arXiv:2303.03167v1 [stat.ME] 3 Mar 2023

成功-失败实验可靠性统计的计算

预印本

Sanjay M. Joshi*独立研究者
sanjaymjoshi@iitbombay.org

2023年2月2日

摘要

可靠性是成功-失败实验中的成功概率。随着样本数量增加，可靠性估计值的置信度也会提高。保证度将置信水平设为与可靠性相同，以便用一个数字更方便地传达结果。假设样本服从二项分布，只有置信度可以用闭式表达式计算。带连续性校正的威尔逊得分区间为可靠性提供了近似的闭式表达式。研究发现，布伦特方法能够快速、准确地计算可靠性和保证度。文中提供了针对多种样本数量的图表，并介绍了两个用于计算可靠性、置信度和保证度的开源 Python 库。

关键词 置信度 · 可靠性 · 保证度 · 样本量

1 引言

零件、功能或试验的可靠性，是指其成功或未发生故障的概率。这类成功或失败的试验通常称为伯努利试验或二项试验。结果或样本是服从二项分布的随机变量。我们假设各样本彼此独立且同分布。

本文回顾了置信度和保证度相关的统计参数，并给出了若干样本量下的计算结果。

1.1 可靠性

可靠性是样本的成功概率。如果在 n 个样本中出现 f 次失败，则可靠性估计值可计算为 $\hat{r} = (n - f)/n$。

这只是一个估计值，因为我们只从假设为无限的总体中观察了 n 个样本。总体的实际可靠性可能与估计值不同。因此，我们需要为可靠性估计值附上一个置信水平。

1.2 置信度

对于一次伯努利试验 有 n 个样本时，在可靠性为 r 的情况下， n 个样本中恰好观察到 k 次失败的概率为 [2]

$$Pr(k; n, r) = f(k) = \binom{n}{k}(1 - r)^k r^{n-k}, \tag{1}$$

其中 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ 是二项式系数。

*https://www.linkedin.com/in/sanjaymjoshi/

置信度 c 是实际可靠性至少为 r 的概率。如果在 n 个样本中有 f 次失败，则计算结果为 [1]

$$c = 1 - \sum_{k=0}^{f} \binom{n}{k} (1-r)^k r^{n-k}, \tag{2}$$

零次失 败（$k = 0$），在 n个样本中，计算很简单，且常见于文献 e

$$c = 1 - r^n \tag{3}$$

因此，当我们基于有限数量的样本传达可靠性水平时，也应同时给出相关的置信水平。对于不熟悉置信度和可靠性概念的人来说，使用这两个数值可能会令人困惑。接下来将介绍一个更简单的保证度概念。

1.3 保证度

保证度这一概念由 Fulton [3] 提出，方法是直接将可靠性设为置信度。例如，90% 的保证度意味着在 90% 的置信度下，可靠性为 90%。直观来说，可靠性很高但置信度很低，意义不大。同样，对低可靠性水平有很高的置信度，也没有多大价值。因此，保证度这一概念只用一个数字，就能让沟通更加简便。

在零次失败、 n 个样本的情况下，将保证度 $a = r = c$代入公式 3，得到

$$a = 1 - a^n \tag{4}$$

在 n 个样本中出现 f次失败时，使用公式 2，得到

$$a = 1 - \sum_{k=0}^{f} \binom{n}{k} (1-a)^k a^{n-k} \tag{5}$$

这些方程不容易求解 $n >$，因为根据 n 和 f 计算 a 并没有闭式表达式。下一节将介绍上述所有参数的计算方法。

2 计算

2.1 可靠性

在上一节中，我们看到，给定 r 和 n 时，计算 c 只涉及闭式表达式。反过来，假设已知 c 和 n ，需要求出 r。例如，已知 10 个样本中没有失败，想要了解可靠性的 95%置信区间。

The 零失效情况确实很简单。将公式 3 改写后，可得一个闭式表达式

$$r = (1-c)^{1/n} \tag{6}$$

然而，存在非零失败次数时，没有与公式 2 对应的闭环公式。由于这些样本服从二项分布，求解 r 的问题等同于在目标错误率 $\alpha = 1 - c$下求置信区间的下界。

一种较好的近似方法是“威尔逊得分区间”，尤其是在 n [4]的值较小时。该区间的下界计算如下

$$r = \frac{1}{1 + \frac{z^2}{n}} \left(\hat{r} + \frac{z^2}{2n} \right) - \frac{z}{1 + \frac{z^2}{n}} \sqrt{\frac{\hat{r}(1-\hat{r})}{n} + \frac{z^2}{4n^2}}, \tag{7}$$

其中，$\hat{r} = (n - f)/n$，即优质样本数占样本总数的比例。参数 z 是与目标错误率 α相对应的标准正态分布的（$1 - \alpha/2$）分位数。

“连续性校正”修改 [4] 在上述方程中用 $\max(\hat{r} - \frac{1}{2n}, 0)$ 替换 $\hat{r}$。在考察这些方法的准确性之前，我们先回顾另一种选择。

可以使用数值方法计算可靠性。方程 2 可以改写为 a s

$$c + \sum_{k=0}^{f} \binom{n}{k} (1-r)^k r^{n-k} - 1 = 0 \tag{8}$$

给定 n、f 和 c，可以通过数值方法求解左侧表达式，得到 r。

这些计算使用 Python 实现，并已作为开源库发布 [5]。我们使用 Google Colaboratory 的默认计算环境（一台 2.0 GHz Intel Xeon）对性能进行了测试，并通过可公开访问的 Jupyter 笔记本进行测试 [7]。在此计算配置下，计算仅耗时几分之一秒。

对于数值方法，采用了 Brent 优化法 [6] 方法（SciPy 库中的函数 brentq），容差设为 0.1%。

为评估威尔逊得分区间近似的质量，n 设为 40。失败样本数 f 从 1 增加到 39。对于由此得到的每个 r 值，使用公式 3 计算置信度 c。使用该 c 值，通过两种威尔逊得分法计算 $\hat{r}$，并计算差值 $\hat{r} - r$ 以获得准确度。

这些方法的准确度图如图 1 所示。从图中可以看出，数值方法的表现要好得多。

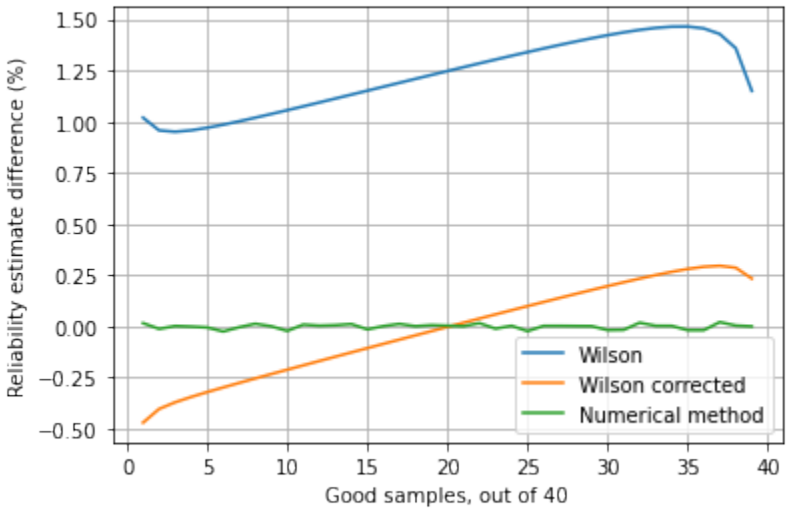

图 1：不同可靠性计算方法的准确度

为观察固定置信度下可靠性的变化，将样本数从 10 调整为 100。置信度设为 95%，失败次数从 0 增加到 5。采用数值方法得到的结果图见图 2，具体数值见表 1

	样本总数									
	10	20	30	40	50	60	70	80	90	100
0	74.1	86.1	90.5	92.8	94.2	95.1	95.8	96.3	96.7	97.0
1	60.6	78.4	85.1	88.7	90.9	92.3	93.4	94.2	94.8	95.3
2	49.3	71.7	80.5	85.1	87.9	89.9	91.3	92.3	93.2	93.8
3	39.3	65.6	76.1	81.7	85.2	87.6	89.3	90.6	91.6	92.5
4	30.3	59.9	72.0	78.6	82.6	85.4	87.4	88.9	90.1	91.1
5	22.2	54.4	68.1	75.5	80.1	83.3	85.5	87.3	88.7	89.8

表 1：95%置信度下不同失败次数对应的可靠性百分比

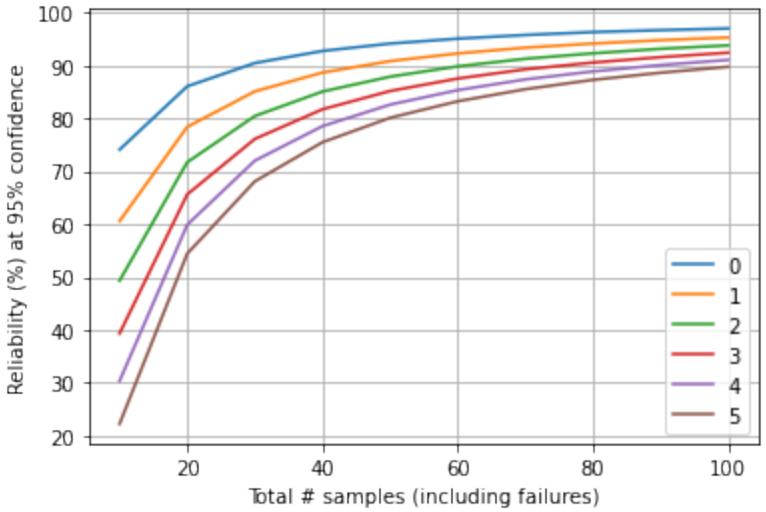

图 2：95%置信度下的可靠性

2.2 置信度

公式 2 是闭式的，在 n和 f取值较小时效果良好。随着 n增大，尤其是超过 1000 时， $\binom{n}{k}$ 的计算会变得数值不稳定。为解决这一问题，我们将公式 2 改写为 $c = 1 - F(k)$，其中$F(k)$是 $f(k)$的累积密度函数（CDF）。

该库在 [5] 这些计算中使用 SciPy 模块中 binom 对象的“生存函数”方法。
图 1 展示了在可靠性水平固定的情况下，置信水平如何随失败次数 f和样本数 n变化。数值见表 2。

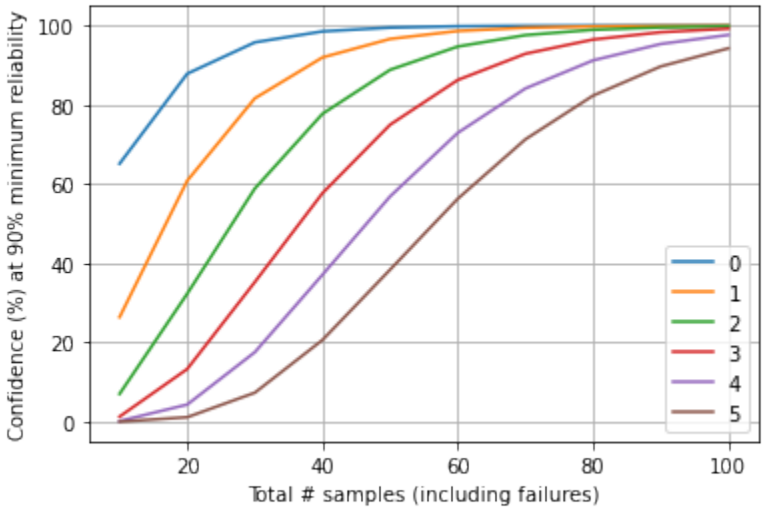

图 3：不同失败次数对应的置信水平

2.3 保证度

为了计算 a 给定 n，我们将方程 4 改写为

$$a^n + a - 1 = 0 \tag{9}$$

	样本总数									
	10	20	30	40	50	60	70	80	90	100
0	65.1	87.8	95.8	98.5	99.5	99.8	99.9	100.0	100.0	100.0
1	26.4	60.8	81.6	92.0	96.6	98.6	99.5	99.8	99.9	100.0
2	7.0	32.3	58.9	77.7	88.8	94.7	97.6	98.9	99.5	99.8
3	1.3	13.3	35.3	57.7	75.0	86.3	92.9	96.5	98.3	99.2
4	0.2	4.3	17.5	37.1	56.9	72.9	84.1	91.2	95.3	97.6
5	0.0	1.1	7.3	20.6	38.4	56.3	71.3	82.3	89.7	94.2

表 2：可靠性为 90% 时，不同失败次数对应的置信百分比

然后，问题归结为求上述方程的实根。对于更一般的形式，将方程 5 改写为

$$a + \sum_{k=0}^{f} \binom{n}{k} (1-a)^k a^{n-k} - 1 = 0 \tag{10}$$

该库 [5]使用布伦特方法进行这些计算。

图 4 展示了保证水平如何随失败次数 f 和样本数 n变化。相应数值见表 3。

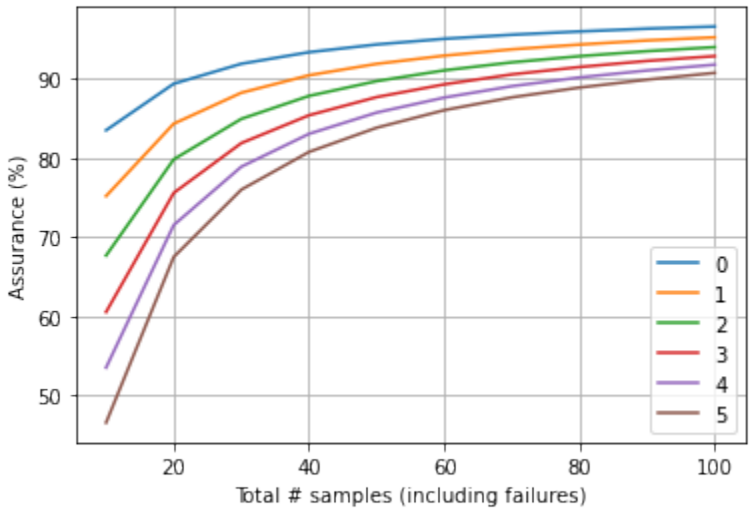

图 4：不同失败次数下的保证度百分比

3 结论

我们回顾了成功-失败实验中的可靠性，即实验成功的概率；此类实验的结果服从二项分布。根据有限数量的样本计算可靠性估计值时，应同时说明该估计值的置信水平。我们还回顾了保证度的概念，它将可靠性和置信度结合为一个数值，以便于沟通。

我们探讨了计算可靠性、置信度和保证度的方法。置信度可以通过闭式表达式计算，但使用二项分布的生存函数参数，即使样本数量很大，也能得到数值稳定的结果。计算可靠性时，带连续性校正的威尔逊得分区间通过闭式表达式通常可提供足够的精度，误差一般低于 5%。使用布伦特方法数值求解方程可以得到更准确的估计值，且计算速度很快。使用布伦特方法也可以高精度地计算保证度。

本文展示了使用新的开源 Python 库通过 Jupyter 笔记本创建的图表和表格。通过简单修改这些图表和表格，即可获得可靠性、置信度和保证度的统计数据。

	样本总数														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	50.0	61.8	68.2	72.4	75.5	77.8	79.7	81.2	82.4	83.5	84.5	85.2	86.0	86.6	87.2
1	0.0	38.2	50.0	57.2	62.3	66.1	69.1	71.5	73.5	75.2	76.7	77.9	79.1	80.0	81.0
2	0.0	0.0	31.8	42.8	50.0	55.3	59.4	62.7	65.4	67.7	69.6	71.3	72.9	74.2	75.4
3	0.0	0.0	0.0	27.6	37.7	44.7	50.0	54.2	57.6	60.5	63.0	65.1	66.9	68.6	70.1
4	0.0	0.0	0.0	0.0	24.5	33.9	40.6	45.8	50.0	53.5	56.4	59.0	61.2	63.2	65.0
5	0.0	0.0	0.0	0.0	0.0	22.2	31.0	37.3	42.4	46.5	50.0	53.0	55.6	57.9	59.9

	样本总数														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0	87.7	88.2	88.6	89.0	89.4	89.7	90.1	90.3	90.6	90.9	91.1	91.3	91.6	91.8	91.9
1	81.8	82.5	83.2	83.8	84.3	84.9	85.4	85.8	86.2	86.6	87.0	87.3	87.7	88.0	88.3
2	76.4	77.4	78.3	79.1	79.8	80.6	81.2	81.8	82.3	82.8	83.3	83.8	84.2	84.6	85.0
3	71.4	72.6	73.7	74.7	75.6	76.4	77.2	77.9	78.6	79.3	79.8	80.4	80.9	81.4	81.9
4	66.5	68.0	69.2	70.4	71.5	72.5	73.4	74.3	75.1	75.8	76.5	77.2	77.8	78.4	78.9
5	61.7	63.4	64.9	66.2	67.5	68.6	69.7	70.7	71.6	72.5	73.3	74.0	74.7	75.4	76.0

	样本总数														
	31	33	35	37	39	41	43	45	47	49	51	53	55	57	59
0	92.1	92.5	92.8	93.1	93.3	93.5	93.7	93.9	94.1	94.3	94.5	94.6	94.8	94.9	95.0
1	88.5	89.1	89.5	89.9	90.3	90.7	91.0	91.3	91.6	91.8	92.1	92.3	92.5	92.7	92.9
2	85.3	86.0	86.6	87.2	87.6	88.1	88.5	88.9	89.3	89.6	89.9	90.2	90.5	90.8	91.0
3	82.3	83.1	83.9	84.5	85.1	85.7	86.2	86.7	87.1	87.5	87.9	88.3	88.6	88.9	89.2
4	79.4	80.4	81.2	82.0	82.7	83.4	84.0	84.5	85.1	85.5	86.0	86.4	86.8	87.2	87.5
5	76.6	77.7	78.7	79.6	80.4	81.1	81.8	82.4	83.0	83.6	84.1	84.6	85.0	85.5	85.8

	样本总数														
	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480
0	95.1	96.4	97.1	97.6	97.9	98.1	98.3	98.5	98.6	98.7	98.8	98.8	99.0	99.0	99.0
1	93.0	94.9	95.9	96.6	97.0	97.4	97.7	97.9	98.1	98.2	98.3	98.5	98.5	98.6	98.7
2	91.1	93.5	94.9	95.7	96.3	96.7	97.1	97.4	97.6	97.8	97.9	98.1	98.2	98.3	98.4
3	89.3	92.3	93.9	94.9	95.7	96.2	96.6	96.9	97.2	97.4	97.6	97.7	97.9	98.0	98.1
4	87.7	91.1	93.0	94.2	95.0	95.6	96.1	96.4	96.8	97.0	97.2	97.4	97.6	97.7	97.8
5	86.1	89.9	92.1	93.4	94.4	95.0	95.6	96.0	96.4	96.6	96.9	97.1	97.3	97.5	97.6

表 3：不同失败次数对应的保证度百分比。数字已四舍五入。零失败时，样本数达到 459 个即可超过 99% 的保证度。

参考文献

[1] 国际标准化组织 (ISO) *ISO 16269-6:2014* 数据的统计解释 — 第 *6* 部分：统计容许区间的确定，第 2 版，2014[2] A. Papoulis。《概率、随机变量与随机过程》（第 2 版），第 57–63 页，麦格劳-希尔，纽约，1984[3] S.N. Luko。《属性可靠性与成功连串：综述》，DOI:10.4271/972753。[4] S.A. Wallis。《二项置信区间与列联检验：数学基础及替代方法的评估》，《定量语言学杂志》，20:3，第 178-208 页。DOI:10.1080/09296174.2013.799918[5] S.M. Joshi。《可靠性统计库》，https://github.com/sanjaymjoshi/relistats，访问于 2023 年 2 月 23 日。[6] R.P. Brent。《第 4 章：一种用于求函数零点的保证收敛算法》，无导数最小化算法，新泽西州恩格尔伍德克利夫斯：普伦蒂斯-霍尔，ISBN 0-13-022335-2，1973[7] S.M. Joshi。《可靠性统计笔记本》，https://github.com/sanjaymjoshi/relistats_笔记本，可在线访问：Google Colab 上的 relistats_binomial.ipynb，访问于 2023 年 2 月 26 日。