

消失的精确度，一个单次实验的精确度检验方法

——以医药产业中的质检过程为例

王雷

西藏大学 财经学院，西藏 拉萨 850000

DOI:10.61369/ME.2026050004

摘 要： 样本一阶矩与二阶矩是同等重要的信息，分别代表了检验的准确度与精确度。检验人员以实验得出的样本均值代替总体真实值的做法是以大数定律作为理论支撑的，保证了实验准确度，然而在单次实验中却没有精确度的判定准则，本文以简单随机抽样作为前提，从理论上推理了一个单次实验精确度检验方法，该方法为纯数学上的统计检验，无需大费周章做实验获取数据，仅需查阅历史检验结果，构建不稳定性标准。该方法操作简单、应用广泛、实用性强、成本小，值得企业以及官方机构进行积极尝试和运用。

关 键 词： 精确度；质检；简单随机抽样；大数定律；F检验

Lost Precision: A Method for Single-Experiment Precision Verification — Taking Quality Inspection in the Pharmaceutical Industry as an Example

Wang Lei

School of Finance and Economics, Xizang University, Lasa, Xizang 850000

Abstract： The first-order and second-order moments of a sample are equally important pieces of information, representing the accuracy and precision of the test, respectively. The practice of testers using the sample mean obtained from experiments to substitute for the true population value is theoretically supported by the law of large numbers, ensuring experimental accuracy. However, there is no criterion for determining precision in a single experiment. Based on simple random sampling, this paper theoretically deduces a method for testing precision in a single experiment. This method is a purely mathematical statistical test that does not require extensive experimentation to obtain data. Instead, it only requires consulting historical test results to construct an instability criterion. This method is easy to operate, widely applicable, highly practical, and cost-effective, making it worthy of active attempts and applications by enterprises and official institutions.

Keywords： precision; quality inspection; simple random sampling; law of large numbers; F-test

引言

现代医药产业中的检验还是非常依赖人工操作，解文欣等（2025）在论文中强调了 AI 在中药质量应用过程中存在模型解释性较差的缺点^[1]，这种情况在短时间内无法改变。就算很多阶段都引入了机器替代人工，但也做不到 100% 替代，试剂的调配、环境的检测、以及检验过程都离不开人的身影。质检不是炒菜做饭，医药检验领域对检验过程有着严格的要求，就算变更一个检验材料的厂商，也要走复杂的变更程序，评估该变更对后续检验的影响，也要做验证，证明本次变更对检验没有影响，陈振东等（2026）在论文中强调了药品制备过程中哪怕一个微小的工艺调整，都要报备审批，走变更程序^[2]。每一个检验都有对应的标准操作规程（SOP），一旦某一个步骤与标准操作规程相违背，都要走偏差流程，过程同样繁琐，当食品药品监督管理局发现厂商的偏差调查过多，也会投入资源，更多的关注该厂家，所以，每一个厂商都尽量避免该现象发生，便对检验过程有了更严格的要求，这是傻傻的机器手臂无法完成的，只有进化了几千年的人类的手才能对检验过程有着更精细的把控。但是这也给检验人员带来了巨大压力，只要是人就会犯错，所以需要建立规程以保护人的权力，并且应该建立评审标准，用来评估每个人对于不同工作的适配程度。

有些人适合精细操作，有些人适合脑力劳动。让适合的人处于适合的位置上是，现代职业经理人的主要目标之一。需要建立更加精细化的评价标准，以确保每个人都能找到适合自己的位置。对于检验员来讲，还需要一个新的资质评价方法，以确保该检验员对该资质是合适的。本文提出的精确度检验方法是一种除准确度外的另一种方法，可以用来判定检验过程是否偏差过大，可以作为另一种判定检

验员资质的办法。

《中国药典》中规定应采用适宜的国家或国际标准品或参考品对每批原液和成品进行效价测定（饶春明等，2015）^[3]。在单次检验过程中，检验员对样品中的目标物含量是未知的，检验员通过对样品检验的多个结果取平均值代替真实值从而判定目标物的含量，这当然是正确的也是有理论支持的，质检部门得到的该平均数往往关系到生产部门进行后续操作，无论是下料还是配比等多个步骤都与该平均值息息相关。然而这也导致人们过分关注样本一阶矩了，对于检验过程是否存在过大的波动，质检部门和检验人员的关心程度是不够的，检验人员更关心的是样品的准确度而不是精确度，现实中精确度是与准确度同样重要的信息，然而隐藏在检验结果二阶矩中的信息未能引起检验人员的重视。

本文提供的精确度检验方法可用于单次实验样品检验结果的判断，若单次检验结果的离散程度太高，理应推翻本次检验结论，即不应用本次检验的样本平均数替代总体均值。这也需要建立针对检验方法的不稳定性标准。

一、创新性与不足

（一）创新性

现在的传统质检中也是有精确度要求的，也可以说是精密度。精密度被认为是多次测定之间的接近程度，包括重复性、中间精密度、再现性等。可以看出现有精密度的测定往往是非常耗时的，因为不只包括一次检验，而是多次检验结果接近程度的判定。而且现有传统的精密度检验往往只在方法验证，标准品标定等大型验证过程中才进行，对于单次样品的检验往往不进行精密度检验验证。有一些针对单次实验精密度检测的数学原理也是不严谨的，比如孙文秀等（2020）在论文中提到实验精密度要求平行双样测定，结果在允许偏差控制范围之内才能采用，要求在重复性条件下获得的两次独立测定结果的差值不得超过算数平均值的20%^[4]。该方法确实是一个衡量精密度的手段，但是缺乏理论依据，比如为什么是20%？缺乏理论基础。而且平行双样检测实际上就是两次实验，还是非常耗费资源！

本文提出的精确度检验方法是针对一次实验的检验方法，给出严谨的统计学推导公式。根据实验结果即可进行精确度检验，是基于统计学方法的数学验证，可以说是不耗时，零成本。最重要的是给出了单次检验过程中的精确度判定方法，补充了相关研究。

（二）不足

因作者已不在生物医药领域工作，再加上质检数据都是保密文件，所以作者没有数据对该方法的实际应用情况进行验证把控，只是进行了理论推导。

二、研究背景

（一）准确度与精确度

准确度对应的是样本平均数距离总体真实值的接近程度，精确度对应的是检验结果的离散程度。由下图1所示，图中靶1靶2展现的是两名士兵分别打10枪的打靶示意图，该两名士兵在靶图上打出了相同的子弹孔分布，但是一名士兵打在了靶心左右，另一名士兵打在了4环左右，可以认为靶图1士兵比靶图2士兵具有更高的准确度，但却不能认为靶图1士兵比靶图2士兵具备更高

的打靶精确度！而且由此判断两人的射击准度不一样也是不准确的，因为射击靶2的士兵有可能是狙击步枪存在问题。

下图中靶3靶4展现的是另外两名士兵射击6枪的结果，射击靶3的士兵打出了9环、8环、8环、6环、6环、4环共41环的成绩，射击靶4的士兵打出了9环、8环、8环、8环、5环、3环共41环的成绩，从表面上看，二者成绩相同，可以说二者具有相同的准确度，但是二者的精确度却不相同，射击靶3的士兵本次射击标准差约为1.84环，射击靶4的士兵本次射击标准差约为2.32环，很显然射击靶4士兵的不稳定性更高，在高压环境下，射击靶3的士兵更有能力完成任务，这是精确度给该士兵带来的优势。

做检验时每一个检验结果都是简单随机抽样，与士兵打靶是一样的，准确度与精确度一样重要，不能只关心其中一样。在确定检验员资质以及单次样品检验时需要同时考虑这两个标准。

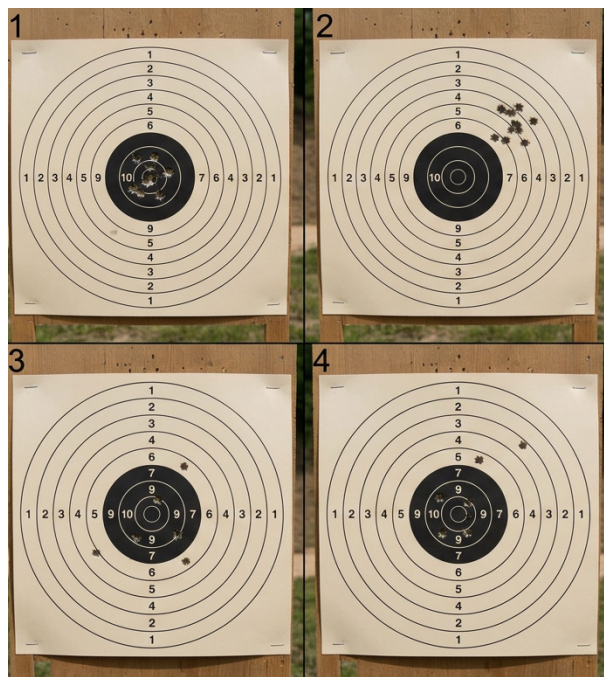


图1 准确度与精确度示意图

（二）现有做法及缺陷

李延玲（2013）在论文中提到，在实验室获得检验资质后应对对检验人员、技术管理人员及辅助人员进行针对性的培训及考核，为考核合格的人员颁发培训合格证，之后才能从事相关工

作^[5]。检验员资质具体是如何确定的呢？一般都是让该名检验人员检验某个确定标识量的样品，如果该名检验人员检验得到的结果置信区间（默认正态分布）包含有该标识量，则认为该检验员达到了要求，可以拥有该资质。这其实忽略了一个重要内容，就是该检验员的稳定性，也就是精确度问题，用置信区间判定的办法其实是做了一个统计检验，可以认为该检验员检测的准确度没有问题，也就是只考虑了样本一阶矩没有考虑样本二阶矩的信息！

很多检验本身就是不稳定的，比如 Elisa 检验，该检验大多数由人来完成，移液器、温箱、酶标仪等都是辅助，所以说该检验本身就有很大不稳定性，其次，不同的实验室，不同的温度湿度、不同的人员都会给该试验带来不稳定性，所以，需要为该不稳定性制定一个标准。该标准是多个已经获得资质的检验人员，在一般条件下检测，得到的针对该方法的不稳定性标准。在移液器、已经获得资质的检验人员、温箱等全部的外部条件确定后，那么这些全部因素加在一起后肯定会产生一定的不稳定性，而且不同的公司以及不同的实验室是不一样的，因为这些不同的外部环境意味着不同的不稳定性，每个实验室或者每个公司将该不稳定性固定下来，就形成了衡量该环境下针对该检验方法的不稳定因素标准。

在实际操作过程中，检验人员是如何根据样品确定总体的真实值的，即用样本检测均值代替总体真实值。

（三）正确做法

那么应该怎么样确定一名检验人员的稳定性是符合要求的！即他对某个样品做出的检验是符合准确度以及精确度要求的呢！首先还是应该做准确度检验，就是看该名检验人员本次检验结果的置信区间中是否含有真实值。其次就是针对检验员稳定性（精确度）的检验，针对样本二阶矩中隐藏的信息进行检验，要求该检验员在资质确认时的样本离散程度在规定范围内，这个规定范围就是该实验室的不稳定标准。

同理已经获得资质的检验人员在进行单次检验时，也要进行精确度检验，以衡量该次检验的离散性是否过大！毕竟影响人的因素包括压力、心情、血糖含量等是非常多的，而在实际操作过程中存在非常多的像 Elisa 检验方法这种人工操作占比非常大的试验，所以尽管该检验人员已经获得了资质，只能证明他在当时的检验过程中没有问题，而本次检验仍然需要进行精确度验证。

三、理论与方法推理

本文推理基于两个认知前提。第一，某些实验方法本身就是存在不稳定性的。第二，人是不稳定性的最大来源。所以对于单次实验也应该进行精确度判定，尤其是针对人为操作占比较大的实验，不能仅凭借该检验人员当初通过了资质检验的验证，就认为检验员本次实验操作的精确度没有问题。

（一）前提

检验人员的每一个检测值都相当于简单随机抽样。在医药领域中很多样品为液体，取样人员在取样之前都会充分摇匀总体，

然后取出一部分作为总体的代表进行检验，即使不是液态样本，《中国药典》中规定，供检验用药材或饮片样本的取样方法为四分法，应遵循四分法取样，寿纪纲等（2022）采用蒙特卡洛模拟等办法进行研究得出结论，四分法与直接随机抽样没有明显区别^[6]，取样人员遵循随机原则，抽样完全随机，无主观选择，个体之间也是相互独立的。所以检验人员得到的每一个结果都相当于一次简单随机抽样，因此，检验人员试验后得到的一组数据是服从正态分布的。这是检验员资质确定时进行准确度检验的前提，也是本文提出的进行精确度 F 检验的前提。

（二）单次检验的准确度理论基础

检验人员检验得出的样品均值为什么能够代表总体真实值？前文说过，每一个试验结果都相当于一次简单随机抽样，则实验结果为随机变量 X ，且 $X_n \sim N(\mu, \sigma^2)$ ，其中 μ 为总体真实值， σ^2 为试验方差，该方差是一个固定值，它是由检验方法本身的不确定性带来的，若是该实验方法是 100% 稳定的，那么就不用多次抽样取结果均值了，一次就能测出准确的值！

为什么可以用样本均值代替总体真实值，原理来自大数定律，由辛钦大数定律，假设 $X_1, X_2, X_3, \dots, X_n$ 独立同分布，存在有限的总体均值 $\mu = E(X_i)$

记样本均值为： $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ 则， $\lim_{n \rightarrow \infty} P\{\bar{X}_n - \mu < \varepsilon\} = 1$

大数定律告诉我们，当样本量足够大的时候，样本均值依概率收敛于总体均值，这便是检验员使用一次检验多次抽样的样本平均数代替总体真实值的理论依据。

大数定律提供了单次检验中样本代替总体的准确度保证，但是精确度却没有保证。

（三）单次实验的精确度检验原理及公式

前提中提到，每一个样品的检验结果相当于一次简单随机抽样记为随机变量 X 。随机变量 X 服从正太分布 $X_n \sim N(\mu, \sigma^2)$ ，现实中 μ 和 σ^2 都是未知的但却都是固定的，实际检测中以 $\bar{x}$ 代替 μ ，有大数定律作为理论基础，而 σ^2 是由试验本身的不稳定性决定的，是个固定值。

将正态分布标准化，
则 $\frac{X_i - \mu}{\sigma} \sim N(0,1)$ ，则 $(\frac{X_i - \mu}{\sigma})^2 \sim \chi^2(1)$
又因为每一个试验结果是独立随机抽样，

则 $\sum_{i=1}^n (\frac{X_i - \mu}{\sigma})^2 \sim \chi^2(n)$ ，

因为 σ 是常数，由试验方法本身决定，上式变形

则 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$

现实中 μ 是未知的，以 $\bar{x}$ 代替，需损失一个自由度

则 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{x})^2 \sim \chi^2(n-1)$

其中 $\frac{1}{\sigma^2}$ 是固定值， $\sum_{i=1}^n (X_i - \bar{x})^2$ 是随机变量，由随机变量的构成可知，该随机变量展现的是抽样离均差平方和，与方差的概念很相似，均是展现抽样的不稳定程度，其中该不稳定程度是由实验环境、试验人员当天的状态等多种因素共同决定的，该不稳定程度不可以不受控制，一些细微的不受控制的变化可能不被记录到试验记录中，但是多种因素复合在一起，就可能引起这个不稳定因素过大，导致试验精确度受损，因此导致本次试验的准确度

结果也是不能接受的。尽管这可能是检验员的资质测试，并且结果通过了准确度测试。

我们知道实验本身都是存在着不稳定因素 σ^2 的，那么怎么衡量针对某一次的试验不稳定性 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{x})^2$ 是过大的还是可以接受呢？

若两个卡方分布是相互独立的，即 $X \sim \chi^2(n_1)$ 与 $Y \sim \chi^2(n_2)$ 相互独立，

$$\text{则 } F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$

我们完全可以构建一个Y，这个Y就是不稳定性的标准，一旦检验员本次检验产生的不稳定性与该不稳定标准的商超过了某个固定值，说明本次试验的不稳定性超标了！精确度不够，这个标准可以是F分布的10%、5%、或者1%尾部值，具体标准可以由具体的公司确定或者由中检院等机构确定。而这个Y是包含了由试验本身带来的、人员的、环境的、试验器材批号的不同等等一些列不确定性的叠加。不同公司的Y肯定是不一样的，因为不同公司的人员、机器等不确定性因素也都是不同的。而这个Y与本次试验是在不同时间进行的，不会互相产生影响，所以说是相互独立的。由此，单次实验的精确度检验理论检验模型如下。

$$\frac{\frac{1}{\sigma^2} \sum_{i=1}^{n_1} (X_i - \bar{x})^2 / (n_1 - 1)}{\frac{1}{\sigma^2} \sum_{i=1}^{n_2} (Y_i - \bar{y})^2 / (n_2 - 1)} = \frac{\sum_{i=1}^{n_1} (X_i - \bar{x})^2 / (n_1 - 1)}{\sum_{i=1}^{n_2} (Y_i - \bar{y})^2 / (n_2 - 1)} \sim F(n_1 - 1, n_2 - 1)$$

σ^2 是固定值是由试验方法本身带来的，故可以约掉， X_i 是具体某一实验中抽样的随机变量，而分母部分的 $\sum_{i=1}^{n_2} (Y_i - \bar{y})^2$ 看起来复杂，而在实际操作中其实是个固定值，只不过这个标准值应该是具备资质的检验人员操作的有代表性的一次或几次的检验形成的不稳定性标准值，由此这个不稳定性标准就是该公司独特的针对该方法的不稳定性标准，与别的公司有所区别。推荐该不稳定性标准定为方法验证时或者标准品标定时的那次检验的不稳定性。

一般检验时的 n_1 是确定的， n_2 也是确定的，此外若公司确定了显著性水平标准，则本次检验就是离均差平方和与某个具体值的比较大小，实际操作非常方便，具体见公式变形，以显著性水平 $\alpha=0.05$ 为例。

实际检验模型（1）

$$\frac{\sum_{i=1}^{n_1} (X_i - \bar{x})^2 / (n_1 - 1)}{\sum_{i=1}^{n_2} (Y_i - \bar{y})^2 / (n_2 - 1)} = F(n_1 - 1, n_2 - 1)$$



检验模型

$$\sum_{i=1}^{n_1} (X_i - \bar{x})^2 = \sum_{i=1}^{n_2} (Y_i - \bar{y})^2 \times F_{0.05}(n_1 - 1, n_2 - 1) \times \frac{n_1 - 1}{n_2 - 1}$$

等号左侧是本次试验的不稳定性随机变量，等号右侧是固定值，若等号左侧的值大于该固定值，则证明本次试验的不稳定性超标，精确度不够，本次实验结果作废。

单独看实际检验模型（1）未变形公式的分子或者分母，不考虑 $\frac{1}{\sigma^2}$ ，因为它是固定值，随机变量 $\sum_{i=1}^n (X_i - \bar{x})^2 / (n - 1)$ 不正是样本的

二阶矩吗！样本二阶矩反应样本离散程度，即精确度，该公式的解释是：“单次实验的不稳定性（精确度）是否超过了针对该环境下的标准不稳定性”。由于该公式的意义非常明显，所以作者在前文中强调在以往的实验结果中，样本一阶矩（平均数）被过分关注了，隐藏在样本二阶矩中的信息（精确度）没有被验证，它是可以推翻实验结论的重要信息！

四、增加精确度检验的意义

第一，对检验员的要求增加，使检验人员必须经历多次练习，减少不稳定性才能通过精确度检验，增加了检验资质的价值。同时，本文提出的精确度检验相当于增加了获取检验资质的难度，本条规则若真的导致检验人员无法通过资质测试，则证明其确实暂时不适合在检验员的位置上，应该增加练习或者对其调岗，以保证他能够以最大效率来工作，具体标准由具体公司制定规章制度决定。

第二，很多批检验和稳定性检验都是单次检验，检验员做检验时，样品的抗原或者抗体等目标物的含量是未知的，仅凭借检验员资质，以检验员检验得出的样品均值作为总体均值，那么本次检验是没有评价标准的，仅仅依靠试验记录中记录的检验员的操作未产生偏差就可以用样本均值代替总体真实值，这样做的信服度较低，因为影响实验精确度的因素非常多，检验人员获得检验资质时的心态与现在的心态完全不同，这是两次独立的实验。如果有了该检验方法的精确度标准，采用一次F检验用来证明本次检验的精确度没有问题，再加上用试验记录来证明未产生偏差，结果的信服度将大大增加。

第三，对于产品来说增加了标准，更好的保证了产品质量。

五、结论及建议

（一）结论

文章首先利用士兵打靶的例子，诠释了准确度和精确度的意义，并且表明了准确度和精确度在抽样过程中均非常重要。第二，解释了检验员的操作原理，为什么能用样本平均数代替总体，背后的原理是大数定律。第三，也是文章主要的表达内容，建立了一个单次实验精确度检验方法。该方法的适用领域非常广泛，涵盖生物、化学、物理等多个领域，可以说只要检验方法是有波动的就能用该方法进行精确度检验，只不过本文多是以生物医药领域的质检过程做例子。

（二）建议

本文提供的单次实验精确度检验方法非常简单，仅仅是统计学上的方差分析检验，简单方便好操作，检验成本几乎为零，不仅仅适用于生物医药产业，应用前景非常广泛。

第一，一些企业需要检查自身的质检领域是否含有基于大数定律的检验方法，或者含有抽样不稳定性的检验方法。需立即采取行动，以方法验证或者标准品标定时期的不稳定性作为标准，运用本文提出的精确度检验方法，衡量每次检验员操作的不稳定性

是否超出了该范围。将该检验方法写到质量标准和标准操作规程中。

第二，官方机构也可以采用本方法进行质检尝试，给出针对检验方法，涵盖实验室及操作人员等各种不确定性的不稳定性标准，该标准甚至同标准品标识量同样重要，分别代表标准品检验的一阶矩和二阶矩的标准信息，需要重视。企业根据该标准构建自身的检验方法，若自身的标准是不超过官方机构标准范围的。则可制定自身标准作为针对自身企业、实验室、检验人员的不稳定性标准。官方机构可将操作过程，不稳定性标准等写进相关规

程中，指导企业进行操作。

本文提出的精确度检验方法理论和实操均非常简单，现实中几乎不需要大费周章的去重新试验，只需拿出历史数据构建不稳定性标准，以后便可针对每一次检验都进行精确度合规判定。该方法实用性非常强，而且具备较好的现实意义，如果作者通过本篇论文对质检领域产生了一点点贡献都是作者的荣幸。

利益冲突

作者声明：无相关财务或非财务利益冲突。

参考文献

[1] 解文欣, 张紫萱, 刘越, 等. 人工智能在中药质量中的应用及研究进展 [J]. 中草药, 2025, 56(15): 5616–5631.

[2] 陈振东, 林秀玲. 原料药供应商变更的审评考量与质量一致性策略 [J/OL]. 医药导报, 1–15[2026–05–30]. <https://link.cnki.net/urlid/42.1293.r.20260528.1916.005>.

[3] 饶春明, 王军志. 2015年版《中国药典》生物技术药质量控制相关内容介绍 [J]. 中国药学杂志, 2015, 50(20): 1776–1781.

[4] 孙文秀, 柳玉晓. 食品理化检验室的质量控制技术探讨 [J]. 中国卫生标准管理, 2020, 11(18): 6–9.

[5] 李延玲. 如何做好食品检验机构实验室资质认定评审工作 [J]. 中国卫生检验杂志, 2013, 23(4): 1028–1029.

[6] 寿纪纲, 霍志鹏, 王玉, 等. 基于随机模拟探讨《中国药典》四分法取样的注意事项 [J]. 化工与医药工程, 2022, 43(1): 34–40.