arXiv:2611.29512v1 [stat.CO] 22 Nov 2026

精确、窄且快速：使用 Gaffke 离散有界变量均值的置信区间

Louis Abraham

摘要

调查评分、测试分数和许多其他数据都只取若干个已知值，我们常常需要为其均值计算置信区间。对于小样本，或均值接近最小值或最大值的情况，常用区间覆盖真实均值的频率远低于其声称的水平。Gaffke 区间将 Clopper–Pearson区间推广到任意有限值集合；近期对Gaffke猜想的证明表明，对于任何样本量和任何分布，其覆盖率都至少达到名义水平。然而，该区间的端点是狄利克雷变量加权和的分位数，且没有闭式形式。

本文有三项贡献。

- 一种快速且精确的端点算法。沿鞍点所在直线进行傅里叶反演积分，可在相对误差较小的情况下求得尾概率；在几何分段上应用高斯–勒让德求积公式，只要奇点相同，收敛速度也相同；对尾部末端的对数距离进行牛顿迭代，平均每个端点需要 2.6 次计算。对于 k 个值和 N 个求积节点，计算成本为 $O(kN)$ 次运算，且 N 不随样本量增长：无论 n 如何，计算一个区间都不到一毫秒。
- 与已发表区间的实验比较。 对于 $n \leq 30$，通过枚举所有结果计算覆盖率时，Gaffke 区间是在保持置信水平的区间中最短的：比Waudby-Smith和Ramdas的投注区间短17%至20%，比Anderson的区间短19%至37%。较短的区间（Wald、Student、轮廓似然、贝叶斯）在某些分布下对均值的覆盖率低于50%。
- 经过测试的*Python*和*R*软件包，以及一个网页。 这些端点与精确参考值（有理数运算中的B样条公式、高精度积分）一致，覆盖值域的 10^{-12}，尾概率的相对误差低于 10^{-10}，最低可达 $\alpha = 10^{-12}$。

因此，该区间精确、窄且快速，我们推荐使用它。

1 引言

问题。 一组包含 n 次试验的样本，每次试验的结果为 k 个已知值中的一个，其未知均值为 μ。典型例子包括1至5量表上的评分、分数，以及已知类别大小的各类项目数量。用于 μ 的常规区间（Wald、Student）基于正态近似；当 n较小时，或当 μ 接近值域端点时，这种近似会失效。无分布假设区间（Hoeffding、经验伯恩斯坦、

Anderson，投注）能维持置信水平，但区间较宽。

已知内容。Gaffke [5]提出了非负变量均值检验，并猜想该检验是有效的。Learned-Miller和Thomas [9]研究了该检验为有界数据给出的置信区间；对于伯努利数据，该区间就是Clopper–Pearson区间。Vlassis 和 Thomas [14]

最近证明了这一猜想，因此该区间具有有保证的覆盖率；Ming 等人 [11] 表明，其宽度渐近等于正态区间的宽度。

贡献。

1. 一种算法，可在 O 次（kN）运算中计算 10^{-12} 个数值的值域端点，且求积节点数量 N 不随 n 增长（第5节）。其关键在于鞍点直线，它使尾概率的相对误差较小；以及几何分段，在这些分段上，高斯–勒让德求积公式的收敛速度不受数据影响。2. 在 $n \le 30$ 个随机问题中，将 Gaffke 区间的覆盖率和宽度与八种已发表的区间进行实验比较（第4节）。3. 提供 Python 和 R 软件包，二者均名为 gaffke，并通过精确参照值和 40,000 个随机问题进行测试（第6节）；此外还提供一个可在浏览器中计算该区间的网页。

Sec第2节定义该区间，第3节回顾其覆盖率为何能够得到保证 .

2 区间

数据。每次 n 次独立试验都会得到 k 个已知值 $v_1, \dots, v_k$ 中的一个，其概率 $p_1, \dots, p_k$ 未知。计数服从 $X \sim \mathrm{Mult}(n, p)$，我们希望为均值构造一个置信区间

$$\mu = v^\top p = \sum_i v_i p_i.$$

我们将 $v_{\min}$ 和 $v_{\max}$ 分别记为最小值和最大值，并将 e_i 记为第 i个单位向量。

Gaffke 检验。设 $Z_1, \dots, Z_n$ 为相互独立的非负变量。为检验它们的均值是否均不超过 1，Gaffke [5] 在样本中加入一个零值，并使用从均匀狄利克雷分布 $\mathrm{Dir}(1, \dots, 1)$中抽取的随机权重 $D = (D_0, D_1, \dots, D_n)$ 对这 $n+1$ 个值求加权平均。当该随机平均值不超过 1 的情况很少见时，检验以 α 的显著性水平拒绝原假设：

$$\text{reject when} \quad G(z) = \mathbb{P}_D\Big\{\sum_{i=1}^{n} z_i D_i \le 1\Big\} \le \alpha.$$

Gaffke 猜想该检验是有效的。Vlassis 和 Thomas [14] 最近证明了这一点：如果 Z_i 相互独立、非负且 $\mathbb{E}Z_i \le 1$，那么对于每个 α，$\mathbb{P}\{G(Z) \le \alpha\} \le \alpha$ 。

从检验到区间。对于候选均值 $m > v_{\min}$，将检验应用于 $Z_i = (v_{I_i} - v_{\min})/(m - v_{\min})$，其中 I_i 是试验 i所属的类别。同一类别中的试验具有相同的值，因此它们的狄利克雷权重会相加；权重 D_0 归于 $v_{\min}$。因此，当且仅当 m 高于 $v^\top P_-$的 $\alpha/2$分位数时，检验才不会在 $\alpha/2$ 的显著性水平下拒绝 m；其中 P_- 服从以各类别计数为参数、并在一个最小值上额外增加一个单位的狄利克雷分布。对$v_{\max} - v$ 采用相同的论证即可得到上端点。置信水平为 $1 - \alpha$ 的区间为

$$L = Q_{\alpha/2}\big(v^\top P_-\big), \quad U = Q_{1-\alpha/2}\big(v^\top P_+\big), \qquad P_- \sim \mathrm{Dir}(X + e_{i_{\min}}), \quad P_+ \sim \mathrm{Dir}(X + e_{i_{\max}}),$$

其中 Q_q 是 分位数函数，且 $i_{\min}$、$i_{\max}$ 分别是最小值和最大值的索引 .

3 精确

此处 精确 一词的含义与 Clopper–Pearson区间中的含义相同：无论样本量是多少，覆盖率都有保证，而不仅仅是在极限情况下。根据 Vlassis 和 Thomas 的结果，对于每个 p 和每个 n，该区间以至少 $1-\alpha$ 的概率覆盖 μ。在这一证明之前，Learned-Miller 和 Thomas [9] 已证明该结论适用于伯努利数据和半伯努利数据。该区间是 Clopper–Pearson区间 [2] 的推广：对于取值 $v = (0,\ 1)$ 且在 n 次试验中有 x 次成功的情形，$v^\top P_- \sim \text{Beta}(x,\ n-x+1)$ 和 $v^\top P_+ \sim \text{Beta}(x+1,\ n-x)$，因此

$$L = I^{-1}_{\alpha/2}(x,\, n-x+1), \qquad U = I^{-1}_{1-\alpha/2}(x+1,\, n-x),$$

即 Clopper–Pearson界（I^{-1} 是正则化不完全Beta函数的逆函数）。在我们的实验中（第6节），一些单侧覆盖率等于 $1-\alpha/2$，因此一般来说无法改进该界。

4 窄

任何区间都可以通过将其设得足够宽而变得有效；问题在于，一个有效区间必须有多宽。表1给出了三种取值 $v = (0, 1\,2, 1)$ 和 $\alpha = 0.05$下已发表区间的覆盖率：在单纯形的一个包含边界的3321个点的格点 p 上取最小值，以及在该格点上对期望宽度取均值。覆盖率通过枚举所有结果计算得出。线以上的区间保持了水平（投注区间的值0.946来自图注中给出的近似值）：其中Gaffke的区间最短；与Waudby-Smith和Ramdas的投注区间相比短17%至20% [13]，与Anderson的区间相比短19%至37% [1]。线以下的区间短5%至30%，但对于某些 p，其覆盖率远低于95%。对于取值 $(0, 0.1, 1)$以及五种取值，结果相同。

对于较大的 n，该区间与正态区间一样窄：$\sqrt{n}$ 其宽度与 $2\sigma z_{1-\alpha/2}$的比值几乎必然趋于 [11]。对于每一种结果，它并不是最短的有效区间：Ming 等人构造了一个在某些情况下严格更短的有效区间，但在我们的比较中，没有已发表的区间比它更短。

5 快速

端点没有闭式形式，而像 $1-10^{-6}$ 这样的置信水平需要以较小的相对误差计算 5×10^{-7}的尾概率。本节介绍该方法（第 5.1–5.4 节）及其计算成本（第 5.5 节）。每个端点都是以下分布的一个分位数：

$$Y = w^\top P, \qquad P \sim \mathrm{Dir}(a), \qquad (w, a) = (v,\ X + e_{i_{\min}}) \text{ or } (v,\ X + e_{i_{\max}}).$$

5.1 约化

删除带有 $a_i = 0$ 的类别，合并相同的值，并将这些值缩放至 $[0, 1]$:

$$a'_j = \sum_{i:w_i = u_j} a_i, \qquad w'_j = \frac{u_j - w_{\min}}{w_{\max} - w_{\min}},$$

	最小覆盖率			平均宽度		
区间	$n=5$	10	30	5	10	30
Gaffke	0.976	0.963	0.954	0.648	0.472	0.268
投注*	0.991	0.982	0.946	0.797	0.592	0.323
Anderson	0.991	0.992	0.991	0.802	0.648	0.426
Hoeffding	0.991	0.992	0.991	0.878	0.728	0.470
经验伯恩斯坦	1.000	1.000	1.000	1.000	1.000	0.847
Student t	0.061	0.118	0.314	0.619	0.451	0.251
Wald	0.061	0.118	0.314	0.454	0.378	0.237
轮廓似然	0.446	0.477	0.493	0.474	0.377	0.236
贝叶斯，Jeffreys先验	0.000	0.000	0.000	0.490	0.377	0.234

表1：覆盖率和期望宽度的均值，$v=(0,\tfrac{1}{2},1)$，$\alpha=0.05$（名义覆盖率 0.95）。Wald：均值 $\pm z\hat{\sigma}/\sqrt{n}$；Hoeffding [7]；经验伯恩斯坦 [10]；Anderson：经验分布函数上的单侧 DKW带 [1]；使用 χ^2_1 截断值的轮廓似然；贝叶斯：先验为 Dir($\tfrac{1}{2},\ldots\ldots,\tfrac{1}{2}$) 的等尾可信区间。*投注区间取决于试验的顺序；其覆盖率是对 8 种随机顺序取平均后，在步长为 10^{-3} 的网格上计算得出的。

其中，$u_1<\cdots<u_{k'}$ 是不同的取值（狄利克雷坐标之和服从狄利克雷分布）。所有容差均相对于 $w_{\max}-w_{\min}$ 而言。对于 $k'=1$，$Y=w_{\min}$；对于 $k'=2$，

$$Y=w_{\min}+(w_{\max}-w_{\min})\,B,\quad B\sim\mathrm{Beta}(a_2',a_1'),\qquad Q_q(Y)=w_{\min}+(w_{\max}-w_{\min})\,I_q^{-1}(a_2',a_1'),$$

与 Clopper–Pearson界一样。下面，$k'\ge 3$ 并省略撇号。

5.2 将尾概率表示为积分

对于相互独立的 $\Gamma_i\sim\mathrm{Gamma}(a_i,1)$，$P=\Gamma/\sum_i\Gamma_i$，因此对于 $0<y<1$

$$\mathbb{P}(Y\le y)=\mathbb{P}(S\le 0),\qquad S=\sum_i\lambda_i\Gamma_i,\quad \lambda_i=w_i-y,$$

$$M(s)=\mathbb{E}\,e^{sS}=\prod_i(1-s\lambda_i)^{-a_i},\qquad \kappa(s)=\log M(s),\qquad \frac{1}{\lambda_{\min}}<s<\frac{1}{\lambda_{\max}}.$$

For 此区间内的实数 $c\neq 0$，直线 $\mathrm{Re}\,s=c$ 上的反演公式给出 [6]

$$\mathbb{P}(S>0)=\frac{1}{\pi}\int_0^\infty \mathrm{Re}\,\frac{M(c+it)}{c+it}\,dt\ \ (c>0),\qquad \mathbb{P}(S\le 0)=-\frac{1}{\pi}\int_0^\infty \mathrm{Re}\,\frac{M(c+it)}{c+it}\,dt\ \ (c<0).\tag{1}$$

我们计算较小的尾概率：若 $\mathbb{E}S=\sum_i a_i\lambda_i>0$，则为 $\mathbb{P}(S\le 0)$；否则为 $\mathbb{P}(S>0)$。我们在鞍点处取 c [4, 8]，

$$\kappa'(c)=\sum_i\frac{a_i\lambda_i}{1-c\lambda_i}=0,$$

通过牛顿法结合二分法求得；当 $|\kappa'(c)|\le 0.1\sqrt{\kappa''(c)}$ 时停止，并在 $|c|<\min(1,|c_{\mathrm{pole}}|/2)$ 时使其远离 0（按下文所述进行缩放）。当 $p_i=1-c\lambda_i>0$ and $x_i(t)=t\lambda_i/p_i$ 时，

$$M(c+it)=e^{\kappa(c)}\,r(t)\,e^{i\theta(t)},\qquad r(t)=\prod_i(1+x_i^2)^{-a_i/2},\qquad \theta(t)=\sum_i a_i\arctan x_i,$$

因此，式 (1) 中的两个积分具有如下被积函数

$$\frac{e^{\kappa(c)}}{\pi}\, r(t)\, \frac{c\cos\theta(t) + t\sin\theta(t)}{c^2 + t^2}. \tag{2}$$

它在 $t = 0$ 处的大小为 $e^{\kappa(c)}/(\pi|c|)$，即尾概率的大小，因此求积误差相对于积分足够小时，相对于尾概率也足够小。对 y 求导可得到密度，且具有相同的 r 和 θ:

$$f_Y(y) = \frac{e^{\kappa(c)}}{\pi} \int_0^\infty r(t)\Big(\cos\theta(t) \sum_i \frac{a_i/p_i}{1 + x_i^2} - \sin\theta(t) \sum_i \frac{a_i x_i/p_i}{1 + x_i^2}\Big)\, dt. \tag{3}$$

代码返回两个积分的 $\kappa(c) + \log(\cdot)$。所有计算都在变量 s/σ中进行，并使用$\sigma^2 = \sum_i a_i\lambda_i^2$，因此被积函数的变化尺度约为 1。

5.3 求积

奇点　　被积函数在变量 t中的奇点是分支点和极　　e,

$$t = -i\,\frac{p_j}{\lambda_j}\ \ (j = 1, \dots, k), \qquad t = ic,$$

点，都位于虚轴上。我们在 T处截断积分，并将 $[0, T]$ 分成长度逐渐加倍的面板:

$$\tfrac{1}{2}\sum_i a_i \log\!\big(1 + T^2\lambda_i^2/p_i^2\big) = 40, \qquad [0,T] = [0,h] \cup \bigcup_{j\ge 0} [2^j h, 2^{j+1}h], \qquad h = \min\Big(1, \tfrac{1}{2}|c|, \tfrac{1}{2}\min_j \tfrac{p_j}{|\lambda_j|}\Big),$$

使得 $r(T) = e^{-40}$，最后一个面板结束于 T。在每个面板上，$[x_0, x_1]$ 我们使用 12 点高斯–勒让德求积公式,

$$\int_{x_0}^{x_1} \phi(t)\, dt \approx \frac{x_1 - x_0}{2} \sum_{l=1}^{12} \omega_l\, \phi\Big(\frac{x_0 + x_1}{2} + \frac{x_1 - x_0}{2}\xi_l\Big).$$

如果 ϕ 在面板的伯恩斯坦椭圆 E_ρ 内是解析的，并且其值被 B 所界定，则误差至多为[12,定理 19。3]

$$\frac{64}{15}\,\frac{B\,\rho^{-24}}{\rho^2 - 1} \cdot \frac{x_1 - x_0}{2}.$$

对于 $[x, 2x]$，虚轴是归一化变量 $u \in [-1, 1]$中 Re $u = -3$所在的直线，因此每个$\rho < 3 + 2\sqrt{2}$ 都允许，对于每个面板以及奇点的每个位置都是如此：该求积公式在所有面板上都以相同的速率收敛。当参数的大小差异很大时，θ 远离$t = 0$时可能会迅速变快；我们将长度为 ℓ 的面板划分为 $\lceil \ell$个最多 $|\theta'|/6\rceil$ 的等份，其中

$$\theta'(t) = \sum_i \frac{a_i\lambda_i/p_i}{1 + x_i^2},$$

使得 θ 在每段的一半区间内转动不超过 3 弧度。

5.4 分位数

要解尾概率$(y) = q$，令 d 表示从 y 到尾部末端（$d = y$ 或 $d = 1 - y$）的距离，并令$u = \log d$。在 $d = 0$附近，尾概率 $\approx C d^{a}$，因此尾概率的对数关于 u近似呈线性，我们对其使用牛顿法

$$g(u) = \log \operatorname{tail}(y) - \log q, \qquad g'(u) = \frac{d\, f_Y(y)}{\operatorname{tail}(y)}, \qquad u \leftarrow u - \frac{g(u)}{g'(u)},$$

使用来自 (3) 的 f_Y，位于根的 $[d_{\mathrm{lo}}, d_{\mathrm{hi}}]$ 区间内（当某一步超出该区间时采用二分法）。起始点是具有 Y 的均值和方差的贝塔分布的分位数：

$$m = \frac{\sum_i a_i w_i}{a_0}, \quad \operatorname{Var} Y = \frac{\sum_i a_i (w_i - m)^2}{a_0 (a_0 + 1)}, \quad \nu = \frac{m(1-m)}{\operatorname{Var} Y} - 1, \quad y_0 = I_q^{-1}\big(m\nu, (1-m)\nu\big),$$

使用 $a_0 = \sum_i a_i$；对于两个值时结果是精确的。当 $|\Delta d| \le \epsilon$ 或下一步的预测误差很小时，迭代停止，

$$100\, d \left| \frac{g''}{2g'} \right| (\Delta u)^2 \le \epsilon, \qquad g'' \approx \frac{g'(u) - g'(u_{\mathrm{prev}})}{u - u_{\mathrm{prev}}},$$

默认使用 $\epsilon = 10^{-12}$。上端点满足 $\mathbb{P}(Y > y) = \alpha/2$，因此从不构造 $1 - \alpha/2$。

5.5 计算成本

设 E 为尾概率和密度的计算次数，I 为一次鞍点搜索中的牛顿步数，N 为一次计算所需的求积节点数。在相位分裂之前使用 $J = \lceil \log_2(T/h) \rceil$ 个面板时，

$$\text{cost of one endpoint} = O\big(k \log k + E\,(I + N)\,k\big), \qquad N = 12\,(J + 1) + 12 \times (\text{splits}).$$

样本量 n 仅通过参数 a产生影响，而计算成本不会随之增加。当a_0 较大时，Y 接近正态分布，被积函数在缩放变量中接近 $e^{-t^2/2}$，而 T、h 和N 趋于常数：在我们的测试中，$N = 60$（五个面板）对应于 $n \ge 10^6$。参数较小时，被积函数衰减得更慢，因此需要更大的 T，但只需增加 $\log_2 T$个面板。我们尚未证明 N的上界；在 $k \le 20$为 5 到 10^9 且 n 的随机问题中，N 的中位数为 60 到 192，最大值为 456。在我们的测试中，E 平均为 2.6，最多为 7，而 I 只需几步，因为每次搜索都从上一个鞍点开始。

作为比较，具有相对误差 δ 的尾概率 q 的蒙特卡洛估计需要大约$1/(q\delta^2)$个样本，这对于 $q = 5 \times 10^{-7}$ 和 $\delta = 10^{-10}$而言是 2×10^{26}。对于 $k \ge 3$ 个值，一个区间需要 0.15 到 0.3 毫秒，而对于两个值大约需要 0.02 毫秒，其中几乎全部是 Python 的开销。

6 精度

我们没有误差界的证明：面板中的论证给出了收敛速率，而测试则测量误差。我们将代码与以下参考资料进行比较：

- 整数参数的精确分布。Y 服从一个 B样条分布，其节点为 w_i，重数为a_i [3]，且 $m = a_0 - 1$，

$$\mathbb{P}(Y > y) = \sum_{j : w_j > y} \operatorname{Res}_{t = w_j} \frac{(t - y)^m}{\prod_l (t - w_l)^{a_l}},$$

我们使用精确有理数运算进行计算。

- 高精度。两个取值时的贝塔分布，以及在参数较大或差异悬殊时精确到 40 位的反演积分。

- 模拟与性质。蒙特卡罗样本；对取值的仿射映射、排列以及类别拆分具有不变性；关于 α 的单调性；以及 $n \le 1$、$n = 10^{12}$ 和 $\alpha = 10^{-15}$ 等边界情况。

- 覆盖率。通过枚举所有结果，计算三个取值和 $n \le 13$ 时区间的覆盖率。

在所有测试 中，包括 40,000 个随机问题，取值最多为 20 个，且 n 最多为 3×10^6 ：

- 端点与取值的值域相距不超过 10^{-12} ；
- 尾概率的相对误差低于 10^{-10}，最低可达 $\alpha = 10^{-12}$;
- 每个单侧覆盖率至少为 $1 - \alpha/2$。

软件

- Python：pip install gaffke，https://pypi.org/project/gaffke/
- R：install.packages("gaffke")，https://CRAN.R-project.org/package=gaffke
- 网页：https://louisabraham.github.io/gaffke/，通过 JavaScript 移植版（gaffke.js）在浏览器中计算区间，其结果与 Python 软件包一致，达到值域的 10^{-15}

两者都提供 gaffke.ci(counts, values, alpha) 以及 $w^\top P$ 的分布律（分布函数、密度、分位数）。R 包是 Python 包的逐行移植版，且其结果与 Python 包的结果相差不超过取值范围的 10^{-10} 。源代码、测试、第4节的实验以及本文均可在 https://github.com/louisabraham/gaffke 获取，采用 MIT 许可证。

参考文献

[1] T. W. Anderson. 任意有界连续分布随机变量期望值的置信限。国际统计学会公报，43:249–251，1969。[2] C. J. Clopper 和 E. S. Pearson. 以二项分布为例说明置信限或枢轴限的使用。*Biometrika*，26(4):404–413，1934。[3] H. B. Curry 和 I. J. Schoenberg. 关于 P´olya 频率函数 IV：基本样条函数及其极限。数学分析杂志 ´数学，17:71–107，1966。[4] H. E. Daniels. 统计学中的鞍点近似。《数理统计年刊》，25(4):631–650，1954。[5] N. Gaffke. 关于非负变量均值的非参数单侧假设的三种检验统计量。《统计学数学方法》，14(4):451–467，2005。[6] J. Gil-Pelaez. 关于反演定理的注记。*Biometrika*，38(3–4):481–482，1951。[7] W. Hoeffding. 有界随机变量之和的概率不等式。《美国统计学会杂志》，58(301):13–30，1963。

[8] C. W. Helstrom. 雷达和光通信中检测概率的近似计算。*IEEE*航空航天与电子系统汇刊，14(4):630–640，1978。[9] E. Learned-Miller 和 P. S. Thomas. 有界随机变量均值的新置信区间。arXiv:1905.06208，2019。[10] A. Maurer 和 M. Pontil. 经验伯恩斯坦界与样本方差惩罚。载于*COLT*论文集，2009。[11] J. Ming、A. Ramdas、Y. Shen、R. Wang 和 I. Waudby-Smith. Gaffke对有界数据均值的置信区间不可容许，但渐近有效。arXiv:2607.18661，2026。[12] L. N. Trefethen. 逼近理论与逼近实践。SIAM，2013。[13] I. Waudby-Smith 和 A. Ramdas. 通过投注估计有界随机变量的均值。英国皇家统计学会杂志，*B*辑，86(1):1–27，2024。[14] N. Vlassis 和 P. S. Thomas. 非负随机变量均值的精确分布无关检验。arXiv:2607.08415，2026。