LC Introduction to Probability and Statistics Lecture Notes

MSci Physics w/ Particle Physics and Cosmology
University of Birmingham

Year 1, Semester 1

Ash Stewart

Lectures Index

Lecture 1: Introdution and Descriptive Statistics 1
Lecture 2: Population Statistics 2
Lecture 3: Error Propogation and Combinations of Variables 3
Lecture 4: Covariance and Correlation 5
Lecture 5: Distributions 6
Lecture 10: Introduction to Probability 8

Wed 01 Oct 2025 12:00

Lecture 1 - Introdution and Descriptive Statistics

0.1 Course Welcome

- First half of the semester: Statistics
- Second hald the of semester: Probability
- All slides and notes on Canvas.

Why Descriptive Statistics? If we want to share an interesting bit of data, sharing the whole data is going to be confusing. Instead, we can share a small number of stats which describe and summarise the data.

0.2 Sample Statistics

One of the most simple is the number of samples (N), and the sample mean:

$$\text{Sample Mean: } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

We can also calculate the sample standard deviation as the average of mean squared error across the points in the sample:

$$\text{Sample STDev: } s_n^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

We can also use median or mode as measures of central tendancy. The mode is a poor estimator however (as it massively depends on how binning is done, for a continuous measurement), while the median is more resistant to outliers.

Thu 02 Oct 2025 09:00

Lecture 2 - Population Statistics

0.1 Accuracy and Precision

We usually take measurements to determine some kind of true value. Usually, we can't actually know what this true value is, but if we could there are two bits of terminology that is particularly important:

Accuracy: Accuracy is the 'closeness' between our value and the 'true' value.

Precision: Precision is the 'closeness' between our measurements, i.e. how spread out are our various measurements.

0.2 Error

Random Error: is uncertainty related to the fact that our measurements are only a finite sample, so is not going to be immediately representative of the true value. The smaller this error, the more precise the measurement is.

Systematic Error: is related to some kind of issue with the measurement or the equipment. This shifts all values, and negatively affects accuracy (but leaves precision unchanged)

Taking many repeat measurements decreases the effects of random error, but the effects of systematic error are much harder to combat.

Ideally, we want to be both precise and accurate, however accuracy is arguably more important. This is because a value which is precise, but not accurate may lead to false conclusions around the inaccurate value.

Wed 08 Oct 2025 12:02

Lecture 3 - Error Propogation and Combinations of Variables

Office Hours: 11:00 to 13:00 Thursdays, Physics West Rm 122

1 Types of Error

Broadly two types of error: Statistical/Random Error (resulting from low precision) and Systematic Error (from Low Accuracy).

Random error widens the distribution, while systematic error shifts the whole distribution up or down, meaning no matter how many repeats you take and how precise you think you are, the value is still nonsense as all datapoints have been equally shifted (i.e. by a poor experimental setup).

For example, you are trying to measure the length of an object using a ruler that has been unknowingly stretched. You cannot get a true value no matter the number of repeats or degree of precision.

1.1 Accuracy vs Precision

High accuracy is preferable to high precision - having high precision but low accuracy can lead to false conclusions (as an incorrect value appears confidently correct). Accuracy is more difficult to improve - precision can be improved by gathering more data, while higher accuracy can only be improved by a better experimental design.

2 Error Propogation

If we take a distribution, and add a constant value to all points, the distribution is shifted up/down without changing the variance.

$$\langle x+k \rangle = \langle x \rangle + k$$

$$Var(x+k) = Var(x)$$

If we multiply by a constant value, the mean is multiplied by this value, but the distribution becomes stretched and the variance grows:

$$\langle xk \rangle = k\langle x \rangle$$

$$Var(kx) = k^2 Var(x)$$

Or taking the natural log:

$$\langle \ln x \rangle \approx ln\langle x \rangle$$

$$Var(\ln x) \approx \frac{Var(x)}{x^2}$$

As this is a non-linear operator, these become good approximations rather than strict rules of equivalence.

And another example:

$$\langle e^x \rangle \approx e^{\langle x \rangle}$$

$$Var(e^x) \approx (e^x)^2 Var(x)$$

Note here, even though our underlying distribution is Normal and symmetric, the new distribution after e^x is neither, and these are an even worse approximation than before.

2.1 Combining Operators

We can apply some linear transformation $mx+c$, we can chain these rules together by doing the multiplicative transformation m first, then the linear scale c.

$$\langle mx + c \rangle = m\langle x \rangle + c$$

$$Var(mx + c) = m^2 Var(x)$$

2.2 Multiple Variables

What if we have mutliple distributed variables we want to add?

$$\langle A + B \rangle = \langle A \rangle + \langle B \rangle$$

$$Var(A + B) = Var(A) + Var(B)$$

And multiplying them (again this are now approximations)?

$$\langle AB \rangle \approx \langle A \rangle \langle B \rangle$$

$$Var(AB) \approx \langle B \rangle^2 Var(A) + \langle A \rangle^2 Var(B)$$

$$\frac{Var(AB)}{\langle AB \rangle^2} \approx \frac{Var(A)}{\langle A \rangle^2} + \frac{Var(B)}{\langle B \rangle^2}$$

Or division?

$$\left\langle \frac{A}{B} \right\rangle \approx \frac{\langle A \rangle}{\langle B \rangle}$$

$$Var\left(\frac{A}{B}\right) = \frac{Var(A)}{\langle B \rangle^2} + \frac{Var(B)}{\langle A \rangle^2}$$

3 One Rule to Rule Them All

$$Var(f) \approx \left(\frac{\partial f}{\partial A}\Big|_{A=\langle A \rangle, B=\langle B \rangle}\right)^2 Var(A) + \left(\frac{\partial f}{\partial A}\Big|_{A=\langle A \rangle, B=\langle B \rangle}\right)^2 Var(B)$$

Thu 09 Oct 2025 09:00

Lecture 4 - Covariance and Correlation

Office Hours: Thursday 11am - 1pm, Physics West Rm 222 (b.becsy@bham.ac.uk)

Previously, when looking at two or more variables for error propogation/combinations etc, we assumed that they were independant of one another. Today we look at how to handle multiple variables which may be correlated.

1 Covariance

Covariance is a measure that indicates how much two variables fluctuate together:

$$\text{Cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance matrices represent all combinations of covariance (noting Cov(x, y) = Cov(y, x) and Cov(x, y) = Var(x))

$$\Sigma = \begin{pmatrix} Cov(x, x) & Cov(x, y) \\ Cov(y, x) & Cov(y, y) \end{pmatrix}$$

We can then define correlation:

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

This is bounded between -1 (x = -y), y (x = y) and zero for no correlation. We can again put this in a matrix, noting it is symmetrical:

$$\begin{pmatrix} 1 & Corr(x, y) \\ Corr(y, x) & 1 \end{pmatrix}$$

1.1 Variable Combinations

Now, with correlated variables, we can say:

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$

$$Var(x, y) = Var(x) + Var(y) + 2Cov(x, y)$$

And (noting the mean slightly increases with correlated variables):

$$\langle xy \rangle = \langle x \rangle \langle y \rangle + Cov(x, y)$$

And the one formula to rule them all

$$Var(f) \approx \frac{\partial f}{\partial A}^2 Var(A) + \frac{\partial f}{\partial B}^2 Var(B) + 2\frac{\partial f}{\partial A}\frac{\partial f}{\partial B}Cov(A, B)$$

Wed 15 Oct 2025 12:00

Lecture 5 - Distributions

1 Coin Flips and Probability Recap

Flipping a coin is one of the simplest distributions we can create.

Given a:

$$P(H) = 0.5$$
$$P(T) = 0.5$$

We know that $P(HHHH) = 0.5^4$.

And $P(\text{Three Heads and One Tail}) = P(\text{HHHT or HHTH or THHH or HTHH}) = 4 \times 0.5^4$. We will take 4 coins, A, B, C, D. We denote a single result as A_{Heads} or C_{Tails} etc.

We can also say that the coins are independent, i.e. the probability of one result given another result is equal to just the probability of the first result:

$$P(A_h|B_h) = 0.5$$

The chance of A and B being heads is:

$$P(A_h \text{and} B_h) = P(A_h) \times P(B_h) = P(HH)$$

The chance of A *or* B being heads is (noting *or* excludes the case where both are true):

$$P(A_h \text{ or } B_h) = P(A_h) + P(B_h) - P(A_h \text{ and } B_h)$$

1.1 Discrete Distribution

Lets consider flipping 4 coins and counting the number of heads. This forms a discrete distribution (where only 5 possible values are possible, 0, 1, 2, 3, 4). This distribution must be normalised (sum to 1), so:

$$\sum_r P(r) = 1$$

We can also consider the mean (expected) number of heads:

$$\langle r \rangle = \sum_r rP(r)$$

This function, $P(x)$ is called a *probability mass function*, and the sum of all values must be 1.

1.2 Continuous Distributions

Continuous distributions have similar conditions:

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

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x)\, dx$$

And for the probability of the result lying between a and b:

$$\int_a^b P(x)\,dx$$

We cannot, in a continuous distribution consider the probability of an exact result, i.e. $P(x = a)$, $a \in \mathbb{R}$. As there are infinitely many possible values, the probability of any precise one is not meaningful (always zero). We therefore must always consider the probability of the result lying in some non-zero range.

P(x) in this case is called a *probability density function* and the area under the PDF curve must sum to zero. Note that this means that P(x) at any point may exceed one, so long as the overall area is equal to 1.

2 Binomial Distribution

3 Normal Distribution

4 Poisson Distribution

Fri 07 Nov 2025 11:00

Lecture 10 - Introduction to Probability

What is probability? Probability is the pure mathematical description of randomness.

1 Set Theory

Say we want to group trees into four sets:

- Tall, or not.
- Variegated (has a lighter coloured leaf border) or not.

In a park of 142 trees, we observe: