

PHYSICS FOR ENGINEERING STUDENTS

Probability distributions

Random variables and characterization of distributions

Lecture 6 — Book-style student notes

Learning focus: Discrete and continuous variables · CDF and PDF · Location and dispersion
· Moments

University of Szeged · Institute of Physics

Probability distributions

Repeated measurements under nominally identical conditions usually give different values. Probability theory provides a consistent language for describing the set of possible outcomes, their likelihood, and the numerical features of their distribution. It allows measurement variability to be quantified, models to be compared with data, and predictions to be made without pretending that a future observation can be known exactly.

Learning objectives

- Distinguish discrete and continuous random variables.
- Use probability mass, cumulative distribution, and density functions.
- Interpret mean, median, mode, variance, FWHM, and moments.
- Apply the binomial model to repeated independent trials.

Central principle: Choose, evaluate, and report a probability or uncertainty model from the physical measurement process and its assumptions.

Further interpretation

A probability distribution contains more information than a mean and standard deviation. It specifies how probability is allocated across possible values. Summary parameters are useful because they answer focused questions—where the distribution is located, how widely it spreads, whether it is asymmetric, and how heavy its tails are—but they must be interpreted together with the shape and physical origin of the distribution.

1. Random variables

A random variable assigns a numerical value to each outcome of a random experiment. Examples include photon counts in a fixed time, the number of successful trials, a measured voltage, temperature, component lifetime, particle diameter, or detector pulse height. The word random describes limited predictability of the outcome, not absence of physical causes.

We normally denote a random variable by an uppercase letter such as X and a realized value by lowercase x . A model specifies both the set of possible values and their probabilities. The same physical quantity may require different models under different conditions: photon counts may be Poisson at constant rate, overdispersed when the source fluctuates, or limited by detector dead time at high intensity.

Further interpretation

A random variable assigns a numerical value to each outcome of a random experiment. Examples include photon counts in a fixed time, the number of successful trials, a measured voltage, temperature, component lifetime, particle diameter, or detector pulse height. The word random describes limited predictability of the outcome, not absence of physical causes. A random variable is defined by both its possible values and the experiment that generates them. The same numerical quantity can follow different distributions under different sampling rules, detector settings, or populations. A probability model should therefore begin with the measurement mechanism, not with the visual appearance of a histogram.

2. Discrete and continuous variables

A discrete random variable takes a finite or countably infinite set of values. Counts are discrete because 0, 1, 2, ... events can occur, but 2.4 events cannot. A continuous random variable can take any value in an interval, as idealized for length, time, voltage, temperature, or particle size. Real instruments digitize continuous quantities, but a continuous model remains appropriate when the step is negligible relative to the relevant variation.

The distinction determines how probability is represented. A discrete probability mass function assigns probability to individual values. A continuous probability-density function assigns probability to intervals; an exact point has probability zero. Mixed distributions also occur—for example, an instrument can have a finite probability of zero response plus a continuous range of positive responses.

Further interpretation

A discrete random variable takes a finite or countably infinite set of values. Counts are discrete because 0, 1, 2, ... events can occur, but 2.4 events cannot. A continuous random variable can take any value in an interval, as idealized for length, time, voltage, temperature, or particle size. Real instruments digitize continuous quantities, but a continuous model remains appropriate when the step is negligible relative to the relevant variation. Discrete and continuous models are idealizations tied to resolution. A fundamentally continuous quantity may be recorded in digital steps, while a large count can be approximated continuously for calculation. The choice is adequate when it preserves the probabilities and moments relevant to the measurement decision.

3. Discrete probability distributions

For possible values k_j , the probability mass function is $p_j = P(K=k_j)$. Every probability satisfies $0 \leq p_j \leq 1$, and mutually exclusive exhaustive outcomes sum to one. A stem or bar plot places possible values on the horizontal axis and their probabilities on the vertical axis. The gaps emphasize that intermediate values are not possible.

$$p_j = P(K = k_j)$$

$$\sum_j p_j = 1$$

For ten fair coin tosses, the number of heads K can take values from 0 to 10. The probabilities are not equal even though individual toss sequences are equally likely: many more sequences produce five heads than zero heads. This distinction between elementary outcomes and values of a random variable is fundamental. The probability assigned to $K=k$ is the sum of probabilities of all sequences that lead to k .

Further interpretation

For possible values k_j , the probability mass function is $p_j = P(K=k_j)$. Every probability satisfies $0 \leq p_j \leq 1$, and mutually exclusive exhaustive outcomes sum to one. A stem or bar plot places possible values on the horizontal axis and their probabilities on the vertical axis. The gaps emphasize that intermediate values are not possible. For ten fair coin tosses, the number of heads K can take values from 0 to 10. For a discrete model, normalization and support should be checked before any mean or variance is calculated. Values outside the allowed set must have probability zero, and the listed probabilities must sum to one. These simple checks catch many indexing and transcription errors in counting models.

4. Continuous distributions and density

A probability density $f(x)$ is nonnegative and normalized so that its total integral equals one. The probability that X lies between a and b is the area under f between those limits. Density is not probability at a point and can exceed one; its unit is the reciprocal of the unit of X , while an integrated probability is dimensionless.

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

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

For a continuous model, $P(X=x_0)=0$ for every exact value. This does not mean exact readings cannot appear on an instrument; it means an ideal mathematical point has zero width. A displayed value

represents an interval determined by resolution and rounding. Thus the probability of observing that display code is an interval probability, not the probability of a dimensionless point.

Further interpretation

A probability density $f(x)$ is nonnegative and normalized so that its total integral equals one. The probability that X lies between a and b is the area under f between those limits. Density is not probability at a point and can exceed one; its unit is the reciprocal of the unit of X , while an integrated probability is dimensionless. For a continuous model, $P(X=x_0)=0$ for every exact value. A density value is not itself a probability and may exceed one when concentrated over a narrow interval. Probability is obtained from area, so units of a density are the inverse units of the measured variable. Any change of measurement unit rescales the density while preserving interval probabilities.

5. The cumulative distribution function

The cumulative distribution function is $F(x)=P(X\leq x)$. It accumulates probability from the far left up to x and completely specifies the distribution. Every CDF lies between zero and one, is non-decreasing, approaches zero as $x\rightarrow-\infty$, and approaches one as $x\rightarrow+\infty$. Discrete CDFs contain jumps; continuous CDFs are continuous when no point masses are present.

$$F(x) = P(X \leq x)$$

$$P(a < X \leq b) = F(b) - F(a)$$

Interval probabilities follow directly from differences: $P(a < X \leq b) = F(b) - F(a)$. One-sided probabilities are $P(X \leq b) = F(b)$ and $P(X > a) = 1 - F(a)$. Endpoint conventions matter for discrete variables because a point can carry nonzero probability. For a continuous distribution, changing $<$ to $\leq$ does not change the result because point probability is zero.

Further interpretation

The cumulative distribution function is $F(x)=P(X\leq x)$. It accumulates probability from the far left up to x and completely specifies the distribution. Every CDF lies between zero and one, is non-decreasing, approaches zero as $x\rightarrow-\infty$, and approaches one as $x\rightarrow+\infty$. Discrete CDFs contain jumps; continuous CDFs are continuous when no point masses are present. Interval probabilities follow directly from differences: $P(a < X \leq b) = F(b) - F(a)$. One-sided probabilities are $P(X \leq b) = F(b)$ and $P(X > a) = 1 - F(a)$. The cumulative distribution is often the most stable object for comparison because it avoids arbitrary histogram bins. Quantiles, medians, and interval probabilities can be read directly from it. Empirical and theoretical cumulative curves also reveal systematic differences across the full range rather than at selected bin boundaries.

6. Relationship between CDF and PDF

When F is differentiable, the density is its derivative $f(x)=dF/dx$. Conversely, the CDF is the integral of the density from $-\infty$ to x . The density shows how rapidly probability accumulates, whereas the CDF shows the accumulated total. A high density over a narrow region produces a steep CDF; zero density produces a flat segment.

$$f(x) = dF(x)/dx$$

$$F(x) = \int_{-\infty}^x f(t)dt$$

The CDF exists for every random variable, including discrete and mixed cases, while an ordinary PDF need not. This makes the CDF a universal description. In data analysis, an empirical CDF can be constructed without choosing histogram bins and is valuable for comparing distributions, estimating quantiles, and identifying tail differences.

Further interpretation

When F is differentiable, the density is its derivative $f(x)=dF/dx$. Conversely, the CDF is the integral of the density from $-\infty$ to x . The density shows how rapidly probability accumulates, whereas the CDF shows the accumulated total. A high density over a narrow region produces a steep CDF; zero density produces a flat segment. The CDF exists for every random variable, including discrete and mixed cases, while an ordinary PDF need not. Differentiating a smooth cumulative model yields its density, while integrating the density reconstructs the cumulative probability. At discontinuities the cumulative function may jump, representing point probability. This distinction prevents an inappropriate derivative-based treatment of discrete mass or mixed distributions.

7. Mean, median and mode

The mean $\mu=E[X]$ is the probability-weighted average and represents a centre of mass of the distribution. The median divides probability into equal halves, while the mode is a value where probability mass or density is maximal. For a symmetric unimodal distribution they often coincide, but for a skewed distribution they can differ strongly.

$$\mu = E[X] = \sum_j x_j p_j \quad (\text{discrete})$$

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{continuous})$$

The mean is sensitive to rare extreme values and may fail to exist for heavy-tailed distributions. The median remains defined in many such cases and is robust to extremes. The mode describes the most probable local region but can be non-unique in multimodal data. Selecting a ‘typical’ value is therefore a physical and statistical decision, not a purely cosmetic choice.

Further interpretation

The mean $\mu=E[X]$ is the probability-weighted average and represents a centre of mass of the distribution. The median divides probability into equal halves, while the mode is a value where probability mass or density is maximal. For a symmetric unimodal distribution they often coincide, but for a skewed distribution they can differ strongly. The mean is sensitive to rare extreme values and may fail to exist for heavy-tailed distributions. Mean, median, and mode answer different questions. The mean balances all magnitudes, the median divides probability in half, and the mode identifies the most likely region. Their separation is informative about asymmetry, but no universal ordering holds for every multimodal or irregular distribution.

8. Particle size as an example of weighting

A particle number size distribution often appears as $dN/d\log D_p$. Its mode identifies the diameter range containing the largest number concentration per logarithmic interval, the median divides the particle count into equal halves, and the mean is a number-weighted average. A long tail of large particles can pull the mean above the median even when most particles are small.

$$dS \propto D_p^2 dN$$

$$dV \propto D_p^3 dN$$

The same aerosol has different typical diameters when weighted by number, surface area, volume, or mass. For spherical particles, surface weighting introduces approximately D_p^2 and volume weighting D_p^3 , so rare large particles dominate mass even when small particles dominate number. Every reported mean, median, and mode must therefore state the weighting and distribution basis.

Further interpretation

A particle number size distribution often appears as $dN/d\log D_p$. Its mode identifies the diameter range containing the largest number concentration per logarithmic interval, the median divides the particle count into equal halves, and the mean is a number-weighted average. A long tail of large particles can pull the mean above the median even when most particles are small. Particle-size averages depend on whether particles are weighted by number, surface, volume, or mass. Larger particles receive rapidly increasing influence under geometric weighting. The averaging convention is therefore part of the measurand and must be stated before datasets or instruments can be compared.

9. Variance and standard deviation

Variance is the expected squared deviation from the mean: $\sigma^2 = E[(X - \mu)^2]$. It measures spread and is nonnegative. The standard deviation σ is its positive square root and has the same physical unit as X . Two distributions can share mean and variance while having very different skewness, tails, or multiple peaks.

$$\sigma^2 = E[(X - \mu)^2]$$

$$\sigma = \sqrt{\text{Var}(X)}$$

Variance combines the probabilities of all deviations and heavily weights extreme values through squaring. It may not exist for sufficiently heavy-tailed distributions. For transformed variables, variance does not generally transform by simply applying the function to σ . A local approximation uses sensitivity coefficients, while exact propagation requires the full distribution or higher moments.

Further interpretation

Variance is the expected squared deviation from the mean: $\sigma^2 = E[(X - \mu)^2]$. It measures spread and is nonnegative. The standard deviation σ is its positive square root and has the same physical unit as X . Two distributions can share mean and variance while having very different skewness, tails, or multiple peaks. Variance combines the probabilities of all deviations and heavily weights extreme values through squaring. It may not exist for sufficiently heavy-tailed distributions. Variance is sensitive to rare extreme values because deviations are squared. A small variance indicates concentration around the mean only when the mean is itself representative. For heavy-tailed or mixed data, quantile ranges and robust summaries should supplement standard deviation rather than being replaced by it.

10. Full width at half maximum

The full width at half maximum is the distance between the two points where a unimodal curve falls to half its peak value. It is widely used for spectral lines, laser pulses, beam profiles, and experimental distributions because it can be read geometrically. FWHM is a width of the plotted function, not automatically a standard deviation.

$$\text{FWHM}_G = 2\sqrt{(2\ln 2)}\sigma \approx 2.355\sigma$$

The relation between FWHM and σ depends on distribution shape. For a Gaussian, $\text{FWHM} = 2\sqrt{(2\ln 2)}\sigma \approx 2.355\sigma$. For a Lorentzian with half-width parameter γ , $\text{FWHM} = 2\gamma$, while the variance is undefined. Quoting FWHM without naming the line-shape model can therefore prevent comparison with statistical uncertainty.

Further interpretation

The full width at half maximum is the distance between the two points where a unimodal curve falls to half its peak value. It is widely used for spectral lines, laser pulses, beam profiles, and experimental distributions because it can be read geometrically. FWHM is a width of the plotted function, not automatically a standard deviation. The relation between FWHM and σ depends on distribution shape. For a Gaussian, $\text{FWHM} = 2\sqrt{(2\ln 2)}\sigma \approx 2.355\sigma$. FWHM is a shape descriptor, not a universal substitute for standard deviation. The familiar Gaussian conversion applies only to a Gaussian profile. For asymmetric, multimodal, or Lorentzian peaks, the same width can correspond to very different tail behaviour and uncertainty in fitted parameters.

11. Moments, skewness and kurtosis

The s -th raw moment is $m_s = E[X^s]$. The first raw moment is the mean. Central moments use deviations from μ : the second is variance, the third measures signed asymmetry, and the fourth responds strongly to tails. Dimensionless skewness $\gamma_1 = E[(X - \mu)^3]/\sigma^3$ is positive for a longer right tail and negative for a longer left tail.

$$m_s = E[X^s]$$

$$\gamma_1 = E[(X - \mu)^3]/\sigma^3$$

$$\gamma_{2,excess} = E[(X - \mu)^4]/\sigma^4 - 3$$

Kurtosis $\gamma_2 = E[(X - \mu)^4]/\sigma^4$ is often reported as excess kurtosis $\gamma_2 - 3$ so that a normal distribution has zero excess. It should not be interpreted only as peak sharpness; it is strongly influenced by tail weight. Sample estimates of high moments are unstable because a few extremes dominate them, so large datasets and physical interpretation are required.

Further interpretation

The s -th raw moment is $m_s = E[X^s]$. The first raw moment is the mean. Central moments use deviations from μ : the second is variance, the third measures signed asymmetry, and the fourth responds strongly to tails. Dimensionless skewness $\gamma_1 = E[(X - \mu)^3]/\sigma^3$ is positive for a longer right tail and negative for a longer left tail. Kurtosis $\gamma_2 = E[(X - \mu)^4]/\sigma^4$ is often reported as excess kurtosis $\gamma_2 - 3$ so that a normal distribution has zero excess. Higher moments amplify the influence of distant observations and therefore require adequate sample size and stable data quality. Skewness and kurtosis should be interpreted with plots and mechanism-based reasoning. A single outlier can change them dramatically without demonstrating a genuine change in the parent distribution.

12. Bernoulli trials and the binomial model

A Bernoulli trial has two outcomes: success with probability p and failure with $q=1-p$. Repeating n independent trials with constant p and counting successes gives a binomial random variable K . The probability of exactly k successes combines the probability $p^k q^{n-k}$ of one arrangement with the number of arrangements $C(n,k)$.

$$P(K = k) = C(n, k)p^k(1 - p)^{n-k}$$

$$E[K] = np$$

$$\text{Var}(K) = np(1 - p)$$

The binomial mean is np and variance $np(1-p)$. As n grows, the absolute standard deviation increases like $\sqrt{n}$, but relative fluctuation decreases approximately as $1/\sqrt{n}$. The independence and constant-probability assumptions are essential. Detector dead time, changing efficiency, common environmental influences, or depletion can make real trials dependent or give different p values.

Take-home message: The mathematical expression is useful only together with the physical assumptions that justify it and the uncertainty statement that limits its interpretation.

Further interpretation

A Bernoulli trial has two outcomes: success with probability p and failure with $q=1-p$. Repeating n independent trials with constant p and counting successes gives a binomial random variable K . The probability of exactly k successes combines the probability $p^k q^{n-k}$ of one arrangement with the number of arrangements $C(n,k)$. The binomial mean is np and variance $np(1-p)$. A binomial model requires fixed trial count, two outcomes, common success probability, and independence. Departures such as changing efficiency, clustering, or depletion lead to overdispersion or underdispersion. Comparing observed variance with the binomial prediction is therefore an important model check in measurement applications.

13. Integrated measurement workflow and worked applications

The following compact studies connect the definitions and equations to realistic interpretation, calculation, and experimental design decisions. An integrated distribution analysis begins by defining the observation, its support, and the sampling rule. The mechanism then suggests a discrete mass function, a continuous density, or a mixed model. Before fitting parameters, verify normalization, units, and physically impossible values. Summaries should be chosen for the decision: the mean for balance or expected total, the median for a central quantile, the mode for the most probable region, and variance or quantile width for dispersion. Next compare the empirical distribution with the model

using more than one representation. A histogram reveals local structure but depends on bin choice; an empirical cumulative distribution avoids bins and makes quantiles visible; probability or residual plots can expose tail and shape discrepancies. Weighting conventions must be recorded whenever observations represent particles, intensities, volumes, or masses rather than equally weighted experimental units. Finally, use the selected model to calculate the requested interval probability or uncertainty and test its assumptions against independent features of the data. For a binomial process, compare variance with $Np(1-p)$; for a continuous model, inspect boundary effects, asymmetry, multimodality, and tail frequency. Report the model, fitted parameters, estimation method, sample size, and any departures that limit interpretation. This workflow keeps probability calculations connected to the physical experiment rather than treating a familiar distribution as an automatic description of every dataset.

Discrete detector output

A particle counter records 0, 1, 2, ... pulses during a gate interval, so the count is discrete even though arrival time is continuous. A probability mass function assigns probability to each count. Dividing count by gate time produces a rate estimate with many possible rational values, but it remains derived from a discrete count. At high counts a continuous approximation may be convenient, yet the underlying discreteness matters for low-count uncertainty and confidence intervals. This example shows that classification depends on the random variable being modelled, not only on the physical phenomenon.

CDF interval calculation

Suppose a continuous variable has $F(1.5)=0.60$ and $F(-0.5)=0.20$. Then $P(-0.5 < X \leq 1.5)=0.40$. The calculation uses only endpoint values and does not require differentiating the CDF. If the same distribution were discrete and had a point mass at -0.5 , changing the lower endpoint from $<$ to $\leq$ would change the probability. Endpoint notation is therefore mathematically relevant even when it is visually insignificant on a graph. Writing the event explicitly before substituting numbers prevents many sign and complement errors.

Density above one

Consider a uniform variable on an interval of width 0.2 mm. Its density is $1/0.2=5 \text{ mm}^{-1}$, which is greater than one but perfectly valid. The total area is $5 \text{ mm}^{-1} \times 0.2 \text{ mm}=1$. A subinterval of width 0.03 mm has probability 0.15. Confusing density height with probability would incorrectly reject the model. Units provide a useful check: density carries inverse units, whereas probability is dimensionless. Histogram density must be interpreted in the same way—bar area, not height alone, estimates probability.

Mean, median and skewness

In a right-skewed particle distribution, many observations lie near small diameters while a few much larger particles extend the tail. The mode locates the highest concentration, the median divides particle number equally, and the mean is pulled toward the large-diameter tail. Thus $\text{mode} < \text{median} < \text{mean}$ is common. Reporting only the mean can suggest that a “typical” particle is larger than most particles. A physical interpretation should specify whether typical means most frequent, 50th percentile, or arithmetic average, and whether the distribution is number-, area-, or mass-weighted.

Aerosol weighting

Assume two size classes: 100 particles of diameter 50 nm and one particle of diameter 500 nm. Number weighting is dominated by the 50 nm class. For spheres, surface per particle scales as D^2 , so the single large particle has the surface of roughly 100 small particles. Volume and mass scale as D^3 , making the large particle equivalent to roughly 1000 small particles in mass. The same population can therefore have a small number median but a much larger mass median. This is why aerosol size statistics are incomplete unless the weighting basis is explicitly stated.

Gaussian FWHM

A Gaussian peak has measured $\text{FWHM} = 4.71$ nm. The corresponding standard deviation is $\sigma = \text{FWHM} / 2.355 \approx 2.00$ nm. This conversion is valid only if the line shape is Gaussian. If the peak is Lorentzian, $\text{FWHM} = 2\gamma$ and the ordinary standard deviation does not exist. Reporting “ $\sigma = 2.00$ nm” solely from a measured half-maximum width assumes a model that must be justified by the broadening mechanism and residuals. Instrument response can also broaden the observed peak, requiring deconvolution or a convolution model.

Moment sensitivity to outliers

Consider values 1, 1, 1, 1, and 6. The mean is 2, the median and mode are 1, and the large deviation of 6 strongly increases variance, skewness, and kurtosis. Replacing 6 by 60 changes high-order moments dramatically. This is not a defect in the formulas; it reflects that higher powers intentionally emphasize tails. Sample skewness and kurtosis are therefore noisy for small datasets. They should be accompanied by plots, robust summaries, and a physical examination of extreme observations.

Binomial measurement efficiency

A detector identifies each of $n=100$ independently presented particles with efficiency $p=0.80$. The detected count has mean 80 and standard deviation $\sqrt{(100 \times 0.8 \times 0.2)} = 4$. A count of 70 is 2.5 standard deviations below the mean and would be unusual under the model. Before concluding that efficiency changed, verify independence, constant p , and the known number presented. Coincidence

loss or a fluctuating threshold can create dependence and overdispersion, invalidating the simple binomial calculation.

Integrated interpretation

Across all examples, the distribution is part of the measurement model rather than a decorative curve fitted after data collection. The support identifies which outcomes are possible, normalization ensures that probability is conserved, and the sampling mechanism determines how observations are weighted. Location measures then answer different questions: the mean predicts balance or expected total, the median locates the halfway quantile, and the mode identifies the most probable region. Dispersion and shape descriptors add information, but they must be matched to the tails and sample size. A sound analysis therefore combines numerical summaries with cumulative plots, histograms under several reasonable bin choices, and checks derived from the generating mechanism. Discrete detector outputs require probability masses; continuous variables require densities whose areas, not heights, are probabilities. Particle-size averages must state whether number, surface, volume, or mass is weighted. Binomial conclusions require fixed independent trials with constant success probability. When these conditions are documented, probability becomes an operational language for predicting repeated measurements, setting decision thresholds, and communicating uncertainty. When they are omitted, even a formally correct calculation may answer a different experimental question from the one intended.