

Statistics and the Treatment of Experimental Data aka Counting Statistics and Error Analysis

FYS4505 – H2025

A. Read

UiO • University of Oslo

Introduction

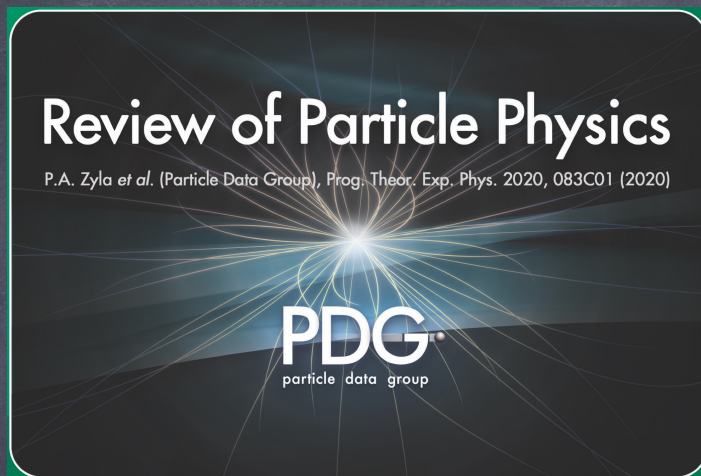
Relevant learning outcome

- «You can apply statistical methods and computational tools in the interpretation of results from measurements»
- and
- « have sufficient knowledge about measurement techniques to assess detector usage in different situations. »

Typical cases

- Best estimate and uncertainty of the mean of a data set (uniform or different errors/data point)
- Counting rates, with uncertainty
- **Lower (!)** limit of a half-life (or lifetime) with a null result
- Efficiency of a detector, with uncertainty
- Propagating uncertainties through functions of observed results
 $x \pm \delta x \rightarrow f(x) \pm \delta f(x, \delta x) \rightarrow f(\vec{x}) \pm \delta f(\vec{x}, \delta \vec{x})$
- Parametric fit to data, i.e., «curve-fitting»
- Is the data consistent with a statistical model of the data?

PDG reviews (<https://pdg.lbl.gov/index.html>)



- Very compact summaries of essential methods and tools in HEP and NP!
- Leo chapter 4 a good complement

Experimental Methods and Colliders

31. Accelerator physics of colliders (rev.)	521
32. High-energy collider parameters (rev.)	529
33. Neutrino beam lines at high-energy proton synchrotrons (rev.)	534
34. Passage of particles through matter (rev.)	535
35. Particle detectors at accelerators (rev.)	551
36. Particle detectors for non-accelerator phys. (rev.)	589
37. Radioactivity and radiation protection (rev.)	612
38. Commonly used radioactive sources (rev.)	618

Mathematical Tools

39. Probability (rev.)	621
40. Statistics (rev.)	626
41. Monte Carlo techniques (rev.)	642
42. Monte Carlo event generators	646
43. Monte Carlo neutrino event generators (rev.)	657
44. Monte Carlo particle numbering scheme (rev.)	661
45. Clebsch-Gordan coefficients, spherical harmonics, and d functions	665
46. $SU(3)$ isoscalar factors and representation matrices	666
47. $SU(n)$ multiplets and Young diagrams	667

Typical inference process

- Define what has been measured
- «Let's write down the **likelihood**», i.e construct the statistical model for the data x of our experiment: $\mathcal{L}(x)$
- Identify the parameters of interest (POI): μ
- Identify model parameters we need but aren't of primary interest
 - Quantify additional sources of uncertainty and associate them to probability distributions and parameters
 - Nuisance parameters (NP): θ
- Maximize the likelihood w.r.t the parameters of the model $\mathcal{L}(x|\mu, \theta)$
- Estimate uncertainties («confidence intervals» or «confidence limits»)

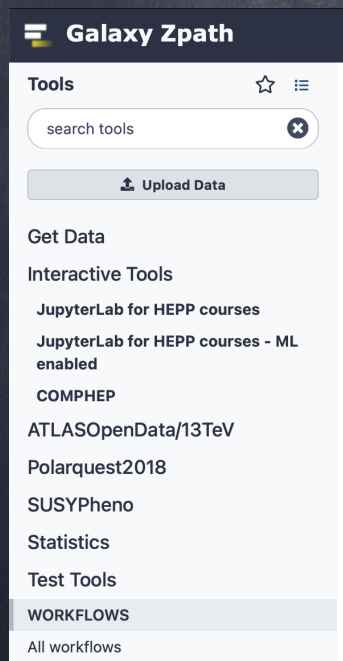
Maximum likelihood

- Maximizing the likelihood $\mathcal{L}$ (simply the probability distribution of the data given the statistical model-parameters) gives **efficient** (and in the large-sample limit, **unbiased**) point-estimates of the model parameters.
- It turns out to be **very** convenient to minimize $-2\ln\mathcal{L}$
- Normally distributed data will lead us naturally to χ^2 —minimization and least-squares (used e.g. in fitting a curve to data points with or without specified uncertainties, respectively)
 - Uncertainty/point specified: χ^2 —minimization
 - No specified uncertainty per point: least-squares
- More reading: [STK1000](#)

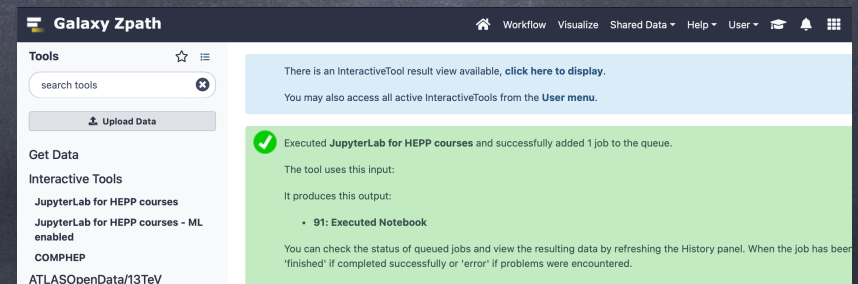
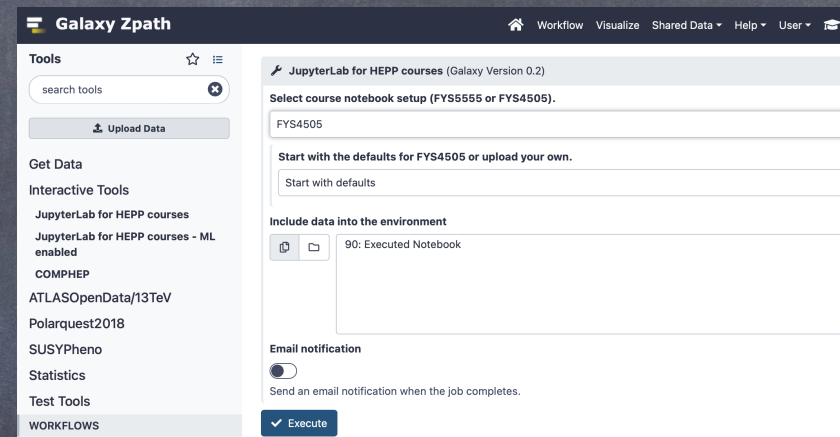
Jupyter (python) notebooks

- Nearly all of the figures here with “data” were made in Jupyter notebooks that are available on <https://github.uio.no/FYS4505/data-analysis/tree/main/python>
- Depend on scipy, numpy, matplotlib, and LaTeX commands such as dvipng – they must be in the PATH
- All this is set up on galaxy-hepp.hpc.uio.no

galaxy-hepp.hpc.uio.no

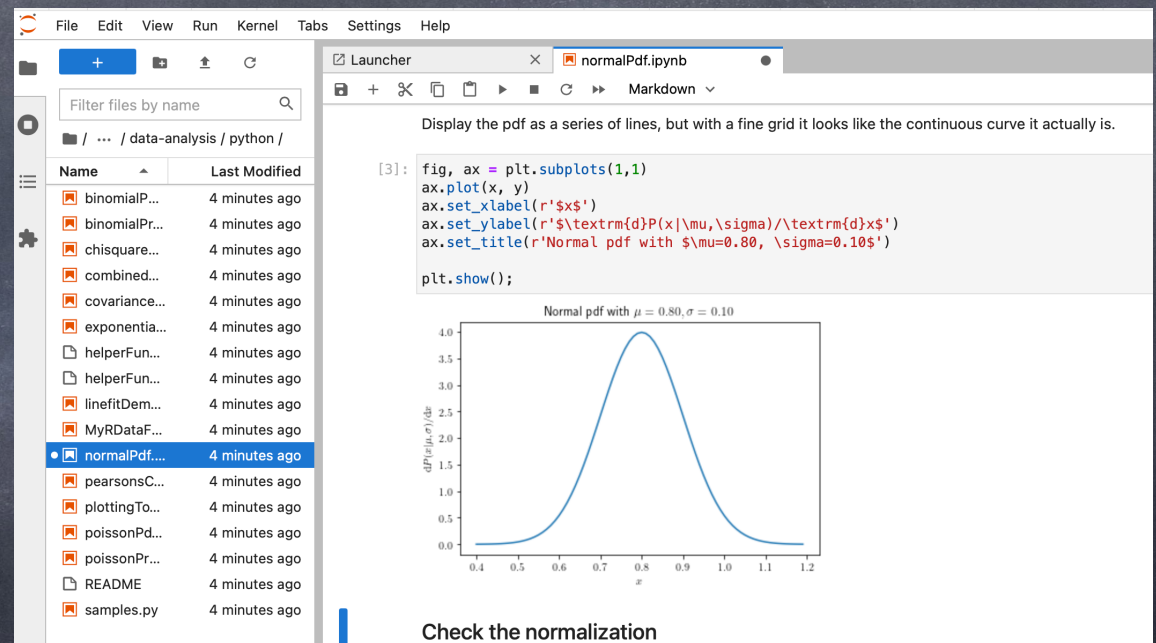


- Go to <https://galaxy-hepp.hpc.uio.no>
- Create your own user
- Click on **Interactive tools** in the menu on left
- Click on **JupyterLab for HEPP courses**
- Select the FYS4505 course notebook
- Click on **Execute**
- After ~30 seconds "click here to display" [InteractiveTool result] becomes available – click on that



jupyter-lab on galaxy-hepp

- Drill down to FYS4505/data-analysis/python
- Double-click on any file (a notebook or README) to get started
- Browse, edit, and execute the notebook cells
- Can download as .ipynb or .py



Part I

Part I

- Binomial **distribution**
- Gaussian **distribution**
- Confidence intervals
- Poisson **distribution**
- Combining multiple measurements
- Exponential **distribution**

Binomial distribution

- A fixed number of «events» sorted into two classes
 - A coin toss giving «heads» or «tails»
 - A charged particle crosses a detector and is either detected or not
- In experimental physics the fraction of occurrences that is detected is called the efficiency ϵ .
 - For example 12/15 detections gives $\epsilon = 80\%$.
 - But what is the uncertainty?!

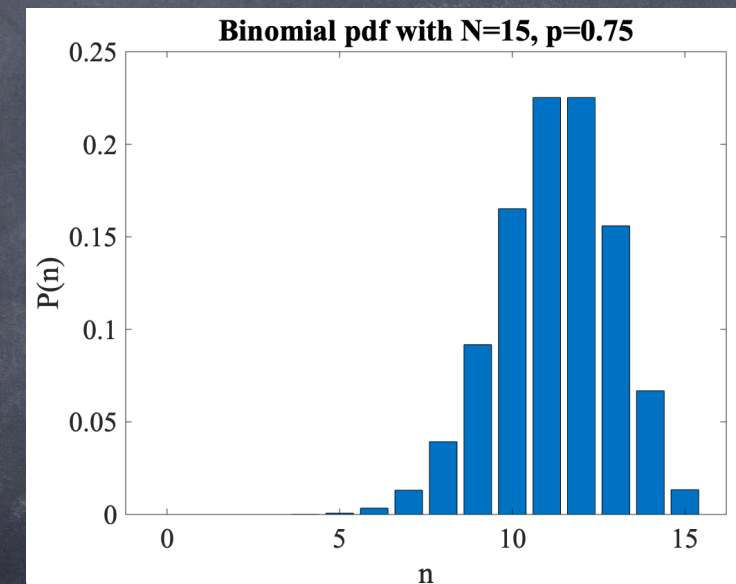
Binomial distribution

- The probability density function (pdf, sometimes prob. distribution func.) describes the theoretical distribution of outcomes for an infinitely large sample, normalized to 1.0.

- For N trials and n «successes» with probability of success per trial of p we have

$$P(n|N, p) = \frac{N!}{n! (N - n)!} p^n (1 - p)^{N-n}$$

- Question: What should $\sum_{n=0}^N P(n|N, p)$ be?
- What distribution do we expect/predict for $N = 15, p = 0.75$?
- In our example we observed 12, which seems compatible with the distribution, but how do we quantify this?



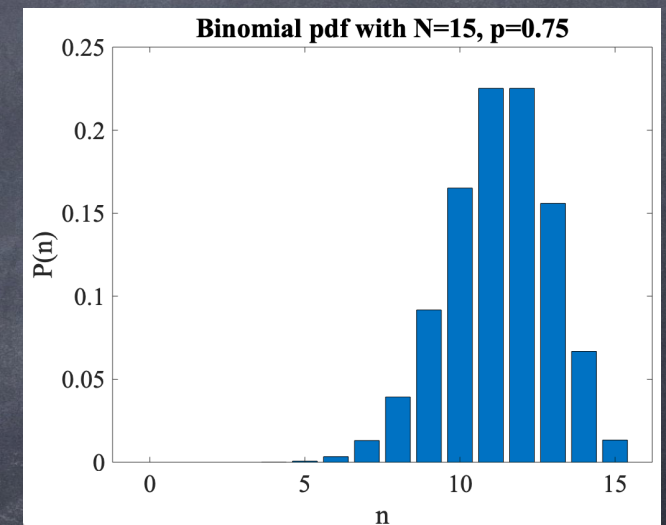
Binomial distribution

- What is the expectation value of n , $E(n)$?

- $$E(n) = \sum_{n=0}^N nP(n) = Np = 11.25$$

- This is simply a weighted average

$$E[f(n)] = \frac{\sum_i f(n_i) \cdot P(n_i)}{\sum_i P(n_i)}, \text{ with } \sum_i P(n_i) \equiv 1$$



Binomial distribution

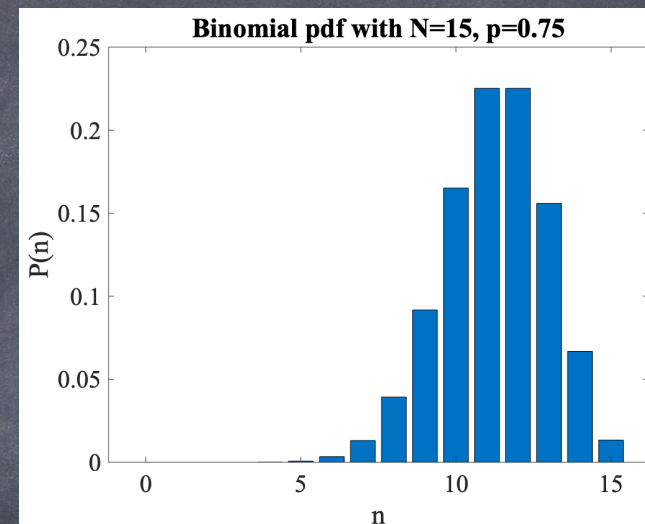
- The spread of the distribution is described by the variance $var(n) = E[(n - pN)^2]$ or simply the standard deviation $\sigma_n = \sqrt{var(n)}$

- The variance of the binomial distribution is the weighted average of the squared deviations w.r.t. the mean:

$$\sigma_n^2 = \sum_{n=0}^N (n - Np)^2 P(n) = Np(1 - p)$$

- In our measurement we don't know p , we are trying to measure it!
- However, it is a pretty good approximation for large N (apart from $n \sim 0$ and $n \sim N$) to use the observed number $\hat{p} = \frac{n}{N}$.
- Since N is fixed we have $\sigma_p = \frac{\sigma_n}{N} = \sqrt{p(1 - p)/N}$

17



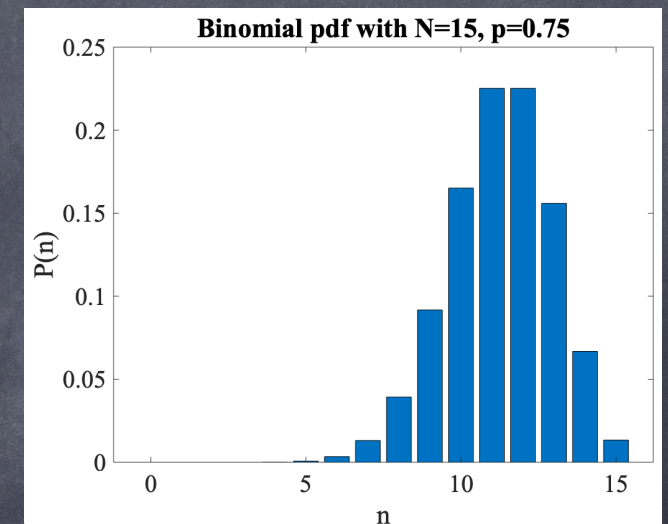
Binomial distribution

- So, coming back to our example, we could quote

$$\epsilon \simeq \frac{12}{15} \pm \sqrt{\frac{\frac{12}{15} \left(1 - \frac{12}{15}\right)}{15}} = 0.80 \pm 0.10$$

- Leo recommends to require $N > 30$ and $p > 0.05$

- Question: What about $p > 0.95$?!

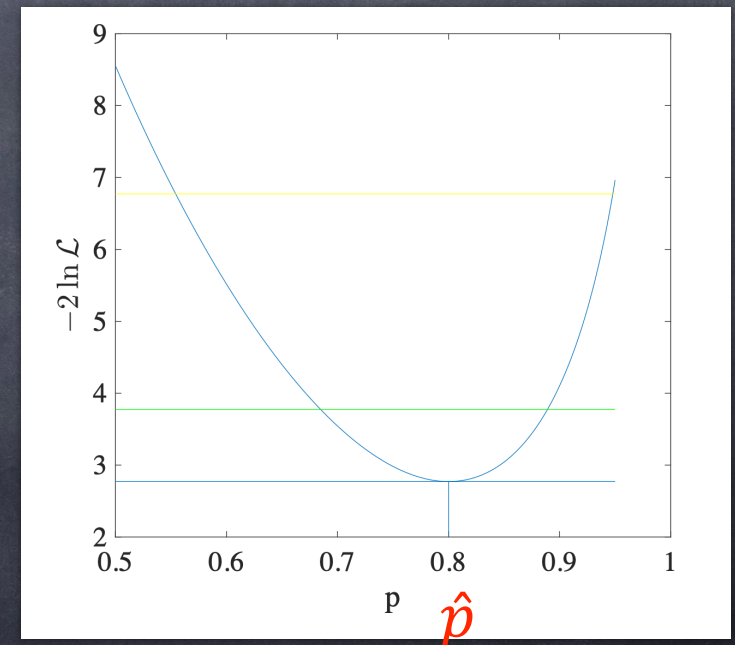


Maximizing binomial likelihood

- We assumed that the «best» $\hat{p}$ (i.e. $\hat{e}$) to quote was $\frac{n}{N}$. This can be justified by application of Maximum Likelihood (ML).
- $\mathcal{L}(n|p, N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$
- Question: What does this $\mathcal{L}$ represent?
- So, we are going to find the $\hat{p}$ that maximizes $\mathcal{L}$ given our data N and n

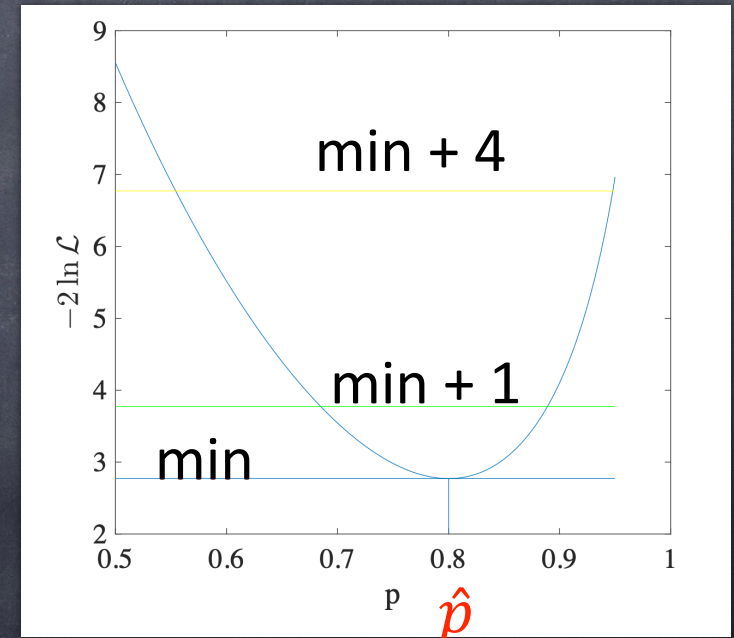
Maximizing binomial likelihood

- $-2\ln\mathcal{L}(n|p, N) = -2 \left[\ln \left(\frac{N!}{n!(N-n)!} \right) + n\ln p + (N-n)\ln(1-p) \right]$
- Minimize: $-2 \frac{d\ln\mathcal{L}}{dp} = 0 = -2 \left(\frac{n}{\hat{p}} - \frac{N-n}{1-\hat{p}} \right)$
- ...a few lines later... $\hat{p} = \frac{n}{N}$
- In perfect agreement with numerical minimization



Bonus: "MINOS" errors

- Intersections of the likelihood profile with $\min(-2\ln\mathcal{L}) + 1(4,9)$ correspond much closer to $\pm 1\sigma(2\sigma, 3\sigma)$ confidence intervals.
- These so-called «MINOS» errors are the most advanced error estimation programmed in MINUIT (a CERN statistics package built into ROOT).
- Note the asymmetry of the likelihood profile – this will lead to asymmetric errors, i.e. $\hat{p} = 0.80^{+0.09}_{-0.11}$ instead of our initial estimate $\hat{p} = 0.80 \pm 0.10$
- Question: Why is the likelihood skewed like this?



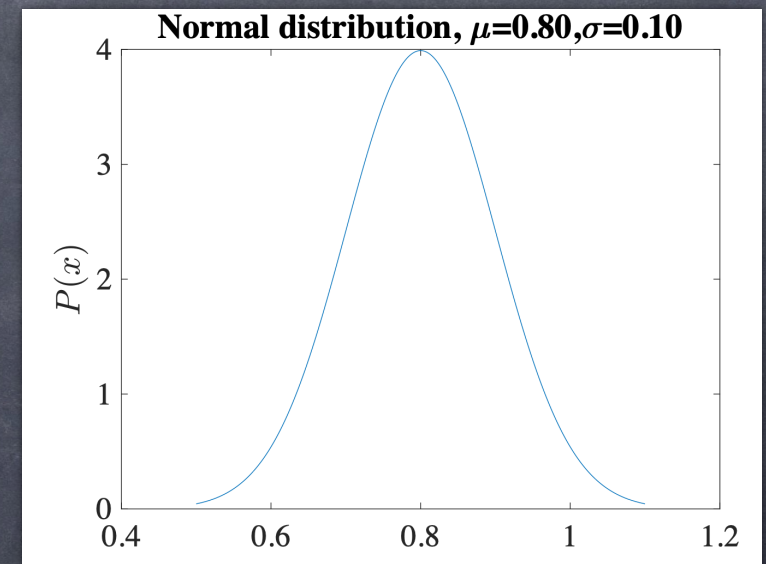
Gaussian (or normal) distribution

- We haven't yet defined what we mean by a confidence interval of " $\pm 1\sigma$ " (or $2\sigma, 3\sigma$) !!

- The pdf of the Gaussian (or normal) distribution is

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Many measurement errors in NP and HEP are approximated very well by the normal distribution.
- A peculiarity of our fields is that we tend to have good estimates of σ before looking at our data. In many fields they have to estimate both σ and μ .



Note that unlike the binomial, this distribution is continuous and symmetric

Gaussian distribution

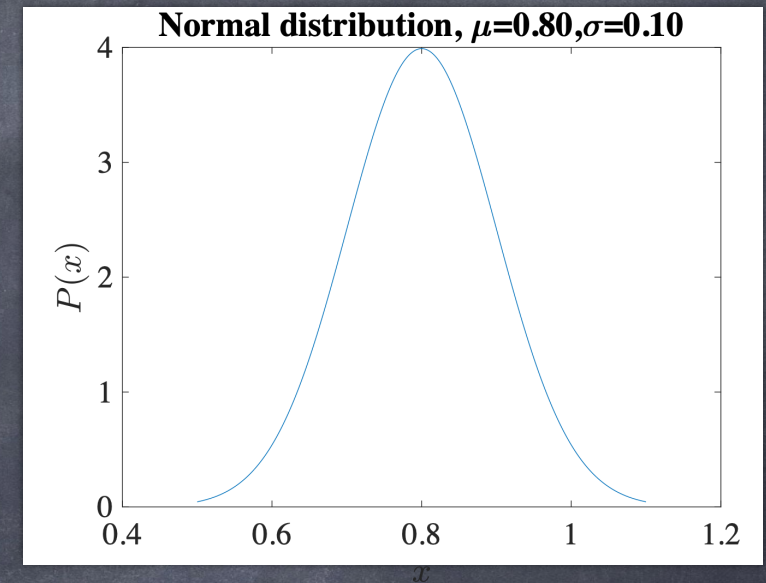
- Most experimental results have several or many sources of uncertainty. Even if some of them are obviously not described by a normal distribution, the **Central Limit Theorem** tells us that their combination tends toward a normal distribution!

Gaussian distribution

- pdf $P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- mean $E[x] = \int_{-\infty}^{+\infty} xP(x)dx = \mu$
- variance

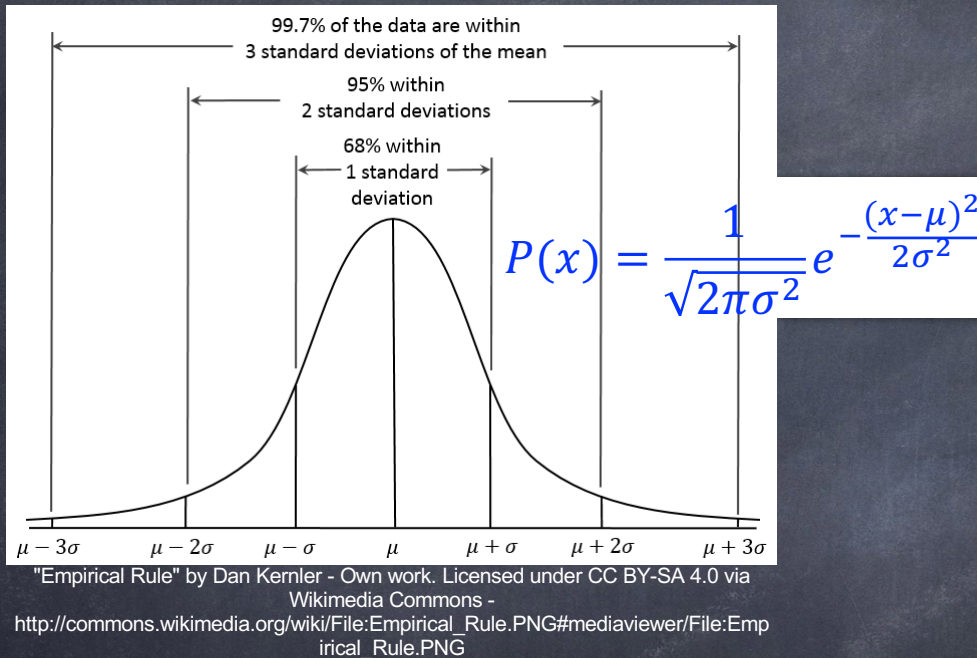
$$E[(x - \mu)^2] = \text{var}(x) = \int_{-\infty}^{\infty} (x - \mu)^2 P(x) dx = \sigma^2$$

- Easy to show that Full-Width, Half-Maximum (FWHM) is $FWHM = 2\sqrt{2\ln 2}\sigma \simeq 2.35\sigma$



P.S. The expectation values have a factor $1/\int_{-\infty}^{+\infty} P(x)dx$ but this is 1 by definition

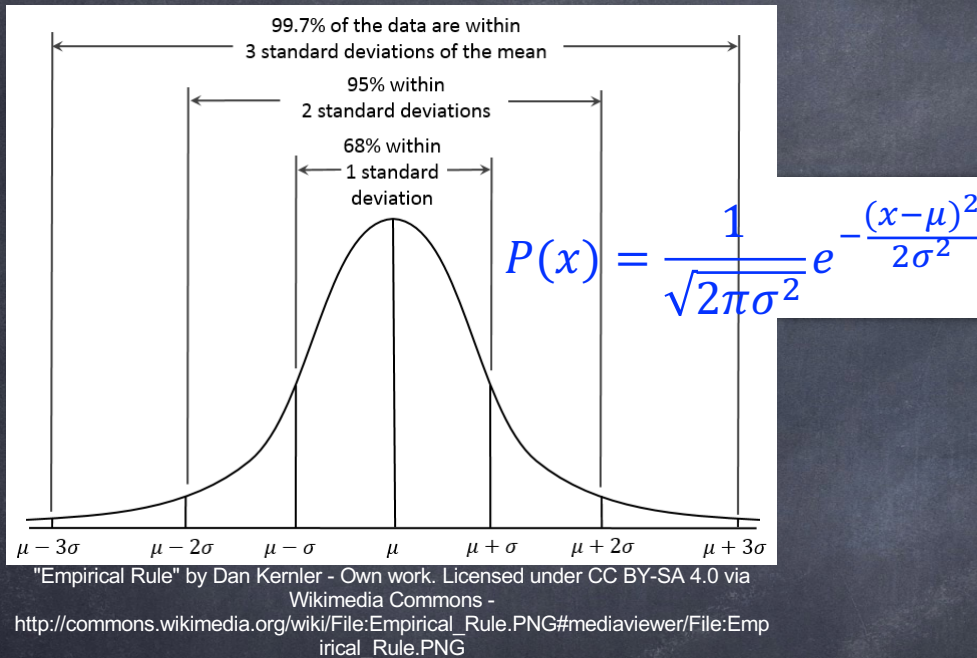
Confidence intervals



- For a classical or frequentist statistician a 1σ interval means that the interval comes from an ensemble of intervals that cover the (unknown) true value 68% of the time.
- For a Bayesian statistician a 1σ credible interval (Bayesians have no confidence in confidence intervals ;-)) means that the probability that the true value is within the interval is 68%.
- Let's not worry too much about this distinction in our course. Most physicists anyway switch (too) smoothly between the two points of view.

- The integrated areas (probabilities) within $\pm N\sigma$ of the normal distribution are used to define what we mean by $N\sigma$ confidence intervals.

Confidence intervals



$N\sigma$	$P(\mu - N\sigma < x < \mu + N\sigma)$	$P(x > \mu + N\sigma)$
1	0.6826	0.1587
2	0.9545	0.0228
3	0.9973	$1.35 \cdot 10^{-3}$
4	0.9999	$3.17 \cdot 10^{-5}$
5	≈ 1	$2.87 \cdot 10^{-7}$

- The integrated areas (probabilities) within $\pm N\sigma$ of the normal distribution are used to define what we mean by $N\sigma$ confidence intervals.

5σ is the mythical threshold for “discovery” in particle physics

Maximizing Gaussian likelihood

- $\mathcal{L}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $-2\ln\mathcal{L} = \frac{(x-\mu)^2}{\sigma^2} + \ln 2\pi\sigma^2$
- Assume that we know σ
- $-2 \frac{d\ln\mathcal{L}}{d\mu} = 0 = -2 \frac{(x-\hat{\mu})}{\sigma^2} \Rightarrow \hat{\mu} = x$
- It's straightforward to show that $-2\ln\mathcal{L}$ increases by 1 above the minimum at $x \pm \sigma$ (and by N^2 at $x \pm N\sigma$), i.e. we can quote $\hat{\mu} = x \pm \sigma$
- If we differentiate one more time we get a very powerful result: $-\frac{d^2\ln\mathcal{L}}{d\mu^2} = \frac{1}{\sigma^2}$
- For any likelihood which is sufficiently well described by a parabola w.r.t. a parameter near the minimum, second-order differentiation gives us a good estimate of the uncertainty of the parameter!!

Poisson distribution

- Relevant any time you are counting the total number of something (rather than a fraction)
 - e.g. how many radioactive nuclear decays in a certain time
 - e.g. number of Higgs boson candidates in an LHC data set
- Can be derived as a limiting case of the binomial distribution for N .
- Can also be derived from a constant average rate $\frac{dN}{dt} = -\lambda N$ and assuming that in each microscopic time interval there are either zero or one occurrences.
- [Wikipedia lists two additional approaches.](#) ("Poisson limit theorem»)

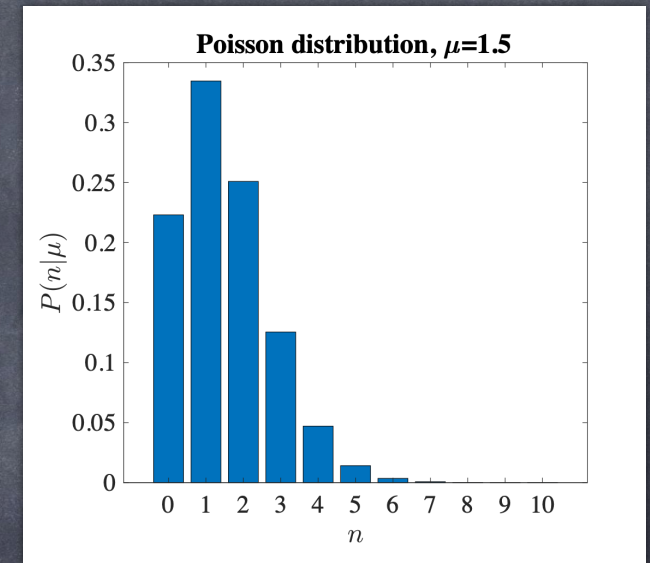
Poisson distribution

- pdf $P(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$, and $\sum_{n=0}^{\infty} P(n|\mu) = 1$

- mean $E[n] = \sum_{n=0}^{\infty} nP(n|\mu) = \mu$

- variance $E[(n - \mu)^2] = \sum_{n=0}^{\infty} (n - \mu)^2 P(n|\mu) = \mu$

- Note that, unlike for the binomial distribution, n goes to ∞ (but in practice the probabilities over a few times μ are tiny)



Maximizing Poisson likelihood

- $\mathcal{L}(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$, $\ln\mathcal{L} = -\mu + n\ln\mu - \ln n!$
- $-2 \frac{d\ln\mathcal{L}}{d\mu} = 0 = -2 \left(-1 + \frac{n}{\hat{\mu}}\right) \Rightarrow \hat{\mu} = n$
- $-\frac{d^2\ln\mathcal{L}}{d\mu^2} = \frac{n}{\hat{\mu}^2} = \frac{1}{\hat{\mu}} = \frac{1}{\sigma^2} \Rightarrow$ Consistent with the variance calculation!
- Question: If we observe 25 counts of some process, what is our estimate of μ , with uncertainty?

Poisson process in pieces

- We want to compare a single measurement of μ with N measurements of μ/N
- For the single measurement we just add the individual measurements $\sum_{i=1}^N n_i = n$ and have $\hat{\mu} = n \pm \sqrt{n}$.

- The likelihood for the N measurements is the product of the likelihoods:

$$\mathcal{L} = \prod_{i=1}^N \frac{e^{-\mu/N} (\mu/N)^{n_i}}{n_i!} = e^{-\mu} \prod_{i=1}^N \frac{(\mu/N)^{n_i}}{n_i!}$$

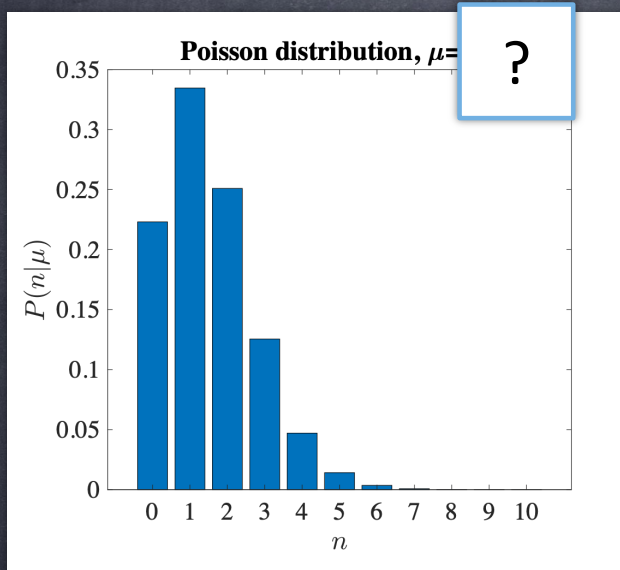
- $\ln \mathcal{L} = -\mu + \sum_i n_i (\ln \mu - \ln N) - \ln n_i!$

Poisson process in pieces

- $\ln \mathcal{L} = -\mu + \sum_i n_i (\ln \mu - \ln N) - \ln n_i!$
- $\frac{d \ln \mathcal{L}}{d \mu} = -1 + \frac{1}{\hat{\mu}} \sum_i n_i = 0 \Rightarrow \hat{\mu} = \sum_i n_i,$
the same as above!
- $-\frac{d^2 \ln \mathcal{L}}{d \mu^2} = \frac{n}{\mu^2}$ again, so $\hat{\mu} = \sum_i n_i \pm \sqrt{\sum_i n_i},$
the same as above!

If you know you are making N measurements of the same Poisson process, you might as well add them together!

Poisson process in pieces



- If you suspect your process may not be Poisson, then you may break your data into pieces and see if the distribution of the n_i 's is Poisson.
- For example, if you find $\hat{\mu} = 150$, you could break the data into 100 equal time intervals and see if your data is distributed like the distribution on the left.
 - Question: What μ have I used to make the plot on the left?

Mean of multiple Gaussian distributed measurements

- Suppose you have N Gaussian-distributed measurements of some quantity μ , but they all have different uncertainties. We want to find a best estimate for μ and its uncertainty.
- Later, we want to test if the data are randomly distributed according to our model...or perhaps not!
- The likelihood is again the product of the individual

likelihoods:
$$\mathcal{L} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}}$$

Mean of multiple Gaussian distributed measurements

- $\mathcal{L} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}}$
- $-2\ln\mathcal{L} = \sum_i \frac{(x_i - \mu)^2}{\sigma_i^2} + \sum_i \ln(2\pi\sigma_i^2) \equiv \chi^2 + \sum_i \ln(2\pi\sigma_i^2)$
- $\frac{d(-2\ln\mathcal{L})}{d\mu} = \frac{d\chi^2}{d\mu} = 0 = -2\sum_i \frac{(x_i - \hat{\mu})}{\sigma_i^2}$
- $\hat{\mu} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}$, the inverse error squared weighted mean

- $\frac{1}{\sigma_\mu^2} = -\frac{d^2(\ln\mathcal{L})}{d\mu^2} = \sum_i \frac{1}{\sigma_i^2}$

- $\hat{\mu} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2} \pm \frac{1}{\sqrt{\sum_i 1 / \sigma_i^2}}$

- Uncertainty in mean decreases with additional measurements unless $\sigma_{i+1}^2 \gg \sigma_\mu^2$

We come back to the very important χ^2

Mean of multiple Gaussian distributed measurements

- Suppose we make N independent measurements all with the same σ

- $$\hat{\mu} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2} \pm \frac{1}{\sqrt{\sum_i 1 / \sigma_i^2}} \rightarrow \hat{\mu} = \frac{\sum_i x_i}{N} \pm \frac{\sigma}{\sqrt{N}}$$

- Uncertainty in the mean decreases as $\sqrt{N}$ increases
- Question: Suppose we have a statistics-dominated measurement and want to reduce the uncertainty by a factor 10. How much more data should we collect?

Exponential distribution

- Decay rates of nuclei or particles ("radioactive decay law")

- $\frac{dN}{dt} = -\lambda N(t)$

- Absorption of radiation

- $\frac{dN}{dx} = -\mu N(x)$

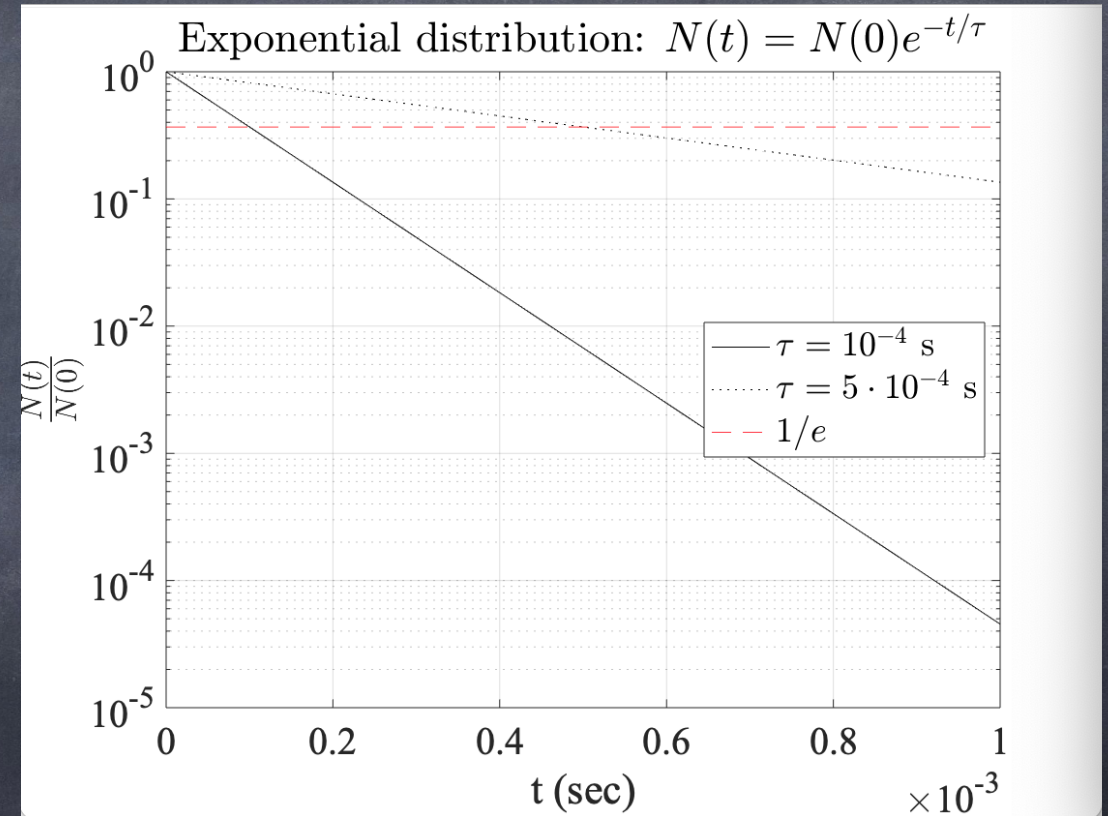
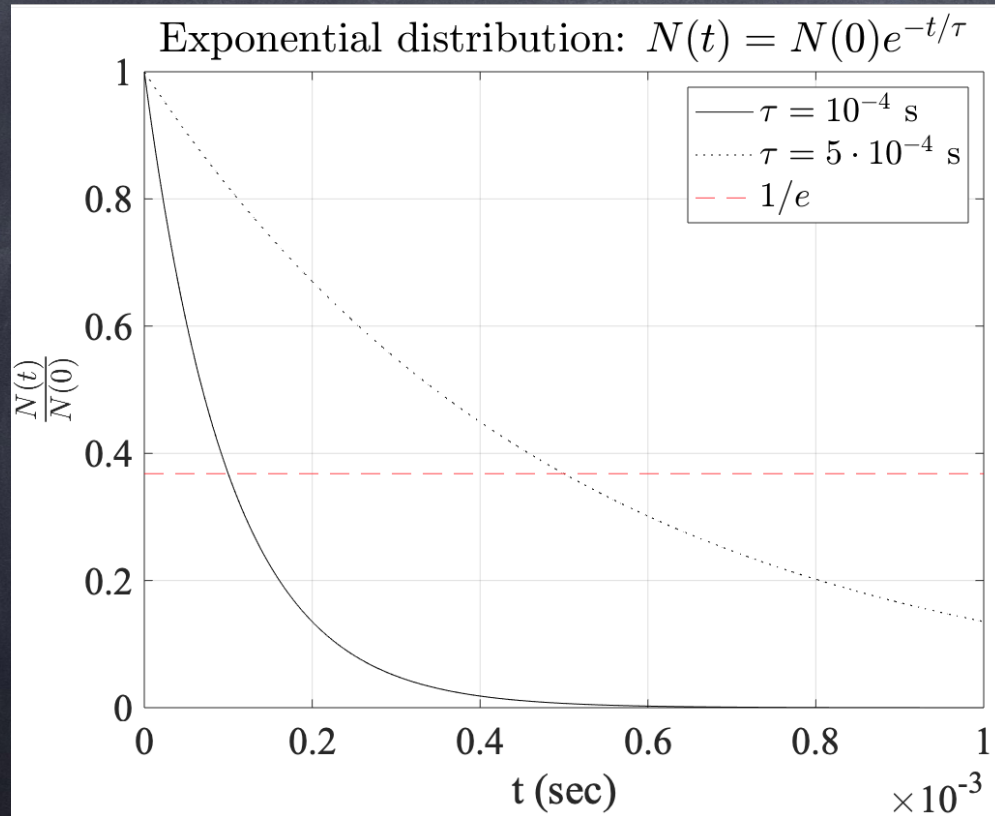
- What are λ and μ ?

- $\lambda = -\frac{1}{N} \frac{dN}{dt}, \mu = -\frac{1}{N} \frac{dN}{dx}$

- λ is average decay rate per unit time or "decay constant"

- μ is average absorption per unit distance

Exponential distribution



Exponential distribution

- Average time of decay $\langle t \rangle \equiv \frac{\int_0^\infty t N(0) e^{-\lambda t} dt}{\int_0^\infty N(0) e^{-\lambda t} dt} = \dots = \frac{1}{\lambda}$
- The average lifetime $\langle t \rangle$, often written τ , is the inverse of the average decay rate per particle: $\tau \equiv 1/\lambda$
- Normalized pdf: $\lambda e^{-\lambda t}$ or $\frac{1}{\tau} e^{-t/\tau}$
- Variance = $\tau^2 = \frac{1}{\lambda^2}$