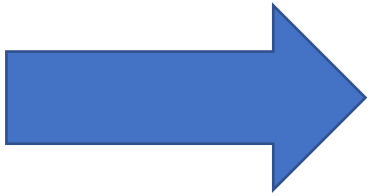
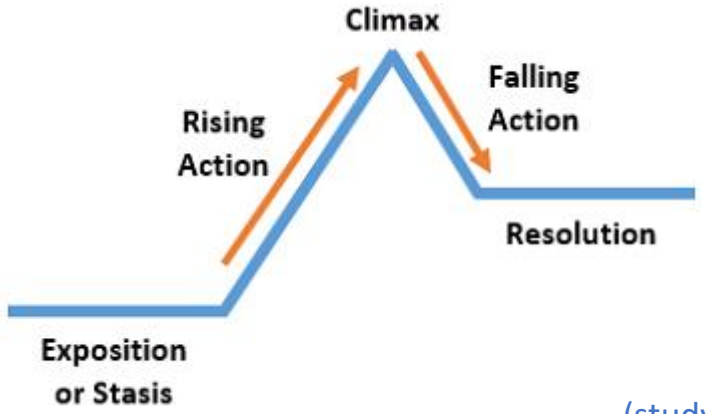
Lecture 4: How to write a report

10101101010100010111001110011010101010101010101010101010101010010100000011111001101010101010101010101





(study.com)

Data

Story

Anton Souslov A.Souslov@bath.ac.uk

Report marking scheme

See Moodle for details

Presentation (25 marks): scientific writing style, correct terminology, clear descriptions and explanations, logical structure, well designed and chosen figures and tables

Background and context (10 marks): physical background, clear objectives, connection to real-life applications

Achievements (35 marks): quality and quantity of results, clear presentation of results, creativity

Analysis and scientific understanding (30 marks): analysis of the results, comparisons to other works, errors and uncertainties, discussion on the characteristics of the techniques

Outline

1. Strategy and tactics for scientific writing

2. How to analyse and present data

Lecture 4: How to write a report

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline, may be included in the first paragraph if the editor considers that the accessibility of the paper is significantly enhanced by their inclusion. Under these circumstances, the length of the paragraph can be up to 300 words. (This example is 190 words without the final section, and 250 words with it).

During cell division, mitotic spindles are assembled by microtubule-based motor proteins[1,2]. The bipolar organization of spindles is essential for proper segregation of chromosomes, and requires plus-end-directed homotetrameric motor proteins of the widely conserved kinesin-5 (BimC) family[3]. Hypotheses for bipolar spindle formation include the 'push–pull mitotic muscle' model, in which kinesin-5 and opposing motor proteins act between overlapping microtubules[2,4,5]. However, the precise roles of kinesin-5 during this process are unknown. Here we show that the vertebrate kinesin-5 Eg5 drives the sliding of microtubules depending on their relative orientation. We found in controlled *in vitro* assays that Eg5 has the remarkable capability of simultaneously moving at ~20 nm s^{-1} towards the plus-ends of each of the two microtubules it crosslinks. For anti-parallel microtubules, this results in relative sliding at ~40 nm s^{-1}, comparable to spindle pole separation rates *in vivo*[6]. Furthermore, we found that Eg5 can tether microtubule plus-ends, suggesting an additional microtubule-binding mode for Eg5. Our results demonstrate how members of the kinesin-5 family are likely to function in mitosis, pushing apart interpolar microtubules as well as recruiting microtubules into bundles that are subsequently polarized by relative sliding. We anticipate our assay to be a starting point for more sophisticated *in vitro* models of mitotic spindles. For example, the individual and combined action of multiple mitotic motors could be tested, including minus-end-directed motors opposing Eg5 motility. Furthermore, Eg5 inhibition is a major target of anti-cancer drug development, and a well-defined and quantitative assay for motor function will be relevant for such developments.

nature.com

study.com

Story

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline, may be included in the first paragraph if the editor considers that the accessibility of the paper is significantly enhanced by their inclusion. Under these circumstances, the length of the paragraph can be up to 300 words. (This example is 190 words without the final section, and 250 words with it).

During cell division, mitotic spindles are assembled by microtubule-based motor proteins[1,2]. The bipolar organization of spindles is essential for proper segregation of chromosomes, and requires plus-end-directed homotetrameric motor proteins of the widely conserved kinesin-5 (BimC) family[3]. Hypotheses for bipolar spindle formation include the 'push–pull mitotic muscle' model, in which kinesin-5 and opposing motor proteins act between overlapping microtubules[2,4,5]. However, the precise roles of kinesin-5 during this process are unknown. Here we show that the vertebrate kinesin-5 Eg5 drives the sliding of microtubules depending on their relative orientation. We found in controlled *in vitro* assays that Eg5 has the remarkable capability of simultaneously moving at ~20 nm s^{-1} towards the plus-ends of each of the two microtubules it crosslinks. For anti-parallel microtubules, this results in relative sliding at ~40 nm s^{-1}, comparable to spindle pole separation rates *in vivo*[6]. Furthermore, we found that Eg5 can tether microtubule plus-ends, suggesting an additional microtubule-binding mode for Eg5. Our results demonstrate how members of the kinesin-5 family are likely to function in mitosis, pushing apart interpolar microtubules as well as recruiting microtubules into bundles that are subsequently polarized by relative sliding. We anticipate our assay to be a starting point for more sophisticated *in vitro* models of mitotic spindles. For example, the individual and combined action of multiple mitotic motors could be tested, including minus-end-directed motors opposing Eg5 motility. Furthermore, Eg5 inhibition is a major target of anti-cancer drug development, and a well-defined and quantitative assay for motor function will be relevant for such developments.

The main story arc

This story arc repeats in:

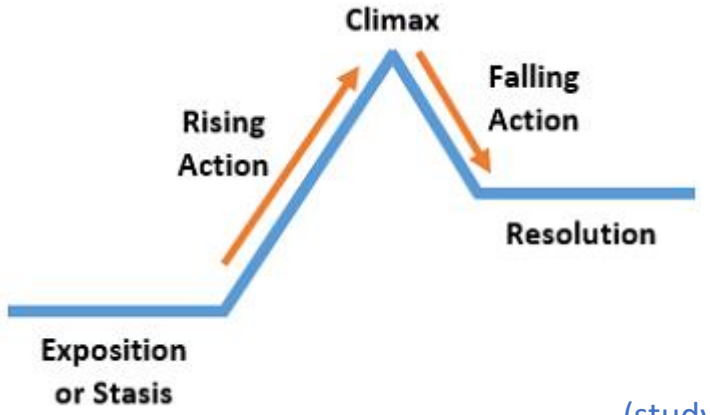
- The abstract
- The introduction
- The report as a whole

study.com

Story arcs throughout the report

This story arc repeats in:

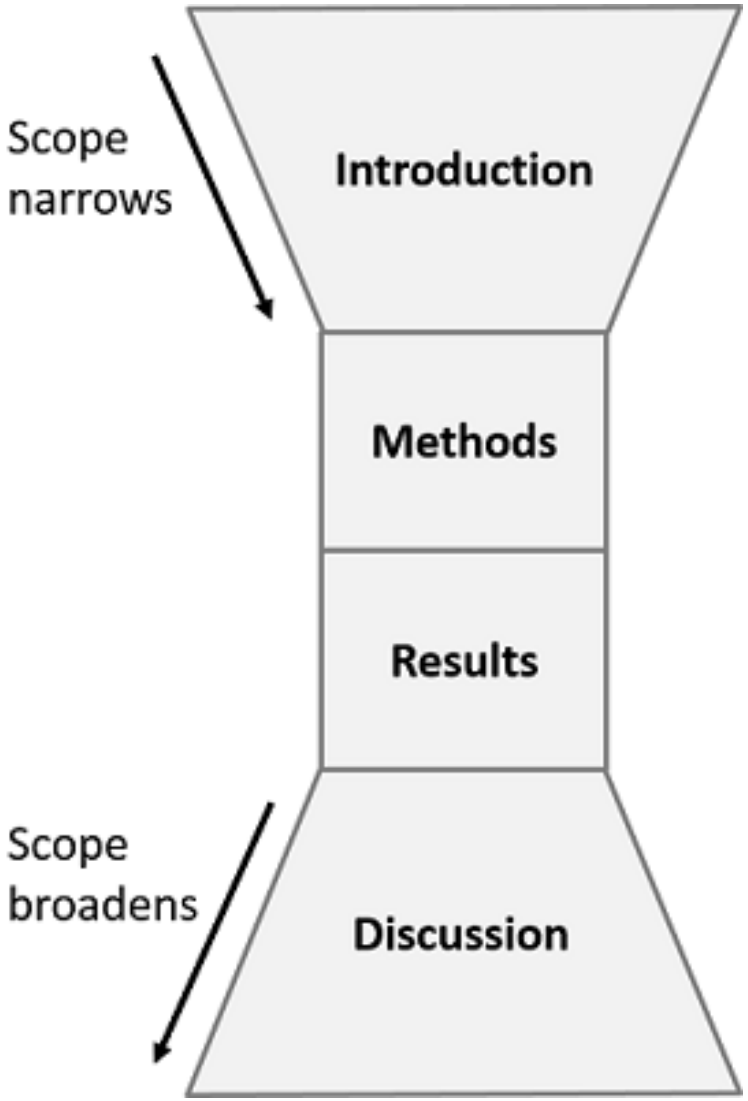
- The abstract
- The introduction
- The report as a whole



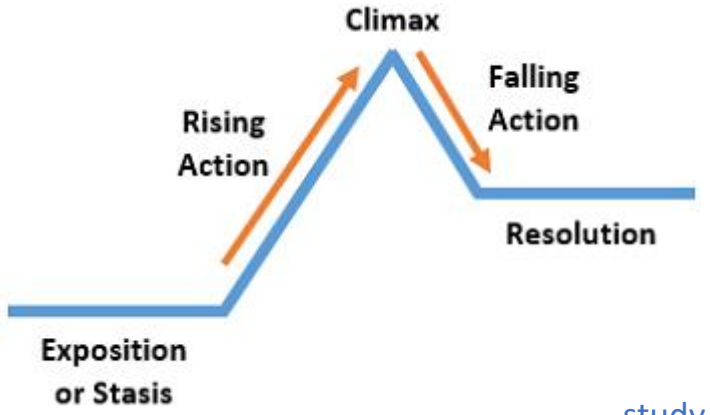
(study.com)

A story arc occurs in every paragraph

Funnel



miamioh.edu



study.com

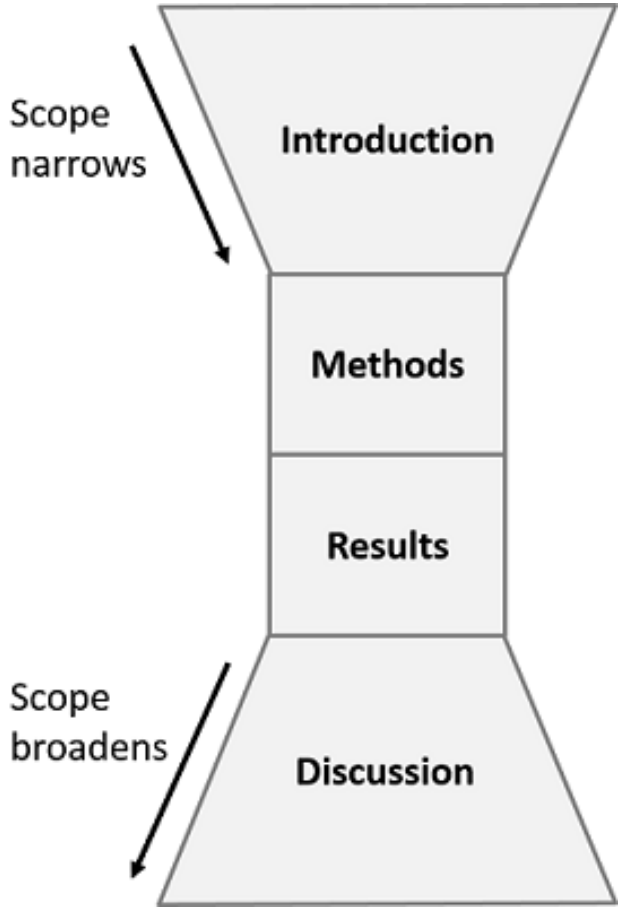
Story

Structure of a report

A standard structure of a report:

1. Title
2. Abstract
3. Introduction
4. Methods
5. **Results**
6. **Discussion**
7. Conclusion
8. Bibliography
9. (Possibly Acknowledgements and Appendices)

(You are not forced to use this structure)



miamioh.edu

5. Results

What are the main findings?

Good and clear visualization is very important:
If the reader has a brief look at the figures, she/he should get a quick idea what you have done

Figure guidelines

Label your axis and what you are presenting

Show your data with points

Include error bars (or state that they are "comparable to symbol sizes")

Either connect adjacent points with straight line segments *or* do not connect them at all

If you are going to **fit** your the data to a model then you should do this only to test a theory of hypothesis. You should justify the fitting function that you use.
(Do not draw "polynomial trend lines"!)

To compare results with different parameters, it often helps to plot them all on the same graph

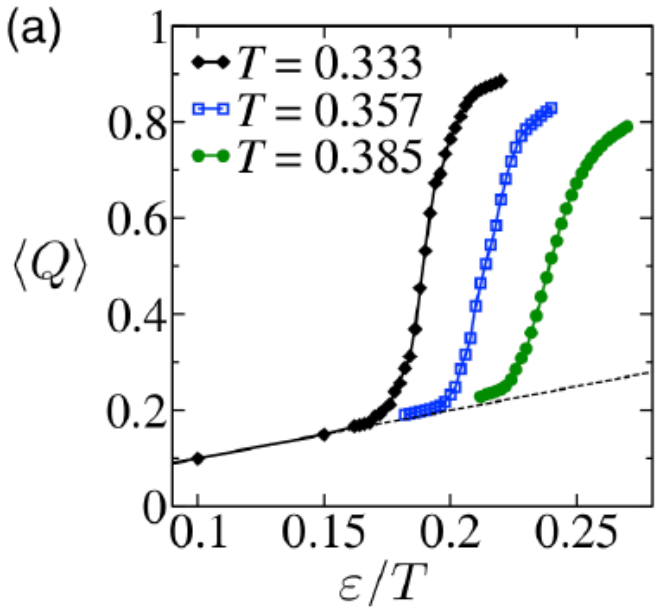


Fig 1a: Sample graph, reproduced from [Jack and Garrahan, Phys Rev Lett 116, 057702 (2016)]. Adjacent data points are connected by straight lines (these lines could also be omitted). The dashed line is a theoretical prediction that is valid for small ε. Error bars are similar to the symbol sizes so they have not been plotted in this case.

Figure guidelines

Write a self-explanatory caption

Graphs should convey a clear message and be easy to understand

Font sizes in all labels should be comparable to the main text in the report

If you copy a figure from a paper or website you need to write, for example, “Reproduced from [1]” in the caption, where [1] is the reference. (A simple citation would be enough in case you only got the idea from a paper but you made your own figure.)

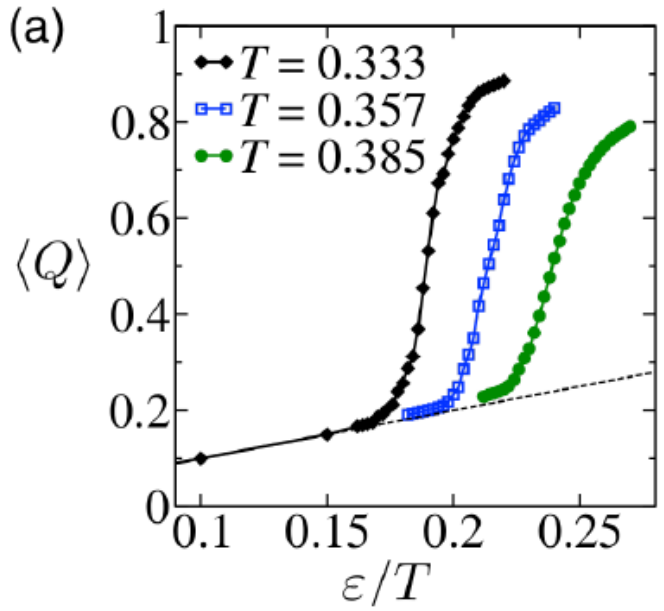


Fig 1a: Sample graph, reproduced from [Jack and Garrahan, Phys Rev Lett 116, 057702 (2016)]. Adjacent data points are connected by straight lines (these lines could also be omitted). The dashed line is a theoretical prediction that is valid for small ε. Error bars are similar to the symbol sizes so they have not been plotted in this case.

6. Discussion

What do the results mean (physical explanations)?

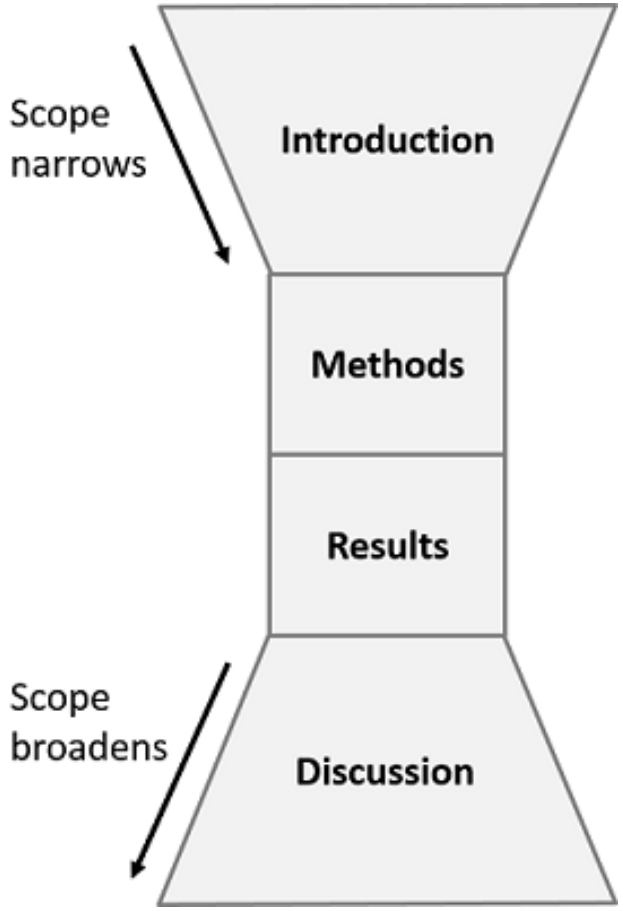
How good and reliable are the used methods?

Comparisons to other (earlier) works

Limitations of the analysis or the methods

Possible improvements in the future

Can be the most important section



miamioh.edu

1. Title

Title should be a short and catchy description of the work.
Where could the following titles appear?

“Fractals in nature”

“Investigating fractal properties of diffusion-limited aggregates in 2D”

“Fractal dimension of dielectric breakdown”

“Modelling and simulation of nanoparticle aggregation in colloidal systems”

1. Title

Title should be a short and catchy description of the work.
Where could the following titles appear?

“Fractals in nature”
(title in Cosmos magazine)

“Investigating fractal properties of diffusion-limited aggregates in 2D”
(title of a DLA course work report)

“Fractal dimension of dielectric breakdown”
(title of a research article in Physical Review Letters)

“Modelling and simulation of nanoparticle aggregation in colloidal systems”
(title of a Master’s thesis)

7. Conclusion and 3. Introduction

Conclusion:

- Highlight the most important findings (“the take-home message”)
- Overall significance of the work

Introduction:

- Background of the physics
- How the work is related to earlier works (brief literature review)
- How the work is related to real-life applications (significance)
- Motivation (why this study is needed) and objectives of the project (how the current know/how is improved)
- What was done in the project

8. Bibliography and 9. Appendices

Bibliography:

Remember to cite all the sources that you have used

Appendices:

Material that is relevant but it does not exactly fit in the main text

Material that makes the main text difficult to read (e.g., details of Methods)

For example, selected pieces of the used C++ codes, additional data, tables or figures

General guidelines on reports

Writing a good report is not easy, and we are fairly demanding when we assess third-year units.

For background and context, **explain all technical terms** that you use.

Try to be as **precise** as possible. If you read a sentence at random from the middle of the report, is it a true statement? If you are making an argument, does each sentence follow logically from the previous one?

Be careful with **units** and dimensionless variables

Use **scientific language**, not informal terms like *wiggly*, *clumps, "this seems suspicious", "the results look correct"*

Be careful with phrases like *this behaviour is not realistic.*
(Why? This usually makes sense only if you compare with a specific example).

Maximum is 7 pages: better to use less and have sharper, more refined message

Concentrate on the physics

(i) Describe and explain the behaviour of the model system you have been studying

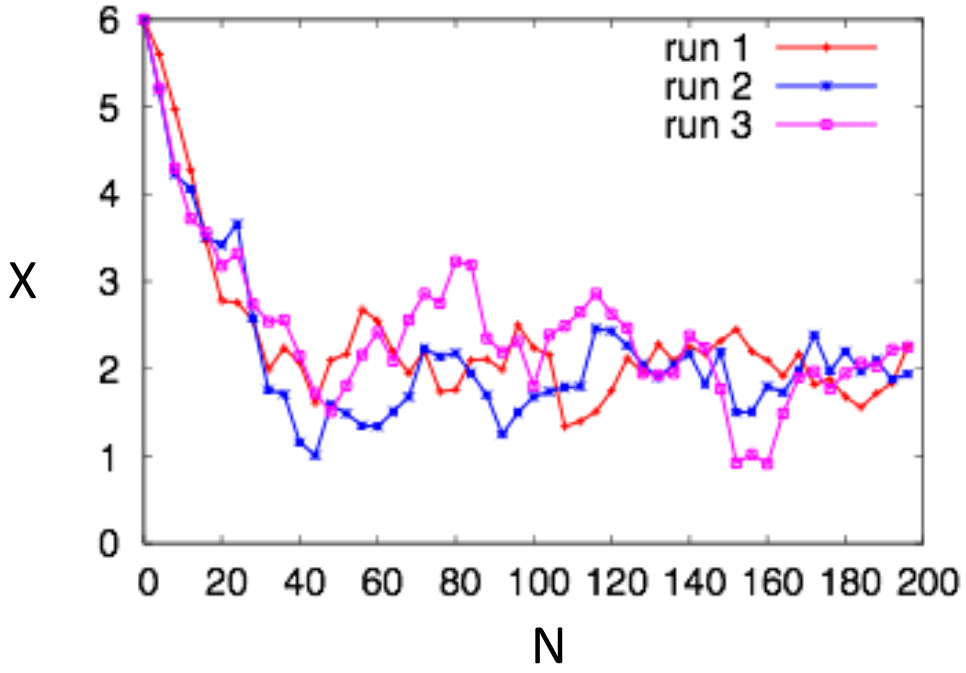
(ii) Explain how your results are related to the properties of physical materials, or the outcomes of experiments

Make sure you **explain what the model represents**.
For example, if you write about DLA, explain how your simulation of particles is related to any experimental examples that you mention.

Analysis of errors is important. If you have not estimated your errors then your conclusions are not justified. (This is especially true in simulations where random numbers are used).

Errors and uncertainties

Illustration of error analysis with computational data

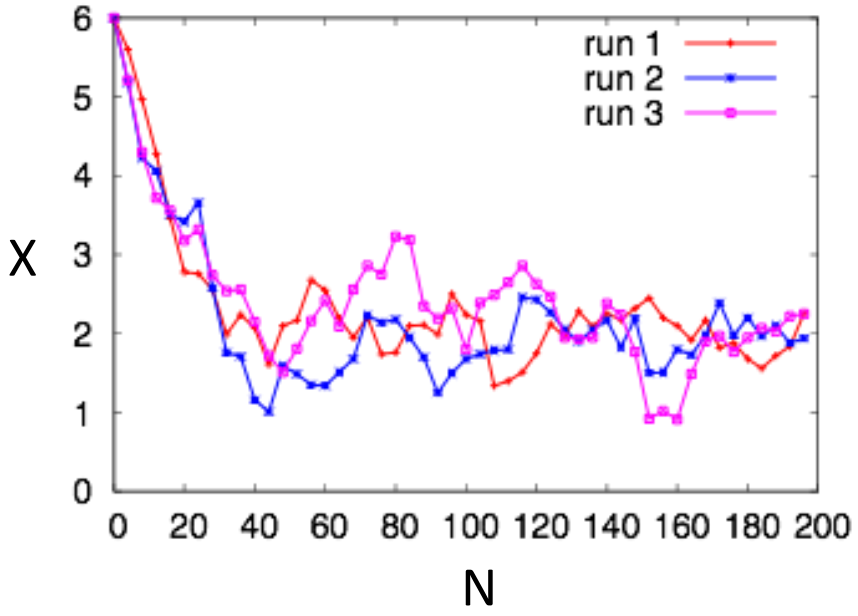


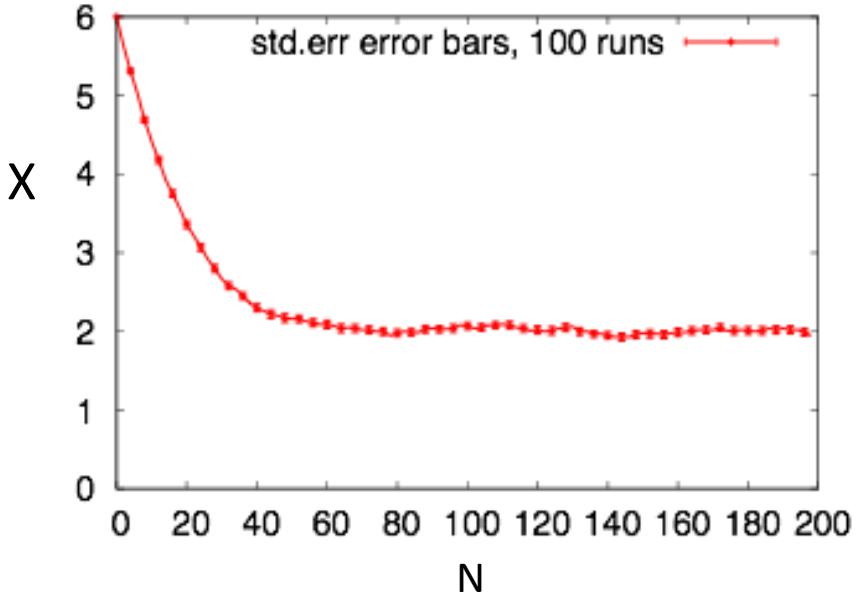
If you run your programs several times, you typically get different results.

How should we think about errors here? What graphs should we plot? What numbers should be calculate?

Do we want to estimate the mean behaviour? Uncertainty of this mean? Or to describe the range of variation?

Errors and uncertainties





Average over runs $i = 1 \ldots n$, at fixed time:

$$\overline{X(t)} = \frac{1}{n} \sum_i X_i(t)$$

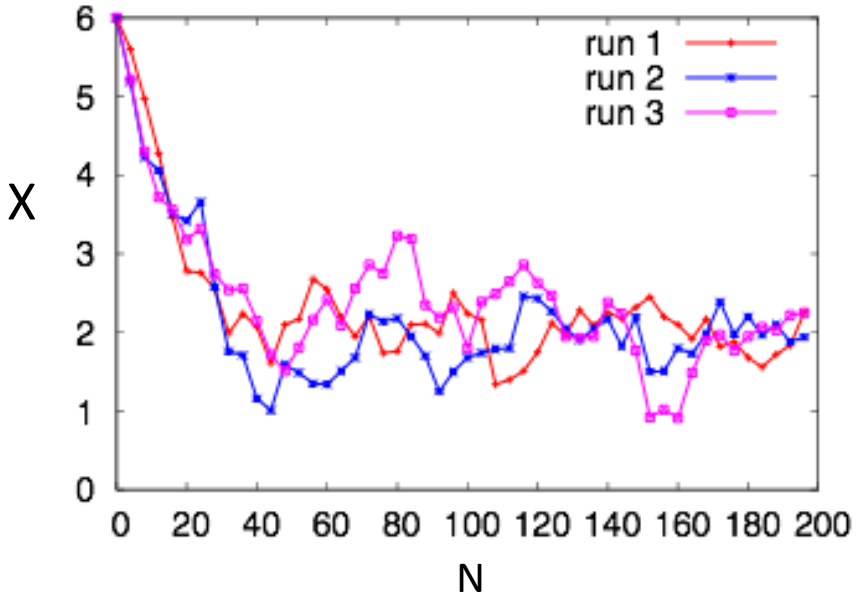
where $X_i(t)$ is the measurement of X in run i at time t.

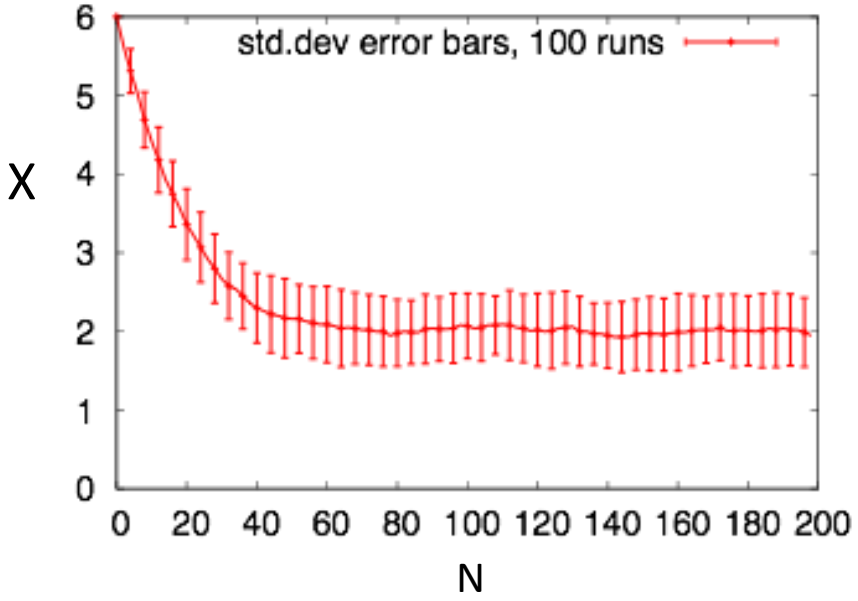
Standard error:

$$\sigma_{\bar{X}} = \frac{1}{\sqrt{n-1}} \left[\overline{X(t)^2} - \overline{X(t)}^2 \right]^{1/2}$$

Standard error is an estimate of how far your answer is from the "true" average value. The more simulations you run (*n* gets larger), the smaller the standard error gets.

Errors and uncertainties





Average over runs $i = 1 \ldots n$, at fixed time:

$$\overline{X(t)} = \frac{1}{n} \sum_i X_i(t)$$

where $X_i(t)$ is the measurement of X in run i at time t.

Standard deviation:

$$\sigma_X = \left[\overline{X(t)^2} - \overline{X(t)}^2 \right]^{1/2}$$

Standard deviation tells you how much variation there is between different runs of the simulation. It does not get smaller with *n*.

Errors and uncertainties

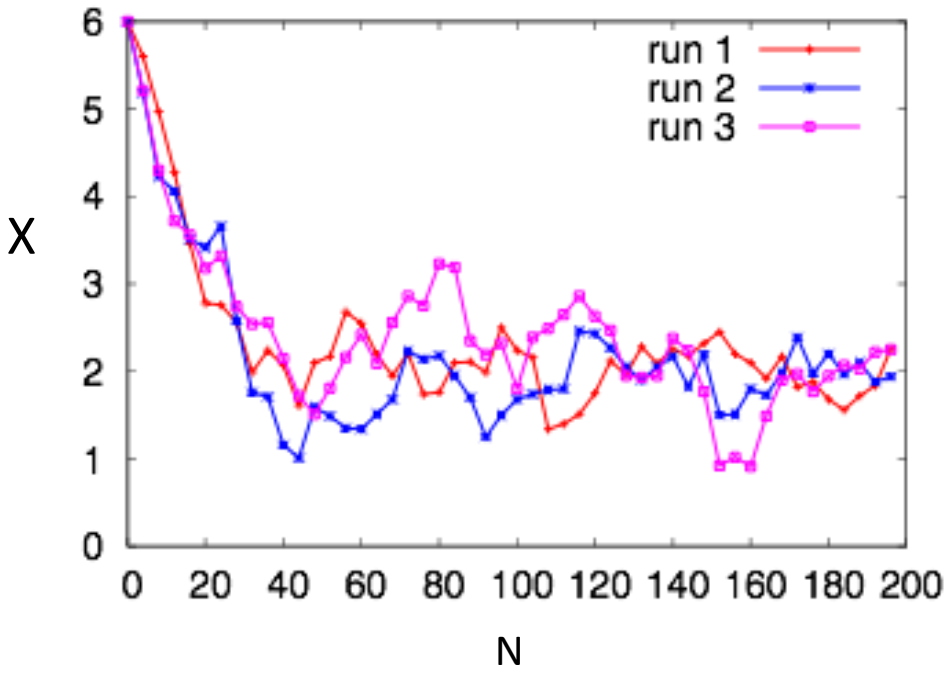
The *standard error* tells you how uncertain you are about the estimated average. This is the usual quantity to use for an error bar.

The *standard deviation* tells you how different are the repeated runs of the simulation or experiment. (This can also be used for an error bar in some cases, but you need to make it clear that it shows a *range of behaviour*, not an *uncertainty*.)

Note: the standard error formula assumes that all measurements are *independent*. This is true for independent runs of the same simulation, but this is not always the case.

Distributions and histograms

Sometimes we really care about the *range* of behaviour in our system

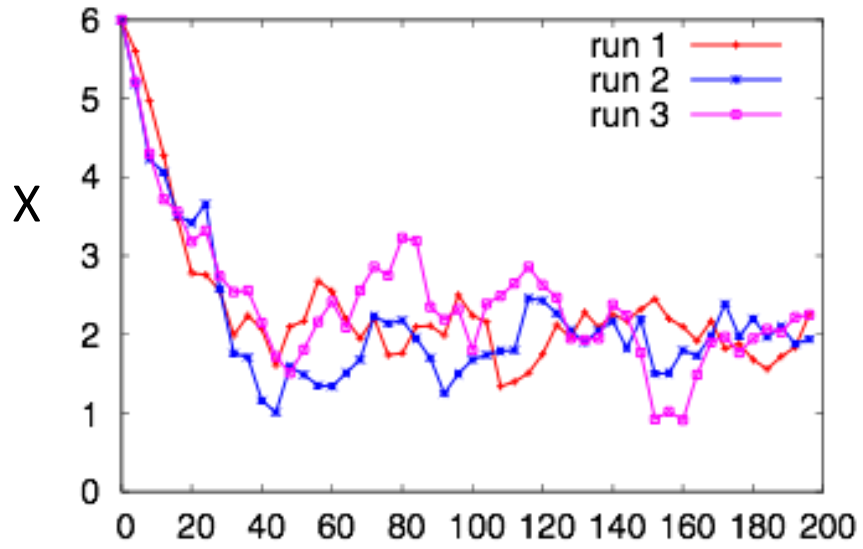


The standard deviation gives an idea of the spread of data

But can we analyse this in more detail? For example, can we say how often particular values appear?

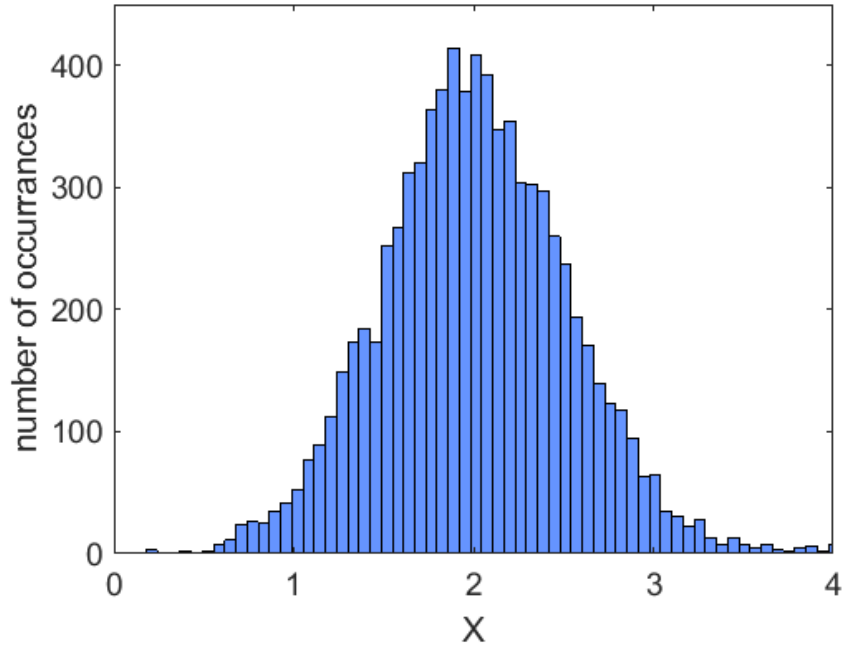
For this purpose, we can draw a **histogram**. Divide the range of values of *X* into n "bins" and count the number of times *X* falls within each bin.

Distributions and histograms



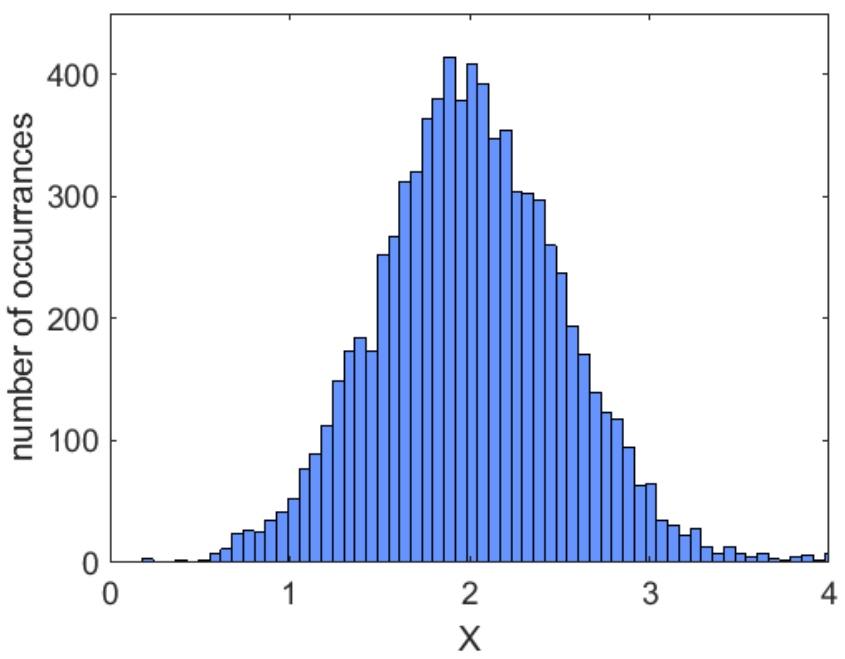
Let's take data from many runs…

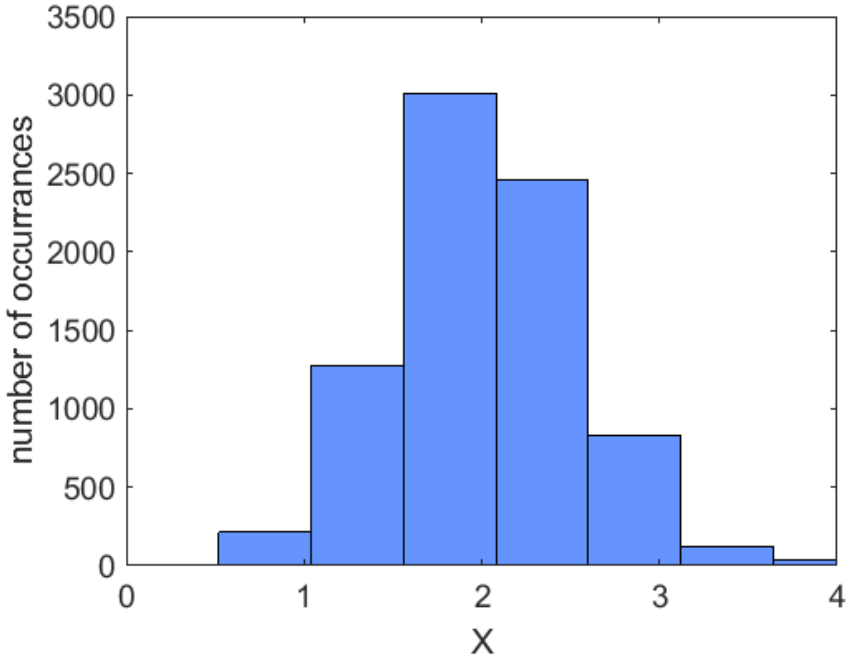
We divide the interval between 0 and 4 into 80 "bins" of width 0.05, and count (with some software) how many times *X* occurs within each "bin".



From the histogram, we can see that the most likely value of *X* is about 2, and the "spread" is about ± 0.5 (which is inline with the standard deviation)

Distributions and histograms





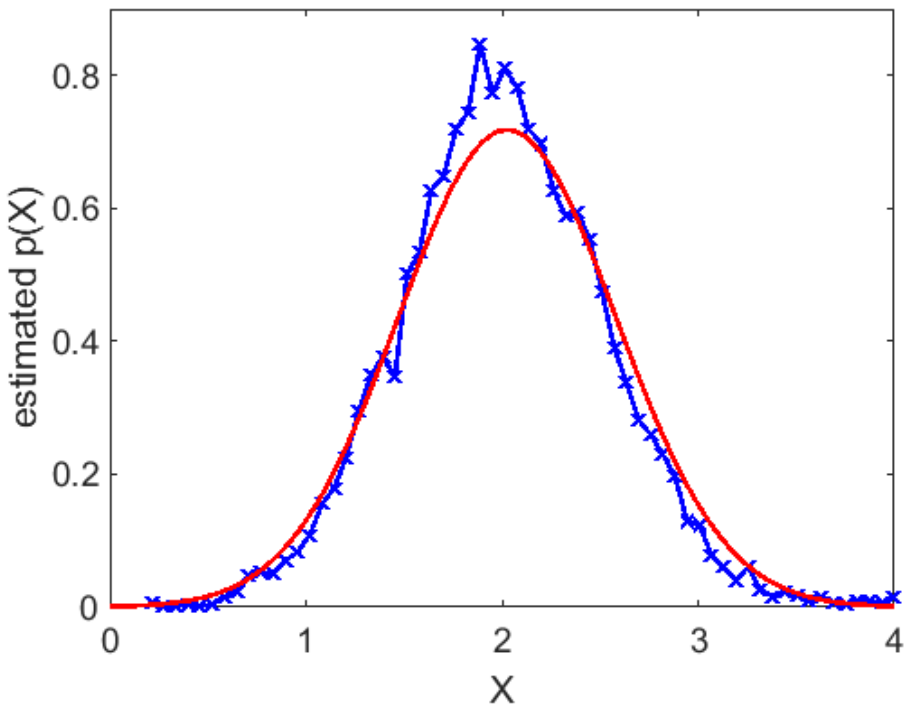
Note how the number of bins affects the visualization. The figure on the left has 80 bins and the figure on the right 10.

Distributions and histograms

A *probability density function* $p(X)$ is a positive function that tells me the fraction of observations are expected to fall in a given range.

If I make many measurements of X, I expect that the fraction falling between X_1 and X_2 should be

$$P(X_1 < X < X_2) = \int_{X_1}^{X_2} p(X)\mathrm{d}X$$



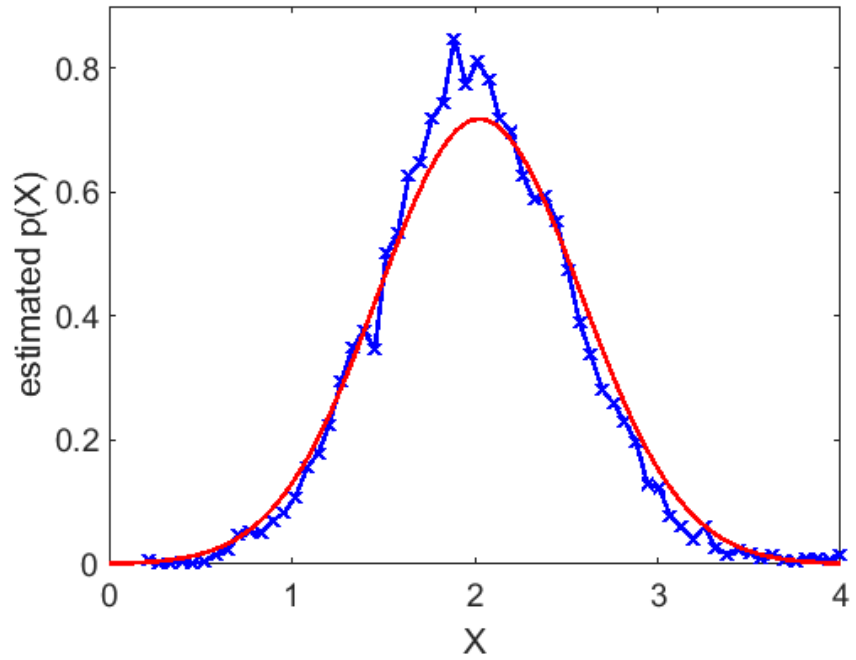
$p(X)$ can be estimated from a histogram of X by dividing the number of occurrences in each bin by the width of the bin, and then dividing by the total number of samples.

In this case $p(X)$ fits well to a Gaussian. . .

Distributions and histograms

Remember the fraction of measurements expected between X_1 and X_2 is

$$P(X_1 < X < X_2) = \int_{X_1}^{X_2} p(x)\mathrm{d}x$$

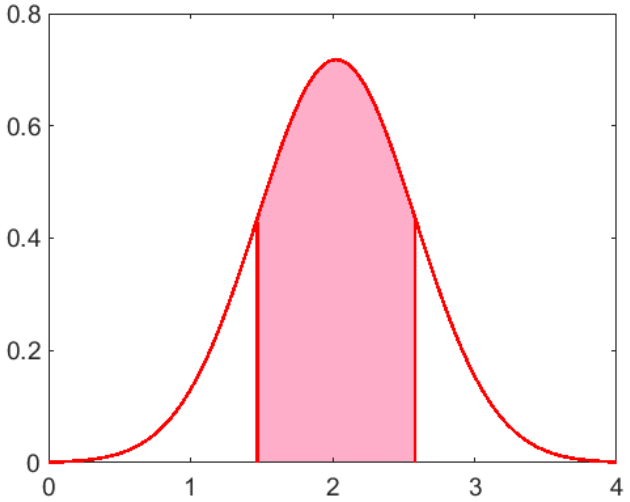


The fraction of measurements in any range corresponds to the *area under the curve*, within that range

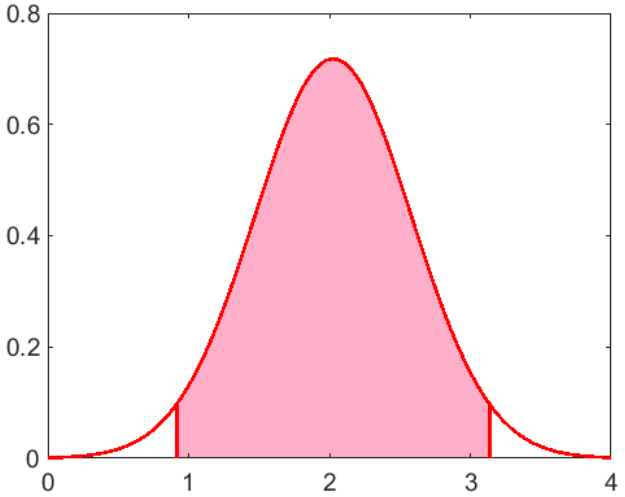
The total area under the curve must be 1 (due to normalisation)

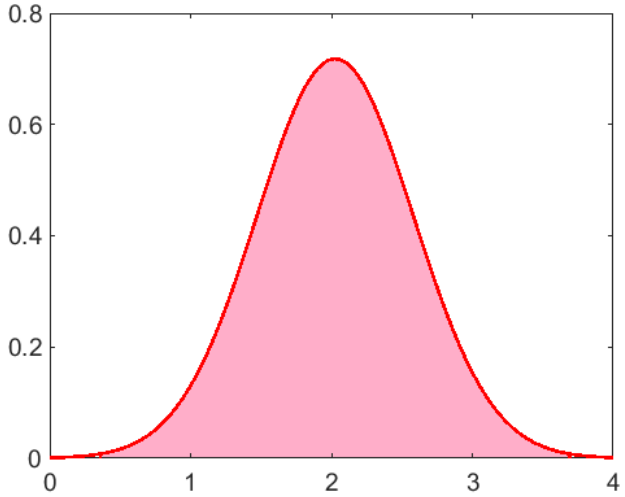
Estimating distributions in this way is a powerful tool when characterising how random observables behave in a system (for example, DLA cluster radius with fixed number of particles…)

Distributions and histograms

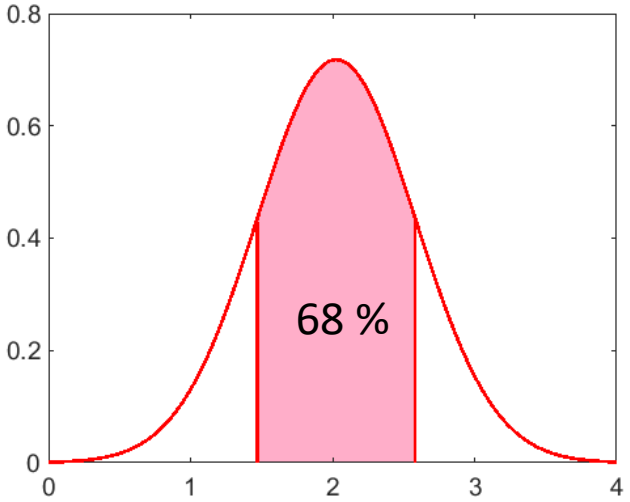


For a Gaussian probability density function, do you remember what percentage of samples falls within the 1, 2 and 3 standard deviation ranges around the mean?

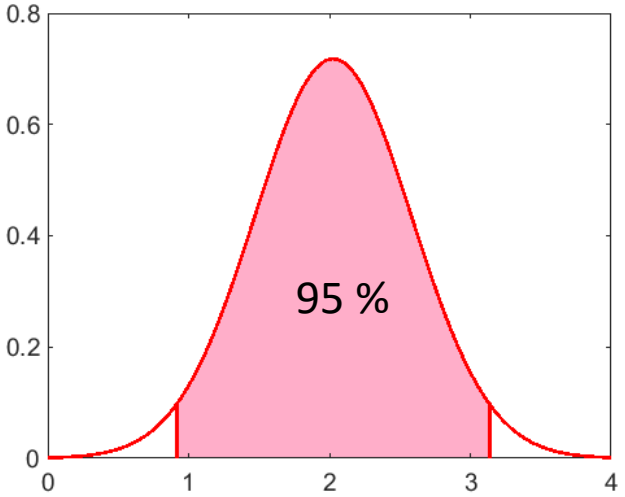


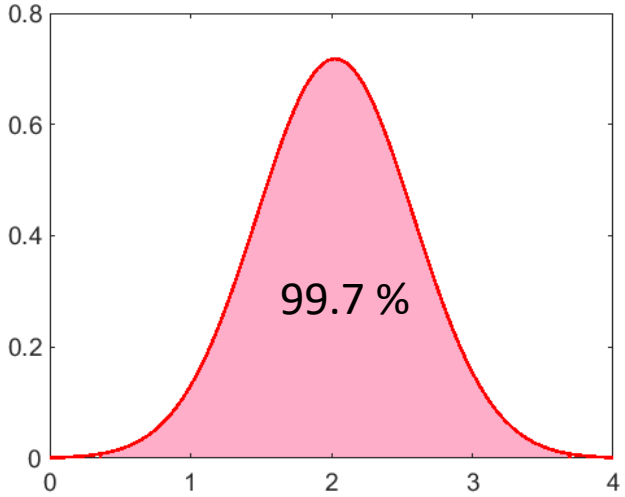


Distributions and histograms



For a Gaussian probability density function, do you remember what percentage of samples falls within the 1, 2 and 3 standard deviation ranges around the mean?





Summary

- Try to use the data from your simulations to tell a story
- A clear report should be engaging, precise, scientifically accurate, and accessible to a third-year student not taking this unit
- Figures can be one of the main pieces of the report, if relevant and well designed. (They do take effort!)
- Simulations are numerical experiments, where errors and uncertainties are important.

There are some guidelines in the report template.
Reading scientific papers can also show you what kind writing and analysis is expected.