 $p(x \mid \theta_t)$ 和 $p_m(x) := \int \mathrm{d}\theta\, \pi(\theta) p(x \mid \theta)$。现在显而易见，为了在原假设 $p(x \mid \theta_t)$ 下高效生成具有较小 p 值的伪实验，只需从备择假设 $p_m(x)$ 生成伪实验，这正是混合-FC 方法所采用的做法。

实际上，使用文中讨论过的关于 $\{\theta\}_S$ 的离散“先验”会更容易，因为除非似然是高斯的，否则边缘化所需的数值积分通常会增加计算成本和复杂度。尽管如此，剖面化与边缘化之间的相似关系仍可用于为 $\{\theta\}_S$ 值选择理想的间距。在众所周知的客观先验中，杰弗里斯先验 [8] 可产生一种先验：若存在某种参数化，使似然呈高斯形式，那么在该参数化下此先验将是均匀的。由于对高斯似然使用均匀先验进行剖面化和边缘化，除相差一个常数偏移量外，会得到相同结果，因此杰弗里斯先验可视为选择生成伪实验的 $\{\theta\}_S$ 值的理想候选。例如，在前文讨论过的 CP 破坏分析中，更适合在 $\sin\delta_{\rm CP}$ 而非 $\delta_{\rm CP}$ 上均匀选择参数值，并使 $\cos\delta_{\rm CP}$ 的正负号具有相同概率，因为主要约束来自事件总数 $N \sim$泊松分布$(\lambda = A + B\sin\delta_{\rm CP})$，其中 A 和 B 为常数，因此 $\sin\delta_{\rm CP}$ 上的似然近似呈高斯分布。

C. 干扰参数

由于混合-FC 方法显著降低误差的优势，利用了 $\Delta\chi^2(\theta_t \mid x)$ 统计量与生成伪实验的分布之间的特定关系，因此不能假定所有特性都能直接应用于包含干扰参数（在物理学中通常称为“系统参数”）的分析。尤其是对于常用的剖面 FC [9] 或后验 Highland-Cousins 方法 [10], ，用于生成伪实验的干扰参数空间会根据观测实验数据施加的约束而大幅缩小；在这种情况下，直接应用混合 FC 方法可能无法像式 36 所示那样，指数级减少估计临界值时的误差。因此，不应依赖这些结果来估算所需的伪实验次数。

在相对一般的情形下，如果目标分布本身就是混合分布的一部分（即 $C = 1$），可以证明，即使在最坏情况下，临界值的方差与传统方法相比也只会略微增加，增幅为 $1/(1 - P(y))$（见附录）。考虑到当置信水平较高时，$P(y) \ll 1$，这一增幅可以忽略不计。权重的上限也受到类似界限的约束，这对于确保重要性抽样行为定义良好十分重要。因此，仍值得尝试将混合 FC 方法直接应用于 Feldman-Cousins 置信区间。事实上，在某些情况下，误差可能会近乎呈指数级减少，如同不存在干扰参数时一样；但由于缺乏理论保证，建议仔细研究权重分布，以及自助法误差估计在这种情况下的可靠性。

由于在存在干扰参数的情况下无法保证误差呈指数级减少，因此在这种情况下，插值估计临界值的能力会更有意义。在这里，重要的是，在相邻的 θ_s 值之间（以及合适的干扰参数取值下）生成的伪实验，在伪实验空间中有足够的重叠。否则，伪实验生成过程与检验统计量背后的统计模型之间的不匹配，可能很快导致权重值的离散程度很大，使估计的临界值及其误差估计都不可靠。这是因为存在干扰参数时，即使检验统计量的取值相近，伪实验仍可能在更多维度上有所不同。

然而，在一种特定情形下，尽管存在干扰参数，前文讨论的所有性质仍可直接适用。这种情形是将先验 Highland-Cousins 方法与边缘化-$\Delta\chi^2$ 统计量结合使用，其中关键是两种情况下都对干扰参数使用相同的先验分布 $\pi(\eta)$ η 。这是因为在此情形下，干扰参数的影响完全被生成伪实验所用的概率模型吸收了，其含义是 $p(x \mid \theta) = \int \mathrm{d}\eta\, \pi(\eta)\ p(x \mid \theta,\ \eta)$；因此，就混合-FC 方法而言，不存在干扰参数。

关于伪实验抽样分布的更多详细讨论、示例及可能的修改，将在另一篇出版物中讨论。

D. 与类似统计推断技术的关系

非常相似的重要性抽样技术已用于在具有似然比统计量的原假设下计算 p 值。例如，Woodroofe [11] 讨论了对感兴趣参数采用连续先验的情形。用我们的记法，

$$p_{\text{sample}}(x) = \int \mathrm{d}\theta\, \pi(\theta) p(x \mid \theta) \tag{41}$$

只给出了权重的下界

$$w(x \mid \theta_t) := \frac{p(x \mid \theta_t)}{p_{\text{sample}}(x)} \geq \frac{p(x \mid \theta_t)}{p(x \mid \hat{\theta}(x))} = \exp\left[-\tfrac{1}{2}\Delta\chi^2(\theta_t \mid x)\right] \tag{42}$$

而非给出一个上界；后者对于证明估计的 p 值的误差较小至关重要。此外，还给出了使用鞍点法推导的权重渐近公式。

文献 [12]（第 5.6 节）介绍了一种在 ATLAS实验 搜寻希格斯玻色子过程中开发的方法 [13]。他们指出，在 Woodroofe 的方法中，对连续先验进行积分存在困难，因此改用一组离散点 $\{\theta\}_S$，其中包括 θ_t；我们在混合 FC 方法中也采用了这种做法（使用 $C = 1$）。不过，权重函数的选择有所不同：只有当

$$\omega_s(x \mid \theta_t) := \frac{p(x \mid \theta_t)}{p(x \mid \theta_s)} = \exp\left[-\tfrac{1}{2}\left(\Delta\chi^2(\theta_t \mid x) - \Delta\chi^2(\theta_s \mid x)\right)\right] \tag{43}$$

生成伪实验时所依据的参数值 θ_s 处的值，是 $\{\theta\}_S$ 中所有其他值里最小的，才会使用该伪实验——即 $\omega_s(x \mid \theta_t) = \min_{s'} \omega_{s'}(x \mid \theta_t)$ ——否则就将其舍弃。如果使用该伪实验，则赋予其权重 $\omega_s(x \mid \theta_t)$。随后，将从所有 $\{\theta\}_S$ 值抽样得到的伪实验及其权重组合起来，便可得到所需的分布 $p(x \mid \theta_t)$，并提高抽到大 $\Delta\chi^2(\theta_t \mid x)$ 伪实验的概率。由于 $\theta_t \in \{\theta\}_S$，对于表现良好的权重，此过程可确保 $w(x \mid \theta_t) \leq 1$。

作者指出，这种否决技术的一个缺点是，$\{\theta\}_S$ 的间距不能过于密集，否则在否决概率较高时会降低该方法的效率。混合-FC 方法不存在这个问题，因为权重是根据实际抽样概率计算的，而实际抽样概率是对 $\{\theta\}_S$ 上的概率求和所得，因此无需进行否决。虽然否决技术声称的优点是无需精确归一化抽样概率分布——因为只使用概率比值 ω_s ——但混合-FC 方法中权重的选取也具有这一优点，其式（2）中的权重可以写为

$$w(x \mid \theta_s) = \frac{1}{\frac{1}{S}\sum_{s=1}^{S}\left[\omega_s(x \mid \theta_t)\right]^{-1}}. \tag{44}$$

因此，对于利用似然比统计量在原假设下求取 p 值的问题，混合-FC 方法的相关部分可视为对参考文献 [12] 所述方法的微小改进。此外，我们明确证明，在适当条件下（对于典型设置，这要求不存在干扰参数），当检验统计量取较大值时，估计的 p 值的方差会呈指数级降低。

最后，我们指出，在重要性抽样的背景下，计算 p-值与 FC 置信区间构造之间的一些差异。计算 p-值时，我们通常关注单个 原假设下检验统计量的分布。相比之下，FC 方法需要检验统计量在所有合理参数值下的分布，实际操作中通过对有限个参数值 $\{\theta\}_S$ 进行计算并在其间插值来实现。因此，FC 构造能够受益于利用重要性抽样对临界值进行插值；而在计算 p-值时，这一点并不总是重要。此外，构造混合分布所需的、从不同参数值下抽取的伪实验，即使在传统 FC 方法中也已具备，因此过渡到混合 FC 方法十分直接。

VI. 总结

我们提出了一种计算 Feldman-Cousins置信区间临界值的新方法。该方法是对传统方法的简单扩展：只需将针对不同参数值生成的同一组伪实验按适当的权重进行组合。我们表明，这种方法能够大幅减少

临界值的误差；对于高置信水平的临界值，误差会呈指数级减少，而几乎不会增加计算成本。我们还证明，该方法能够在生成伪实验所对应的参数值之间准确插值。我们通过一个中微子振荡分析的简单示例，验证了理论计算所得的性能。虽然目前只有在不含干扰参数的分析中才能保证误差呈指数级减少，但这项通用技术适用于任何采用 Feldman-Cousins 方法的分析。

附录：通用混合分布的临界值方差分析

我们将 θ_t 处伪实验的目标分布记为 $p_t(x)$。在存在干扰参数η 且其概率分布为 $p(x \mid \theta, \eta)$ 的情形下，对于 profile-FC 方法，这可以是例如 $p\big(x \mid \theta_t, \hat{\hat{\eta}}(\theta_t \mid x_{\mathrm{obs}})\big)$；对于后验 HC 方法，则可以是 $\int \mathrm{d}\eta\, \pi(\eta \mid x_{\mathrm{obs}}, \theta_t)\; p(x \mid \theta_t, \eta)$，其中 $\hat{\hat{\eta}}(\theta \mid x_{\mathrm{obs}}) = \arg\min_\eta \chi^2(\theta, \eta \mid x_{\mathrm{obs}})$ 为轮廓最佳拟合值，而 $\pi(\eta \mid x_{\mathrm{obs}}, \theta)$ 是在目标 θ 值条件下、针对观测数据拟合所得的干扰参数后验分布 x_{obs}。其余伪实验从 $p_a(x)$ 中抽样；此处分布未明确指定，但例如可以是对不同 θ 和 η 值的混合分布。对从 $p_t(x)$ 抽样的 N_t 个伪实验和从 $p_a(x)$ 抽样的 N_a 个伪实验进行混合，并按 $w(x) = (N_t + N_a) p_t(x) / \big[N_t p_t(x) + N_a p_a(x)\big]$ 加权，其结果可参照正文并使用这些估计量进行类似评估

$$\hat{P}(y) := \frac{1}{N_t + N_a} \sum_{i=1}^{N_t + N_a} w(x_i) I\big(Y(x_i) \geq y\big) \tag{A.1}$$

$$\hat{P}_{\mathrm{conv}}(y) := \frac{1}{N_t} \sum_{i=1}^{N_t} I\big(Y(x_i^{(t)}) \geq y\big) \tag{A.2}$$

$$\mathbb{E}[\hat{P}(y)] = \mathbb{E}[\hat{P}_{\mathrm{conv}}(y)] = P(y) \tag{A.3}$$

使方差缩减达到

$$\gamma = \frac{\mathrm{Var}[\hat{P}(y)]}{\mathrm{Var}[\hat{P}_{\mathrm{conv}}(y)]} \tag{A.4}$$

$$\leq \frac{\frac{1}{N_t} P(y) - \frac{1}{N_t + N_a} P(y)^2}{\frac{1}{N_t} P(y) - \frac{1}{N_t} P(y)^2} \tag{A.5}$$

$$\leq \frac{1}{1 - P(y)} \tag{A.6}$$

其中

$$x_i \sim \frac{N_t p_t(x) + N_a p_a(x)}{N_t + N_a} \tag{A.7}$$

$$x_i^{(t)} \sim p_t(x). \tag{A.8}$$

致谢

我们谨向 Christophe Bronner 和 Louis Lyons 致谢，感谢他们进行了有益的讨论，并引荐我们认识 Kyle Cranmer；也感谢 Kyle Cranmer 向我们介绍了参考文献 [11, 12]。本研究得到了 JSPS KAKENHI 资助，项目编号为 19J22440。

---

[1] S. S. Wilks，用于检验复合假设的似然比大样本分布，《数学统计年鉴》**9**，60（1938）。[2] G. J. Feldman 和 R. D. Cousins，小信号经典统计分析的统一方法，《物理评论D辑》**57**，3873（1998），arXiv:physics/9711021。

[3] L. Li、N. Nayak、J. Bian 和 P. Baldi，《利用高斯过程进行高效的中微子振荡参数推断》，《物理评论D辑》**101**，012001（2020），arXiv:1811.07050 [physics.data-an]。[4] B. Efron，《自助法：再谈刀切法》，统计年鉴**7**，1（1979）。[5] R. G. Miller，《刀切法——综述》，Biometrika**61**，1（1974）。[6] K. Abe *et al.* (T2K)，《T2K实验》，核仪器与方法 A**659**，106（2011），arXiv:1106.1238 [physics.ins-det]。[7] A. M. Ferrenberg 和 R. H. Swendsen，《优化的蒙特卡洛分析》，物理评论快报**63**，1195（1989）。[8] H. Jeffreys，《估计问题中先验概率的一种不变形式》，伦敦皇家学会会刊：A辑，数学与物理科学**186**，453（1946），https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1946.0056。[9] M. A. Acero *et al.* (NOvA)，《存在干扰参数时用于置信区间构造的轮廓化 Feldman-Cousins 方法》，（2022），arXiv:2207.14353 [hep-ex]。[10] R. D. Cousins 和 V. L. Highland，《将系统不确定性纳入上限》，核仪器与方法 A**320**，331（1992）。[11] M. Woodroofe，《重要性抽样与错误概率》，http://www.sfu.ca/~lockhart/richard/banff2010/ woodroofe.pdf（2010），班夫发现研讨会。[12] K. Cranmer，《大型强子对撞机实用统计学》，载于 *2011 European School of High-Energy Physics*（2014），第267–308页，arXiv:1503.07622 [physics.data-an]。[13] G. Aad *et al.*(ATLAS)，《利用大型强子对撞机上的ATLAS探测器搜寻标准模型希格斯玻色子并观测到一种新粒子》，物理快报 B**716**，1（2012），arXiv:1207.7214 [hep-ex]。