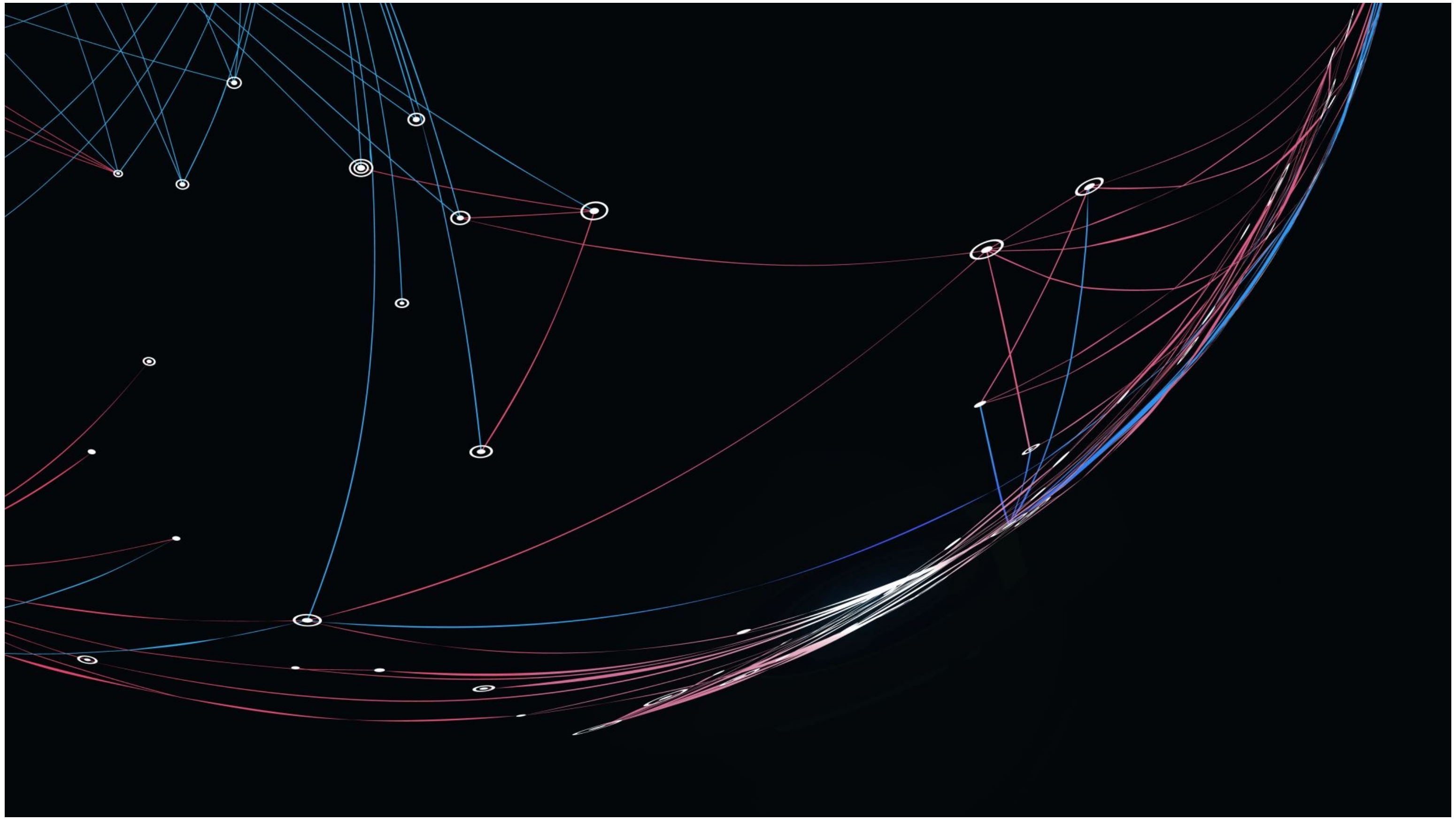
Statistics in particle physics



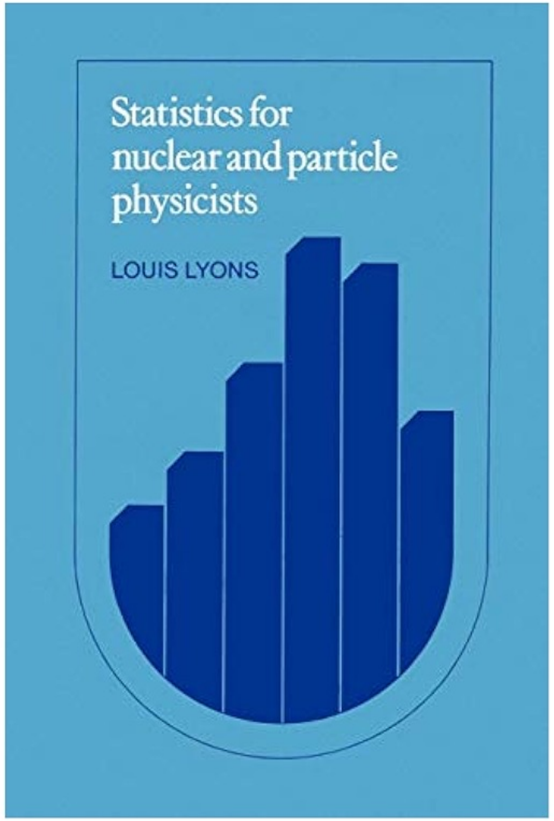
Pin-Jung Chiu

NCTS Particle Physics
Summer Undergraduate Student
Program 2025

July 1st – 3rd

What will you learn in these lectures?

- Experimental errors
- Probability and statistics
- Distribution
- Interpretation of probability



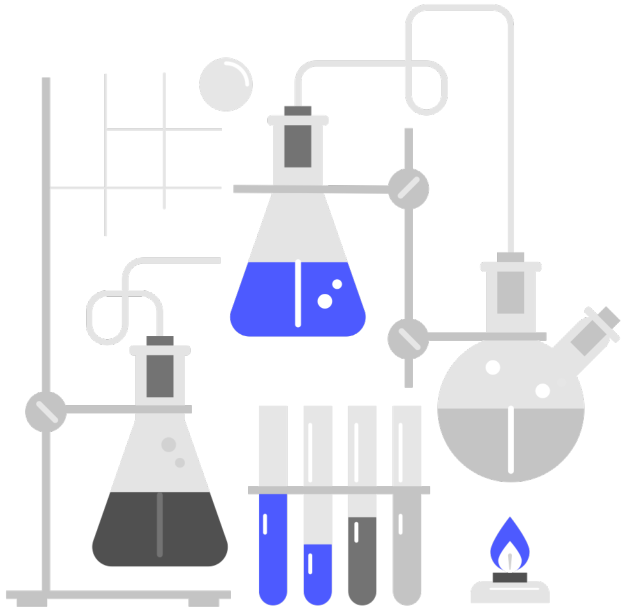
- *Statistics for nuclear and particle physics, L. Lyons,* Cambridge University Press

Why do we do experiments?

- Parameter determination: determine the numerical values of some physical quantities

- Hypothesis testing: test whether a particular theory is consistent with our data

Experimental errors



Why estimate errors?

- When performing parameter determination experiments, we are concerned not only with the answer but also with its accuracy

- E.g., suppose a determination of the velocity of light yields $c = (3.09 \pm 0.15) \times 10^8\ \mathrm{m}/s$, how does this compare to the previously accepted value of $2.998 \times 10^8\ \mathrm{m}/s$?
  - With an error of $\pm\ 0.15 \times 10^8\ \mathrm{m}/s$, it is consistent with the old value
  - With an error of $\pm\ 0.01 \times 10^8\ \mathrm{m}/s$, it is inconsistent with the old value → evidence of increase in c
  - With an error of $\pm\ 2 \times 10^8\ \mathrm{m}/s$, it is consistent with the old value, but the accuracy is very low
  - If we only determine $c = 3.09 \times 10^8\ \mathrm{m}/s$ → unable to judge the significance of this result

- Whenever you determine a parameter, estimate the error or the experiment is useless

What is “error”?

- Error is used in different ways:
  - In everyday language: error is used to describe a mistake
  - In statistics: error is used as both
    - Mistake (doing something wrong) as in “type-I error” and “type-II error”
    - Discrepancy (deviation of single value from true)

- Physicists do not talk about the discrepancy of single measurements but on the overall uncertainty
- Still, physicists often use “error” when they actually mean “uncertainty”,
  - E.g., error bars, errors on the result, error analysis, error propagation, …

Types of errors

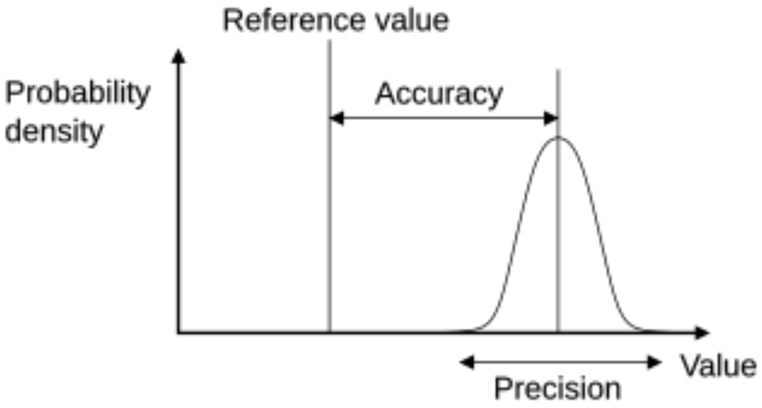
- E.g., prisoner in court
  - Hypothesis H_0: he is guilty
- Type-I error = reject true case
  - Reject H_0, if H_0 is in fact true
  - Let the prisoner free, when he is in fact guilty
- Type-II error = accept false case
  - Accept H_0, if H_0 is in fact wrong
  - Convict the prisoner, when he is in fact innocent

Precision or accuracy?

- Accuracy
  - How close the measured value is to the reference nominal value
- Precision
  - How reproducible is the measurement under identical conditions

Precision and Accuracy in Physics

By Pekaje at English Wikipedia



Random and systematic uncertainties (1)

Aspect	Random uncertainty	Systematic uncertainty
Definition	Unpredictable variations in measurements	Consistent, repeatable error in the same direction
Cause	Uncontrolled or unknown fluctuations (e.g., thermal noise, statistical variation)	Faulty equipment, calibration issues, or biased procedures
Effect on Results	Causes scatter in data; averages out over many measurements	Shifts all results in the same direction; does not average out
Detectability	Can be detected by repeating measurements	Harder to detect without external reference or control
Correction	Reduced by taking more measurements (statistics)	Must be identified and corrected through calibration or modeling
Visualization	Scattered data points around the true value	Data consistently offset from the true value

Random and systematic uncertainties (2)

- Consider an experiment involving counters to determine the decay constant λ of a radioactive source

$$-\frac{\mathrm{d}n}{\mathrm{d}t}=\lambda n$$

- **Random errors**:
  - Statistical error in counting random events
  - Timing of the period for which the decays are observed
  - Uncertainty in the mass of the sample
- **Systematic errors**:
  - Efficiency and/or location of the counter → counting rate < true decay rate
  - The counter sensitive to other particles → counting rate > true decay rate
  - Impurity of the source → number of nuclei capable of decaying < deduced number from the mass
  - Calibration errors in the clock (time) and balance (mass)

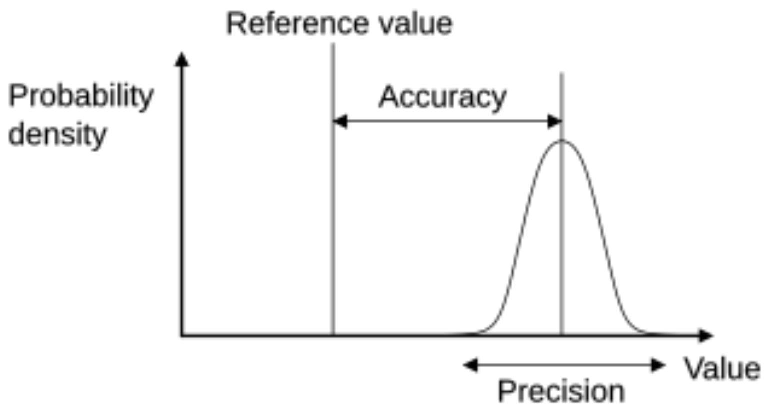
A good experimental physicist minimizes and realistically estimates the random error of his/her apparatus and reduces the effects of systematic errors to a much smaller level.

~ *Statistics for nuclear and particle physics, L. Lyons*

What is the meaning of the uncertainty?

- Related to the spread of values obtained from a set of repeated measurements
  - Distributions
  - The mean and variance of a distribution
  - Combining errors → error analysis

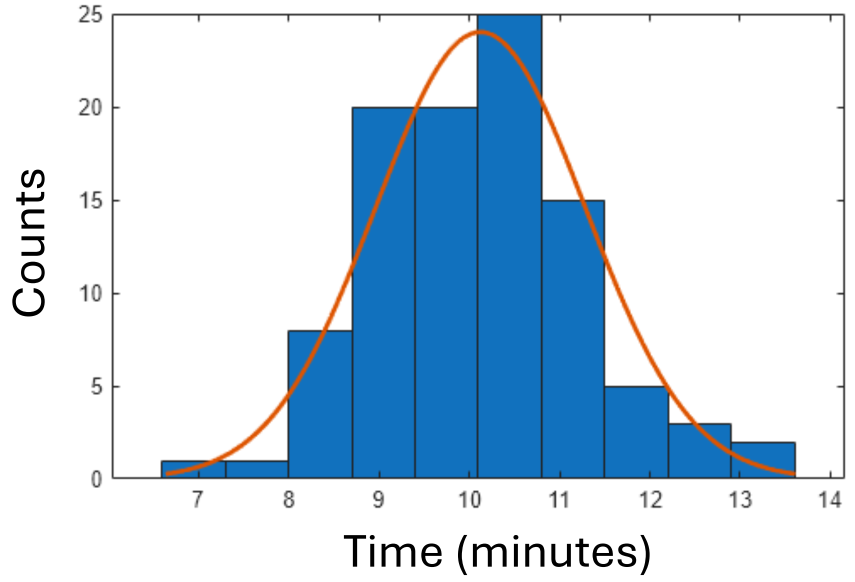
By Pekaje at English Wikipedia



Distribution

- A distribution $n(x)$ describes how often a value of the variable x occurs in a defined sample

- Variable x can be continuous or discrete
  - E.g., x is the day of the week (1 to 7)
    → number of marriages on day x (discrete)
  - E.g., commuting time from the bus station to the classroom (continuous)



- Histogram
  - Bin size/width
  - A large number of observations and a small bin size: histogram → continuous distribution

Mean and variance (1)

- To describe the distribution, we need measures of the value x at which the distribution is centered, and how wide the distribution is → true mean (μ) and variance (σ^2) ($\sigma = \sqrt{\sigma^2}$: standard deviation)
- For a set of *N* measurements, sample (arithmetic) mean (average) and variance are

$$\bar{x} = \frac{\sum x_i}{N}$$

$$s^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum (x_i - \bar{x})^2}{N-1} = \frac{N}{N-1}\left(\overline{x^2} - \bar{x}^2\right), \text{ where } \overline{x^2} = \frac{\sum x_i^2}{N}$$

- The variance of the mean is

$$u^2 = s^2/N$$

- With increasing N, s^2 will not change, while u^2 decreases

Mean and variance (2)

- If the measurements are grouped together so that at the value x_j there are m_j events

$$\bar{x} = \frac{\sum m_j x_j}{\sum m_j}$$

$$s^2 = \frac{\sum m_j \left(x_j - \bar{x}\right)^2}{\sum m_j - 1}$$

$$u^2 = \frac{\sigma^2}{\sum m_j}$$

- For the continuous distribution

$$\bar{x} = \frac{\int n(x) x \mathrm{d}x}{N}$$

$$s^2 = \frac{\int n(x)(x - \bar{x})^2 \mathrm{d}x}{N}$$

$$N = \int n(x) \mathrm{d}x$$

- Assume N is large, so the usual $(N-1)$ in the denominator of s^2 is replaced with N

Mean and variance (3)

- If individual measurements have different efficiencies ($1/w_k$)

$$\bar{x} = \frac{\sum w_k x_k}{\sum w_k}$$

$$s^2 = \frac{\sum w_k (x_k - \bar{x})^2}{\sum w_k} \times \frac{n_{\text{eff}}}{n_{\text{eff}} - 1}$$

$$u^2 = \frac{s^2}{n_{\text{eff}}}$$

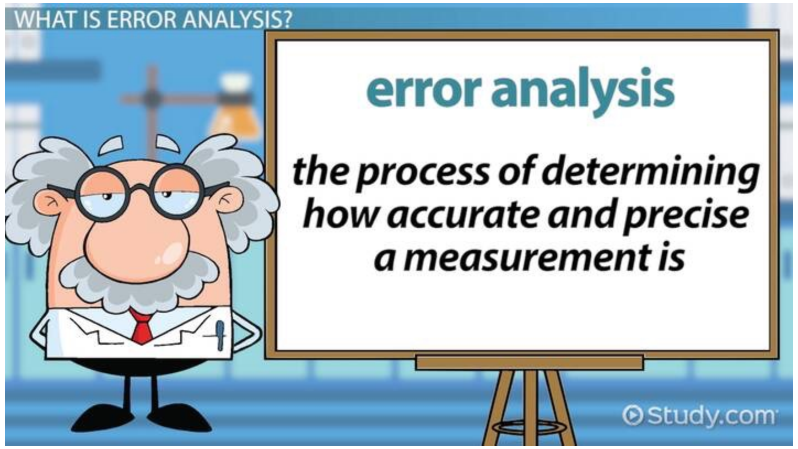
- n_{eff} is the effective number of events. If the total number of events is $T \pm \delta$,

$$n_{\text{eff}} = \frac{T^2}{\delta^2} = \frac{(\sum w_k)^2}{\sum w_k^2}$$

- E.g., With 4 ± 2 real events with a constant detection efficiency of 4%
  - The corrected total number of events is 100 ± 50
  - $n_{\text{eff}} = \frac{4^2}{2^2} = 4$
- A sample of events with a very low detection efficiency, $n_{\text{eff}} \sim 1$

Combining errors

- We are often facing situations where the experimental result is given in terms of two (or more) measurements
- We want to know the error on the final answer in terms of the errors on the individual measurements
- → Error analysis



Linear situations (1)

- Consider $a = b - c$
- For maximum possible errors $\delta a = \delta b - \delta c$
- Root mean square deviation (provided that b and c are uncorrelated)

$$\sigma_a^2 = \sigma_b^2 + \sigma_c^2$$

- N.B., if in an experiment, we know that the measurements on b and c were incorrect by δb and δc, we can correct for it. However, we often don't know these values but only know their mean square values σ^2 over a series of measurements

Linear situations (2)

Derive $\sigma_a^2 = \sigma_b^2 + \sigma_c^2$

$$\sigma_a^2 = \langle (a - \bar{a})^2 \rangle$$
$$= \left\langle \left[(b - c) - \left(\bar{b} - \bar{c}\right) \right]^2 \right\rangle$$
$$= \left\langle \left[\left(b - \bar{b}\right) - (c - \bar{c}) \right]^2 \right\rangle$$
$$= \left\langle \left(b - \bar{b}\right)^2 \right\rangle + \langle (c - \bar{c})^2 \rangle - 2\left\langle \left(b - \bar{b}\right)(c - \bar{c}) \right\rangle$$
$$= \sigma_b^2 + \sigma_c^2 - 2\mathrm{cov}(b, c)$$

$$\mathrm{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$$
$$= \langle xy \rangle - \langle x \rangle \langle y \rangle$$

Covariance of b and c, which has to do with whether the errors are correlated or not

Linear situations (3)

- We know that it is better to take the average of several independent measurements of a single quantity than just to measure once
- For n measurements q_i each with an uncertainty σ and whose average is $\bar{q}$,

$$n\bar{q} = \sum_i q_i$$

- The statistical uncertainty on the mean u is given by

$$n^2u^2 = \sum_i \sigma^2 = n\sigma^2$$

- Thus,

$$u = \frac{\sigma}{\sqrt{n}}$$

The error on the mean is known more accurately than the error characterizing the distribution by a factor of $\sqrt{n}$
→ useful to average

Linear situations (4)

- E.g., measure $a = b + b$
  - From $\sigma_a^2 = \sigma_b^2 + \sigma_c^2$, we get

$$\sigma_a = \sqrt{2\sigma_b^2} = \sqrt{2}\sigma_b$$

Incorrect!

  - We know $a = b + b = 2b$, where the two variables on the R.H.S. are correlated
  - Hence,

$$\sigma_a = \sqrt{4\sigma_b^2} = 2\sigma_b$$

$\sigma_a^2 = \sigma_b^2 + \sigma_c^2$ is valid only if b and c are uncorrelated

Non-linear situations (1)

- Consider $a = b^r c^s$, where r and s are known constant
- The fractional error (assuming that the errors on b and c are uncorrelated)

$$\left(\frac{\sigma_a}{a}\right)^2 = r^2\left(\frac{\sigma_b}{b}\right)^2 + s^2\left(\frac{\sigma_c}{c}\right)^2$$

- When correlation are present

$$\left(\frac{\sigma_a}{a}\right)^2 = r^2\left(\frac{\sigma_b}{b}\right)^2 + s^2\left(\frac{\sigma_c}{c}\right)^2 + 2rs\,\frac{\mathrm{cov}(b,c)}{bc}$$

Non-linear situations (2)

- E.g., consider a measurement of the cross-sections σ_i for two different processes in a given experiment, for instance, producing two or four charged particles in an interaction, respectively
- For thin targets, the cross-sections are

$$\sigma_i = \frac{n_i}{tB}$$

n_i: numbers of the observed interactions
t: thickness of the target
B: beam intensity

- Consider the ratio of the cross sections $q = \sigma_1/\sigma_2$
- If the major contribution to the error in σ_i arises from the statistical uncertainty in n_i, then the errors $\delta\sigma_1$ and $\delta\sigma_2$ are independent, the error on the ratio is

$$\left(\frac{\delta q}{q}\right)^2 = \left(\frac{\delta\sigma_1}{\sigma_1}\right)^2 + \left(\frac{\delta\sigma_2}{\sigma_2}\right)^2$$

- If the main uncertainty in σ_i is due to, for instance, a poorly determined beam flux, then the errors $\delta\sigma_1$ and $\delta\sigma_2$ are correlated, and the error on the ratio is much smaller than that given in the above equation

Linear vs. non-linear situations

- For the linear case, $\sigma_a^2 = \sigma_b^2 + \sigma_c^2$ applies whatever the magnitudes of the individual errors
- For the non-linear case, $\left(\frac{\sigma_a}{a}\right)^2 = r^2\left(\frac{\sigma_b}{b}\right)^2 + s^2\left(\frac{\sigma_c}{c}\right)^2$ applies only when the magnitudes of the individual errors are small
- E.g., let

$$a = \frac{b}{c} = \frac{100 \pm 10}{1 \pm 1} = 100 \pm ?$$

- The error calculated from the above equation gives

$$\left(\frac{\sigma_a}{a}\right)^2 = 1^2\left(\frac{\sigma_b}{b}\right)^2 + (-1)^2\left(\frac{\sigma_c}{c}\right)^2 = \left(\frac{10}{100}\right)^2 + \left(\frac{1}{1}\right)^2 = 1.01$$

$$\sigma_a = \sqrt{1.01} \times 100 \approx 100.$$

- Is 100 ± 100 realistic?

Combining results of different experiments (1)

- When several experiments measure the same physical quantity and give a set of answers a_i with different errors σ_i, then the best estimates of a and its uncertainty σ are

$$a = \frac{\sum \frac{a_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}$$

$$\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$$

- N.B., σ_i here is the true variance

Combining results of different experiments (2)

- E.g. (false situation), measure a counting rate where the rate stays constant. Assume that we measure 1 ± 1 counts in the first hour and 100 ± 10 counts in the second hour. What is the average counting rate?
- With the weighted formulae

$$a = \frac{\sum \frac{a_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} = \frac{\frac{1}{1^2} + \frac{100}{10^2}}{\frac{1}{1^2} + \frac{1}{10^2}} = \frac{2}{1.01} \approx 2$$

$$\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2} = \frac{1}{1^2} + \frac{1}{10^2} \rightarrow \sigma \approx 1$$

2 ± 1 counts per hour

- The value is close to 1 due to the large uncertainty (lower weights) on the second measurement

Combining results of different experiments (3)

- We made the mistake by taking the sample variance as the weight but not the true variance
- Since we assume that the particle flux and the apparatus do not vary over the two hours, the true counting rates are the same, and so are the true variance
- Hence, the correct way of averaging is the simple arithmetic average

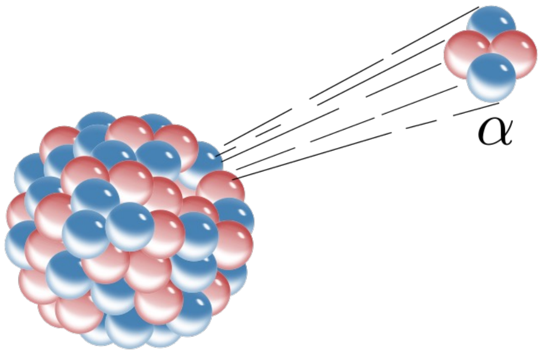
$$a = \frac{\sum a_i}{N} = \frac{1 + 100}{2} = 50.5$$

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{N^2}} = \sqrt{\frac{1^2 + 10^2}{2^2}} \approx 5$$

50.5 ± 5 counts per hour

Combining results of different experiments (4)

- Be careful in combining experiments on, for example, rare decay modes of nuclei or particles
- Likely to be dominated by statistical uncertainties associated with the limited number of observed decays
- Essential to impose a constraint: the expected decay branching ratio is the same for all experiments
- This enable us to assign sensible weights to each experimental observation



Combining results of different experiments (5)

- E.g., a source emitting particles is completely surrounded by two hemispherical counters, one of 100% efficiency and the other of 4%
- The observed numbers of counts in these detectors are 100 ± 10 and 4 ± 2. After correcting for counting inefficiency, the latter becomes 100 ± 50
- Each counter sustends a solid angle of 2π as seen from the source. The average rate is

$$a = \frac{\sum \frac{a_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} = \frac{\frac{100}{10^2} + \frac{100}{50^2}}{\frac{1}{10^2} + \frac{1}{50^2}} = 100$$

$$\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2} = \frac{1}{10^2} + \frac{1}{50^2} \rightarrow \sigma = 9.8$$

Combining results of different experiments (6)

- The total decay rate over 4π is 200 ± 19.6
- The corresponding effective number of events is

The knowledge of extra information (relative solid angles) results in improved precision

$$n_{\text{eff}} = \frac{T^2}{\delta^2} = \frac{200^2}{19.6^2} = 104$$

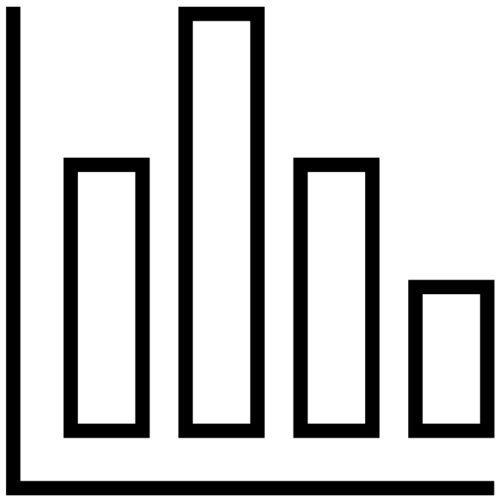
- Consisting of 100 observed events from the first and 4 from the second counter

- If we do not know the relative solid angles of the two counters, we will estimate the total counts by simply summing the two results

$$(100 + 100) \pm \sqrt{10^2 + 50^2} = 200 \pm 51$$

- This is of considerably lower precision

Probability and statistics



Probability and statistics

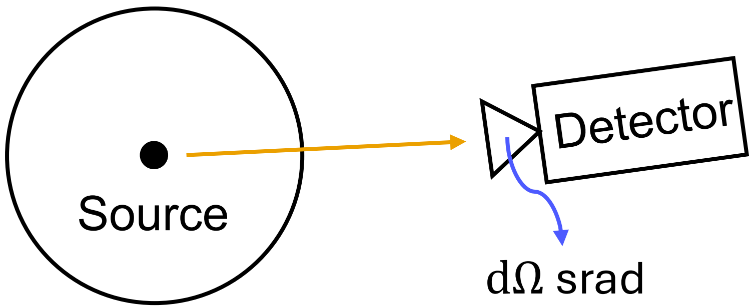
- **Probability**: from theory to data
  - The essential circumstances are kept constant, yet repetitions of the experiment produce different results
  - Start with a well-defined problem and calculate all possible outcomes of a specific experiment
  - → This corresponds to predictions of experiments
- **Statistics**: from data to theory
  - Try to solve the inverse problem
  - Using the data to deduce what are the rules or laws relevant to our experiment
  - → Analyzing experimental data

Probability

- The probability p of obtaining a certain specified result on performing one of these measurements is

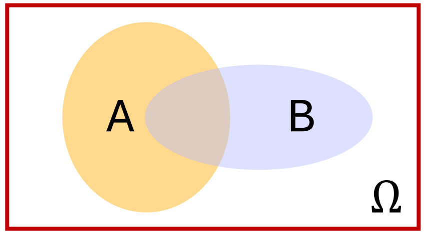
$$p = \frac{\text{number of occations on which that result occurs}}{\text{total number of measurements}}$$

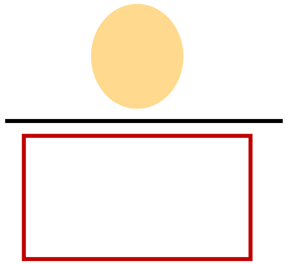
- E.g., a radioactive source decays isotropically in space. A counter which detects one of the decay products sustends a solid angle $\mathrm{d}\Omega$ srad as seen by the source. The probability that in any given decay, the decay product will pass through the detector is $\mathrm{d}\Omega/4\pi$.

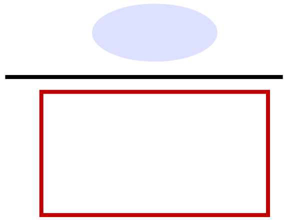


Example of two events (1)

- Let *A* and *B* be two events in a sample space Ω



$p(A) =$ 

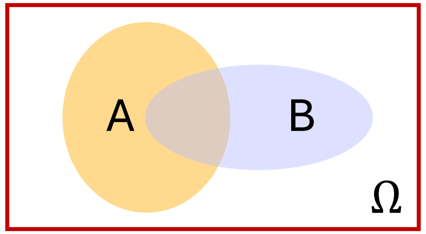
$p(B) =$ 

- Conditional probability $p(A|B) =$ probability of *A* being true, given *B* is true. Similar for $p(B|A)$.

$p(A|B) =$

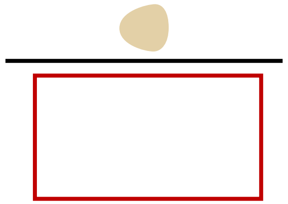
$p(B|A) =$

Example of two events (2)



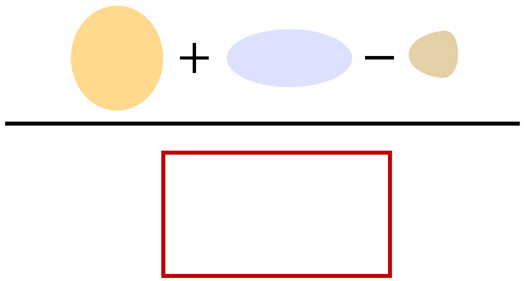
- Probability of both *A* and *B* be true

$$p(A \text{ and } B) = p(A \cap B) =$$



- Probability of either *A* or *B* be true

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) =$$



Rules of probability (1)

- Rule 1
  - The probability of any particular event occurring is

$$0 \leq p \leq 1$$

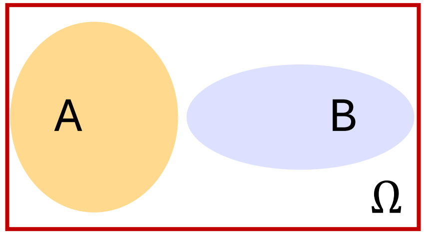
  - $p = 0$ implies that this event will never occur
  - $p = 1$ implies that this event will always occur

- E.g., for a die, the probability of
  - throwing a seven is zero
  - obtaining any number less than 10 is unity
  - obtaining an even number is $1/2$

Rules of probability (2)

- Rule 2
  - The probability that at least one of the events *A* or *B* occurs is

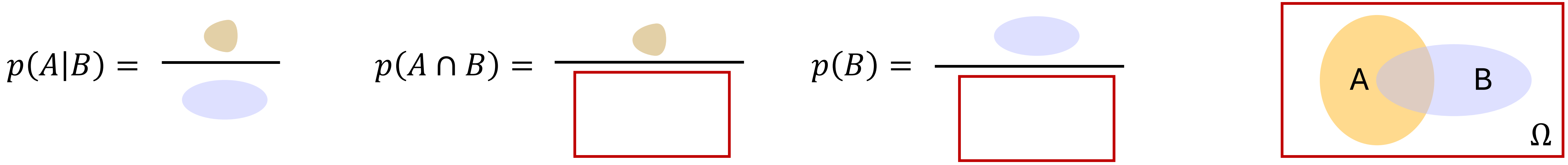
$$p(A \text{ or } B) = p(A \cup B) \le p(A) + p(B)$$

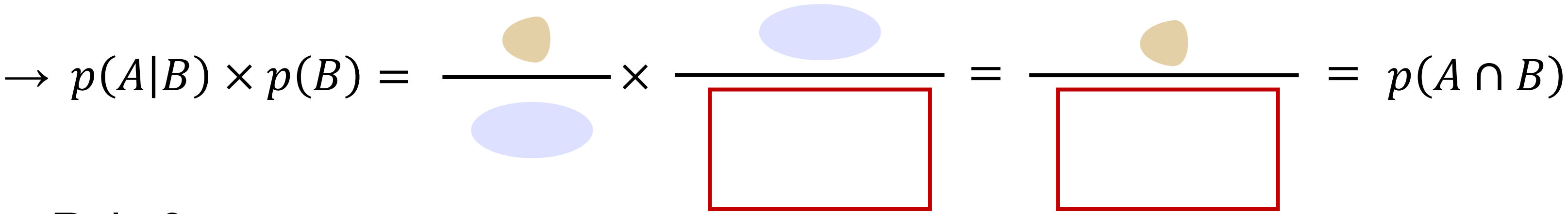


  - The equality applies if the events *A* and *B* are exclusive

- E.g., for a die, the probability of
  - throwing a three or an even number is $p(3 \cup \text{even}) = p(3) + p(\text{even}) = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$
  - obtaining a number below 3.5 or even is $p(< 3.5 \cup \text{even}) = \frac{5}{6} < p(< 3.5) + p(\text{even}) = \frac{1}{2} + \frac{1}{2} = 1$

Rules of probability (3)





- Rule 3
  - The conditional probability of *A* is obtained by dividing the number of times that both *A* and *B* are observed together by the total number of times that *B* occurs

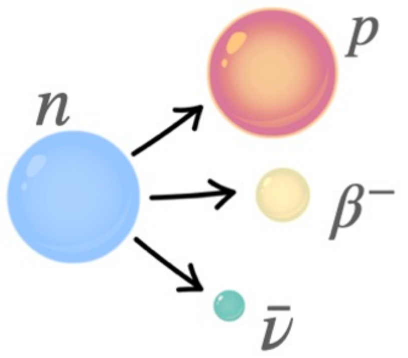
$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Rules of probability (4)

- Rule 3
  - $$p(A|B) = \frac{p(A \cap B)}{p(B)}$$
  - If the occurrence of *B* does not affect whether or not *A* occurs (*A* and *B* are independent), then

$$p(A|B) = p(A) \rightarrow p(A \cap B) = p(A)p(B)$$

- E.g., in a beta decay, $n \rightarrow p + e^- + \bar{\nu}$, the decay energy is shared between e^- and $\bar{\nu}$. We can obtain the probability of e^- having a certain high fraction of the available energy. We can do a similar calculation for $\bar{\nu}$. But the probability of having both of them having high energies is zero, as this is constrained by energy conservation.



Probabiltiy density function (PDF)

- The probability to measure a value x in the interval $[x, x + \mathrm{d}x]$ is given by the probability density function

$$f(x) = \lim_{\mathrm{d}x \to 0} \frac{P(x \leq \mathrm{result} \leq x + \mathrm{d}x)}{\mathrm{d}x}$$

- P is a measure of how often a value of x occurs in a given interval

$$P(x_1 \leq x \leq x_2) = \int_{x1}^{x2} f(x)\mathrm{d}x$$

- $f(x) \geq 0$ and is normalized to 1

$$\int_{x_{\mathrm{min}}}^{x_{\mathrm{max}}} f(x')\mathrm{d}x' = 1$$

Cumulative distribution function (CDF)

- The probability that in a measurement of a variable x', the value is less than x is given by the cumulative distribution function.
- It is related to the probability density function by

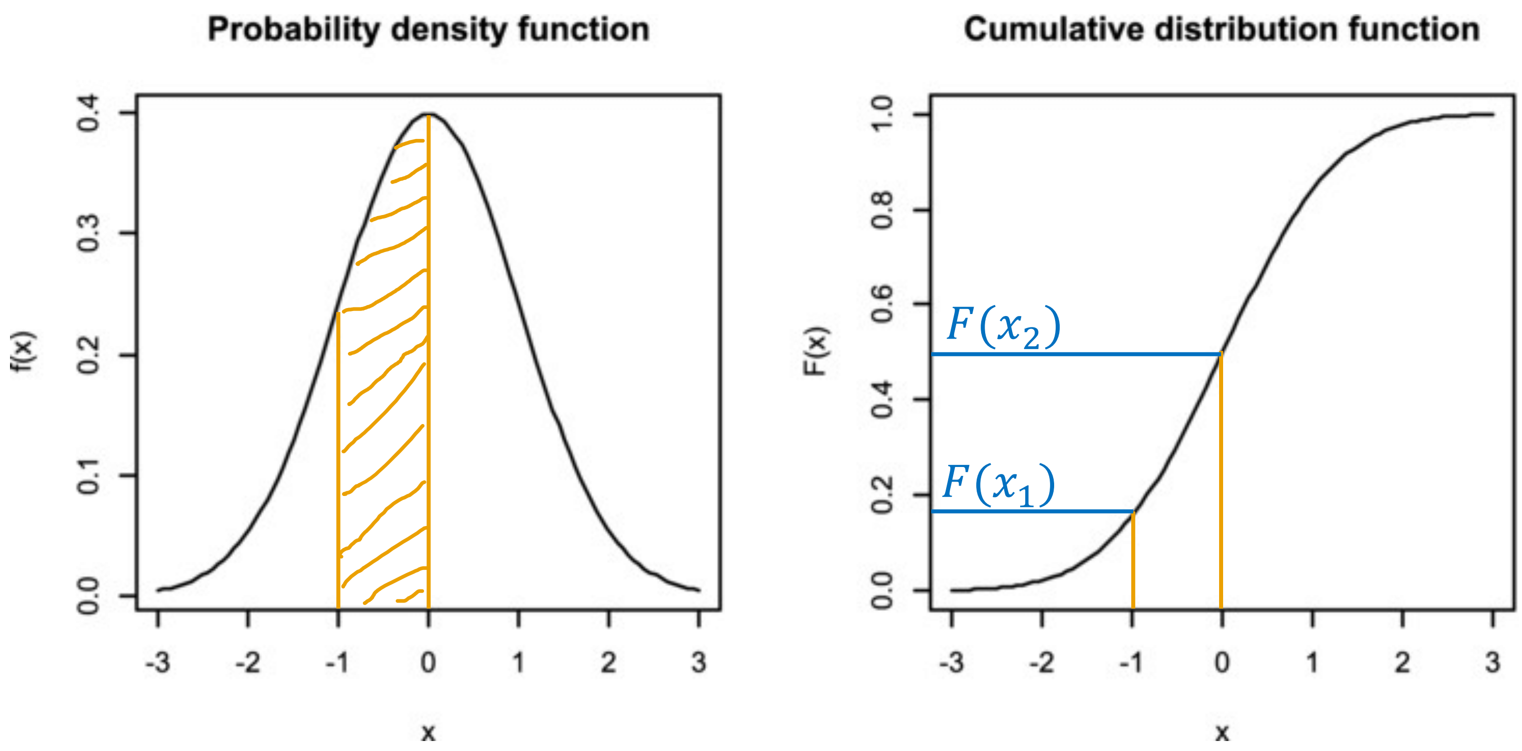
$$F(x) = \int_{x_{\min}}^{x} f(x')\mathrm{d}x'$$

- $F(x)$ is a continuous, non-decreasing function
- $F(-\infty) = 0$ and $F(+\infty) = 1$

Relation between PDF $f(x)$ and CDF $F(x)$

$$f(x) = \frac{\partial F(x)}{\partial x}$$

$$F(x) = \int_{x_{\min}}^{x} f(x')\mathrm{d}x'$$



$$P(x_1 \le x \le x_2) = \int_{x1}^{x2} f(x')\mathrm{d}x'$$

$$= F(x_2) - F(x_1)$$

Colin Leverger Thesis for PhD

Expectation value

- Expectation value represents the average outcome of a random variable or a physical quantity if an experiment or measurement is repeated many times

- For a discrete random variable X with values x_i and probabilities $p(x_i)$, the expectation value is

$$\langle X \rangle = \sum_i x_i p(x_i)$$

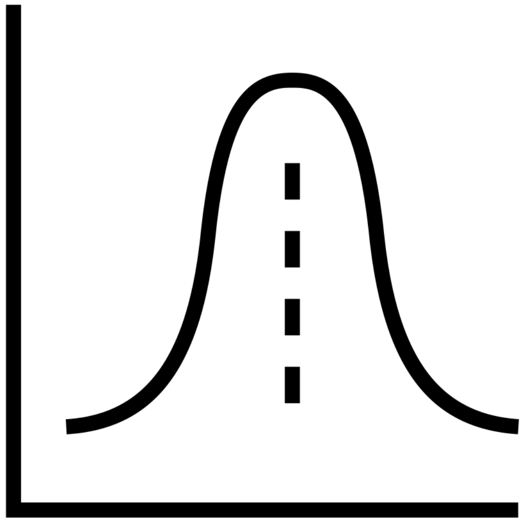
- For a continuous random variable X with a probability density function of $p(x)$, the expectation value is

$$\langle X \rangle = \int_{-\infty}^{+\infty} x p(x) \mathrm{d}x$$

Statistics

- Why do we do experiments? How to summarize our data efficiently?
  - **Parameter determination**: determine the numerical values of some physical quantities
    - Determine the value of a parameter and its uncertainty in an unbiased and efficient way
  - **Hypothesis testing**: test whether a particular theory is consistent with our data
    - E.g., we are trying to answer the question "Is the shape of the electron's energy spectrum as observed in a beta decay process in agreement with the Fermi theory?"
    - The answer is not just yes or no, but will contain a statement of how confident we are
  - Parameter determination and hypothesis testing often coexist in a real-life situation

Distributions

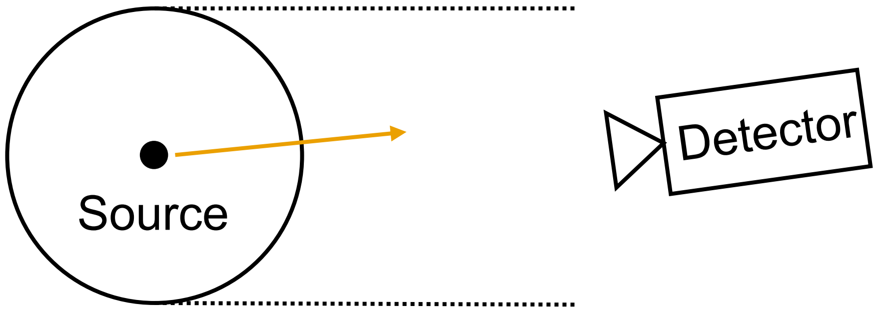


1. Binomial distribution (1)

- Fixed number of trials N
- Only two outcomes: success (with a probability p) or failure (with a probability of $1-p$)
- The probability of obtaining r successes is

Probability mass function (PMF) $P(r) = \frac{N!}{r!\,(N-r)!} p^r (1-p)^{N-r}$, where $r \subset [0, N]$

- E.g., the angles that the decay products from a given source make with a fixed axis are measured. If the expected distribution is known (e.g., isotropically), what is the probability of observing r decays in the forward hemisphere from a total sample of N decays?



1. Binomial distribution (2)

- Expectation value of the number of success r

$$\langle r\rangle = \bar{r} = \sum rP(r) = Np$$

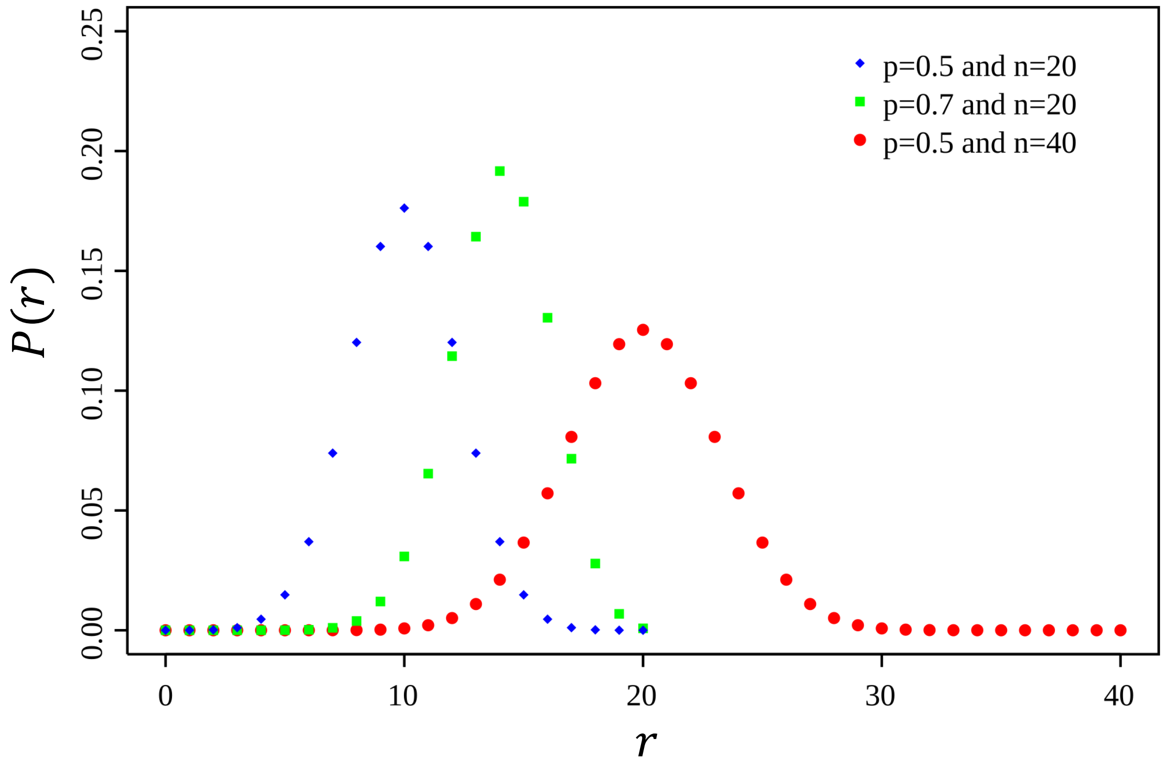
- Variance of the distribution

$$\sigma^2 = Np(1-p)$$

- When p is unknown, we can estimate with

$$s^2 = \frac{N}{N-1} N\frac{\bar{r}}{N}(1-\frac{\bar{r}}{N})$$

By Tayste - Own work, Public Domain



2. Poisson distribution (1)

- Limit of the binomial distribution as
  - $N \longrightarrow \infty$
  - $p \longrightarrow 0$
  - $Np = \text{constant} = \mu t = \lambda$
- The probability of observing r independent events in a time interval t, when the counting rate is μ and the expected number of events in the time interval is λ

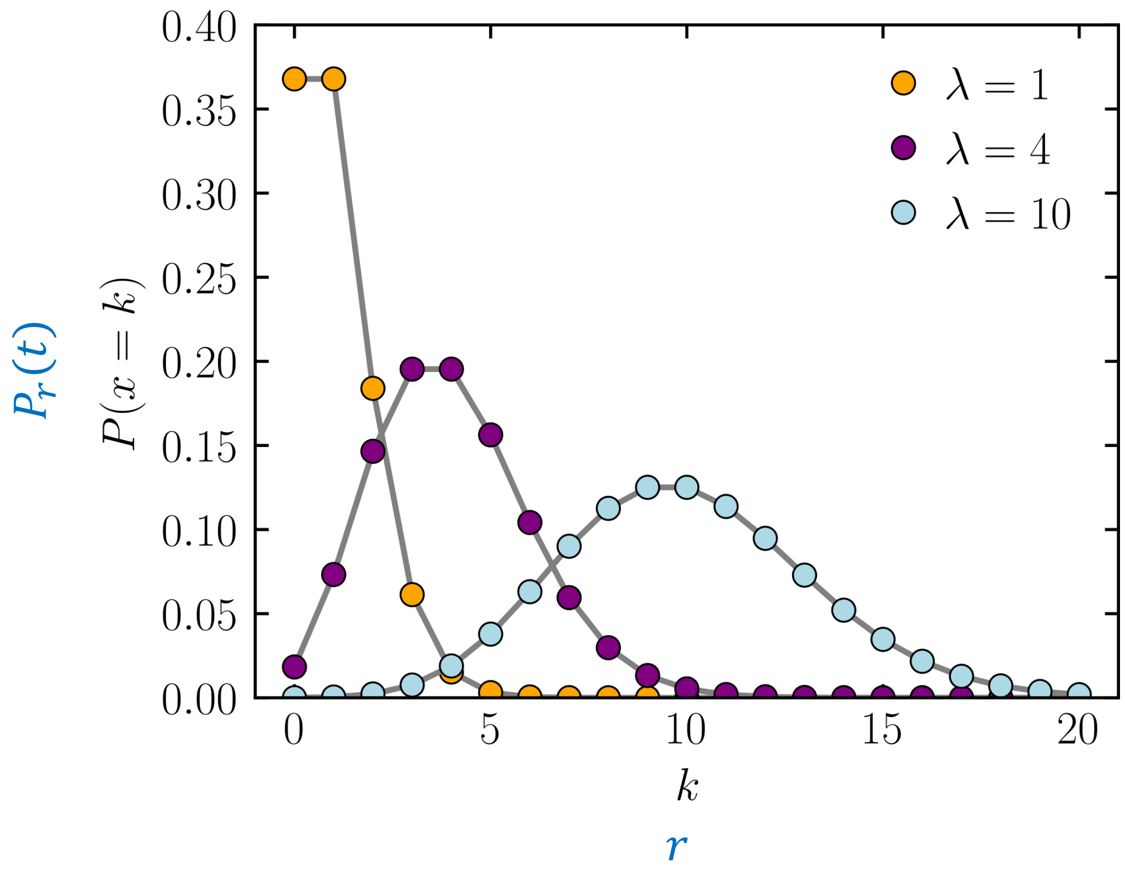
Probability mass function (PMF) $$P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t} = \frac{\lambda^r}{r!} e^{-\lambda}$$

- E.g., the number of particles detected by a counter in a time t, in a situation where the particle flux ϕ and the detector efficiency are independent of time, and where counter dead-time τ is small such that $\phi\tau \ll 1$

2. Poisson distribution (2)

- The mean (μ) and variance (σ^2) of a variable r following a Poisson distribution are both λ

By Skbkekas - Own work, CC BY 3.0

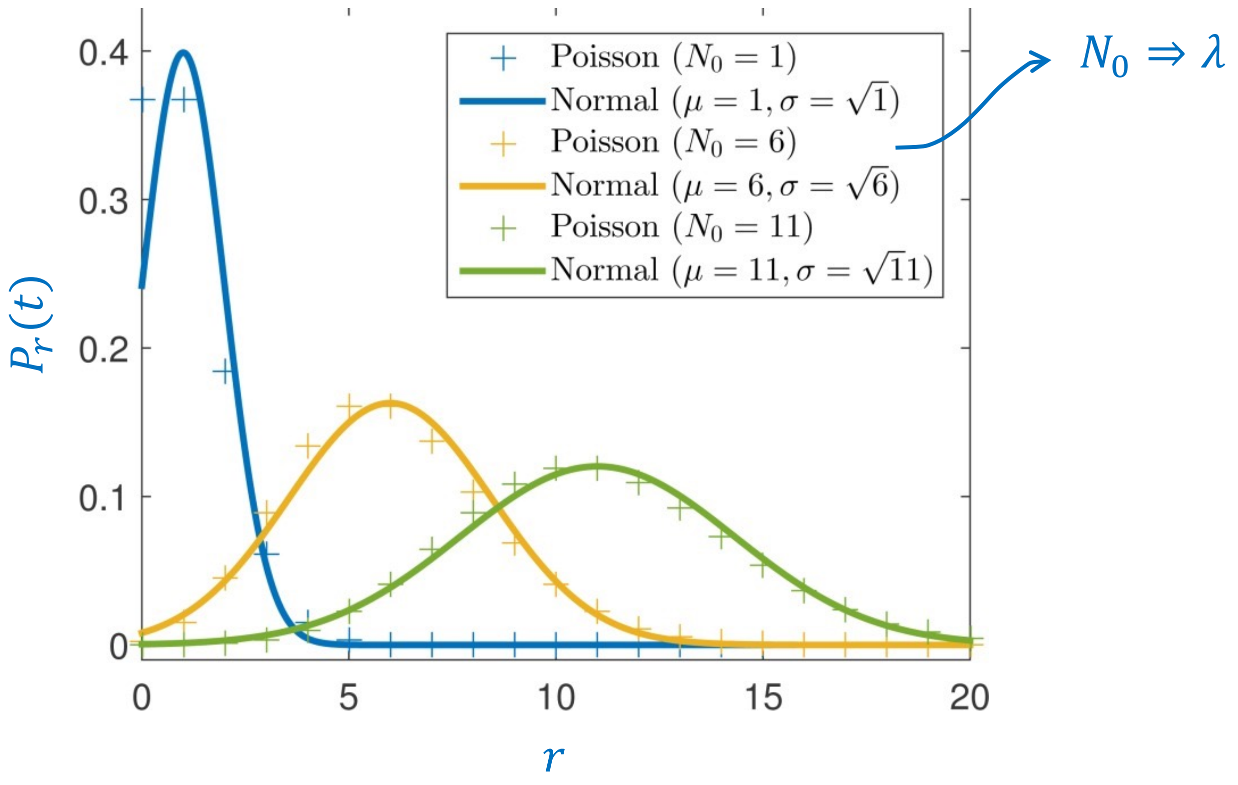


2. Poisson distribution (3)

- As $\lambda \longrightarrow \infty$, the Poisson distribution tends to a Gaussian (normal) one
- $\lambda \geq 5$ is usually a good approximation

- Poisson distribution is defined at non-negative integers
- Gaussian distribution is continuous and extends down to $-\infty$

Medical Imaging Systems: An Introductory Guide [Internet]



2. Poisson distribution (4)

- E.g., take a book with 500 pages that contains 50 typos in total
- Question: what is the probability that on a randomly chosen page, you'll find zero, one, or two typos?
- Answer: use Poisson PDF

$$P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t} = \frac{\lambda^r}{r!} e^{-\lambda}$$

  - Average number of typos per page is $\lambda = \frac{50}{500} = 0.1$
  - $p(0, \lambda) = \frac{1}{1} \times \exp(-0.1) = 90.5\%$
  - $p(1, \lambda) = \frac{0.1}{1} \times \exp(-0.1) = 9.05\%$
  - $p(2, \lambda) = \frac{0.01}{1} * \exp(-0.1) = 0.45\%$

3. Gaussian distribution (1)

- The distribution of a variable x with a mean μ and a standard deviation σ is

Probability density function (PDF) $$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

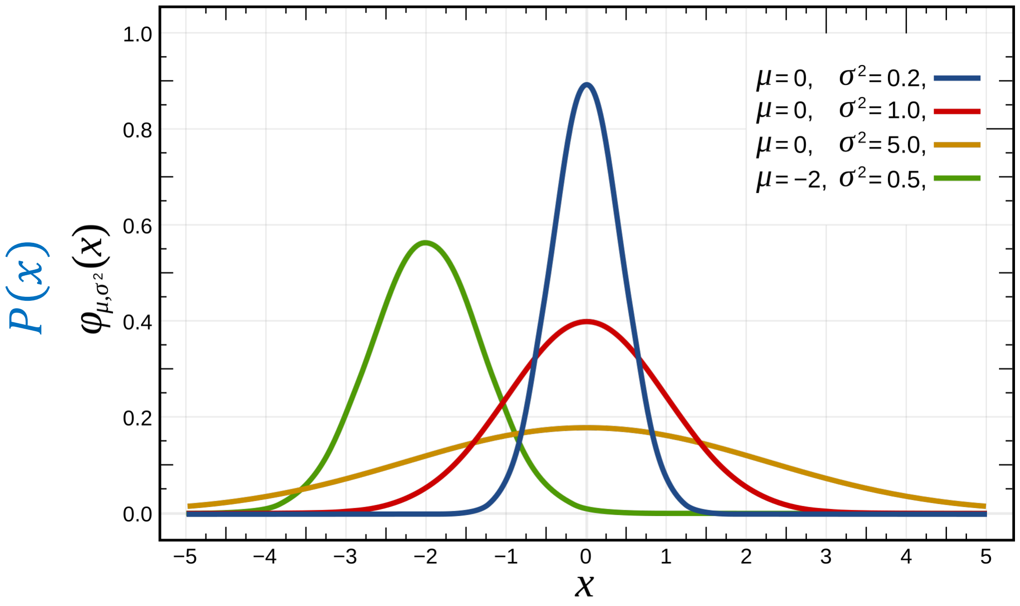
- The distribution is symmetric about $x = \mu$ at which point $P(x)$ has its maximum value
- σ characterizes the width of the distribution
- The factor $\left(\sqrt{2\pi}\sigma\right)^{-1}$ ensures that

$$\int_{-\infty}^{+\infty} P(x)\mathrm{d}x = 1$$

3. Gaussian distribution (2)

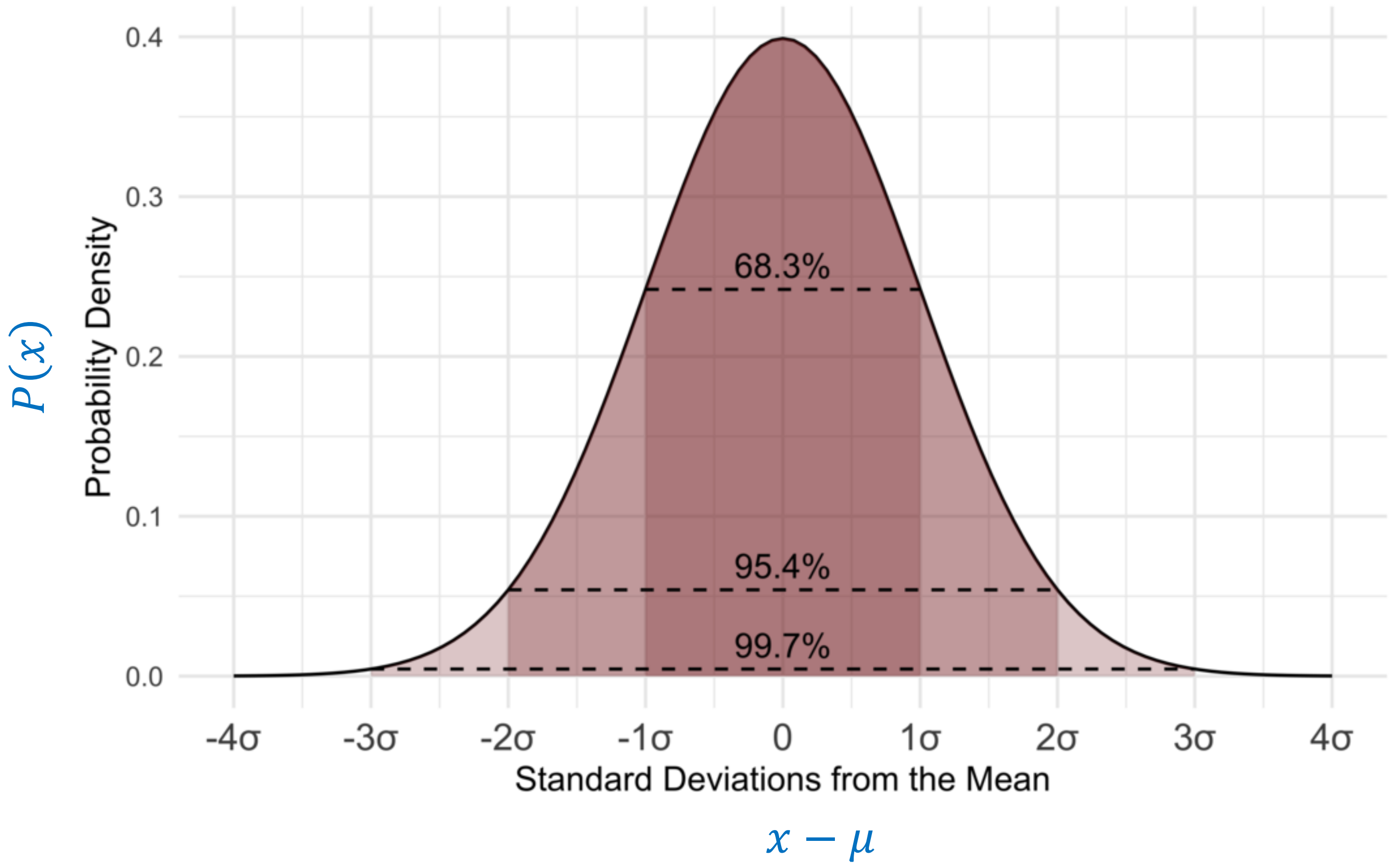
- The mean square deviation from its mean is σ^2
- The height of the curve at $x = \mu \pm \sigma$ is $1/\sqrt{e}$ of the max.
- The height of the curve at its maximum is $1/\left(\sqrt{2\pi}\sigma\right)$
  → the smaller the σ, the narrower the distribution, and the higher the peak
- Full-width at half-max. $\mathrm{FWHM} = 2.355\sigma$

By Inductiveload - Own work, Public Domain



3. Gaussian distribution (3)

- The fractional area underneath the curve and with $\mu - \sigma \leq x \leq \mu + \sigma$ is 0.68



Bell curve

[Normal Distribution (Bell Curve): Definition, Word Problems]

3. Gaussian distribution (4)

- E.g., we measure the life-time of the neutron as $950 \pm 20\ s$ in our experiment. A certain theory predicts that the life-time is $910\ s$. To what extent are these numbers in agreement?
- Define

$$t = \frac{x - \mu}{\sigma}$$

- This example gives $t = \frac{950-910}{20} = 2$. The area in the tails of the Gaussian corresponds to 4.6%.
- Thus, if 1000 experiments of the same precision as ours were performed, and if the theory is correct, and if the experiments are bias-free, then we expect about 46 of them to differ from the predicted value by at least as much as ours

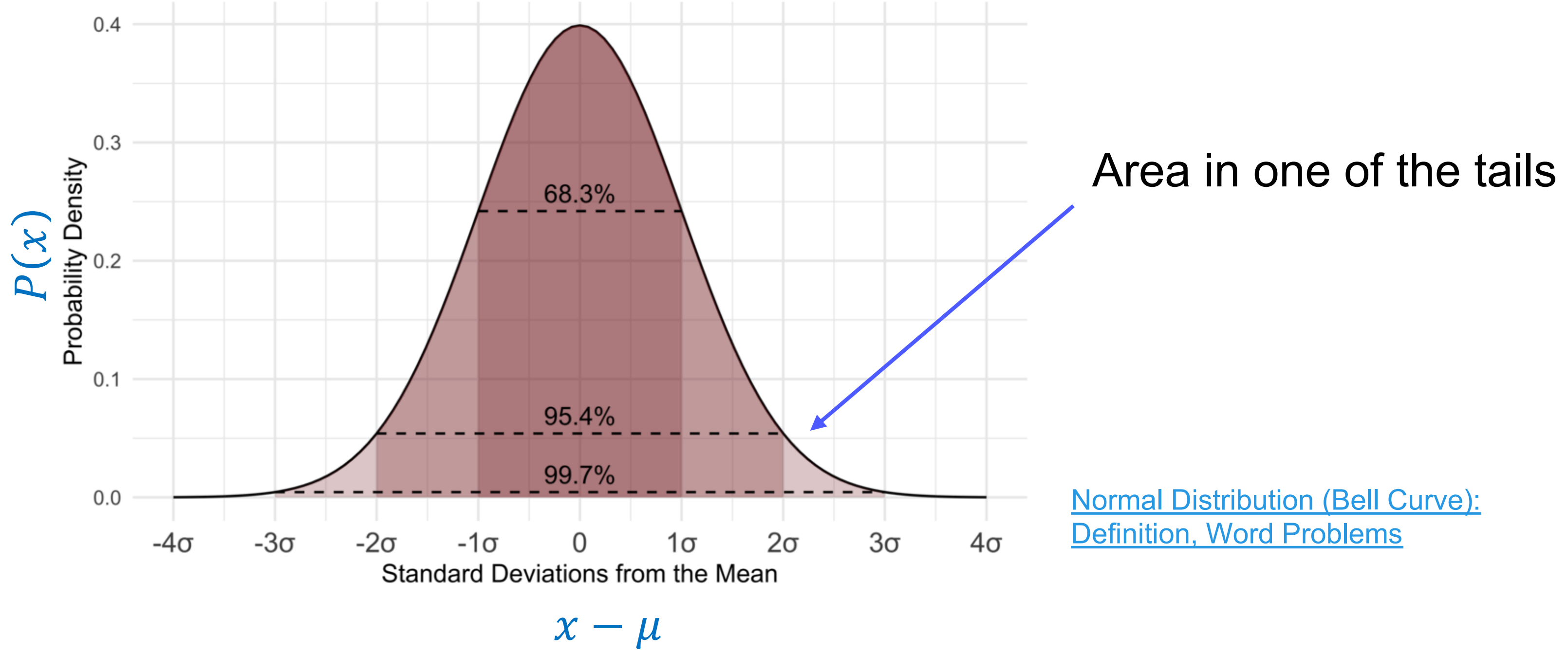
3. Gaussian distribution (5)

- σ in the definition of t is supposed to be the true standard deviation. In some cases, we simply estimate this from the observed spread of a repeated set of measurements
- In most cases, the theoretical estimate y_{th} will have an uncertainty σ'. With our measured value being $y_{\text{ob}} \pm \sigma$, we redefine

$$t = \frac{y_{\text{ob}} - y_{\text{th}}}{\sqrt{\sigma^2 + \sigma'^2}}$$

3. Gaussian distribution (6)

- Sometimes, we are interested in the sign of possible deviations from predicted value
- E.g., a nuclear reactor will explode if the neutron production rate is greater than a certain value λ_c. We measure the rate at $\lambda \pm \sigma$ where λ is slightly smaller than λ_c
- We are interested in knowing when the true rate is greater than λ_c



Normal Distribution (Bell Curve): Definition, Word Problems

Central limit theorem

- Consider the sum X of n independent variables x_i, each taken from a (possibly different) distribution with expectation value μ_i and variance σ_i^2
- The distribution for $X = \sum x_i$ has the following properties:
  - Its expectation value is $\langle X \rangle = \sum \mu_i$
  - Its variance is $V[X] = \sum \sigma_i^2$
  - It becomes Gaussian distributed for $n \to \infty$
- N.B., special case: if all x_i are from an identical PDF
  - $\langle X \rangle = n\mu$
  - $V[X] = n\sigma^2$
- If x_i are not independent, then only the expectation value $\langle X \rangle = \sum \mu_i$ is true; not the variance

Interpretation of probability

Frequentist and Bayesian approaches

- Frequentist
    - Probability is interpreted as the long-run relative frequency of an event occurring in repeated identical trials

- Bayesian
    - Probability is interpreted as a degree of belief or subjective confidence in a particular event or hypothesis, which can be updated with new evidence

Frequentist approach

- Emprical definition: frequency of occurrence
    - Perform experiment N times in identical trials. Assume event E occurs k times

$$P(E) = \lim_{N\to\infty} k/N$$

    - Intuitive interpretation in particle physics (many repetitions of events)
- Useful, but has some limitations
    - Cannot be proven that it converges. How large is N?
    - What does repeatable under identical conditions mean?
- E.g., “It will probably rain tomorrow.”
    - There is only ONE tomorrow. We can do this only ONCE.

Bayesian approach

Thomas Bayes

- Probability is seen as a degree of believe → credibility of a statement
- Deals with the probability of a hypothesis or a theory

$$P(\text{hypothesis } E) = \text{degree of belief that } E \text{ is true}$$

- Bayesian probability is a state of our information/knowledge

$$P(E) = P(E|I) \text{ probability of } E \text{ given } I$$

- Includes our "prior" knowledge about the theory, situation, etc…
- E.g., "It will probably rain tomorrow."
  - We believe this stronger, if it has been raining for several days, and is still in forecast

Bayes' theorem (1)

- From the conditional probability

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \quad \& \quad p(B|A) = \frac{p(B \cap A)}{p(A)} \quad \rightarrow \quad p(A|B) \times p(B) = p(B|A) \times p(A) = p(A \cap B)$$

- Bayes' theorem

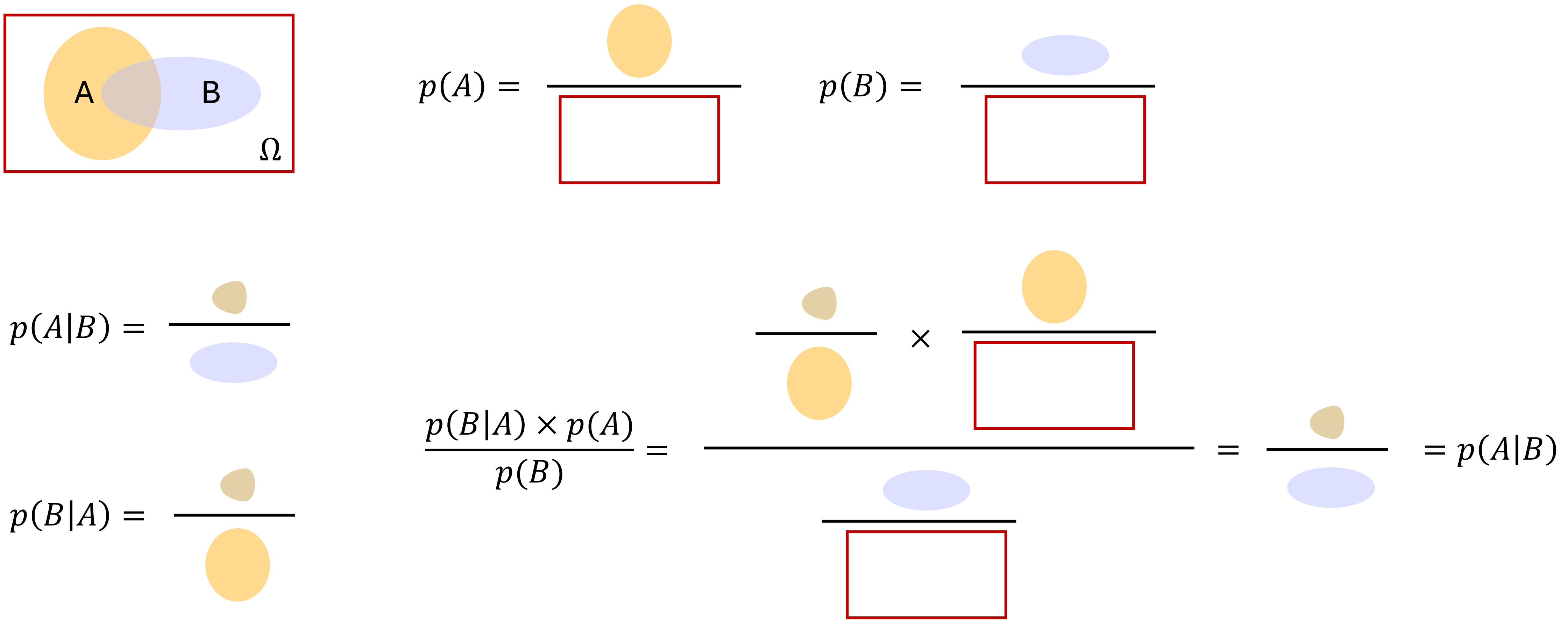
$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

Probability of A being true, given B is true

- This further depends on model and information $p(A|B) = p(A|B; \text{assuming Model } M, \text{Information } I)$

Bayes’ theorem (2)

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$



Bayesian inference

- Set $A = \text{theory}$, $B = \text{data}$

$$p(\text{theory}|\text{data}) = \frac{p(\text{data}|\text{theory}) \times p(\text{theory})}{p(\text{data})} = p(\text{theory}|\text{data}; M, I)$$

- $p(\text{data}|\text{theory})$ called likelihood: how likely it is to measure data, assuming theory is correct and known
- $p(\text{theory}|\text{data})$ called posteriori probability: probability of the theory (or its parameters) to be true given the set of the measured data (+ model M, information I) → "our degree of believe in the theory, given these data"
- $p(\text{theory})$ called prior probability, needs to be known; completely independent of the measured data
- $p(\text{data})$ called evidence: probability of data assuming any model

Bayesian probability as "learning"

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

- Before the observation B, our degree of belief of A is $p(A)$ → prior probability
- After observing B, our degree of belief becomes $p(A|B)$ → posterior probability
- Strengthen our "degree of belief" by subsequent observations
  - E.g., combine experiments, i.e., multiple probabilities → process of learning
- Consider $p(B)$ as a normalization factor

$$p(B) = \sum_i p(B|A_i) \times p(A_i) \quad \text{if} \quad \bigcup_i A_i = \Omega \;\; \text{and} \;\; \bigcap_i A_i = 0$$

- Bayes' theorem is reformuated to

$$p(A_i|B) = \frac{p(B|A_i) \times p(A_i)}{\sum_j p\left(B\middle|A_j\right) \times p(A_j)}$$

Bayes' theorem applied (1)

- Consider systems, where we only have two possible states, e.g., digital decision [0, 1]
- Rewrite $p(B)$ as $p(not\ A)$
  - $p(A) + p(not\ A) = 1$
  - $p(not\ A) = 1 - p(A)$
  - $p(B) = p(B|A)\times p(A) + p(B|not\ A)\times p(not\ A) = p(B|A)\times p(A) + p(B|not\ A)\times\left(1 - p(A)\right)$

- Bayes' theorem can be rewritten as

$$p(A|B) = \frac{p(B|A)\times p(A)}{p(B|A)\times p(A) + p(B|not\ A)\times\left(1 - p(A)\right)}$$

Bayes' theorem applied (2)

- E.g., influenza virus has infected 0.1% of the swan colony. A new influenza test has 98% probability to detect virus is birds are infected; detection error (false positive) is 3%.
  - $p(flu) = 0.001$ and $p(no\ flu) = 1 - 0.001 = 0.999$
  - $p(+|flu) = 0.98$ (infected; test is positive)
  - $p(-|flu) = 1 - 0.98 = 0.02$ (infected, test is negative)
  - $p(+|no\ flu) = 0.03$
  - $p(-|no\ flu) = 1 - 0.03 = 0.97$
- Question: what is the probability that one (arbitrarily picked) swan really has influenza, if the test reacts positive?
  - $p(flu|+) = \frac{p(+|flu)\times p(flu)}{p(+|flu)\times p(flu)+p(+|no\ flu)\times p(no\ flu)} = \frac{0.98\times 0.001}{0.98\times 0.001+0.03\times 0.999} = 0.03 = 3\%$
  - The influenza is present only in 3% of the cases in which the test is positive

Bayes' theorem applied (3)

- E.g., use a Cherenkov counter to detect pions and kaons. Suppose that the counter is 95% efficient for pions but also has a 6% probability to trigger on kaons
  - $p(+|\pi) = 0.95$
  - $p(+|k) = 0.06$
- Question: if the counter triggers, what is the probability that it was a pion and what that it was a kaon?
  - $p(\pi|+) = \frac{p(+|\pi)\times p(\pi)}{p(+)}$; $p(K|+) = \frac{p(+|K)\times p(K)}{p(+)}$
  - A convenient way is through ratios: $\frac{p(\pi|+)}{p(K|+)} = \frac{p(+|\pi)\times p(\pi)/p(+)}{p(+|K)\times p(K)/p(+)} = \frac{p(+|\pi)}{p(+|K)} \times \frac{p(\pi)}{p(K)}$
  - $p(+)$ cancels in the ratio (removes systematic errors)
  - No need to consider all possible hypotheses, which are in practice often unknown
  - Ratio of $p(\pi)/p(K)$ plays an important role

With these analysis methods, we can perform…

Parameter determination: determine the numerical values of some physical quantities

Hypothesis testing: test whether a particular theory is consistent with our data

Parameter determination (1)

- Determine the underlying distribution for a set of measurements $\{x_i, y_i\}$ with known uncertainties σ_i
- The y_i are assumed given by a function $y = f(x|a)$ with parameters a → estimate parameters a
- E.g., determine the number of events in a decay

Parameter determination (2)

- Assume N independent measurements of a random variable x, which is distributed according to an unknown PDF $f(x)$
  - Want to infer the properties of $f(x)$ from the measurements of x
  - Determine the underlying distribution: start with hypothesis for $f(x|a)$ and find the optimal parameters a in $f(x|a)$ from the given set of measuremts x_i → parameter estimation
  - → Want to find the best estimate $\hat{a}$ for the true parameter a
- Estimation leads to an imprecise result whose imprecision is known

Hypothesis testing (1)

- Make a statement about how well the observed data stand in agreement (accept) or not (reject) with a given predicted distribution, i.e., a hypothesis
- Formulate the hypothesis → collect data → test the data against the hypothesis → accept or reject
- An hypothesis is a statement that can be proved experimentally
- Define null hypothesis H_0 (typically the background only hypothesis) to be the hypothesis under consideration (vs. alternative hypothesis H_1, describes the presence of some signal)
- One cannot meaningfully accept a hypothesis; one can only reject it → you always check that a hypothesis is not consistent with data

Hypothesis testing (2)

- To quantify the agreement between the observed data and a given hypothesis → construct a function of the measured variables x and the given hypothesis H

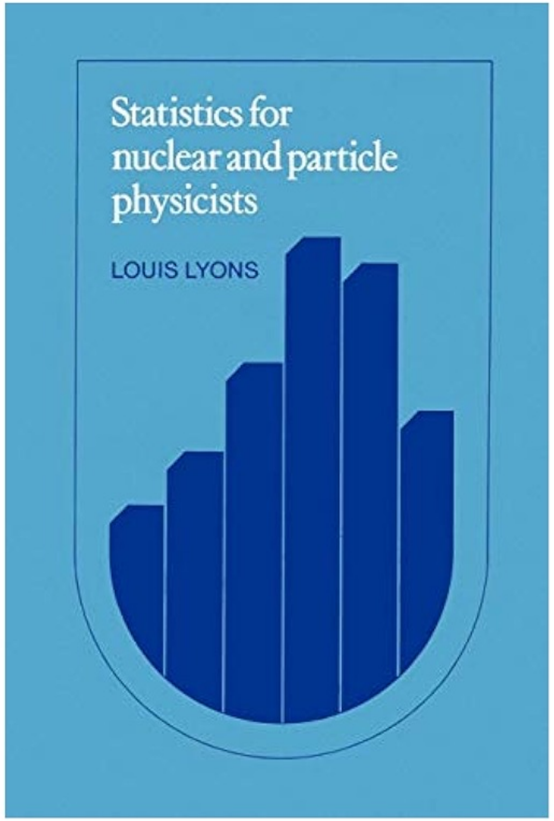
$$t(x|H)$$

x: test statistics
H: hypothesis
$t(x|H)$: distribution of the test statistics under the hypothesis H

- Typical test statistics, e.g., the integral of the data above a certain value ($\text{energy} > 10\ \text{GeV}$), the data itself
- The choice of the test statistic $t(x)$ depends on the particular case. Different test statistics, given the same data, will be distributed differently

What you learned in these lectures?

- Experimental errors
- Probability and statistics
- Distribution
- Interpretation of probability



- For topics related to applying analysis methods to perform parameter determination and hypothesis testing, refer to the textbook:
  - *Statistics for nuclear and particle physics, L. Lyons,* Cambridge University Press

Thank you very much.

Questions?