第一章　物理实验误差及统计分布

§1.1　实验误差及概率统计性质

物理学是一门以实验为基础的科学，特别是近代物理学中的每一个重大进展都是和实验研究工作的推进分不开的．随着科学技术水平的提高，实验工作的难度和精度愈来愈高．为了在高难度的条件下尽可能地提高实验精度，以便最有效地获得实验所可能获得的科学结果，在实验中愈来愈多地运用现代数学工具．

对于初接触实验物理的人来说，谈到实验的数学处理时，往往自然想到的是误差和有效数字，即在实验做完后对数据进行处理，并且主要是进行误差和有效数字的处理

但是在现代物理实验中，数学处理远不只是这些内容，它包括实验前、实验进行中和实验后的全部数学处理过程．换言之,它表现在实验前的实验设计和论证，实验过程中的控制和监视，实验后的数据处理和结果分析．至于误差和有效数字的知识，作为基础知识的一部分，在这些过程中都是不可缺少的，但不只是在实验后的数据处理中才有用．正因为实验的数学处理是现代物理学发展中的重要一环，现代实验物理学家都需要具有这方面的较高素养，同时在一个大的实验研究组内，也有专门从事这方面工作的成员．

例如，现在国际上重大的高能物理实验课题研究组一般由 30 至 40 人组成,其中有实验物理学家、工程师,还有专门做数学处理工作的理论物理学家，他们从实验的设计开始到实验全部结束一直是在不停地紧张工作．

实验的数据处理中任何一个环节上的失误都会影响实验结果的可靠性和科学意义．实验进行过程中的控制和监视对实验结果的影响是比较明显的，而实验的设计和论证以及实验后的数据处

理及其对结果分析的影响则往往被忽视．近代物理学发展中许多例子可以说明，在某些重要的实验研究中，这些数学处理甚至起了关键作用．

在大学物理实验中，我们不可能对实验的数学处理作系统的介绍，仅对常用的误差理论及数学处理方法进行讨论．

在实验中所得的测量结果，因受到被测对象、所用仪器、周围环境(如气温、气压、风向、湿度及光照等等)以及实验者本人情况的影响，会偏离真值而产生误差．由于影响测量结果的因素很多，它们又以各自不同的方式变动，所以对某次具体的测量来讲，很难确定测得的值对真值偏离究竟有多大，是偏大还是偏小，这样就使得每次测量值的误差大小与正负都带有随机性．而且在同样的实验条件下，使用同一种仪器，采取同一种测量程序和方法多次测量同一物理量的数值，各次所得结果往往不同．这种情况表明，尽管这些测量在科学上的含义是确定的，但是每次测量的结果又具有随机性．

测量结果的随机性的来源是多方面的，实际上在多数情况下往往是几个方面同时起作用．大体说来，有下述几个方面：

(1) 测量的偶然误差：在确定的实验条件下，总有不能完全控制的偶然因素，造成仪器性能的不稳定性和辨别率上的统计涨落以及观察者辨别能力的涨落．随着科学技术水平的提高，后者已可能用自动记录设备来避免．但是前者总是存在的，不可能完全消除．现代一些重要的实验物理成果，往往在实验设计上花了很大力量来分析各种可能造成误差的偶然因素，并设法采取适当的实验措施来尽可能地减少这些偶然因素，或在实验结果中通过适当的处理方法加以扣除．

(2) 物理量本身的统计涨落性质：许多物理量实际上是建立在统计基础上的，这样的量本身就具有统计涨落这种随机性质不能简单地靠提高仪器精确度来改变．

(3) 某些物理量就是作为物理现象某种随机性质的统计描述引入的，对这些物理量的测量必须通过对多次测量结果进行统计

处理来实现．这种随机性质当然也不能靠提高仪器的精度来改变．

尽管测量结果随机性的来源是多方面的，但从提高实验结果的精度水平来看，多次测量总是有利的．考虑到现代重大的精密实验要花费巨大的人力、资金和时间，针对各种不同的情况，在实验设计和结果处理方法上采取相应的措施是很重要的．

§1.2 误差的分类

物理实验中所涉及到的误差有如下几类：

一、统计误差（又称偶然误差）

包括测量的偶然误差以及物理过程或物理量的统计所反映在测量上的涨落性，都归为统计误差．通常实验的误差实际上主要是指统计误差，它包括直接测量的误差和间接推算出来的误差．

偶然误差就某一次来讲，其误差值的大小和正负都有随机性，难以事先确定，但它在大量的重复测量中又遵循一定的统计规律．

二、系统误差

在实验中由于不可控制的因素造成测量结果偏离正确值，这样形成的误差称为系统误差．在物理实验研究中往往要做许多工作来找出并消除系统误差，特别是在某些重要的精密实验中，对系统误差的分析处理，对于整个工作的科学意义和水平起十分重要的甚至是决定性的作用．

系统误差就其来源有以下几个方面：

(1) 仪器误差

这是由于仪器本身制造及校准的不完善，以及仪器没有调整到理想使用状态(如不垂直、不水平、零点没有对准等)所引起的误差．

(2) 环境误差

由于各种环境因素与要求的标准状态不一致引起的误差．如

要求在 20℃ 温度下使用的标准元件，在 30℃温度下使用．磁电式仪表附近有强磁场存在等均会引起环境误差．

(3)方法误差

方法误差是由于测量过程中实际起作用的一些因素，在测量结果的表达式中没有得到反映所造成的误差．例如线路误差 在实验中由于没有把接线电阻、接触电阻、仪表内阻以及交流电路实验中的分布电容、分布电感等因素反映在测量结果的公式中而产生的．如大家熟悉的伏安法测电阻实验，若不考虑电表内阻影响，一定会引起方法误差．

(4) 人员误差

人员误差是由于观察者本人感觉器官不完善或心理特点造成的．记录某一信号时，观测者有滞后或超前的趋向；对准标志物读数时总是习惯偏左或偏右等等．

在实验中研究系统误差有着十分重要的意义，系统误差的存在会影响测量结果的准确性．特别对于随机误差较小（精密度较高)的测量更明显．对系统误差的分析研究就可使测量结果更接近真值，同时还可从中发现某些新问题，这在科学发展史上有不少典型事例．

例如：1907～1917 年密立根(R. A. Millikan)巧妙地用油滴法第一次精确证明了任何物体所带电荷都是最小电荷的整数倍，这个基本电荷就是电子电荷，同时也精确地测定了电子电荷的数值．他在漫长的研究过程中，经历了不少的艰辛和曲折，其中关于测定电子电荷 e 的数值就是一例．

最初他所测得的电子电荷值为

$$e_1 = (1.590 \pm 0.003) \times 10^{-19}\mathrm{C}$$

而用阿伏伽德罗常数 N 和法拉第常数 F，根据 $F = Ne$ 这一关系也可以求得 e 的数值，其中 N 值，根据 X 射线的衍射，测量某种物质的晶格常数，由密度和相对原子质量算出摩尔原子数；法拉第常数是由测量一定时间内电解析出银的质量而求得．用这种方法求得电子电荷值为

$$e_2 = (1.6007 \pm 0.0003) \times 10^{-19}\text{C}$$

两种方法所得的电子电荷 e 值之差为

$$e_2 - e_1 = 1.07 \times 10^{-21}\text{C}$$

而此差值的标准误差为

$$\delta(e_2 - e_1) = \sqrt{\delta^2(e_1) + \delta^2(e_2)} = 3.02 \times 10^{-22}\text{C}$$

显然两种方法所得 e 值之差几乎为标准误差 $\delta(e_2 - e_1)$的 4 倍.这表明两个结果之一或全部存在系统误差.经过分析研究发现，是原先计算电荷时所用的空气黏滞系数不正确，造成系统误差很大，改用精确的黏滞系数后，便得到正确的结果.

又例如,瑞利(Rayleigh)曾由不同来源和用不同方法制氮气，测得氮气密度如下：

(1) 用化学方法制取的氮气密度为

1）由 NO 制取的为 2.30143 2.29890 2.29816 2.30182.

2 由 制取的为

）由 制取的为

4）由 NO_2 制取的为 2.30074 2.30054.

从大气中提取的氮气密度为

）空气与 作用得

）空气荷电灼热铁上处理得

）空气于灼热铁上处理得

根据以上数据算得二种不同方法的氮气密度及其标准误差为

用化学方法的 $p_1 = 2.29971, \delta_1 = 0.00042$

由大气中提取的 $p_2 = 2.31022, \delta_2 = 0.00020$

两种结果的平均值之差为

$$p_2 - p_1 = 0.01051$$

而其标准误差则为

$$\delta(\bar{p}_2 - \bar{p}_1) = \sqrt{\delta_1^2 + \delta_2^2} = 0.00047$$

可见两种方法的平均值之差已大到其标准误差 $\delta(p_2 - p_1)$的 20 多倍. 在当时，各种制氮气的技术已使随机误差很小了. 瑞利并未设法使两者之差减小，而是认为两种方法之间存在着系统误差. 由于强调了两种方法的可能差别，导致后来拉姆齐（Ramsay)发现了空气中的惰性气体.

在一个物理实验的设计和数据处理的实验分析过程中，设法找出并消去系统误差是极其重要的，如果有一部分系统误差未被注意到和考虑进去，则会夸大实验的精确性，并往往使得实验结果在科学上最终垮台. 正因为如此，在对实验工作进行审查时，很重要的一环是审查其对于可能出现的系统误差是否都考虑到了. 有些系统误差虽然认识其存在，但无法消除，则在实验结果的报道中应该包含进去. 在与其他实验结果比较时，系统误差与统计误差的影响没有差别，因此在报道实验结果时，常把系统误差的影响合并到统计误差中去，并以统计误差的形式报道出来，这就是我们在看到实验结果的报道时，往往只看到记载了统计误差.

近年来，随着精密实验的发展，人们日益感到在某些实验结果的报道中，有必要将尚未消除的系统误差与统计误差区别开来，以便在需要区分时采用. 例如，在电磁相互作用的研究中，有一个精细结构常数 α，强相互作用也相应地有一个常数 α_δ. 1980 年，在联邦德国的一个大实验组测得 α_δ 的值为

$$\alpha_\delta = 0.17 \pm 0.02 \pm 0.03$$

其中 0.02 表示的是统计误差，0.03 表示的是系统误差. 这种表述已在许多科学论文中出现并被固定下来.

三、理论误差

这是近年来随着实验科学的发展而逐渐明朗起来的新概念. 有许多物理实验要测量的物理量并不是直接测量的量，而是由直接测量到的结果再通过适当的理论公式换算出来的. 如果所用的

理论公式有一定近似性，而由取近似所带来的偏差虽能估计其大小量级，但并不是确定的，因而也就难以消除掉，这样就给实验结果带来不确定性，这种不确定性称为理论误差.

理论误差同样也用随机的概率分布形式表述出来，但是它和统计误差以及系统误差有很大的不同.设法改进实验设备，提高实验测量精度和增加测量次数可使统计误差和系统误差有所改善，然而这些措施对于理论误差没有任何影响.

过去的实验工作中很少考虑理论误差，是因为实验的精度没有高到必须考虑理论误差的地步，实际上理论误差往往远小于统计误差，从而可以在实验结果的报道中忽略掉.随着实验技术水平的提高，大大减少了统计误差.这时，理论误差的影响就不应完全忽略掉了.这也就是为什么近年来这个概念逐渐明朗并为人们所重视的原因.

理论误差一般发生在对物理量进行间接测量的结果中，这是因为需要运用一定的理论方法从直接测量的结果出发推算出所需的结果，直接测量的结果则不会有理论误差例如，对反映基本粒子弱相互作用和电磁相互作用的系数（称为 Weinberg 角)进行测量是很重要的，1981 年实验测量的结果为

$$\sin^2\theta_{\mathrm{w}} = 0.229 \pm 0.009\,(\pm 0.005)$$

其中 0.009 是统计误差.而 0.005 则是理论误差.这个结果表明，如果设法提高实验精确度.可以把统计误差逐步减少，但在分析实验过程中所用的理论工具的近似性决定了至少有 0.005 的理论误差存在.

从上述关于误差的分类的讨论可以看到，一个实验从设计和论证到实验结果的处理和分析，都离不开要回答下述问题：

实验中可能出现的系统误差有哪些，其中哪些是可以在实验中设法消除的，哪些是可以从实验结果中扣除的；还剩下哪些系统误差是既难于完全消除也难于完全从实验结果中扣除的，从而必须最后以未能去掉的系统误差的形式保留下来；在实验结果的处理中，是否有理论误差；如何尽可能地减少理论误差，特别是设法

减少到小于统计误差的程度；如何认真估算各种因素造成的统计误差.

在上述列出的问题中，估算实验的统计误差往往是最单纯的一个，亦即有确定的方法和程序可以遵循.但是把估算误差仅仅理解为只估算统计误差是不妥当的.

§1.3 几种常用的统计分布

一、统计分布的物理意义

在实验的数学处理中经常用到的概率分布有多种，一般来说各种统计分布都是用来反映随机变量的某种概率分布规律的，从概率论的角度处理这些分布时并不直接涉及到这些分布的物理来源，但是要能正确和熟练地掌握和运用这些分布，不仅需要对运算规则和公式熟悉，而且还应该对每种分布是与哪种类型的物理机制相联系，以及该分布的物理意义和来源尽可能有感性的了解.

在大量的重复测量中统计误差必将遵守一定的统计分布规律，但其到底遵守什么样的统计规律，则要由所研究的问题的性质来决定.

下面对几种常用的分布进行讨论.

二、二项式分布与泊松分布

最常见的离散型随机变量统计分布是二项式分布与泊松分布.

若某随机事件 A 发生的概率为 p,则 A 不发生的概率为 $q=1-p$,现在考虑 N 次独立试验，其中事件 A 发生 k 次的概率为

$$p(k;N,p)=\frac{N!}{k!(N-k)!}p^k(1-p)^{N-k} \tag{1-1}$$

这个概率表达式可以由二项式展开式

$$(p+q)^N=\sum_{k=0}^{N}\frac{N!}{k!(N-k)!}p^kq^{N-k} \tag{1-2}$$

式中含 $p^k q^{N-k}$项里的 $q=1-p$ 给出，这也就是取名为二项式分布的原因.

二项式分布的物理背景在于要求这 N 次试验是完全独立的.二项式分布的期待值和方差为

$$\langle k\rangle = Np \tag{1-3}$$

$$\delta^2(k) = Np(1-p) \tag{1-4}$$

二项式分布的应用很广，在实际分析实验时经常遇到.值得注意的是，二项式分布中有两个独立参数 N 和 p,它们对应于期望值和方差是独立的.二项式分布中总有 $\delta^2(k) - \ k\rangle<0$,或者说有 $\delta^2(k)/\langle k\rangle<1$.这是二项式分布的特点之一.二项式分布的这一性质，可以作为判断一个离散型随机变量的分布是否是二项式分布的必要性检验.

在二项式分布中考虑试验次数 $N\to\infty$,每次试验中 A 发生的概率 $p\to 0$,并使期望值 $\langle k\rangle = Np\to m$ 为有限值，这时得到的概率函数为

$$p(k) = \mathrm{e}^{-m}\frac{1}{k!}m^k \tag{1-5}$$

这时称为泊松分布，泊松分布是进行无穷多次独立试验的总体结果，与二项式分布相比，要求各次试验是互相独立的，这是两种分布相同之处，但泊松分布解决了试验次数有限的要求.由此可见，一个离散型随机变量满足泊松分布比满足二项式分布所受的物理上的限制要少些，但在数学表述上，由于只出现一个参数 m,期望值和方差都通过这个参数来表述，泊松分布比二项式分布受到的限制实际上更强了.

$$\langle k\rangle = m \tag{1-6}$$

$$\delta^2(k) = m \tag{1-7}$$

换言之,对于泊松分布总有 $\delta^2(k) - \ k\rangle=0$,亦即 $\delta^2(k)/\langle k\rangle=1$,这是泊松分布的特点.

泊松分布的这个特点的原因在于各次试验是相互独立的，正因为如此，这个特点常被利用来作为研究其他离散型统计分布的

标准，主要是用来考虑各次试验之间互相独立的还是有关联的，如果随机变量 k 在期待值周围的分布比泊松分布宽，特别是当 k 很大时，概率函数随 k 的增加而减小得比泊松分布要慢时，这种统计分布往往带有 $\delta^2(k) - \langle k\rangle > 0$ 的特点．在这种情况下，其物理意义常反映为各次试验不是互相独立的，表现为一次试验中如果事件 A 发生，则会增加其他次试验中事件 A 发生的概率．各次事件之间的这种相互影响称为正关联．如果随机变量 k 在期望值周围的分布比泊松分布窄，这种统计分布往往带有 $\delta^2(k) - \langle \bar{k}\rangle < 0$ 的特点．在这种情况下，其物理意义常反映为一次试验中如果事件 A 发生，则会减少其他次试验中事件 A 发生的概率．相应地，各次事件之间的这种相互影响称为负关联．

核物理实验中一块放射性物质在一定时间间隔内的衰变数，一定时间间隔内计数器记录到的粒子数，高能荷电粒子在某固定长度的路径上的碰撞次数等都遵从泊松分布．

三、指数分布

指数分布是一种重要的常见的统计分布．许多物理现象的物理机制决定了其概率分布为指数分布．

放射性衰变现象遵从的是指数分布，它可表述为：粒子在时间 t 内衰变的概率为 $1-e^{-\lambda t}$，其中 λ 为衰变常数，即平均寿命的倒数，因此，其概率密度函数为

$$p(t,\lambda) = \lambda e^{-\lambda t} \tag{1-8}$$

衰变现象遵从指数分布的原因是，单位时间内粒子衰变掉的概率为常数，从而单位时间内衰变掉的粒子数目与该时刻存在的粒子总数成正比．

指数分布的来源和泊松分布联系，如果单位时间发生的事件服从泊松分布，则相邻两事件之间的时间间隔服从指数分布．这里单位时间发生事件数的期望值就是指数分布中期望值的倒数指数分布的这个物理来源表明，如果在所讨论的物理问题中涉及用泊松分布描写的物理现象，常常伴随着有由指数分布所描写的物

理现象.

当指数分布的概率密度函数为

$$p(x,\lambda)=\lambda e^{-\lambda x}\quad(x\geqslant 0)\tag{1-9}$$

时,随机变量 x 的期望值和方差分别为

$$\langle x\rangle=\frac{1}{\lambda},\quad \delta^2(x)=\frac{1}{\lambda^2}\tag{1-10}$$

指数分布有两个值得注意的特点.一个特点是,指数分布的概率密度函数的极大值落于 $x=0$ 处,它与 x 的期望值没有任何联系.另一个特点是,如果考虑期望值附近一个标准误差之内的总概率,其值为 0.865,这个值在常见的各种分布中是偏大的.

和泊松分布类似,指数分布中也只出现一个参数既是指数分布的期望值的倒数又是指数分布标准误差的倒数.期望值等于标准误差是指数分布的一个重要性质,是一个统计分布为指数分布的必要条件,但并不是充分条件.

四、均匀分布

均匀分布是实验处理中常用的分布之一,它是最简单的连续型分布,同时又是许多复杂的连续型分布的基础.

均匀分布是指随机变量 x 的取值在某一给定区间内时,概率密度为一个大于零的常数,在 x 取该区间外的值时,概率密度为零.典型的均匀分布的概率密度函数可表示为

$$p(x)=\begin{cases}\frac{1}{2\Delta} & |x-x_0|<\Delta\\ 0 & |x-x_0|\geqslant\Delta\end{cases}\tag{1-11}$$

即随机变量 x 在区间 $x_0-\Delta,x_0+\Delta$ 内的概率密度函数不为零,在区间外时概率密度函数为零,虽然,x_0 是该区间的中心位置,2Δ 为该区间的宽度,上述均匀分布的期望值和方差分别为

$$\langle x\rangle=x_0;\quad \delta^2(x)=\frac{1}{3}\Delta^2\tag{1-12}$$

在期望值附近一个标准误差以内的概率为

$$P_r\left(|x - x_0| < \frac{1}{\sqrt{3}}\Delta\right) = \frac{1}{\sqrt{3}} = 0.577 \tag{1-13}$$

它和其他分布相比是属于偏小的.

服从均匀分布的随机变量取值对期望值的偏离有个严格的界限,即偏离不能超过 Δ.但是从另一方面来说，均匀分布中在期望值附近一个标准误差的区间内的概率比最常见的正态分布要小，这些是均匀分布特有的应该注意的特点，另外在一些实验中，为了消除某些因素对实验结果的影响，进行实验时特意将某些实验条件在一定范围内随机地变动，这就通过人为的方法实现某种分布.

尽管许多情况下最原始的概率分布是均匀分布，但是多个均匀分布的总合效果却完全和均匀分布不同，为了说明这点，我们看电阻串联的例子，设所取的电阻精度是由均匀分布描写的.这样一个电阻的阻值的概率密度如图 1-1 所示,即是一个均匀分布,取这样的两个电阻串联起来时，其总阻值的概率密度函数就不再是均匀分布,而是如图 1-2所给出的形式.这样的三个电阻串联起来时，其总阻值的概率密度函数如图 1-3 所示，从这个趋势可以看出，多个均匀分布的合效果表现为趋向于正态分布.这个结论虽然对任何分布都对，只不过在均匀分布时这个趋于极限的过程表现得特别明显和迅速.

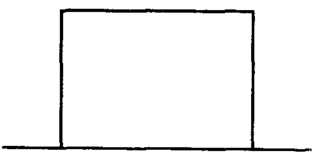

图 1-1

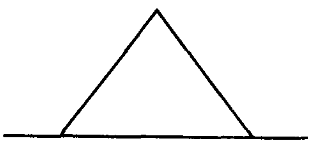

图 1-2

均匀分布在普通物理实验中应用十分广泛，现在与截尾误差为例做些讨论.

在数据处理过程中截去尾数是经常碰到的，我们看到一个数是 1.36 时,很可能它是由 1.361～1.365 舍去第四位数得到的，也

可能它是由 1.355～1.359 将第四位数进入去后得到的.至于它到底是由 1.355～1.365 中的哪一个数得来的就很难肯定了.或者讲，在 1.355～1.365 这个范围内任意一个数，截尾成 1.36 的概率是相同的，很难讲它是由这个范围内的某个数截尾来的可能性更大.故对截尾误差来讲，其概率密度分布显然应如图 1-1 所示.概率密度函数即(1-11)式.显然对截尾误差来讲,公式中的 Δ 是所取数据最末一位数的半个单位例如对 1.36 来说 $\Delta=0.005$.

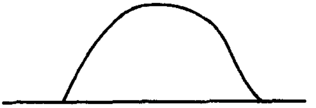

图 1-3

均匀分布求得的标准误差为

$$\delta(x)=\frac{\Delta}{\sqrt{3}}$$

由此式不难看出误差落在 $-\delta\sim\delta$ 范围的概率为

$$P_r(-\delta,\delta)=\frac{\partial\delta}{\partial\Delta}=\frac{\delta}{\Delta}=\frac{\delta}{\sqrt{3}\cdot\delta}=0.577$$

即遵守均匀分布的随机误差的标准误差，其对应的置信概率为 57.7%.

对均匀分布而言，标准误差为 $\delta=\frac{\Delta}{\sqrt{3}}$，其对应的置信概率为 0.577,当取置信概率为 0.683 时（这里，为了在作误差比较时，对应于遵守正态分布的随机误差，在同一概率水平下进行比较）其置信区间为 $0.683\cdot\Delta$ 或 $1.18\cdot\delta$.

对遵守均匀分布的随机误差而言，在截尾过程存在时，单单就数字的截尾这一点来讲，就有 1.18δ 的误差存在，或者说有 0.683Δ 的误差项存在($P_r=0.683$，是以便与遵守正态分布的标准误差项作比较).

例如，某次实验测出对某一个电流的值，表左边是用七位数字电流表测得的结果，表右边是用三位数的指针电流表对同一电流

的测量结果.

从表 1－1 右侧的数据看，每次测量的电流为 7.37A.因此，其平均电压 $\bar{x}=7.37\mathrm{A}$，且由 $\delta_x=\sqrt{\dfrac{\sum V_i^2}{N-1}}$ 可得此时的 $\delta=0$. 由此似乎可以得出结论，此时的误差项等于零.或可以说这个实验没有误差.

而从表 1-1 左侧的数据看，此时因为用的仪表精度较高，它可以反映出被测电流在小数点后第五位及第六位的变化，所以每次测得的数据就不一样了，且此时因 V_i 不等于零，$\sum V_i^2$ 也有一定的大小，由公式可算得其 $\delta_x=4\times10^{-5}\mathrm{A}$.

表 1-1

N	I/A	$V_i\times10^{6}$	$V_i\times10^{12}$	N	I/A	V_i	V_i^2
1	7.368782	−6	36	1	7 37	0	0
2	7 368814	24	576	2	7 37	0	0
3	7 368743	−4 5	2025	3	7 37	0	0
4	7 368764	−24	576	4	7 37	0	0
5	7 368803	15	225	5	7 37	0	0
6	7 368874	14	196	6	7 37	0	0
7	7 368793	5	25	7	7 37	0	0
8	7 368809	21	441	8	7 37	0	0
9	7 368755	−33	1089	9	7 37	0	0
10	7 368847	59	3481	10	7 37	0	0
	$x=7\ 368788$	$\sum V_i^2=8\ 67\times10^{19}$		$\bar{x}=7\ 37$		$\sum V_i^2=0$	
$\delta_x=\sqrt{\dfrac{8\ 67\times10^{1}}{9}}=4\times10^{5}(\mathrm{A})$				$\delta_x=0$			

初看上去，似乎对同一个被测对象，愈粗糙的仪表有时测量的

误差反而愈小．实际上并不是这样，应考虑到较粗糙仪表最后一位数包括截尾误差时，则表右侧测量结果的随机误差项应不小于 $0.683\cdot\Delta = 1.18\delta = 1.18\cdot\Delta\Big/\sqrt{3} = (1.18\times0.005)\Big/\sqrt{3} = 0.004$ (A)，这个误差显然要比左侧用较精细仪表测量所得结果的误差大．

上述例子说明了两个问题：

(1) 要想对随机误差进行分析，所用的仪表精度必须达到能反映数据的变化才行．如对上例来讲，至少要用具有能读五位数字以上的仪表．

(2) 如仪表的精度不够，不足以反映出随机误差的变化，而每次都测得同样大小的数据时，也绝不是说此时随机误差项等于零，实验没有误差．实际上它至少还有因数据截尾带来的误差项．$0.683\cdot\Delta$ 存在(对应 $P_r=0.683$) ,或者说‘任何一个测量结果,其误差项不可能小于 $0.683\cdot\Delta$ 或 $1.18\cdot\delta$”

如果取 Δ 等于最末一位数的半个单位代入时，则截尾误差项为 0.683×0.5(最小单位) $=0.35$(最小单位)．有时为方便起见，也常认为截尾误差就等于 0.5 个最小单位（取它等于 Δ)因此有时认为如对某一个量只测一次（有时只允许测一次），则其随机误差项至少应为最末一位数的半个单位，或为更有把握而取它为一个单位，甚至还可取得更大些．

五、正态分布

正态分布是误差理论中最重要的统计分布．如果随机变量 x 服从正态分布，则其概率密度函数为

$$p(n) = \frac{1}{\delta\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2\right] \tag{1-14}$$

其中 δ 和 μ 为常数,并且 $\delta>0$.这个分布通常用符号 $n(x;\mu,\delta^2)$ 表示，其期待值和方差分别为

$$\langle x\rangle = \mu;\quad \delta^2(x) = \delta^2 \tag{1-15}$$

这也就明显地显示出参数 μ 和 δ 的物理意义.在期望值附近一个标准误差之内的概率为

$$P_{\mathrm{r}}(|x-\mu| \leqslant \delta)=68.3\% \tag{1-16}$$

正态分布又称高斯分布，从下述两个定理可以看出正态分布在误差理论中的特殊的重要地位

定理 1　若 x_i(其中 $i=1,2,\cdots,N$ ）是互相独立的随机变量，随机变量 x 是各 x_i之和，

$$X=\sum_{i=1}^{N} x_i$$

每一个 x_i 对 X 的影响都不大，则当 $N \rightarrow \infty$时，X 渐近地服从正态分布.

定理 2　如果 x_i(其中 $i=1,2,\cdots,N$)是随机变量 x 的 N 次独立测量值(即 x 的容量为 N 的样本)，x 的期望值和方差分别为 μ 和 δ^2，则当 $N \rightarrow \infty$时，x 的样本平均值

$$x=\frac{1}{N} \sum_{i=1}^{N} x_i$$

渐近地服从正态分布

$$n\left(\bar{x};\mu,\frac{\delta^2}{N}\right).$$

定理 1 告诉我们，如果造成某一物理量测量上的不确定性的因素很多，并且没有一个因素是主要的.从而其影响远远超过其他因素，则尽管每一因素的概率分布可以不同，并且甚至是未知的，但最终这个物理量的概率分布近似为正态分布.根据这一点实际上在实验中测量某物理量时，把可能的系统误差扣除之后，该物理量的统计误差往往可以判断是由正态分布描写.下面我们通过具体例子来说明.

用步枪打靶时，中弹位置对靶心形成正态分布.这个熟知的结论实际上可以由上述定理导出.我们将中弹位置与靶心的距离记作 x.x 是一个随机变量，它的值是多个因素影响的结果，例如瞄准时判断上的不准，持枪不稳，子弹飞行过程中局部气流的影响等

等.每一个方面的影响又往往可以更细致地分解为许多部分之和.例如上述子弹飞行过程中局部气流的影响就可以分为许多部分，如果子弹飞行 1000 m 距离，我们可以把每 20 m 距离作为一部分，共分 50 个部分，则局部气流影响可以分为 50 部分之和.这样根据上述定理 1,尽管我们不知道每 20 m 的局部气流影响对 x 的贡献(即某一个 x_i)服从什么分布,但可以判断 x 足够准确地由正态分布描写（因为这样分析时 x_i 的数目 N 相当大）.

接着遇到的一个问题是如果造成某一物理量测量上的不确定性的诸因素中，有一个因素给出比较大的影响，这时应该怎样处理.对于这种情况，首先要分析是否有可能把这个因素的影响更具体地分解为许多部分之和，而各部分之间可以看作是互相独立的，像在上例中对于局部气流影响的处理那样如果能够做到这点，上述定理 1 仍能适用，该物理量仍服从正态分布.如果没有可能做到这点，一个常用的作法是设法研究这个因素的影响所服从的统计分布，将这部分影响和其他因素的影响分开来分别处理.在某些情况下，甚至可以把这个因素的影响作为系统误差而加以扣除.

定理 2 告诉我们，尽管随机变量 X 的概率分布未知，但 X 的样本平均值 $\bar{x}$ 在测量次数 N 足够大时总是服从正态分布的，并且其标准误差会大大减小，这也就是在实验中测量一个物理量时，尽可能进行多次测量求平均的有利之处.

这两个定理说明了在实验的误差分析中经常遇到正态分布的原因.

正态分布是连续型随机变量的一种统计分布.对于离散型的随机变量，如果满足的二项式分布或泊松分布，其特征数字量远大于 1，则离散型也就接近连续型了，这些分布也就渐近地趋于正态分布.可以证明，对于泊松分布，当 $m\to\infty$ 时，$p(k;m)\to n(k;m,m)$ 对于二项式分布,当 $N\to\infty$时，$p(k;N,p)\to n[k,Np,Np(1-p)]$.这说明对于某些离散型随机变量，正态分布也是较好的近似描写.

正因为正态分布的重要性，现在对于实验结果的报道都沿用