详细表述准确度的实用规则

P. J. 坎皮恩
J. E. 布森斯 A. 威廉斯 著

原子能出版社

详细表述准确度的实用规则

P. J. 坎皮恩　J. E. 布森斯　A. 威廉斯　著

李　琳　培　译

原　子　能　出　版　社

内容简介

本书从实用的目的出发，以简练的篇幅阐述了在实验科技领域中广为使用的偶然不确定度和系统不确定度的概念，并论述了在科研成果报告中或科学论文中应如何记录数据以及如何使用、处理和表述不确定度。在附录中简单介绍了几种常用的检验正态性和一致性的方法。

A code of practice for the detailed statement of accuracy

详细表述准确度的实用规则

P. J. 坎皮恩　J. E. 布森斯　A. 威廉斯　著

李琳培　译

原子能出版社出版

（北京 2108 信箱）

北京印刷一厂印刷

新华书店北京发行所发行・新华书店经售

☆

开本 787×1092 1/32・印张 2・字数 43 千字

1979 年 8 月北京第一版・1979 年 8 月北京第一次印刷

印数 001—9,200・定价：0.32元

统一书号：15175・162

原文序言

实验科学工作者在他从事科研工作的初期就要在这方面受到严格的训练：即必须对每一项测量结果进行分析和定出相应的不确定度。在浏览实验科学文献时便可看到这种训练的成果，表现在最常用的数学符号(通常是±号)上。但是，当人们想使用这种方法来说明不确定度时，常常发现这种做法看起来很成功，实际上确不然。例如，很难见到关于不确定度极限推导过程的论述；缺乏充分的资料来说明所报道的结果和某些其它测量结果之间的差异。尤其使人失望的是，甚至给出的资料本身的含义，有时也可能是含混不清的。

本书的主旨是提出如何使不确定度的表示不致含混不清的建议。书中探讨了根据各次测量值估算不确定度的某些方法，包括将不确定度分成随机的和系统的两类的方法；还探讨了将每项系统不确定度合併起来和将每项随机不确定度合併起来时用到的方法；也探讨了定出可以用来表示准确度的各种不同办法。本书既不是统计学方面的手册，也不去讨论仪表说明书中用到的“误差范围”。它主要关心的是，如何详细报道最高水平的科学测量的结果。但是，对任何要求说明不确定度的工作（例如为某些校准证书列出不确定度），本书提出的这些建议都是适用的。然而，对于那些常规的测量，当仪表校准中的不确定度不大时，从使用数据的角度着眼，也可以不必按照本书所要求的细节去说明不确定度。

目　录

第一节　不确定度的种类

某项测量结果的不确定度(uncertainty)通常都是以定量的形式$+x$，$-y$来表示的，最经常用到的是对称形式$\pm x$。这一对数字意味着什么，至少一般说来是清楚的。它表明了最后结果的数值所在的某个范围，被测量的真值就在这个范围之内。这本小册子的主旨是对此概念加以阐明和推广，并对估计和表示x值和y值的方法加以详述。然而在此之前，有必要先决定我们应该怎样称呼这两个表示不确定度的数字。通常的习惯是称它为误差(error)，例如说"估计的系统误差为±1%"。然而还广泛地流行着另一种术语，即"系统误差的估计范围"。二者的差别是很重要的，因为按照平常的含义，误差一词应该是用来表示测量值偏离真值的量；根据这一观点，"误差是±1%"的这种说法是没有意义的。

使用误差这一词来表示不同的东西，通常是科技写作中引起误解的一个原因，因此我们将在附录1中用稍长篇幅对它进行讨论。此段没有必要详细讨论它，因为在本书中将尽可能不用"误差"一词，只有在不会造成含混的情况下才用到它，例如在"平均值的标准误差"(它是今天统计学语言中已得到公认并具有唯一定义的一个术语)这样的表示中才使用，否则就使用"不确定度"这一词。

在本书中，测量值的不确定度照例分为两类，即随机不确定度(random uncertainty)和系统不确定度(systematic uncertainty)。前者在第二节中将要讲到，随机不确定度是

对那些遵守正态分布或其它特定分布*的测量值重复测量多次并通过统计分析来求得的。在第三节中将要讨论到，系统不确定度是通过分析那些预料会影响结果或使结果偏离的物理效应来估计的。在那些重复测量不遵循正态分布或其它一定规律分布的情况下，系统误差是通过分析测量值的分布来估计的。再者，当某个实验的结果要用于以后的实验中时，则先前实验结果的随机不确定度也要纳入后一个结果的随机不确定度之中，对系统不确定度的处理也是一样。

与上述方法不同，经常用到的另一种方法是，当某个实验的结果要用于往后的实验中时，将以前实验结果的随机不确定度纳入后一实验结果的系统不确定度之中。这种处理的根据是前一个结果的任何误差（从对真值的偏差这个含义上来说）都会对后一结果引起一个恒定的偏离量。但是这种说法在逻辑上是否合理是值得商榷的。因为这个偏离量的最佳估计值应当是零，并且也有一个不确定度，这个不确定度的分布就是前一个结果的随机不确定度的分布。

举一个例子就可以将这两种处理方法的差别说得比较清楚些。假定在某一实验中要用到某个物理常数。在上面讨论过的两种方法中，物理常数测量值的随机不确定度都是由确定这个常数时读数的分散度导出的。但是实验中需要用到这个常数时这两种处理方法就不一样了。上述第二种方法认为，物理常数值的“随机”不确定度对本实验结果的“系统”不确定度有贡献，因为常数值的任何误差都会使本实验结果有偏离。我们在本书中采用了前一种方法，那就是认为物理常数值的“随机”不确定度对本次实验结果的“随机”不确定度有

* 只有在观测次数很大时才可能验证观测值是否与某种分布一致，而通常都是假定观测值为正态分布。

贡献。

本书采用的处理方法比较合适，因为它处理不确定度的方法是一致的，在此方法中，由一组读数的分散度导出的随机不确定度在以后的全部分析中仍属于随机的种类，而不与很大程度上取决于主观分析的系统不确定度相混。在此方法中，所有得到的关于随机不确定度的信息仍全部保留。显然，为了充分利用这种方法的优点，不仅在递次合併整个复杂实验的各项不确定度时必须将这两类不确定度分开，而且在最后结果的表述中也要将它们分开。

读数的估计限

可以用仪表读数的估计限作为例子来说明：本来是一个系统不确定度，但通过改变实验方式就可用"随机"不确定度来代替它。由于这个例子包含的基本概念很重要，我们将要较为详细地讨论它。

不管用什么仪表进行测量，读数总是以有限位数的有效数字给出的。对准确度的这种特定的限制就是熟知的读数的估计限。对于数字显示的仪表，估计限通常用末位数的± 1来表示。在模拟装置中，估计限依赖于观测者对标尺的细分度间隔的估计能力及标尺本身的特点。

虽然这种估计限常常是用上述非统计方法估得的，因而是当作系统不确定度来处理的，但是情况并非一定如此。让我们以用等分度的尺子来测量某物体的长度作为例子。假定物体的一端置于标尺的零位上，而另一端的位置最接近标尺的第 n 个细分度。那末此物体的长度位于标尺的第 $n-\frac{1}{2}$ 和 $n+\frac{1}{2}$ 个细分度之间，而且不管测量多少次，其结果总是如

此。在这种情况下估计限就是一项系统不确定度。这里假定了估计限是标尺的一个细分度，但实际上可用目测估计读出比标尺细分度更小的某个间隔；那末估计限就是标尺间隔的一个更细的分度，但照样是系统不确定度。

然而有可能用下面的方法来使估计限“随机化”(randomization)，即使物体的一端不固定在零点，而是随机地放在靠近标尺低端的某点上。假定物体的真实长度在标尺上为 $n+x$ 个分度间隔，这里 x 小于标尺的一个分度间隔。如果把物体置于标尺上的任意位置，我们就会发现它或者与标尺的 n 个或者与 $n+1$ 个分度重叠，读 n 的几率为$(1-x)$，而读$(n+1)$的几率则为 x。假如进行 R 次独立的测量，并且物体每次都是随机地放在标尺上，如果其中 r 次读数得值 n，则$(R-r)$次得值 $(n+1)$，那末比值 $(R-r)/R$ 将给出 x 的无偏估计值。换句话说，

$$\text{物体的长度}=n+\frac{R-r}{R}\text{个标尺分度。}$$

另外，由 R 次测量得到的平均值的方差由下式给出：

$$S^2(n+x)\approx\frac{r(R-r)}{R^3}\approx\frac{x(1-x)}{R}$$

也就是说，在 R 很大的情况下，平均值的标准误差与 $\sqrt{R}$ 成反比(参阅 Jeeffreys, H. Scientific Inference, 2 nd edition, Gambridge Universtty Press,1957, 61)。因此可达到的精度就只受观测次数的限制，于是，这个不确定度就成为一项随机不确定度。如果读数中还有另外一些引起随机变化的因素，那末平均值的标准误差的值就自动会包括所有这些成份。请注意，虽然随机不确定度的降低只受测量次数的限制，但是至少在原则上还会出现象标尺刻度的非线性这类原因引起的

剩余的系统不确定度。在其它模拟装置情况下，可以按如下方式来达到随机化：在每次测量之前，将仪表的零点随机地定在标尺低端的某点上。在某些情况下，有些数字显示仪表的运转过程，本身就包含着随机化技术，这时也可以采用这个概念。

上例说明，如何通过处理方法的改变，利用引进随机不确定度(它的大小依赖于观测的次数)来减少凭主观分析的系统不确定度。在检验某个实验值究竟能达到多大的准确度时，经常值得考虑一下，是否能采用与上面描述相类似的随机化方法将系统不确定度的主要成分降低。

结论

在决定把一个不确定度究竟当做随机的还是当做系统的不确定度时，判别准则应该是看不确定度限制值究竟是由一些测量数据的统计分析导出的呢（假定是正态或其它特定的分布），还是只能通过非统计的分析途径估计出的。由此建议，一旦某项结果的不确定度列为系统性的或随机性的以后，不管以后如何使用这项结果，它总保持它原来的种类。在表示该实验的最后结果时，应当把整个复杂实验的各项不确定度分成两类分别逐项合併，而且在今后的任何其他实验中用到这个结果时，也应将这两类不确定度分开。最后建议，为了降低系统不确定度，只要切实可行，就应在实验过程中引进随机化技术。

第二节　随机不确定度

随机不确定度是指用统计方法处理重复多次的读数得出的不确定度。但是某项实验最后结果的总随机不确定度不仅仅是由该实验过程中测量值的随机涨落引起的。对总随机不确定度的贡献还可以来自该实验中为常数的那些量，因为这些量本身的"数值"通常要经过重复的测量才能得到，根据这些重复测量就可以估计出随机效应的不确定度。显然，仪表的校准因子就是这样一个参数，虽然它的数值在整个实验过程中被当作常数，但是仍然具有随机因素造成的不确定度。

严格地说，在分析重复测量结果的过程中，必须证明测量值符合于正态分布呢还是在一定的情况下符合于其它某种分布。但是只有当观测次数大于 30 时，才足以能从数学上验证是否符合正态分布（见附录三）。在观测次数较少时，符合正态分布只能是假定的，但在此情况下，直观地检查一下数据的分布情况是必要的。我们假定某分布是正态分布的，依据是：在大多数情况下的观测值经过详细地析验，证明确实是正态分布。[按中心极限定理 (Gentral Limit theorem) 预示的那样]。然而，应该承认，这通常是一个假定，并不一定总可以得到证明。在本书其余部分中，除另有明确说明以外，随机不确定度的讨论都是基于正态分布的情形。

分析

假如参量 y 表示某个量的测量值，那末在正态分布情况下，y 值处于 y 和 $y+\mathrm{d}y$ 之间的几率为

$$P(y)\mathrm{d}y=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]\mathrm{d}y,$$

这里 μ 是一常数，它与分布曲线 $P(y)$ 处于最大值时的 y 值相等，σ 是表征上述曲线宽度或分散度 (disperion) 的一个量；量 σ^2 称为该分布的方差 (variance)。下面将要指出，σ 值可由观测结果的分析来估计。这个估计值与自由度数(参阅本节"自由度"一段) 一起用来导出随机不确定度。

如果取 n 次测量 $y_1, y_2, \cdots y_n$ 为样群 (sample)，可以证明，正态分布的常数 μ 的最佳估计值将可由样群的平均值 (mean value) 给出，这里

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i, \qquad [2.1]$$

而正态分布的方差 σ^2 的最佳估计值则由样群的方差 $S^2(y)$ 给出，这里

$$S^2(y)=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2。 \qquad [2.2]$$

量 $S(y)$ 称做样群的标准偏差(standard deviation)；而 $\sum_{i=1}^{n}(y_i-\bar{y})^2$ 则称做残差平方和 (residual sun of squares)。

因为任何一个平均值 $\bar{y}$ 都是由有限的 n 次测量值导出的，因此在反复测量中 $\bar{y}$就会产生一系列不同的值。在n值很大的情况下，中心极限定理指出，$\bar{y}$ 的这些值的分布将接近正态分布，而不管 y 本身是怎样分布的*。这个分布的标准差当然可以由给出的一些 $\bar{y}$ 值去求得，但是也可以在一个单次平均值中根据测量结果的分散度来估计。这个估计值称做平均值的标准误差$S(\bar{y})$(standard error of the mean 简写成 SEOM)，并由下式给出

* 按较为完整的形式，中心极限定理是这样指出的，将大量的其规模和分布形状都是彼此独立的分布叠加起来，只要它们各自的方差值是有限的且有同一量级（也就是说没有一种分布在其中占主导地位），那么总和分布就趋于正态分布。

$$S^2(\bar{y})=\frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{S^2(y)}{n} \qquad [2.3]$$

$S^2(\bar{y})$ 称做平均值的方差 (variance of the mean)。请注意，标准偏差只依赖于所使用的方法和仪表的精密度，只要这些方法和仪表的精密度不变，无论取多少次观测结果，它总不会有显著的变化；可是平均值的标准误差则依赖于样群的观测次数和方法的精密度。因此这两个量表示不同的信息，而在结果中对两者都加以说明有时是适宜的。

在某些情况下，预期的分布不一定是正态分布而是其它分布，例如是二项式分布。假定观测值与所预计的分布一致，那末就应该说明这种分布的类型，并用足够的参量来定量地定出这个分布。在另一些情况下，频数分布就是不接近所预期的分布，这时，就有必要看是否能从数据中找出某种关联或倾向。在这种分布中找不出任何关联时，就应该用图表或其他惯用的形式给出频数分布。在此情况下，由这种因素引起的不确定度不应该加到随机不确定度中去，而应当考虑成系统不确定度。

分组数据

有这样一些情况，某一给定的物理量的测量是成组地进行的，而不是一连串地进行单次测量。有时这样做是为了方便，但经常却是故意使某些系统效应随机化，例如在进行每一组测量之前重新对实验装置进行调整，或者在周围条件可能有所差别的不同日子里来进行各组的测量。

综合这些不同测量组的数据之前，应该检验一下，看看各个组的平均值的分散度是否与单个组中读数的分散度相一致。这有下面三种可能性：第一种情况，这些平均值和分散

度都是内在一致的。第二种情况，这些组的平均值的分散度与这些组内各个测量值的分散度相一致，虽然后一种分散度相互之间相差很大，也就是说某些组的测量要比另外一些组的测量精密一些。第三种情况，这些组的平均值之间的差别比根据测量结果的分散度所预计的要大。

在这三种情况下合成各组数据的方法是各不相同的，这在附录四中作了描述。

测量几个物理量的实验

许多实验只是观测单一的物理量，而有些实验则可能测量几个物理量。在分析最后结果的随机不确定度时，首先需要确定所包含的各个实验观测值的方差，从而确定最后结果的方差。假如某个物理量的值 Y 是通过各别的彼比独立的物理量 $a, b, c, \cdots$ 的测量值由关系式 $Y=f(a, b, c, \cdots)$ 来确定的，而 a, b, c，等的估计方差分别为 $S^2(a), S^2(b), S^2(c)$，那末 Y 的估计方差 $S^2(Y)$ 就可由下式给出：

$$S^2(Y)=\left(\frac{\partial Y}{\partial a}\right)^2 S^2(a)+\left(\frac{\partial Y}{\partial b}\right)^2 S^2(b)+\left(\frac{\partial Y}{\partial c}\right)^2 S^2(c)+\cdots \qquad [2.4]$$

同样，对于 $\overline{Y}=f(\bar{a}, \bar{b}, \bar{c}\cdots)$，以各个量的平均值的方差 $S^2(\bar{a}), S^2(\bar{b}), S^2(\bar{c})$，$\cdots$ 表示的方差 $S^2(\overline{Y})$ 也可以用类似的式子表示，即：

$$S^2(\overline{Y})=\left(\frac{\partial \overline{Y}}{\partial \bar{a}}\right)^2 S^2(\bar{a})+\left(\frac{\partial \overline{Y}}{\partial \bar{b}}\right)^2 S^2(\bar{b})+$$

$$+\left(\frac{\partial \overline{Y}}{\partial \bar{c}}\right)^2 S^2(\bar{c})+\cdots=\sum_i di \qquad [2.5]$$

其中 $i=\bar{a},\bar{b},\bar{c}$ 等，和 $u\bar{a}=\left(\frac{\partial\overline{Y}}{\partial\bar{a}}\right)^2 S^2(\bar{a})$，等等。虽然这里将参量 $a,b,c\cdots$ 当做实际测量值，但在某些典型的实验中，它也可以是某些修正因子和物理常数。

除了忽略了较高次项之外，还假定这些等式中所有各不确定度分量都是相互独立的。虽然通常的情况是这样的，但也有时并非如此。举一简单例子可以说明这个问题。

假定 Y 的测量值是依赖于某装置的读数 r_1 以及该装置的校准因子 c，即 $Y=f(cr_1)$，那末 r_1 和 c 的方差对 Y 的方差的贡献都是独立的。但是如果 Y 的测量值依赖于同一装置的读数的比值 r_1/r_2，而对于这两个读数校准因子 c 又是相同的，即 $Y=f(cr_1/cr_2)$，那末由于 Y 的值与 c 无关，所以 Y 的方差也与 c 的方差无关。这是一个浅显的例子，但它却表明了需要以方程的基本成分来表示该方程的方差时，就要保证各基本成分都是独立的。关于处理含有部分相关成分的方法，读者最好参考统计学的课本。

使用等式 (2.4) 或 (2.5) 可以算出最终量的方差，因此也可以算出它的平均值的标准误差。但是为了使这个方差有使用价值，例如用来将一个结果与另一个结果进行比较，还必须知道自由度数。

自由度

自由度是指残差平方和中独立项的数目，通常用 ν 表示。将 n 次观测值直接平均时，它等于 $n-1$。在用最小二乘法拟合直线时，自由度等于 $n-2$，而在 F 检验的某些应用中，当其中 n 个观测值被分成 r 个组时，自由度等于 $n-r$。

虽然由测量各个量而导出的某个量值的方差中，自由度

数并没有物理意义，但是利用下面的近似公式，有可能从 $S^2(\bar{Y})$ 中估计出"有效自由度数 ν_{eff}"(见 Aspin, A. A. Biometrika, 1949, **36**, 290 和 Welch, B. L. Biometrika, 1947, **34**, 28)，

$$\frac{1}{\nu_{eff}}=\frac{\sum_i \alpha_i^2/\nu_i}{\left(\sum_i \alpha_i\right)^2}=\sum_i c_i^2/\nu_i \qquad [2.6]$$

其中 $c_i=\alpha_i/\sum\alpha_i$，$\alpha_i$ 如等式(2.5) 中所定义。因此 c_i 是第 i 个分量对总随机不确定度贡献的份额，ν_i 是该分量的自由度数。一般情况下等式 (2.6) 给出的不是整数，应将它舍成最近的一个整数。需要强调的是，等式 (2.6) 只是一个近似式，当各个分量的自由度数很小时，这个等式就不大有效。这在附录二中有进一步的讨论。

在给出最后结果的 ν_{eff} 时，原则上也需给出每个分量的自由度数，以便为以后对数据准确地作出显著性检验 (significance test)。但是 ν_{eff} 主要是由具有最大 c_i^2/ν_i 值的分量所决定，在附录二中将会看到，在许多情况下 c_i^2/ν_i 值较小的分量对 ν_{eff} 的影响很小。因此，虽然我们应该记住有些情况要求有详细说明，但对于许多实际情况来说，没有必要引出较小分量的自由度数。具有最大 c_i^2/ν_i 值的分量通常是实验者自己测出的值很分散的结果，但也有可能会来自先前对某常数的测量结果。在这种情况下，显然需要引出这两类成分的自由度数。如果自由度数如此之大，以至实际上可以看做无穷大，那末指出自由度数很大就足够了。

置信限

知道了平均值的标准误差以后，就可以预计一组重复测量的数据落在 $\bar{y}+\delta_1$ 与 $\bar{y}-\delta_2$ 间的几率，P，其中 $\bar{y}$ 是样群的平均值。量 $\bar{y}+\delta_1$ 与 $\bar{y}-\delta_2$ 分别称为上置信限（upper confidence limit）和下置信限（lower confidence limit），区间 $\delta_1+\delta_2$ 称为置信区间（confidence interval），而 P 则称为置信水平（confidence level）。在正态分布时，置信区间在平均值 $\bar{y}$ 两旁对称。可用下述公式算出 δ_1：

$$\delta_1 = t \times (\text{平均值的标准误差}) \qquad [2.7]$$

这里 t 值是相应于所需的置信水平与自由度数的函数，通常称为 Student t 值*，可从表 3 得到。从这个表中可以看到，通常将 3 倍的平均值的标准误差当做置信水平为 99% 时的置信限，在自由度数约低于 8 的情况下是有重大差错的，而且事实上在自由度很大时，这样做多少也是有些问题的。

显著性检验

一个实验的结果常常要与另一个或更多一些同类型的实验结果进行比较。在进行这样的比较时，就需要判断这些结果的差异是否比我们根据实验过程中观测到的随机涨落的效应所预期的要大；附录三中给出了一些用于这种判断的统计学的显著性检验。

在这些检验中要算出这样一个参量，它除了包括其它一些量以外，还要包括每种测量平均值的随机不确定度。但是这些检验所要求的测量值的随机不确定度不一定是根据本书所讨论的判据和方法导出的总的随机不确定度。在这些显著性检验中应当包括的不确定度仅是那些独立地影响每个实验

* Student 是英国一位统计学家的笔名。在一些中文译著中曾将它译为“学生氏”。——译者