3. Measurement error and its representation

3.1 Probability and statistics

(omitted)

3.2 Measurement error and error type

Error is a difference between measured value and true value and can be shown as follows:

$$Error = Measured\ Value - True\ Value.$$

There must arise some error when measuring, and it is impossibility that measure true value. Here, the value with higher accuracy is assumed that it is the true value. And relative error can be expressed by following equation,

$$Relative\ Error = \frac{Error}{True\ Value}.$$

It uses percent to express usually.

There are three types of error, systematic error, random error and mistake.

(1) Systematic Error

For example, use micrometer to measuring object, there may occur error around zero point. This error can be corrected by adjust zero point. Or, if the value of error is known, it can be corrected when read the scale. The error like this, coursed by some reason, and same size with same sign is called Systematic Error. The error can be corrected by look for value of Systematic Error. There are three reasons of Systematic Error.

(2) Random Error

Error occurred by variation, sign and value will have a change.

(3) Mistake

Error is occurred by measurer's carelessness or recording mistake.

3.3 Statistical Handling of Error

Use a length measurement device to measure an object 50 times. The results are being shown in Table.3.1. And Fig.3.1 is a histogram made by using Table.3.1 and shows the approximately of the probability occurred by each measured value. In vertical axis, $f\ [1/\mu m]$ is been written, and f can be calculated by following function.

$$f = \frac{Relative\ Frequency}{Ranges\ of\ Values(Width\ of\ Pillar)}$$

In vertical axis which shows parameter f, each rectangle's area can be expressed by relative frequency of corresponding width. For example, about the measured value of the shaded part 50.00305~50.00325 in Fig.3.1, the occupied percent to all is 0.1×(2.6 + 1.8) = 0.44(44%).

Tab.3.1 Result of length

Measured Value(mm)	Frequency	Relative Frequency(%)
50.0026	2	4
27	3	6
28	2	4
29	10	20
30	4	8
31	13	26
32	9	18
33	4	8
34	1	2
35	2	4
Total	50	100

3.4 Probability Density

When increasing times of measurement, if the

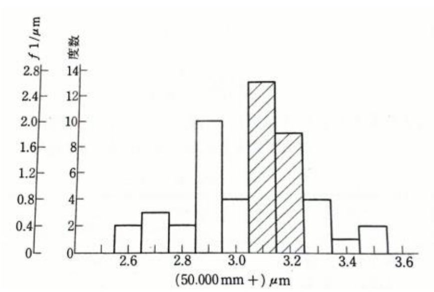

Fig.3.1 Frequency distribution

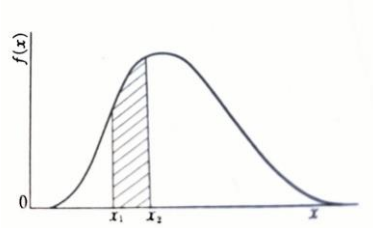

Fig.3.2 Probability density function

times is infinite, the result will approximate a curve like Fig.3.2. The group of measurement value like this called population. If there is a probability density function $f(x)$, The probability between x_1 and x_2 , which is $P(x_1 < x < x_2)$, can be expressed by shaded part in Fig.3.2, and be expressed by following equation. Of course, the whole area in Fig.3.2 is 1.

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)dx \qquad (3.1)$$

(1) Normal Distribution

Normal distribution is a general and useful probability density function, which can be expressed by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (3.2)$$

m is average of population. Normal distribution is symmetric around m, the degree of variation is decided by σ. In Fig.3.3, there are two different σ of normal distribution of probability density function expressed by $f(x)$.

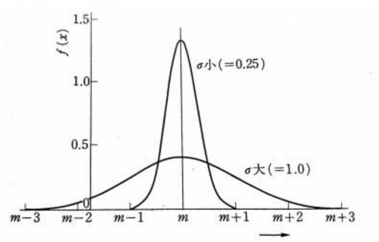

Fig.3.2 Probability density function

Put equation 3.2 into equation 3.1, calculate probability between x_1 and x_2. Select $m \pm \sigma$, $m \pm 2\sigma$, $m \pm 3\sigma$ for x_1 and x_2.

$$P(m - \sigma < x < m + \sigma) = 68.3\%$$

$$P(m - 2\sigma < x < m + 2\sigma) = 95.5\%$$

$$P(m - 3\sigma < x < m + 3\sigma) = 99.7\%$$

In another world, if the probability of x_i is 68.3% and the range is $m \pm \sigma$, in same way, if the probability is 95.5%, 99.7% and the range is $m \pm 2\sigma$, $m \pm 3\sigma$. And if the probability is 95%.

$$P(m - 1.96\sigma < x < m + 1.96\sigma) = 95\%$$

Measured data is not a normal distribution strictly, but in most situations the data can be considered as normal distribution.

(2) Average and Standard Deviation

Average $\bar{x}$ and standard deviation s_n can be calculated from $x_1, x_2, \dots, x_n$.

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

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

s_n^2 is called sample unbiased variance, it is a express of variation like s_n.

Example

In Table 3.2, there are 10 data, and calculate $\bar{x}$, s_{10}, s_{10}^2.

$$\bar{x} = 50.000 + \frac{0.0282}{10} = 50.00282\text{mm}$$

$$s_{10}^2 = \frac{1}{10-1} \times 0.52 = 0.0578(\mu m)^2$$

$$s_{10} = \sqrt{0.0578} = 0.240[\mu m]$$

The $\bar{x}$ and s_n is the estimated value of m and σ. And because of systematic error there is an error between m and true value. This error, in another word, subtracted value of average m and true value is called bias.

Tab.3.2

i	x_i mm	$x_i - \bar{x}$ μm	$(x_i - \bar{x})^2$ $(\mu m)^2$
1	50.0031	+0.3	0.09
2	26	-0.2	0.04
3	24	-0.4	0.16
4	27	-0.1	0.01
5	26	-0.2	0.04
6	30	+0.2	0.04
7	29	+0.1	0.01
8	31	+0.3	0.09
9	30	+0.2	0.04
10	28	0	0
$\sum$	282	0	0.52

(3) Three Axioms of Error

1) Positive error and negative error where are same value are occurred by same possibility. Therefore, the pdf (probability density function) $\Psi(y)$ of error y, which average m is 0, is even function.

$$\Psi(y) = \Psi(-y)$$

But, probability of all error occur is 1.

$$\int_{-\infty}^{+\infty} \Psi(y)\, dy = 1$$

2) Small errors occur more frequently than large errors. Therefore, $\Psi(y)(y \geq 0)$ is a monotone decreasing function. And the only maximal value of $\Psi(y)$ is $y = 0$.

3) Frequency of very large error occur is small. Therefore, it can be assumed that $\Psi(y)(y \geq 0)$ is rapidly decreasing (Exponentially) by increasing of y.

3.4 Accuracy and precision

Generally, the degree of "bias" is defined as accuracy, and a small degree of "variation" is called precision For example, as shown in Fig.3.3:

(a) Accuracy and precision are all good.
(b) Accuracy is bad, but precision is good
(c) Accuracy is good, but precision is bad
(d) Accuracy and precision are all bad.

Display of result: (It is represented by a section that falls within a range which is a certain ratio)

$$x = \bar{x} \pm \sigma$$

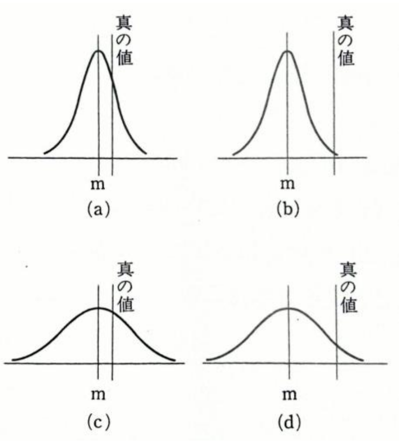

Fig.3.3 Accuracy and Precision

3.5 Significant Figure

(1) Significant figure

Significant figure is a number that meaningful and can be trusted. Numbers are taken to the digit where error first comes in.

3 digits significant number:
12.3, 2.34, 0.0456, 789×10

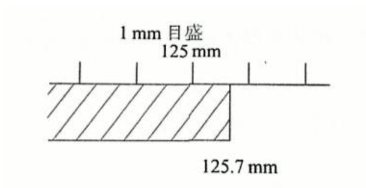

Fig.3.4 Significant Number

These numbers are correct until second digit, and the third digit is error.
In Fig.3.4, measure the object with $1[mm]$ scale ruler. Read $0.7[mm]$ from area between $125[mm]$ and $126[mm]$. The last digit is not $0.7[mm]$, but just express that "closest to $0.7[mm]$". Especially, if read as $125.70[mm]$, it

Table 3.3 Example of Sum

391.6	391.6	391.6
562.35	562.35	562.4
75.154	75.154	75.2
2.7362 (+	2.74 (+	2.7 (+
1031.8402	1031.84	1031.9
(×意味がない)	(◎有効数字を5桁とし, 1031.8とする)	(×有効桁落ち)

is means that after the decimal point first digit, which is $0.7[mm]$ is correct, and after the decimal point second digit is error. It will make a big difference.

(2)Calculate of Significant Figure

Measure True value T_1, T_2, and get measurement value M_1, M_2, the error is ε_1, ε_2.

$$T_1 = M_1 \pm \varepsilon_1$$

$$T_2 = M_2 \pm \varepsilon_2$$

The $\pm$ means true value is a range that $(M_1 - \varepsilon_1) < T_1 < (M_1 + \varepsilon_1)$.

1) Sum and Subtraction

$$T_1 \pm T_2 = (M_1 \pm \varepsilon_1) \pm (M_2 \pm \varepsilon_2)$$

2) Product

$$T_1 \times T_2 = M_1 \left(1 \pm \frac{\varepsilon_1}{M_1}\right) M_2 \left(1 \pm \frac{\varepsilon_2}{M_2}\right)$$

$$= M_1 M_2 \pm (M_2 \varepsilon_1 + M_1 \varepsilon_2)$$

3) Division

$$\frac{T_1}{T_2} = \frac{M_1(1 \pm \varepsilon_1/M_1)}{M_2(1 \pm \varepsilon_2/M_2)}$$

$$= \frac{M_1}{M_2}\left(1 \pm \frac{\varepsilon_1}{M_1}\right)\left(1 \mp \frac{\varepsilon_2}{M_2}\right)$$

$$= \frac{M_1}{M_2}\varepsilon$$

$$\varepsilon = \frac{M_1}{M_2}\left(\frac{\varepsilon_1}{M_1} + \frac{\varepsilon_2}{M_2}\right)$$

ε is called total error, and ε_1, ε_2 is called relative error.

(3)Error in Calculation

Result of subtraction between numbers in same degree will have a big error. In calculation, there are two points must be careful.
1) Make 1~2 extra digit than necessary
2) Devise a calculation method

The significant digit does not become larger than the smallest one of the significant figures of the original number. Calculate extra one digit than original error, sum significant figure later.
1) Rounding error
Rounding off, rounding up, truncating and etc.
2) Truncation error
Finite infinite series
3) Digital loss
Devise a subtraction method

(4)Devise a subtraction method

1) Difficult in sum
If irregular digits for each number or after the decimal point, it will become difficult. See Tab.3.3.
2) Example for subtraction
Do subtraction $\sqrt{2.03} - \sqrt{2}$ in same degree.

$$\begin{array}{rl} \sqrt{2.03} & =1.424781 \\ \sqrt{2} & =1.414214 \quad (- \\ \hline & 0.010567 \end{array}$$

This situation lost 2 digits. Digit loss will decrease accuracy. Digit loss is the main reason of modern computer error. There is another way that won't decrease accuracy.

$$\frac{2.03-2}{\sqrt{2.03}+\sqrt{2}}=\frac{0.03}{1.424781+1.414214}$$

$$=0.01056712\cdots$$

This way will make sure accuracy, but if numbers are far apart, use subtraction generally.

3) Example for product

When measurement value times some number, a number with high probability is computed by taking a large number of digits

（測定値）		（数）		（5桁：1桁余分にとって計算）		（後で丸める）
56.29	$\times$	$\sqrt{2}$	$=$	56.29×1.4142	$=$	79.61

If use $56.29 \times 1.414(4\text{ digits}) = 79.59$, will get a larger error than above equation.