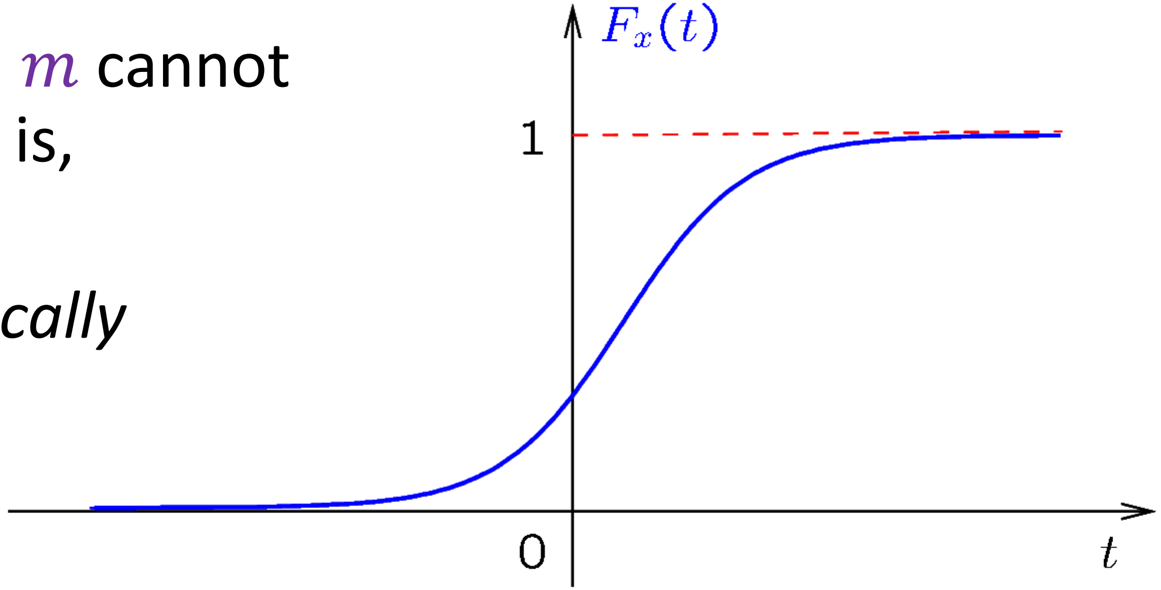
Statistical Evaluation of Measurement Data and Errors

M.V. Iordache, *EEGR2051 Circuits and Measurements Lab*, Fall 2020, LeTourneau University
See https://mviordache.name/EEGR2051 for more information.

Probability Distribution

- In statistics, *probability* is a number between 0 and 1 indicating the likelihood of the occurrence of a random event.
- The probability that a random variable x is less or equal than a number t is called *probability distribution function*.
- Let $F_x(t)$ be the probability distribution function.
- If $m < M$, then the probability of $x \leq m$ cannot exceed the probability of $x \leq M$, that is, $F_x(m) \leq F_x(M)$.
- We conclude that $F_x(t)$ is a *monotonically increasing function*.

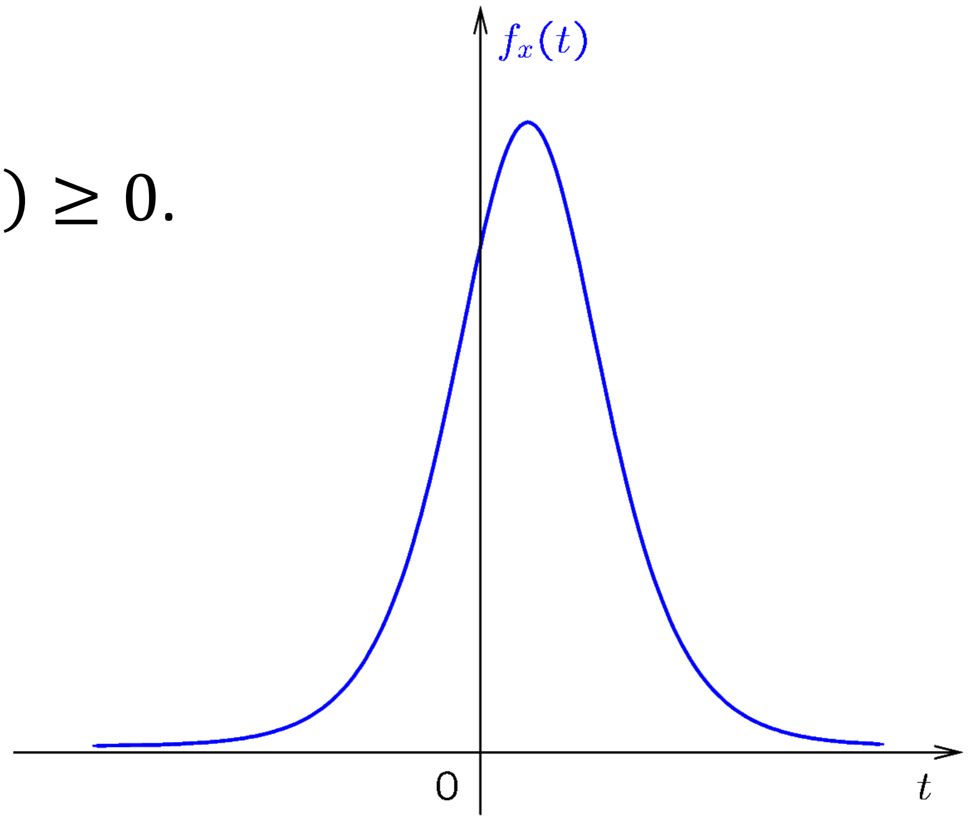


Probability Density

- The derivative of the probability distribution function is the *probability density function*.

$$f_x(t) = \frac{dF_x(t)}{dt}$$

- Since $F_x(t)$ is monotonically increasing, $f_x(t) \geq 0$.
- The area under the curve of $f_x(t)$ equals 1.

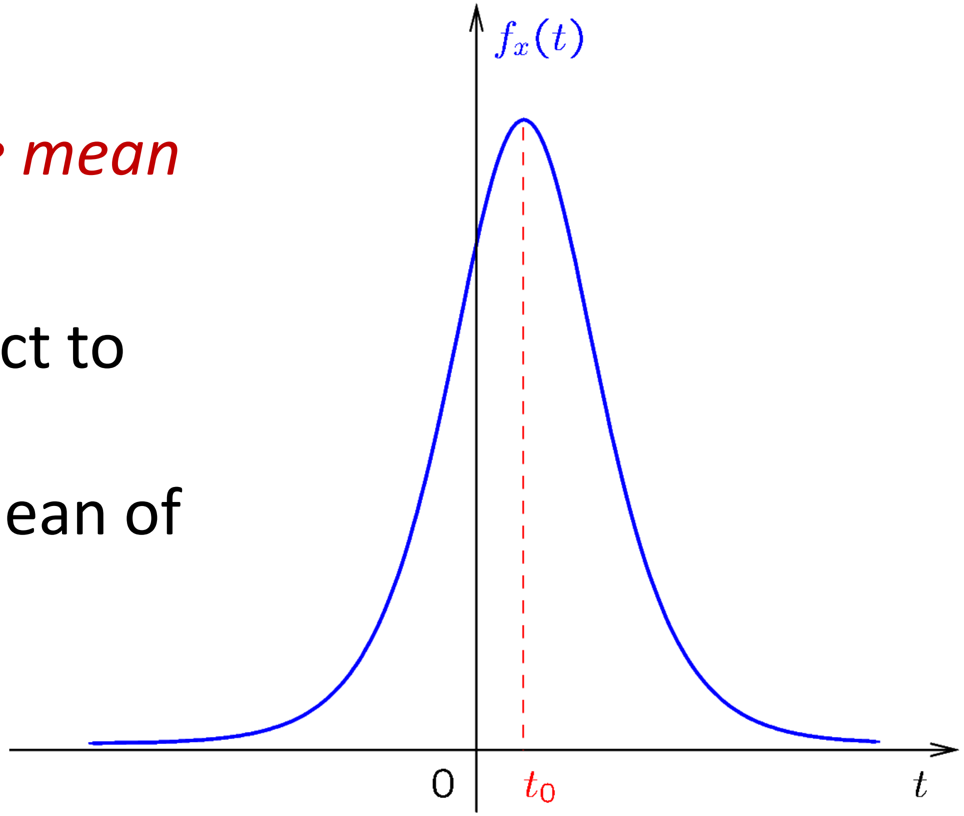


The Mean

- The *expected value* of a random variable x is

$$E(x) = \int_{-\infty}^{\infty} t f_x(t) dt$$

- The expected value of x is also known as *the mean of* x.
- If the graph of $f_x(t)$ is symmetric with respect to the vertical line going through t_0, that is, $f_x(t_0 + z) = f_x(t_0 - z)$ for all z, then the mean of x is t_0.



The Standard Deviation

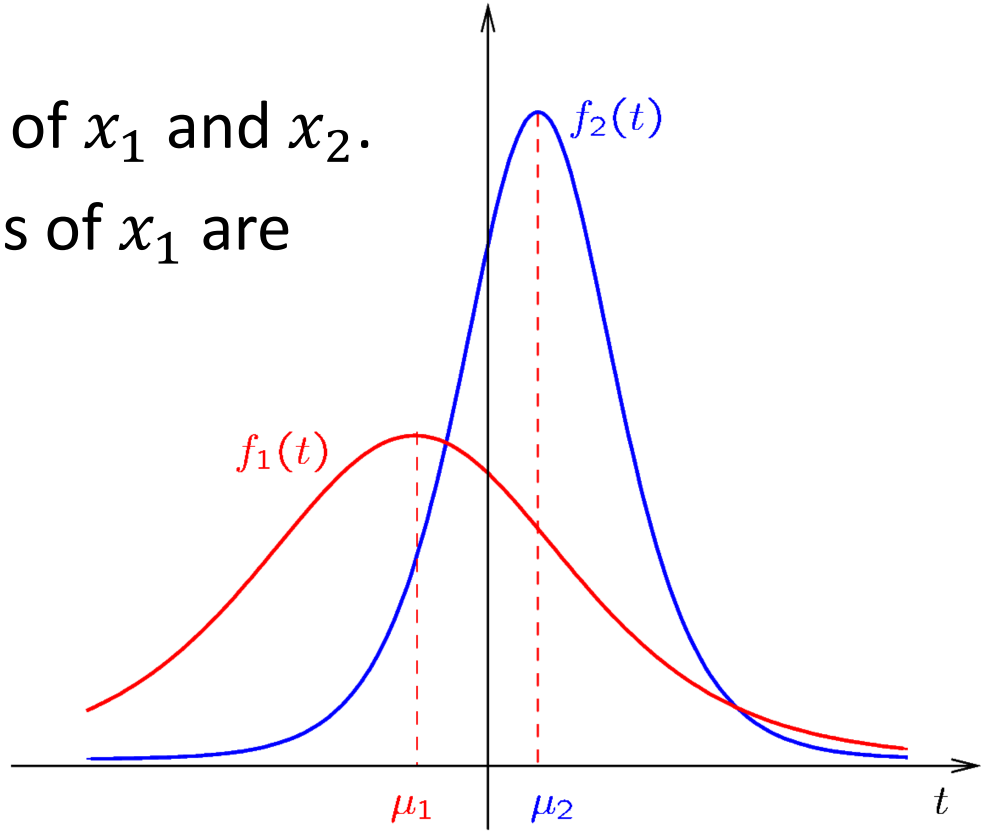
- Let $\mu = E(x)$ be the mean of x.
- The *variance* is the expected value of the square of the deviation of x from its mean μ.

$$V(x) = E[(x - \mu)^2] = \int_{-\infty}^{\infty} (t - \mu)^2 f_x(t) dt$$

- The *standard deviation* is $\sigma = \sqrt{V(x)}$.
- The standard deviation is a measure of the dispersion of the random variable about its mean.

The Standard Deviation

- The standard deviation σ is a measure of the dispersion of the random variable about its mean.
- The figure shows the probability density functions of two random variables x_1 and x_2.
- Let σ_1 and σ_2 be the standard deviations of x_1 and x_2.
- It is clear that $\sigma_1 > \sigma_2$ because the values of x_1 are more spread out.

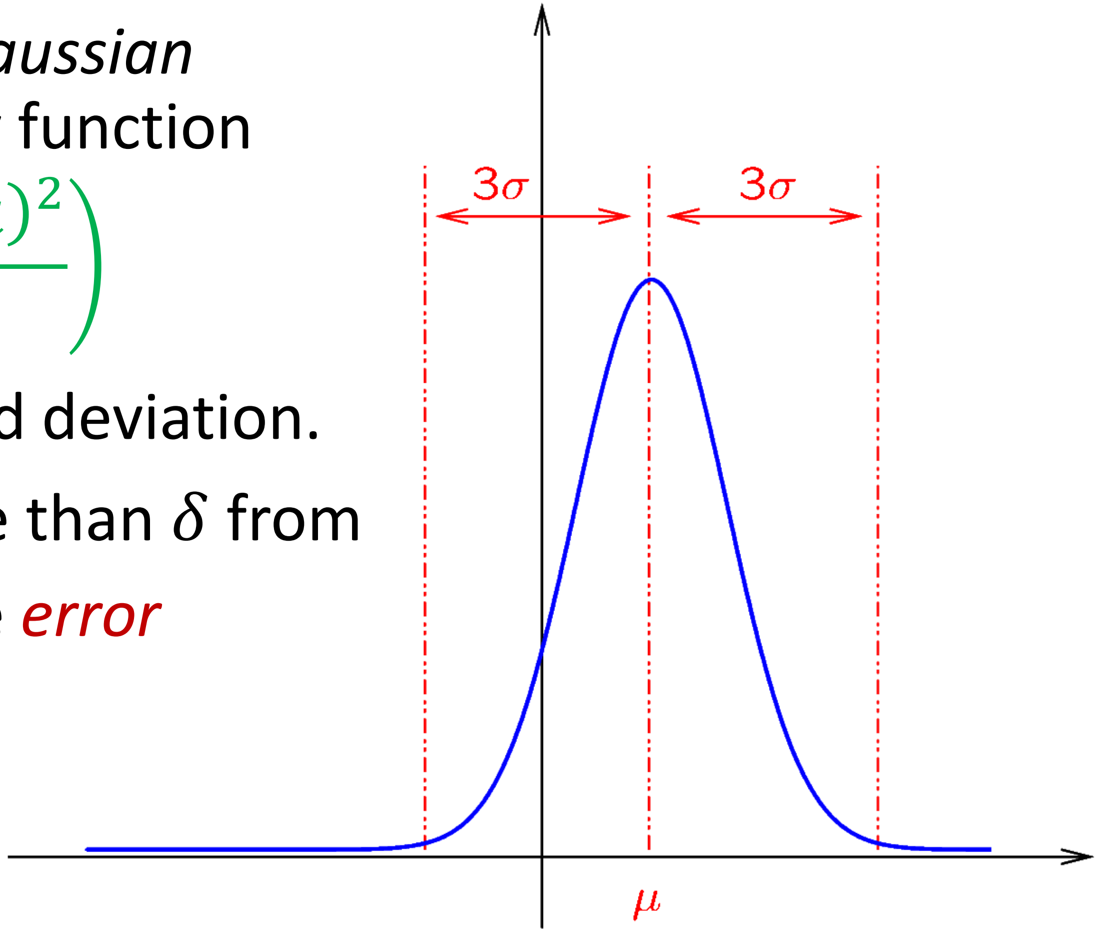


The Normal Distribution

- Measurement errors often have distributions that are approximately *normal*.
- A *normal distribution* (also known as *Gaussian distribution*) has the probability density function

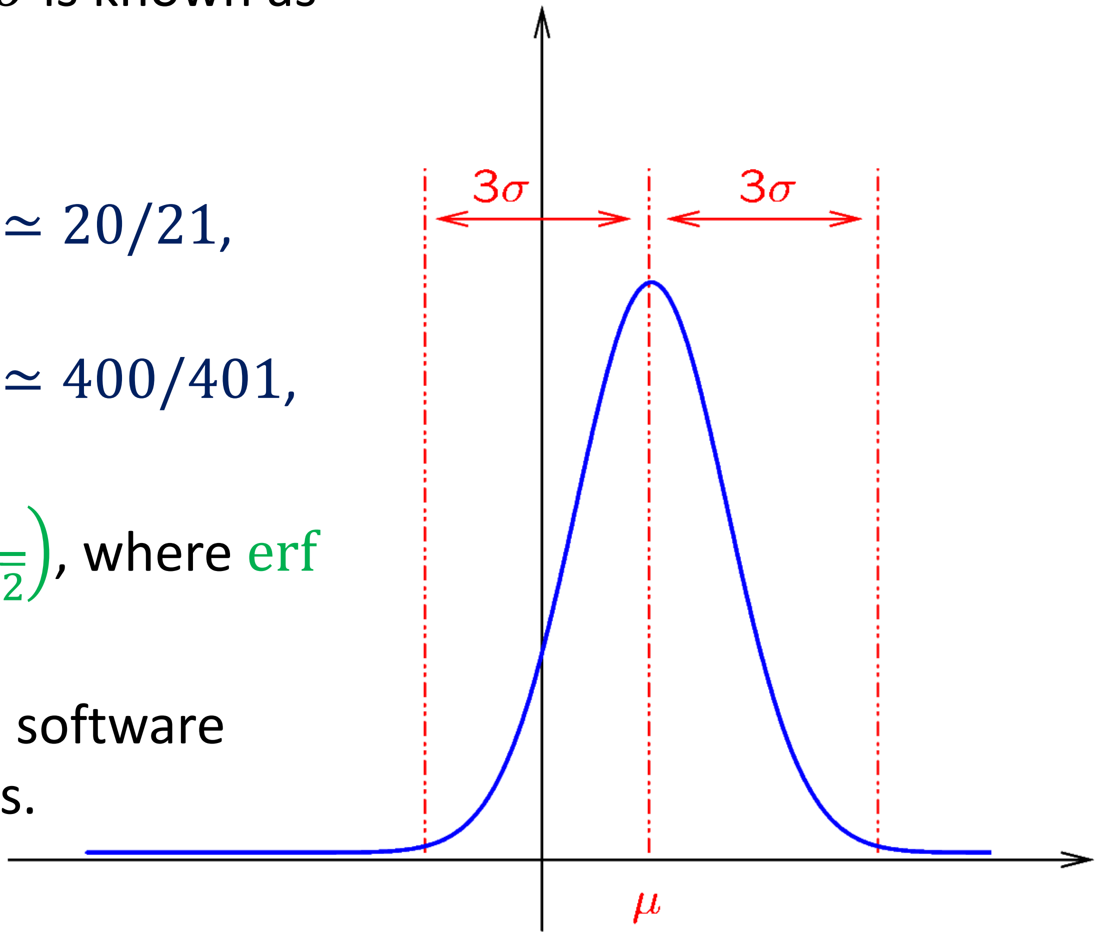
$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

  where μ is the mean and σ the standard deviation.
- The probability that x deviates no more than δ from its mean μ is $\mathrm{erf}\left(\frac{\delta}{\sigma\sqrt{2}}\right)$, where erf is the *error function*.



The Normal Distribution

- The probability that x deviates by no more than δ from its mean μ is
  - 50% when $\delta = 0.675\sigma$. Note that $\pm 0.675\sigma$ is known as the *probable error*.
  - 68.27% when $\delta = \sigma$.
  - 95.45% when $\delta = 2\sigma$. (Note that $95.45\% \simeq 20/21$, that is, $20/1$ odds.)
  - 99.73% when $\delta = 3\sigma$. (Note that $99.73\% \simeq 400/401$, that is, $400/1$ odds.)
  - Can be calculated with the formula $\mathrm{erf}\left(\frac{\delta}{\sigma\sqrt{2}}\right)$, where erf is the *error function.*
  - The error function is implemented in math software (such as MATLAB) and advanced calculators.



Estimating Mean and Standard Deviation

Given n samples $x_1, x_2, \ldots, x_n$ of a random variable x:

- The mean of x can be estimated as the sample mean:

$$\mu \approx \overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- The standard deviation can be estimated with the formula:

$$\sigma \approx S = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots (x_n - \overline{x})^2}{n-1}}$$

Example 1

Five measurements of a DC voltage v_a *are* $0.2, 0.7, 1.4, 2.2, 3.0\ mV$. *Estimate the voltage and the standard deviation of the measurement error.*

- *The mean of the five samples is* $\frac{0.2+0.7+1.4+2.2+3}{5} = 1.5\ mV$. *This is the estimate of* v_a.
- *The standard deviation is*

$$\sigma \simeq \sqrt{\frac{(0.2-1.5)^2+(0.7-1.5)^2+(1.4-1.5^2)+(2.2-1.5)^2+(3.0-1.5)^2}{5-1}} = 877.5\ \mu V.$$

Example 2

Five measurements of a DC voltage v_b *are* $1.5, 1.5, 1.5, 1.4, 1.6\ mV$. *Estimate the voltage and the standard deviation of the measurement error.*

- *The estimate of* v_b *is* $\frac{1.5+1.5+1.5+1.4+1.6}{5} = 1.5\ mV$.
- *The standard deviation is*

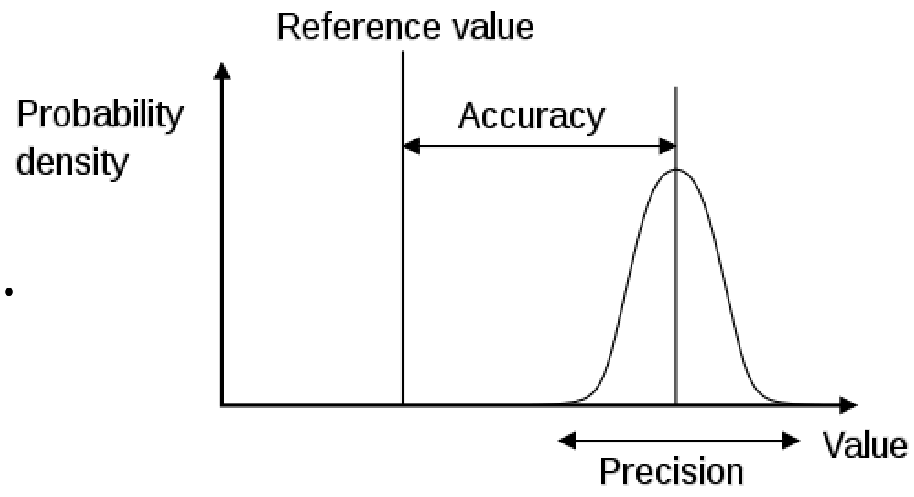
$$\sigma \simeq \sqrt{\frac{(1.5-1.5)^2+(1.5-1.5)^2+(1.5-1.5^2)+(1.4-1.5)^2+(1.6-1.5)^2}{5-1}} = 70.7\ \mu V.$$

Comparing Example 1 and Example 2:

- While the mean is the same, the standard deviation is smaller in Example 2.
- This is because the samples are not as spread out as in the first example.
- *The standard deviation measures the spread of the data about the mean value.*

Accuracy and Precision

- Measurement errors often have distributions that are approximately *normal*.
- Assuming a normal distribution, the probability that x deviates by no more than $\delta = 3\sigma$ from its mean μ is 99.73%.
  - This means that virtually all samples of x will be in the interval $\mu - 3\sigma \ldots \mu + 3\sigma$.
- The *precision* of a measurement system specifies the degree to which repeated measurements show the same results.
- Precision is related to σ, the standard deviation.
- *Accuracy* corresponds to the difference between the true value and the measured value.



The relation between the precision and accuracy, as illustrated in Wikipedia.
https://en.wikipedia.org/wiki/Accuracy_and_precision

Example 3

Two series resistors have the values $1\ k\Omega \pm 2\%$ *and* $2\ k\Omega \pm 5\%$*. Find the uncertainty of the total resistance in ohms.*

- $R_1 = 1\ k\Omega \pm 2\%$, *that is,* $1\ k\Omega \pm 20\ \Omega$.
- $R_2 = 2\ k\Omega \pm 5\%$, *that is,* $2\ k\Omega \pm 100\ \Omega$.
- The total resistance is $R = R_1 + R_2$, *that is,* $1\ k\Omega + 2\ k\Omega \pm 20\ \Omega \pm 100\ \Omega$.
- *Thus,* $R = 3\ k\Omega \pm 120\ \Omega$. *The uncertainty is* $\pm 120\ \Omega$.

Example 4

A resistor R *in series with a* $1\ k\Omega \pm 2\%$ *should result in a resistance of* $3\ k\Omega$ *with an uncertainty of no more than* $\pm 5\%$*. What is the maximum permissible uncertainty of* R*?*

- *Let* $R_2 = 1\ k\Omega \pm 2\%$*, that is,* $1\ k\Omega \pm 20\ \Omega$*.*
- *Let* $R_t = 3\ k\Omega \pm 5\%$*, that is,* $3\ k\Omega \pm 150\ \Omega$*.*
- *Note that* $R = R_t - R_2$*, that is,* $3\ k\Omega - 1\ k\Omega \pm 150\ \Omega \mp 20\ \Omega$*.*
- *Note that when numbers are subtracted, their uncertainties are added.*
- *Thus,* $R = 2\ k\Omega \pm 170\ \Omega$*. The maximum uncertainty is* $\pm 8.5\%$*.*

Example 5

A quantity x can be measured with the uncertainty $\pm\delta_0$. Assuming n measurements of x, determine the uncertainty of the mean of the n measurements.

- *Let $x_1, x_2, \ldots x_n$ be the measured values.*
- *Let x_t be the true value of x.*
- *Note that $x_1 = x_t \pm \delta_0$, $x_2 = x_t \pm \delta_0$, … , $x_n = x_t \pm \delta_0$.*
- *Substituting in the equation $\mu = (x_1 + x_2 + \cdots x_n)/n$,*

$$\mu = \frac{x_t \pm \delta_0 + x_t \pm \delta_0 + \cdots + x_t \pm \delta_0}{n} = x_t \pm \delta_0$$

- It may be surprising to see that the uncertainty of the mean is not smaller.
- This is because we found the *worst-case uncertainty*, not the *likely uncertainty*.

The Kline-McClintock Formula

- The *likely uncertainty* can be found with the Kline-McClintock formula.
- Let z be a variable that can be calculated based on the measurements of n other variables $x_1, x_2, \dots, x_n$.
- Let $z = f(x_1, x_2, \dots, x_n)$ be the equation used to find z, where f is a function.
- Let $\pm\delta_1$, $\pm\delta_2$, …, $\pm\delta_n$ be the uncertainties of the measurements of $x_1, x_2, \dots, x_n$.
- The uncertainty $\pm\delta$ of z is found with the equation

$$\delta = \sqrt{\left(\delta_1 \frac{\partial f}{\partial x_1}\right)^2 + \left(\delta_2 \frac{\partial f}{\partial x_2}\right)^2 + \cdots + \left(\delta_n \frac{\partial f}{\partial x_n}\right)^2}$$

The Kline-McClintock Formula

- The uncertainty $\pm\delta$ of z is found with the equation

$$\delta = \sqrt{\left(\delta_1 \frac{\partial f}{\partial x_1}\right)^2 + \left(\delta_2 \frac{\partial f}{\partial x_2}\right)^2 + \cdots + \left(\delta_n \frac{\partial f}{\partial x_n}\right)^2}$$

- Note that $\frac{\partial f}{\partial x}$ denotes the derivative of f with respect to x.
- If $\pm\delta_1$, $\pm\delta_2$, …, $\pm\delta_n$ have the same odds, then the uncertainty $\pm\delta$ of z has the odds of $\pm\delta_1$, $\pm\delta_2$, …, $\pm\delta_n$.

Example 6

A quantity x can be measured with the 20/1 odds uncertainty $\pm\delta_0$. Assuming n measurements of x, determine the 20/1 odds uncertainty of the mean.

- *Let $x_1, x_2, \dots x_n$ be the measured values.*
- *The mean is calculated with the equation*

$$\mu = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- *Note that $\mu = f(x_1, x_2, \dots, x_n)$, where*

$$f(x_1, x_2, \dots, x_n) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- *The derivatives are* $\frac{\partial f}{\partial x_1} = \frac{1}{n}, \ \frac{\partial f}{\partial x_2} = \frac{1}{n}, \dots, \ \frac{\partial f}{\partial x_n} = \frac{1}{n}.$

Example 6 (continued)

$$\delta = \sqrt{\left(\delta_1 \frac{\partial f}{\partial x_1}\right)^2 + \left(\delta_2 \frac{\partial f}{\partial x_2}\right)^2 + \cdots + \left(\delta_n \frac{\partial f}{\partial x_n}\right)^2}$$

$$= \sqrt{\left(\delta_0 \frac{1}{n}\right)^2 + \left(\delta_0 \frac{1}{n}\right)^2 + \cdots + \left(\delta_0 \frac{1}{n}\right)^2}$$

- *We conclude that the mean has the 20/1 odds uncertainty* $\pm\delta$ *where*

$$\delta = \frac{\delta_0}{\sqrt{n}}$$

Example 7

Assume two blocks of mass $m_1 = 1 \pm 0.02\ kg$ *and* $m_2 = 2 \pm 0.03\ kg$*. Assuming 40/1 odds uncertainties, find the 40/1 odds uncertainty of the total mass.*

- *Note that the uncertainties of* m_1 *and* m_2 *are* $\pm\delta_1$ *and* $\pm\delta_2$*, where* $\delta_1 = 0.02\ kg$ *and* $\delta_2 = 0.03\ kg$*.*
- *The total mass is* $m = m_1 + m_2$*.*
- *Note that* $m = f(m_1, m_2)$*, where* $f(m_1, m_2) = m_1 + m_2$*.*
- *The derivatives are* $\frac{\partial f}{\partial m_1} = 1$ *and* $\frac{\partial f}{\partial m_2} = 1$*.*

$$\delta = \sqrt{\left(\delta_1 \frac{\partial f}{\partial m_1}\right)^2 + \left(\delta_2 \frac{\partial f}{\partial m_2}\right)^2}$$

Example 7 (continued)

- *Substituting* $\frac{\partial f}{\partial m_1} = 1$ *and* $\frac{\partial f}{\partial m_2} = 1$,

$$\delta = \sqrt{\delta_1^2 + \delta_2^2}$$

$$= \sqrt{0.02^2 + 0.03^2}$$

$$= 0.036\ kg$$

- *Note that the* $\delta = 0.036\ kg$ *is considerably smaller than* $\delta_1 + \delta_2 = 0.05\ kg$.

Example 8

A resistance R *is determined as the ratio of a* $1 \pm 0.01\ V$ *voltage and a* $10 \pm 0.05\ A$ *current, both with 20/1 odds uncertainties. Find* R *and its 20/1 odds uncertainty.*

- *The voltage and the current have the uncertainties* $\pm\delta_v = \pm 0.01\ V$ *and* $\pm\delta_i = \pm\ 0.05\ A$.
- *The resistance is calculated with the formula* $R = v/i$.
- *Note that* $R = f(v,i)$, *where* $f(v,i) = v/i$.
- *The derivatives are* $\frac{\partial f}{\partial v} = \frac{1}{i}$ *and* $\frac{\partial f}{\partial i} = -\frac{v}{i^2}$.

$$\delta = \sqrt{\left(\delta_v \frac{\partial f}{\partial v}\right)^2 + \left(\delta_i \frac{\partial f}{\partial i}\right)^2}$$

Example 8 (continued)

- *Substituting* $\frac{\partial f}{\partial v} = \frac{1}{i}$ *and* $\frac{\partial f}{\partial i} = -\frac{v}{i^2}$,

$$\delta = \sqrt{\left(\delta_v \frac{1}{i}\right)^2 + \left(\delta_i \left(-\frac{v}{i^2}\right)\right)^2}$$

$$= \sqrt{\left(0.01 \frac{1}{10}\right)^2 + \left(0.05 \left(-\frac{1}{10^2}\right)\right)^2}$$

$$= 0.00112\ \Omega$$

- *Since* $R = \frac{v}{i} \pm \delta$, *we obtain* $R = 100 \pm 1.12\ m\Omega$.

Example 9

A resistance R is determined as the ratio of a $1 \pm 0.01\ V$ voltage and a $10 \pm 0.05\ A$ current, where the uncertainty of the voltage has 400/1 odds and the uncertainty of the current 20/1 odds. Find R and its 400/1 odds uncertainty. Assume normal distributions.

- *This example could be solved the same way as the previous example once we find the 400/1 uncertainty of the current.*
- *Let $\pm\delta_i' = \pm 0.05\ A$ be the 20/1 odds uncertainty of the current.*
- *Note that 20/1 odds correspond to a probability of 20/21.*

$$\frac{20}{21} = \operatorname{erf}\left(\frac{\delta_i'}{\sigma\sqrt{2}}\right) \Rightarrow \sigma = \frac{\delta_i'}{\operatorname{erf}^{-1}(20/21)\sqrt{2}} \Rightarrow \sigma = 0.0252\ A$$

Example 9 (continued)

- *Let* $\pm\delta_i$ *be the 400/1 uncertainty of the current.*

$$\frac{400}{401} = \operatorname{erf}\left(\frac{\delta_i}{\sigma\sqrt{2}}\right) \Rightarrow \delta_i = \sigma\sqrt{2}\,\operatorname{erf}^{-1}\left(\frac{400}{401}\right) \Rightarrow \delta_i = 0.0763\ A$$

- *As shown in the previous example,*

$$\delta = \sqrt{\left(\delta_v \frac{1}{i}\right)^2 + \left(\delta_i\left(-\frac{v}{i^2}\right)\right)^2}$$

$$= 0.00126\ \Omega$$

- *Since* $R = \frac{v}{i} \pm \delta$*, we obtain* $R = 100 \pm 1.26\ m\Omega$.