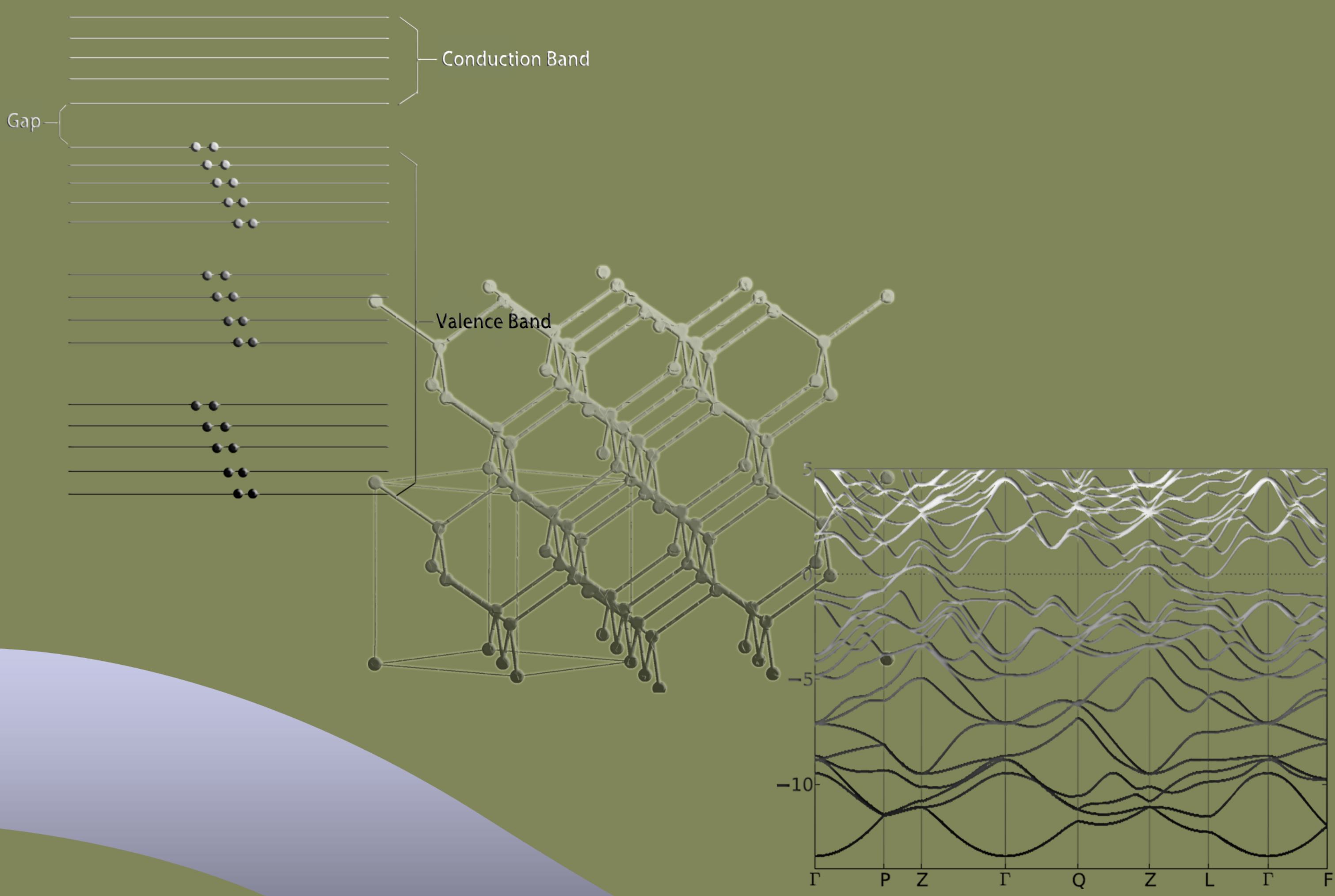
Queen Mary, University of London

School of Physics and Astronomy

误差分析相关总结



一、 误差

误差，指的是实验测量值与真值之间的差，而所谓真值是物理量客观存在但却永远无法真实得到的值，实验上测量无论精度多高都无法完全真实的反映真值，只能是逼近真值，误差定义的是实验测量值与真值之间的差，这就意味着误差只是理论上的一个量，既然实验上无法得到准确的真值，那么也就无法得到准确的误差，所以误差的概念也只能是一个概念而已，它只是表明了实验上得到的测量值与客观真值不同而已。

二、 不确定度

在实际的测量过程中，为了反映测量结果的准确性，就需要根据实验测量量定义一个量来进行表征结果的分散性，于是就引入了不确定度的概念，简单的理解，不确定度就是反映测量结果的实际可能分布区间，对于高斯分布而言，指的是置信概率大于68.3%的置信区间，即$(-\sigma, \sigma)$区间，σ是高斯分布中的标准误差（standard error），在统计学中，简称为标准误（从后面的讨论可以看出，这实际上就是实际测量过程中的A 类不确定度，只不过前者是高斯分布的理论值，后者是根据实验测量量计算确定的值）。

三、 标准差—standard deviation与标准误—standard error

实际测量得到n个测量值，根据这n个测量值可以计算得到一个平均值作为真值的最佳估计值，这也是实验测量的最终目的，平均值的计算公式具有简单的形式：

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

对于多次测量而言，每次测量值与最终计算得到的平均值（最佳估计值）之间的偏差称为偏差，而总体的偏差根据误差理论计算得到应该具有如下的表达形式，称为标准（偏）差（standard deviation）：

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{3-1}$$

这一公式称为贝塞尔公式，由贝塞尔公式计算得到的是测量值的标准（偏）差，这里需要注意标准（偏）差与误差的区别，如第一部分所述，误差是实际得到的测量值与客观真实值之间的偏差，是不可测的（因为客观真值永远无法得到），而这里给出的标准（偏）差，则给出的是实际测量值与多次测量的平均值之间的偏差，是实际可以得到的，也是用于衡量实际测量值分散性的一个重要的量，二者之间的另外一个区别在于函数传递关系的计算式上，这一点将在第四部分中给出。

接下来要给出的就是衡量实验平均值与客观真值之间的偏差，也就是标准（误）差的概念，这里注意区别标准（误）差与上面定义的标准（偏）差，前者是对误差（不可测）的一种估计，而后者则反映实验测量值在多次测量平均值周围的波动，关于标准（误）差的计算公式推导这里

不做讨论，直接给出公式：

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}} \tag{3-2}$$

需要指出，实验测量过程中涉及的A类不确定度的计算公式就是上述标准（误）差的计算公式，二者所指的内容也完全相同。

四、 误差传递

这里给出误差以及相对（偏）差的传递公式，具体的推导过程从略，首先是就对误差的传递公式：

$$\Delta N = |\frac{\partial f}{\partial x_1}|\Delta x_1 + |\frac{\partial f}{\partial x_2}|\Delta x_2 + |\frac{\partial f}{\partial x_3}|\Delta x_3 \tag{4-1}$$

相对误差传递公式：

$$\frac{\Delta N}{N} = |\frac{\partial ln\, f}{\partial x_1}|\Delta x_1 + |\frac{\partial ln\, f}{\partial x_2}|\Delta x_2 + |\frac{\partial ln\, f}{\partial x_3}|\Delta x_3 \tag{4-2}$$

其次是绝对标准（偏）差的传递公式：

$$S_N = \sqrt{(\frac{\partial f}{\partial x_1})^2 S_{x_1}^2 + (\frac{\partial f}{\partial x_2})^2 S_{x_2}^2 + (\frac{\partial f}{\partial x_3})^2 S_{x_3}^2} \tag{4-3}$$

相对标准（偏）差传递公式：

$$\frac{S_N}{N} = \sqrt{(\frac{\partial ln\, f}{\partial x_1})^2 S_{x_1}^2 + (\frac{\partial ln\, f}{\partial x_2})^2 S_{x_2}^2 + (\frac{\partial ln\, f}{\partial x_3})^2 S_{x_3}^2} \tag{4-4}$$

从式——（4-1）至（4-4）可以看出误差传递与标准（偏）差传递计算公式之间的区别，造成这种差别的原因在于一旦得到了实验测量结果，则误差就已经是完全确定的了（即便如此，但是不可实测），而标准（偏）差则需要根据式-（3-1）给出的贝塞尔公式计算得到，因此合成的过程也较为复杂，另外还需要指出，误差合成的意义在于，如果我们能够得到足够逼近客观真实误差的近似值，那么我们就可以按照式-（4-1）或（4-2）给出的计算式合成得到函数的“足够逼近”的误差估计。

五、 More on Standard Error and Standard Deviation

First of all, it should be pointed out the purpose of exploring statistics is to get the information (mean level - what the average level is, vibration range - the distribution around the average level) of some quantity of specific ensemble (technically called population) of individuals. For example, if we want to know about the average altitude of Chinese people, the population here is, of course, all Chinese people, and the 'quantity' here is the 'altitude'. Thus the information about the 'quantity' (i.e. altitude) includes the average altitude of Chinese people, and the distribution of altitude of Chines people (i.e. how significantly the altitude of all Chinese people vibrates around the average level). Now we know our purpose, the next question is: how do we do it? Are we going to measure all Chinese people one by one and then do some basic calculation to obtain the information that we want? If we can, then the calculation process is indeed simple:

$$\bar{x} = \frac{\sum_{i}^{N} x_i}{N} \tag{5-5}$$

$$\sigma_x = \sqrt{\frac{\sum_{i}^{N}(x_i - \bar{x})^2}{N}} \tag{5-6}$$

The $\bar{x}$ and σ here is called the **true mean** and **standard deviation** (or **population standard deviation**) However, is it possible to obtain the HUGE amount of altitude information for every Chinese people? Of course not! So we need to pick up some small amount of people, which we think can be representative of all Chinese people, e.g. we select randomly 1,000 people from each province and then do the measurement and calculation. The task is then doable, but, obviously, what we obtain can only be called 'estimation' (however, this is indeed what we can do in real life). This 'picking up' process is called 'sampling'. Once the population is sampled, we can then obtain our estimation for the **true mean** and **standard deviation**, which is given the name of **sample mean** and **sample standard deviation**, respectively. For sample mean, the calculation method is identical to equation-5-5. However for the sample standard deviation, we have something different, which we call the **corrected sample standard deviation** as an unbiased estimation for the true **standard deviation** based on the Bessel's correction theory. The corresponding formula is given as following:

$$s = \sqrt{\frac{1}{N-1}\sum_{i}^{N}(x_i - \bar{x})^2} \tag{5-7}$$

Now it should be clarified the relationship between the two quantities - **standard deviation** (or **population standard deviation**), which is the true standard deviation of the whole population (however impossible to obtain in real life for large population like all Chinese people); **sample standard deviation**, which is our estimation for the true **standard deviation** based on sampling a relatively small group. The calculation formula was given as equation-5-6 and 5-7, respectively for the true **standard deviation** and the estimation (corrected unbiased **sample standard deviation**). The other thing is then about the sample mean. Imagine we finally obtain all the altitude values of all Chinese people, then the true mean is just a unique value, corresponding to all Chinese people. However, the real case is that through sampling, we finally obtain a **sample mean**, which is only an estimation for the **true mean**. That also means, for different samplings, we should have different **sample mean**, i.e. our **sample mean** should be vibrating around some value, which, ideally, is the **true mean** of the whole population. Imagine, again, we obtain all altitude values from all Chinese people, then definitely we can have the true **standard deviation** - σ. Then we have the formula to calculate the true standard deviation of the sample mean - the deviation of sample mean from the true mean:

$$SD = \frac{\sigma}{N} \tag{5-8}$$

Then one can easily guess what will happen in real life, we have a similar formula to evaluate the standard deviation of sample mean. Remember what we discussed above? Different sampling should give different estimation of the mean level of the population - different sample mean, so there is some deviation there! And the formula given below is the common way to estimate this deviation:

$$SE = \frac{s}{N} \tag{5-9}$$

where s is what we obtain from equation-5-7, which is the sample standard deviation. And what we obtain from equation - 5-9 is called the **standard error** (of sample mean).