

PHYSICS FOR ENGINEERING STUDENTS

Measurement uncertainty

Direct measurements: from resolution to statistical uncertainty

Lecture 4 — Book-style student notes

Learning focus: Resolution intervals · Random fluctuations · Sample statistics · Normal distribution

University of Szeged · Institute of Physics

Measurement Uncertainty in Direct Measurements

No measurement determines an exact true value. The observed indication is influenced by finite resolution, random fluctuations, calibration, environmental conditions, the measurement procedure, and the mathematical model. Measurement uncertainty expresses the remaining dispersion of values that could reasonably be attributed to the measurand.

Learning objectives

- Explain the difference between measured value, error, and uncertainty.
- Identify resolution, random, and systematic contributions.
- Represent instrument resolution as an interval and evaluate its standard uncertainty.
- Describe repeated measurements as a statistical sample.
- Calculate the sample mean, variance, and standard deviation.
- Connect histograms with probability-density functions.
- Recognize the normal distribution and interpret its parameters.
- Combine resolution and repeatability contributions for a direct measurement.

Central principle: Uncertainty quantifies the quality of knowledge about the measurand. It accompanies the result and does not disappear merely because the instrument gives many digits.

Further interpretation

A complete result combines an estimate, an uncertainty statement, a unit, and a definition of the measurand. These elements are inseparable. Two laboratories may report the same numerical estimate but measure slightly different quantities because their temperature, bandwidth, spatial averaging, or correction conventions differ. Uncertainty is therefore conditional on the measurand definition and the measurement model. It does not describe all imaginable doubt; it describes the effects considered relevant under stated conditions. A transparent report names major contributions and assumptions so another reader can judge whether the result is suitable for a new purpose.

1. Measured value, true value and uncertainty

The true quantity value is the value consistent with the definition of the measurand. In practice it is unknown. A measurement produces an estimate x and associated uncertainty $u(x)$, which summarizes the dispersion of plausible values under the stated model and information.

$$X = x \pm U$$

The $\pm$ notation is shorthand and must be interpreted carefully. If U is an expanded uncertainty, the coverage factor and approximate coverage probability should be stated. If u is a standard uncertainty, it behaves like a standard deviation. Neither notation means that the true value is guaranteed to lie in the interval.

Measurement error is conceptually $e = x - X_{\text{true}}$. Since X_{true} is unknown, the exact error is also unknown. Known significant systematic effects should be corrected; uncertainty then describes doubt in the correction and remaining effects.

$$x_{\text{corrected}} = x_{\text{indicated}} + C$$

Define the measurand

Uncertainty evaluation begins with a precise statement of what is measured. “Length of the rod” is incomplete if temperature, contact force, orientation, or reference points materially affect the value.

Language: Do not use “uncertainty” and “mistake” as synonyms. A blunder should be detected and corrected; uncertainty characterizes the result of a valid measurement process.

Further interpretation

Corrections deserve special attention. If a known zero offset is ignored, the result remains biased even when the uncertainty is made large. The proper procedure is to apply the best available correction and assign an uncertainty to that correction. A correction can be zero while its uncertainty is nonzero—for example, when calibration shows no statistically significant offset but has limited precision. Reference values also have uncertainty, so calibration cannot reveal the exact true error of the instrument. The corrected indication is an estimate obtained from a chain of imperfect information, which is precisely why an uncertainty budget is needed.

2. Sources of measurement uncertainty

The lecture groups uncertainty contributions into three practical categories: measurement resolution, random fluctuations, and systematic effects. The categories help organize reasoning, although real contributions can interact.

Contribution	Typical manifestation	Treatment
Resolution	Indication changes in finite steps	Interval model
Random fluctuation	Repeated values scatter unpredictably	Statistical analysis
Systematic effect	Results shift consistently	Correction and uncertainty

For aluminium-foil thickness, resolution may come from the micrometer graduation or digital increment. Random variation may arise from positioning, local surface variation, operator force, vibration, or temperature. Systematic effects may include zero offset, calibration scale error, nonparallel anvils, compression of the foil, or sampling a nonrepresentative location.

A measurement model makes these effects explicit. If thickness t is obtained from an indication d , zero correction c_0 , and thermal correction c_T , then

$$t = d + c_0 + c_T$$

Each input quantity carries uncertainty. The model prevents double counting and shows which effects require control, correction, repeated observations, or additional information.

Cause-and-effect thinking: Build the uncertainty analysis from the physical measurement process, not from a generic list. Every included contribution should have a mechanism.

Further interpretation

Cause-and-effect diagrams are useful for organizing sources without double counting. Starting from the measurand, branches can represent calibration, resolution, repeatability, environment, operator, sampling, and model assumptions. Each branch should connect to an input quantity in the measurement equation wherever possible. Contributions that are already reflected in repeated scatter should not automatically be added again as separate terms. Conversely, a stable effect such as calibration bias will not appear in short-term repeatability and needs an independent evaluation. The classification into random and systematic is descriptive; combination is based on standard uncertainties and correlations, not on simply adding category labels.

3. Instrument and measurement resolution

Instrument resolution Δx_{inst} is the smallest indication increment or distinguishable change provided by the instrument. Measurement resolution Δx_{meas} is the effective resolution achieved in the complete procedure. They may be equal, but reading method, noise, interpolation, display stability, and experimental protocol can make them different.

An analog scale may permit interpolation between graduations when the pointer and scale are sufficiently sharp. Conversely, an unstable digital display may require rounding several fluctuating final digits, so the effective measurement resolution is coarser than the nominal least significant digit.

Situation	Relationship	Reason
Stable digital display	$\Delta x_{\text{meas}} \approx \Delta x_{\text{inst}}$	One count defines the interval
Readable analog scale	$\Delta x_{\text{meas}} < \Delta x_{\text{inst}}$	Defensible interpolation
Noisy indication	$\Delta x_{\text{meas}} > \Delta x_{\text{inst}}$	Final digits are not stable
Coarse protocol	$\Delta x_{\text{meas}} > \Delta x_{\text{inst}}$	Procedure limits repeatability

Resolution should be determined from the actual indication under measurement conditions. Manufacturer specifications, screenshots, and the number of displayed digits are evidence, but they do not by themselves establish effective resolution.

Key distinction: Nominal instrument resolution is a device property; effective measurement resolution belongs to the whole measurement procedure.

Further interpretation

Resolution is often confused with precision because both affect displayed digits. A four-digit display can be highly repeatable but poorly calibrated, and a coarse scale can hide excellent underlying stability. Effective resolution also depends on observation time: averaging a noisy digital signal may reveal changes smaller than one instantaneous count if the quantization is sufficiently dithered, whereas a perfectly stable quantizer cannot reveal sub-count changes without additional information. Interpolation on an analog scale is justified only when pointer width, parallax, scale quality, and observer consistency support it. The claimed resolution should therefore be an empirical property of the procedure, not a cosmetic property of the display.

4. Resolution defines an interval

If a reported value x is rounded to the nearest increment Δx_{meas} , the unresolved value can lie approximately from $x - \Delta x_{\text{meas}}/2$ to $x + \Delta x_{\text{meas}}/2$. With no further information, all positions in this interval are often modeled as equally likely.

$$x - \Delta x_{\text{meas}}/2 \leq X < x + \Delta x_{\text{meas}}/2$$

The corresponding probability model is rectangular (uniform). For an interval with half-width $a = \Delta x_{\text{meas}}/2$, the standard deviation is $a/\sqrt{3}$. Therefore the standard uncertainty due to rounding or resolution is

$$u_{\text{res}} = \Delta x_{\text{meas}}/\sqrt{12} = \Delta x_{\text{meas}}/(2\sqrt{3})$$

Example: a digital ruler increments by 0.1 mm and the stable reading is 24.7 mm. Modeling rounding as rectangular gives $u_{\text{res}} = 0.1/\sqrt{12} = 0.0289$ mm, normally reported as 0.029 mm.

Some contexts define “resolution uncertainty” using the full increment divided by $\sqrt{3}$. That corresponds to assuming a half-width equal to the stated resolution. The convention must therefore be explicit: is Δx the full interval width or its half-width?

Convention: This note uses Δx_{meas} as the full rounding interval width, giving $u_{\text{res}} = \Delta x_{\text{meas}} / \sqrt{12}$. Always verify the definition used in a course, instrument manual, or laboratory protocol.

Further interpretation

The rectangular model is a statement of knowledge, not necessarily a claim that the physical error literally fluctuates uniformly. It expresses that any value within the rounding interval is considered equally plausible when no finer information exists. Different information can justify a triangular, normal, or other distribution and will produce a different standard uncertainty for the same limits. The interval boundaries also depend on the rounding rule. If an instrument truncates rather than rounds, the interval is not centred on the displayed value and a correction may be required. Establishing the actual digital behaviour is therefore part of the uncertainty evaluation.

5. Random fluctuations

Random fluctuations are unpredictable changes in repeated indications under nominally identical conditions. They arise from many small effects whose individual values are not controlled: electronic noise, vibration, microscopic positioning changes, turbulent flow, thermal variation, or observer judgement.

Random error does not mean careless work and does not necessarily shift the expected value. It broadens the distribution of results. Repetition reveals this variability and allows it to be characterized statistically.

Two experimental regimes

- Fluctuations smaller than resolution: repeated indications may be identical, so resolution sets the observable spread.
- Fluctuations larger than resolution: repeated indications vary, and sample statistics characterize the spread.

Higher resolution can reveal variation that a coarse instrument hides. Identical ruler readings do not prove the object is perfectly constant; a micrometer may expose real thickness variation or procedure-related scatter. Better resolution can therefore increase the observed variability while improving knowledge.

Averaging reduces the random uncertainty of the mean when observations are independent and drawn from a stable process. It does not correct calibration bias or drift.

Diagnostic question: If repeated results do not vary, ask whether the process is genuinely stable or whether the instrument is simply unable to resolve the variation.

Further interpretation

Random variation can be reduced at its source through shielding, vibration isolation, temperature control, stable timing, improved positioning, and better experimental procedure. Averaging is helpful only after these physical controls have been considered. Correlated observations contain less independent information than their number suggests; rapid repeated readings of a slowly drifting sensor may not achieve the expected $1/\sqrt{N}$ reduction. Time-series plots and autocorrelation can reveal this problem. Increasing the interval between readings may improve independence but also allow environmental change. Sampling design must balance these effects and should reflect the time scales of both the measurand and the disturbance.

6. Repeated measurements as a sample

A set of N repeated measurements $x_1, x_2, \dots, x_n$ is a sample from the population of possible outcomes generated by the measurement process. The population includes hypothetical repetitions under the specified conditions; it is characterized by parameters such as a mean μ and standard deviation σ .

$$\{x_1, x_2, \dots, x_N\}$$

The observations should be sufficiently independent and obtained under clearly stated repeatability conditions: same method, instrument, operator, location, and short time interval, as appropriate. If the measurand drifts during collection, the sample may combine random variability with a systematic time trend.

Before calculation

- Plot the sequence against measurement number.
- Check units, transcription, and instrument status.
- Investigate outliers physically; do not delete them automatically.
- Record environmental conditions and procedural changes.
- Decide whether the measurand is stable enough for one sample summary.

A histogram reorganizes the sample into bins and displays its empirical distribution. The original sequence should also be retained because a histogram hides time order. Two datasets can have the same histogram but very different drift or correlation patterns.

Statistical validity: Formulas assume a defined data-generating process. A larger sample cannot rescue measurements collected under changing or uncontrolled conditions without an appropriate model.

Further interpretation

A sample should represent the conditions to which the result will be generalized. Repeating one location on a nonuniform sheet estimates local repeatability, not average sheet thickness. Measuring many locations introduces spatial sampling variation, which may be part of the measurand definition rather than measurement noise. Similarly, repeated timing by one operator does not quantify between-operator reproducibility. The measurement plan should specify which factors remain fixed and which are deliberately varied. Statistical summaries are meaningful only relative to this design. Randomization can prevent slow drift from becoming confounded with sample order, while replication allows components of variability to be separated.

7. The sample mean

The arithmetic sample mean is the sum of the observations divided by their number. It estimates the central location or expected value of the measurement distribution.

$$\bar{x} = (1/N) \sum_{i=1}^N x_i$$

For five period measurements 40.12, 40.08, 40.15, 40.10, and 40.05 s, the sum is 200.50 s and the sample mean is 40.10 s. The unit of the mean is the same as the observations.

The mean uses all values and is mathematically convenient, but it is sensitive to outliers and may be misleading for a strongly skewed or multimodal distribution. Median and robust summaries can be useful complements, although they answer slightly different questions.

Mean of repeated readings

When repeated observations estimate one stable measurand, the mean is usually a better estimate than a single reading. Under independence with population standard deviation σ , the standard deviation of the mean is $\sigma/\sqrt{N}$.

$$\overline{\sigma_x} = \sigma/\sqrt{N}$$

The square-root law gives diminishing returns: reducing random uncertainty by a factor of two requires four times as many independent observations.

Limitation: A precise mean can still be inaccurate if the entire sample is shifted by an uncorrected systematic effect.

Further interpretation

The sample mean is optimal under common assumptions of independent observations with a stable mean and finite variance, but those assumptions must be checked. A trend makes early and late readings systematically different; a warm-up transient can shift the first observations; a bimodal sample may represent two operating states whose average describes neither. Robust estimators reduce sensitivity to isolated extremes but do not resolve an unidentified mixture. Plotting measurements in acquisition order, grouping by relevant conditions, and understanding the apparatus should precede automatic averaging. The goal is not always to obtain the smallest uncertainty; it is to estimate the defined measurand without concealing physically meaningful variation.

8. Deviations, variance and standard deviation

The deviation $d_i = x_i - \bar{x}$ describes how far each observation lies from the sample mean. Positive and negative deviations cancel exactly, so their ordinary average cannot measure spread.

$$d_i = x_i - \bar{x}, \quad \sum_i d_i = 0$$

Squaring the deviations prevents cancellation. The unbiased sample variance divides their sum by $N-1$ because the sample mean has already been estimated from the same data.

$$s^2 = [\sum_{i=1}^N (x_i - \bar{x})^2] / (N - 1)$$

The sample standard deviation is the positive square root of the variance.

$$s = \sqrt{s^2}$$

Variance has squared units and is important in statistical theory. Standard deviation has the original measurement unit and is easier to interpret as a typical scale of scatter. Neither should be confused with the standard uncertainty of the mean, which is smaller by $\sqrt{N}$ when the assumptions are satisfied.

$$u_A(\bar{x}) = s/\sqrt{N}$$

Denominator: Use $N-1$ for the sample variance when estimating population variability from a finite sample. Dividing by N describes the observed set itself but is biased as an estimator of σ^2 .

Further interpretation

The $N-1$ denominator accounts for the degree of freedom used to estimate the mean. Once the deviations are formed around $\bar{x}$, only $N-1$ of them can vary independently because their sum is zero. For very small samples, s is itself uncertain and confidence intervals for the mean should use Student's t distribution rather than a normal critical value when normal sampling is assumed. The standard deviation describes individual observations, while $s/\sqrt{N}$ describes the estimated mean.

Reporting one in place of the other can change the apparent precision by a large factor, so the statistic represented by every uncertainty value must be named explicitly.

9. Alternative variance formula and worked example

Expanding the squared deviations gives a computationally equivalent form that uses the sums of x_i and x_i^2 .

$$s^2 = [\Sigma x_i^2 - (\Sigma x_i)^2/N]/(N - 1)$$

For period data 40.12, 40.08, 40.15, 40.10, and 40.05 s, the mean is 40.10 s. The deviations are 0.02, −0.02, 0.05, 0.00, and −0.05 s. Their squared sum is 0.0058 s².

$$s^2 = 0.0058/4 = 0.00145 \text{ s}^2$$

$$s = 0.0381 \text{ s}, \quad u_A(\bar{x}) = 0.0381/\sqrt{5} = 0.0170 \text{ s}$$

A statistically consistent preliminary report is therefore $\bar{T}=(40.100\pm0.017) \text{ s}$ when the uncertainty shown is the standard uncertainty of the mean and only repeatability is included. After rounding the uncertainty to two significant digits, the mean is rounded to the same decimal place.

Direct evaluation of the alternative formula can lose numerical precision when values are large and their spread is very small. Modern software usually uses numerically stable algorithms based on centered or incremental calculations.

Check: Verify that deviations sum to approximately zero and that the reported uncertainty and mean use the same decimal place.

Further interpretation

A worked calculation should retain enough intermediate precision to reproduce the final result. A deviations table is useful for teaching because it exposes signs, squares, and the zero-sum check. In software, the two-pass centred calculation is usually safer than subtracting two nearly equal large sums in the alternative formula. Data entry should also be validated: one misplaced decimal point can dominate a variance. If an observation is suspected, analyse the physical record and show results with and without it rather than deleting it silently. A statistical outlier test is meaningful only when its distributional assumptions and significance criterion were chosen appropriately.

10. From histogram to probability density

As the number of repeated observations increases, a normalized histogram often becomes smoother and reveals the underlying distribution more clearly. If the bin width is decreased appropriately while the sample grows, the histogram can approximate a continuous probability-density function $p(x)$.

$$p(x) \geq 0, \quad \int_{-\infty}^{\infty} p(x) dx = 1$$

Probability over an interval is represented by area under the density curve.

$$P(a \leq X \leq b) = \int_a^b p(x) dx$$

The density $p(x)$ is not itself a probability at one exact value and may exceed one when the unit interval is small; its unit is the inverse of the measured quantity. The integral, not the height, is dimensionless probability.

The limiting-distribution idea separates the finite observed sample from the statistical model that could generate it. A smooth curve should not be chosen solely because it looks attractive. The model must be supported by the measurement mechanism, plots, and diagnostic checks.

Connection: A density histogram estimates probability density: each bar area, not merely its height, estimates the probability assigned to that bin.

Further interpretation

A probability density is a model for potential observations, while a histogram is a finite-data summary. Agreement should be evaluated with more than visual resemblance: quantile plots, cumulative distributions, residual patterns, and knowledge of the generation mechanism all contribute evidence. Increasing N reduces sampling noise, but it can also reveal small departures from a simple model that were previously invisible. No real dataset becomes exactly continuous; the continuous density is an abstraction that supports probability calculations. Instrument resolution may leave visible discretization even for large samples, and that feature should be modeled or acknowledged rather than smoothed away automatically.

11. The normal distribution

Many measurement distributions are approximately normal when the observed deviation is the sum of numerous small, independent effects. The normal distribution is symmetric and determined completely by its mean μ and standard deviation σ .

$$p(x) = [1/(\sigma\sqrt{2\pi})] \exp\{-(x - \mu)^2/(2\sigma^2)\}$$

The mean locates the centre, and the standard deviation controls the width. Approximately 68.27% of probability lies within $\mu \pm \sigma$, 95.45% within $\mu \pm 2\sigma$, and 99.73% within $\mu \pm 3\sigma$.

Interval	Approximate probability	Interpretation
$\mu \pm \sigma$	68.27%	Typical central range
$\mu \pm 2\sigma$	95.45%	Broad central range
$\mu \pm 3\sigma$	99.73%	Extreme observations are rare

Normality is not guaranteed. Resolution effects can create rectangular or discrete distributions; physical lower bounds can cause skewness; mixed operating states can produce multiple peaks; and outliers can produce heavy tails. A normal model should therefore be checked rather than assumed automatically.

Parameter distinction: Population parameters μ and σ are theoretical. Sample statistics $\bar{x}$ and s estimate them from finite data and therefore carry sampling uncertainty.

Further interpretation

The normal distribution arises naturally for sums of many small effects, but the central-limit argument does not require every individual observation distribution to be normal. It concerns the behaviour of sums or means under conditions. Measurement data can depart from normality because of thresholds, saturation, multiplicative noise, bounded quantities, intermittent disturbances, or mixed populations. The familiar 68–95–99.7 percentages should not be applied mechanically to such cases. A normal probability plot can reveal skewness and heavy tails, while a physical explanation is needed to decide whether transformation, robust analysis, or a different probability model is appropriate.

12. Random uncertainty of a direct result

For a direct measurement based on the mean of repeated readings, a practical random standard uncertainty can include repeatability and finite resolution. If these contributions are independent, they are combined by the root-sum-of-squares rule.

$$u_c(\bar{x}) = \sqrt{u_A(\bar{x})^2 + u_{\text{res}}^2}$$

Using the period example, $u_a = 0.0170$ s. If the timer resolution is $\Delta T = 0.01$ s, the rectangular resolution contribution is $u_{\text{res}} = 0.01/\sqrt{12} = 0.00289$ s. The combined random standard uncertainty is $= \sqrt{0.0170^2 + 0.00289^2}$ s, so repeatability dominates.

Systematic contributions such as calibration, time-base accuracy, reaction delay, or model corrections must be added separately to the measurement model and combined according to their sensitivities

and correlations. The simple equation above is not a complete uncertainty budget unless those effects are negligible or already included.

Integrated workflow

- Define the measurand and conditions.
- Determine effective resolution.
- Repeat measurements and inspect sequence and histogram.
- Calculate $\bar{x}$, s , and $s/\sqrt{N}$.
- Evaluate systematic effects and corrections.
- Combine standard uncertainties and report meaningful digits.

Take-home message: Repetition reveals random scatter; resolution defines hidden variation; a defensible uncertainty statement combines both with the remaining model and calibration contributions.

Further interpretation

Combining uncertainties by root sum of squares assumes that the represented input effects are independent or uncorrelated after applying their sensitivity coefficients. If two results share the same calibration reference, their uncertainties may be correlated and can partly cancel in a difference or reinforce in a sum. The general propagation law includes covariance terms. An uncertainty budget lists each input estimate, standard uncertainty, distribution or evaluation method, sensitivity coefficient, and contribution. Expanded uncertainty is then obtained only after choosing an appropriate coverage factor. This structured approach makes clear which improvement would matter: reducing a negligible term changes little, whereas addressing the dominant contribution directly improves the final result.