

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

FYS-KJM3506/FYS4505 – H2026

A. Read

UiO • University of Oslo

Introduction

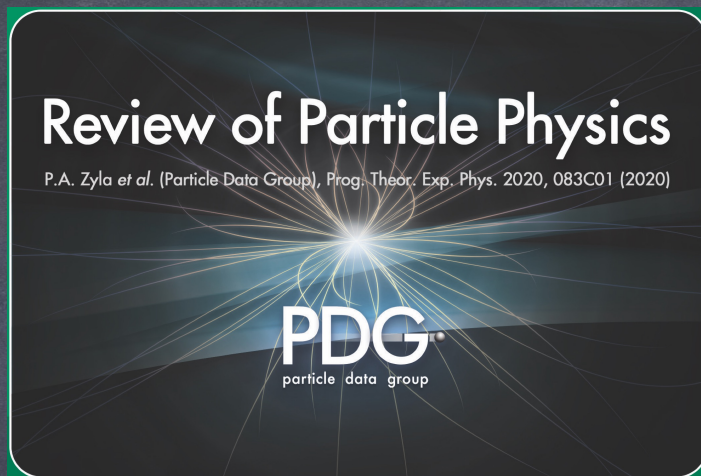
Relevant learning outcome

- «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

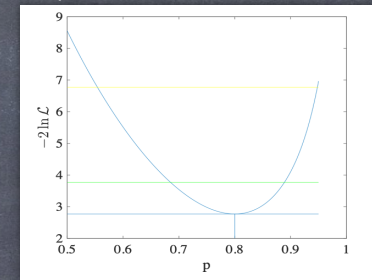
39. Probability (rev.)	621
40. Statistics (rev.)	626
41. Monte Carlo techniques (rev.)	642
42. Monte Carlo event generators	646
43. Monte Carlo neutrino event generators (rev.)	657
44. Monte Carlo particle numbering scheme (rev.)	661
45. Clebsch-Gordan coefficients, spherical harmonics, and d functions	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 (probability) model for the data x of our experiment: $\mathcal{L}(x)$
- Identify the Parameters of Interest (POI): μ
- Identify the Nuisance Parameters (NP): θ
 - Model parameters we need but aren't of primary interest
 - **Quantify additional sources of uncertainty and associate them to probability distributions and parameters (advanced topic)**
- **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

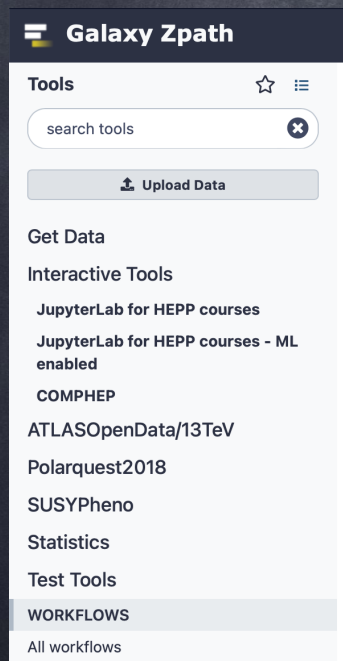
- Maximizing the likelihood $\mathcal{L}$ gives **efficient** (and in the large-sample limit, **unbiased**) point-estimates of the model parameters.
- It turns out to be **very** convenient to minimize $-2\ln\mathcal{L}$
 - $-2\ln\mathcal{L}$ often has a parabolic shape with a clear minimum
- Normally 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
 - Uncertainty per point specified: χ^2 —minimization
 - No specified uncertainty/point: least-squares
- More reading: [STK1000](#)



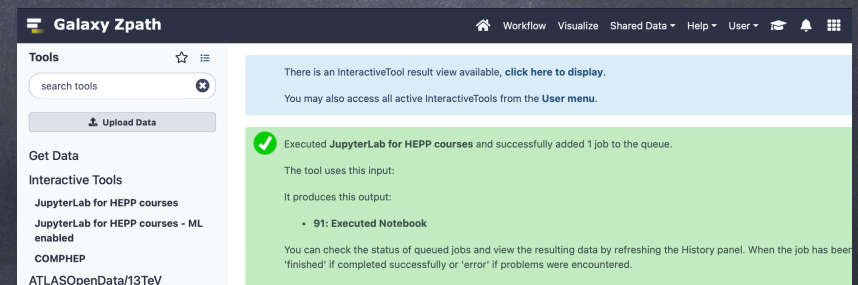
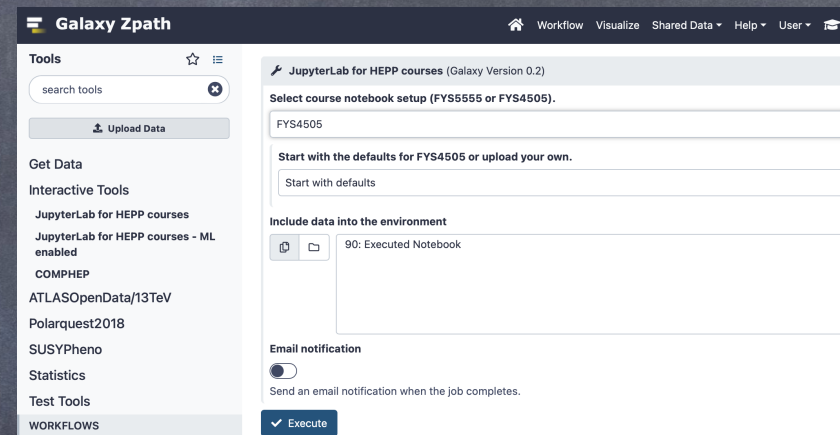
Jupyter (python) notebooks

- 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
- If this is difficult on your laptop, all this is set up on galaxy-hepp.hpc.uio.no

galaxy-hepp.hpc.uio.no

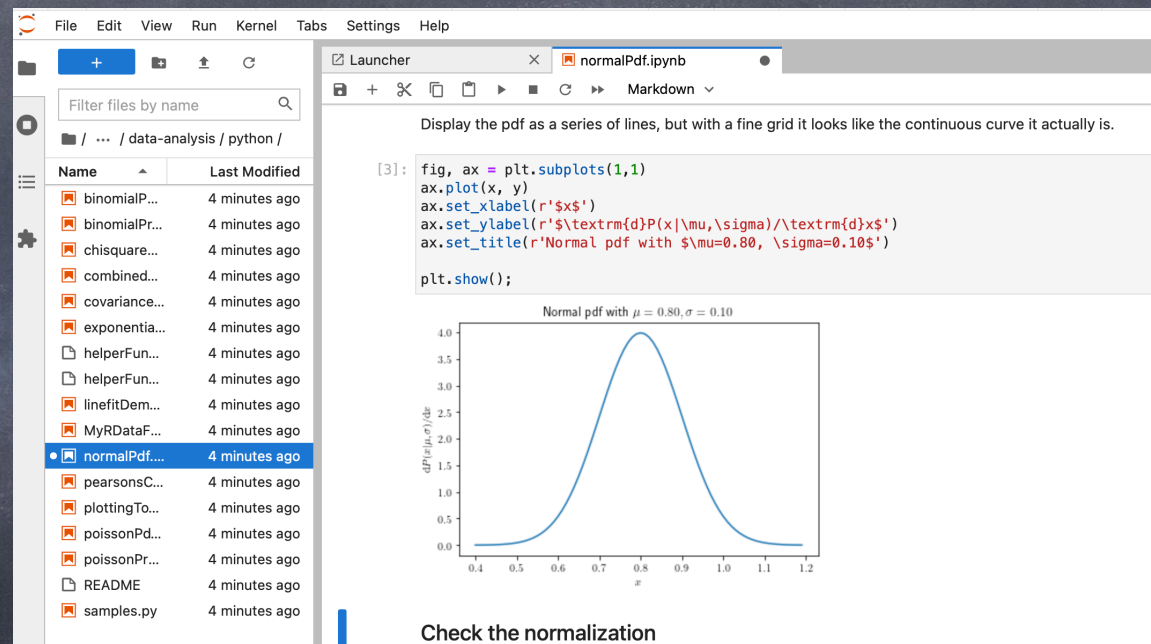


- Go to <https://galaxy-hepp.hpc.uio.no>
- Create your own user
- Click on **Interactive tools** in the menu on left
- Click on **JupyterLab for HEPP courses**
- Select the FYS4505 course notebook
- Click on **Execute**
- After ~30 seconds "click here to display" [InteractiveTool result] becomes available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

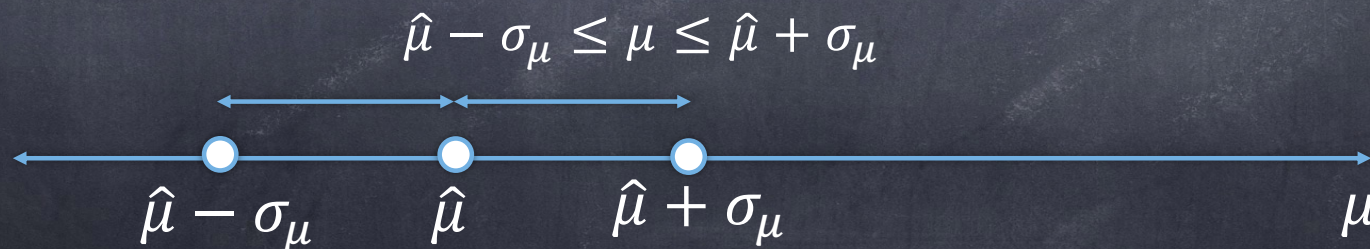
- Review some basics of statistics
- Gaussian **distribution**
- Confidence intervals
- Binomial **distribution**
- Poisson **distribution**
- Exponential **distribution**

We will also:

- maximize some likelihoods
- do some combinations of results
- identify association between errors, confidence intervals, and changes in log-likelihood

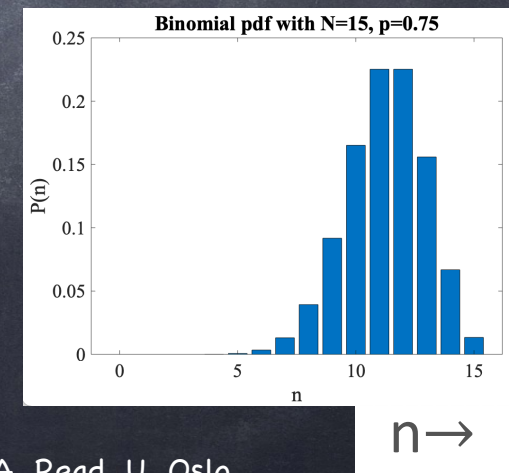
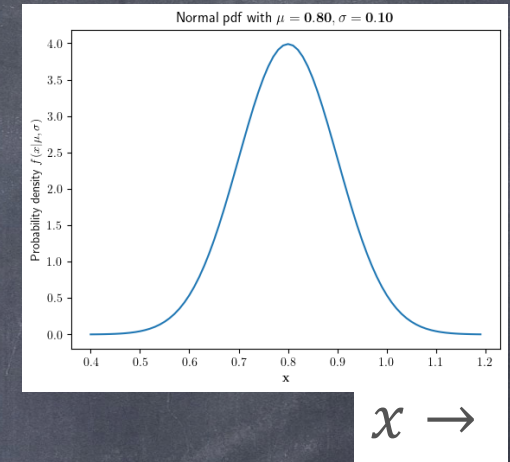
Point-estimate and confidence interval

- Use data to infer a **best estimate** $\hat{\mu}$ of the true value μ_{True} of a physical parameter μ
- Also infer a range of μ that tells us something about how wrong the estimate could be
- Our inferences will generally report $\hat{\mu} \pm \sigma_{\mu}$, i.e., a **point-estimate** $\hat{\mu}$ and an **uncertainty** σ_{μ} or **confidence interval**



Probability distributions

- Given parameter(s) μ , the data x will have a characteristic distribution.
- The **probability density function** $f(x, \mu)$ (pdf, sometimes prob. **distribution** func.) describes the theoretical distribution of outcomes for a specific value of μ for an infinitely large sample, normalized to 1.
- $f(x, \mu)dx$ is the probability in the interval $[x, x + dx]$
- $\int_S f(x, \mu)dx \equiv 1$ (S is the domain of x), i.e., total probability is 1
- If the data are integers n (**counting experiments**), the integral is replaced by a sum $\sum_{S_n} f(n, \mu) \equiv 1$ (S_n is the domain of n)
- Examples: $n = [0, N]$ for N trials, $n = [0, \infty]$ for counting



Properties of distributions: mean

- The **mean**, or **expectation value**: $E[x] \equiv \langle x \rangle = \frac{\int_S x f(x) dx}{\int_S f(x) dx}$,
- where $\int_S f(x) dx \equiv 1$
- For **counting experiments**, the integral is replaced by a sum: $E[n] \equiv \langle n \rangle = \sum_S n f(n) / \sum_S f(n)$,
- where $\sum_S f(n) \equiv 1$
- Also called the **first moment** of the distribution

Properties of distributions: variance

- The **variance** $V[x] = E[(x - E[x])^2] = \int_S (x - \langle x \rangle)^2 f(x) dx = \langle x^2 \rangle - \langle x \rangle^2$
- **Covariance** $V_{xy} = E[(x - E[x])(y - E[y])]$
$$= \int_S (x - \langle x \rangle)(y - \langle y \rangle) f(x, y) dx dy$$
- If $f(x, y) = g(x)h(y)$, what is V_{xy} ? It's 0! (NB! $\langle \langle x \rangle \rangle = \langle x \rangle$)
- For **counting experiments** $V[n] = E[(n - E[n])^2] = \sum_S (n - \langle n \rangle)^2 f(n)$
- The variance is often written σ^2 , so σ is the root-mean-square of the deviations from the mean, i.e., the "spread" of the distribution
- Also called the **second moment** of the distribution

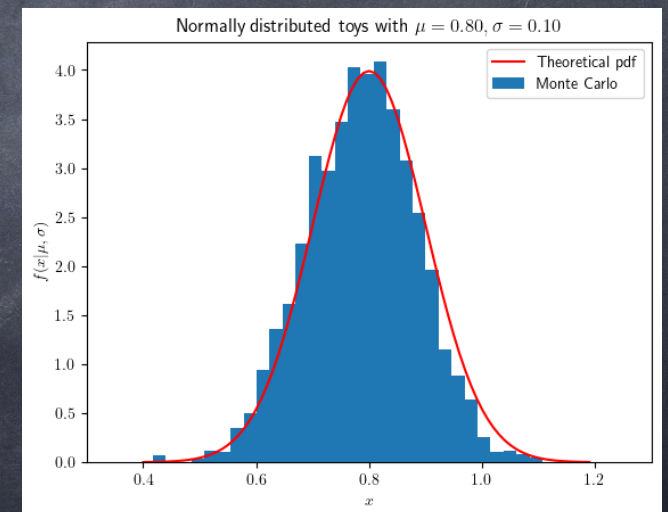
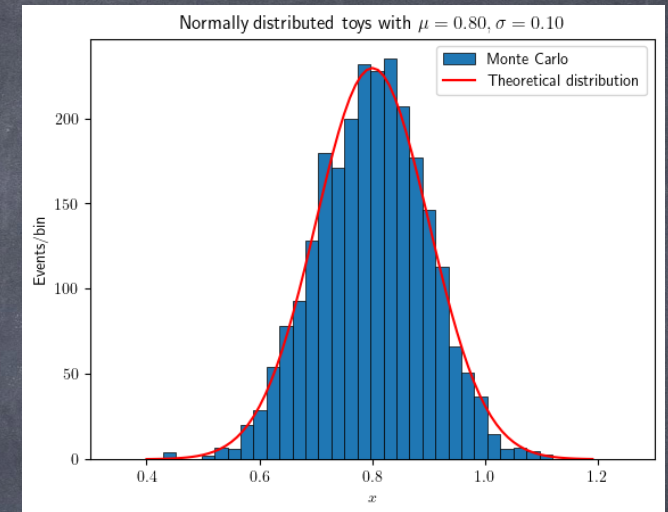
Histograms

- We frequently use histograms to visualize data,
- both continuous and discrete data
- Break up the x-axis into “bins” of width Δx and count how often a measured x ends up in bin i :

$$x_i - \frac{\Delta x}{2} < x \leq x_i + \frac{\Delta x}{2}$$

- If the data are integers n , the smallest Δn is 1
- For a sufficient number of observations, a histogram of many bins whose **sum of areas** is normalized to 1 can be a good approximation of the **probability density**.

$$y'_i = \frac{y_i}{\Delta x \sum y_i}, \text{ where } y_i \text{ is the number of counts in bin } i$$

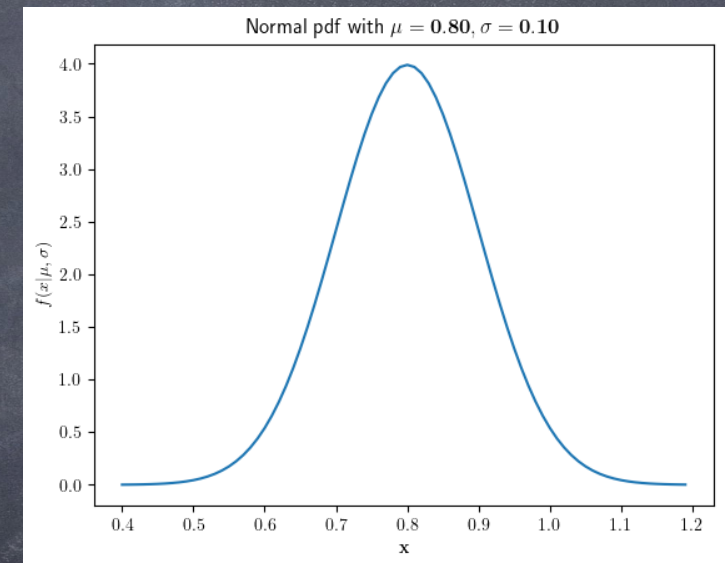


Gaussian (normal) distribution

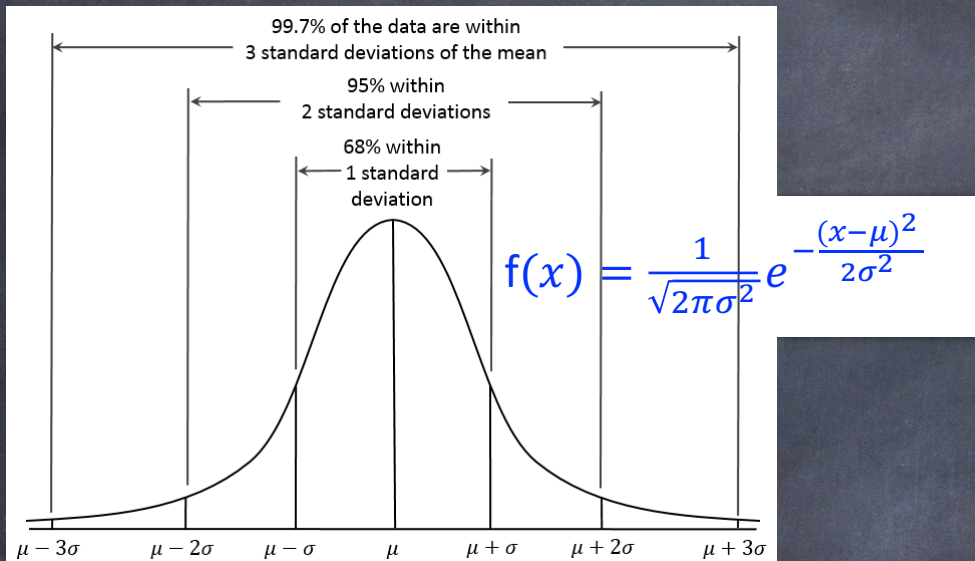
- Most experimental results have several or many sources of uncertainty or fluctuations.
- Even if some of them are obviously not described by a Gaussian distribution, the **Central Limit Theorem** tells us that their combination tends toward a **Gaussian** or **normal distribution**!
- Many measurement errors in NP and HEP are approximated very well by the normal distribution.

Gaussian (normal) distribution

- pdf $f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- mean $E[x] = \int_{-\infty}^{+\infty} x f(x|\mu, \sigma) dx = \mu$
- variance $E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x|\mu, \sigma) dx = \sigma^2$
- 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 .
- Easy to show that **Full-Width, Half-Maximum (FWHM)** is $FWHM = 2\sqrt{2\ln 2}\sigma \simeq 2.35\sigma$



$N\sigma$ Confidence intervals



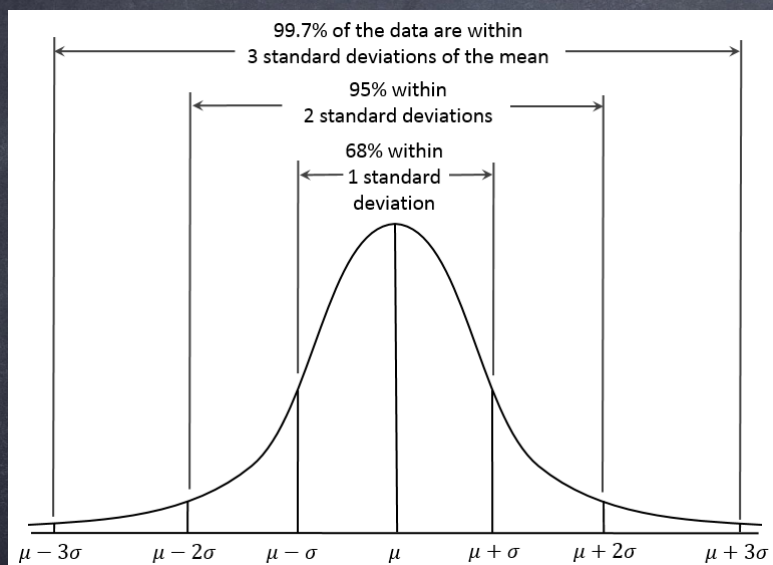
"Empirical Rule" by Dan Kernler - Own work. Licensed under CC BY-SA 4.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Empirical_Rule.PNG#mediaviewer/File:Empirical_Rule.PNG

$N\sigma$	$P(\mu - N\sigma < x < \mu + N\sigma)$ «Confidence level» in %	$P(x > \mu + N\sigma)$
1	68.26	0.1587
2	95.45	0.0228
3	99.73	$1.35 \cdot 10^{-3}$
4	99.99	$3.17 \cdot 10^{-5}$
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.

- « $\pm\sigma$ » implies 68.3% CL
- 5σ is the mythical threshold for “discovery” in particle physics

Confidence intervals



"Empirical Rule" by Dan Kemler - Own work. Licensed under CC BY-SA 4.0 via
Wikimedia Commons -
http://commons.wikimedia.org/wiki/File:Empirical_Rule.PNG#mediaviewer/File:Empirical_Rule.PNG

- A frequentist/classical **confidence interval** expresses the probability of such intervals to contain the true value.
- A Bayesian **credible interval** expresses the probability that the true value is within **this** interval.
- These are not the same thing, but in most practical situations (generally when there is plenty of data), they will be numerically very similar.
- **Let's not worry about the difference in this course!**

Binomial distribution

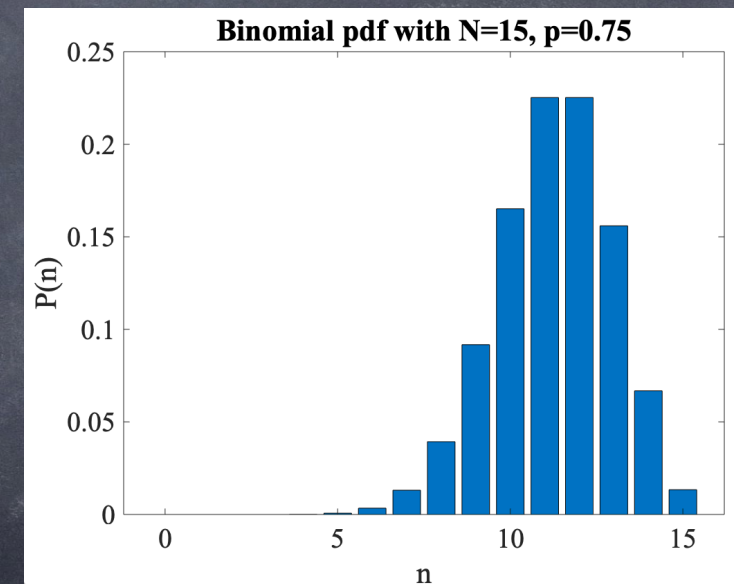
- A fixed number of N «events» sorted into two classes
 - A coin toss giving «heads» or «tails»
 - A charged particle crosses a detector and is either detected or not (Y/N)
 - An electron hitting a detector is identified correctly or not (T/F)
- The fraction of occurrences that is detected by a detector is called the efficiency ϵ .
 - For example, $\frac{n}{N} = \frac{12}{15}$ detections gives $\epsilon = 80\%$.
 - Come back to: But what is the uncertainty σ_ϵ on ϵ ?! (The width of the confidence interval)

Binomial distribution

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

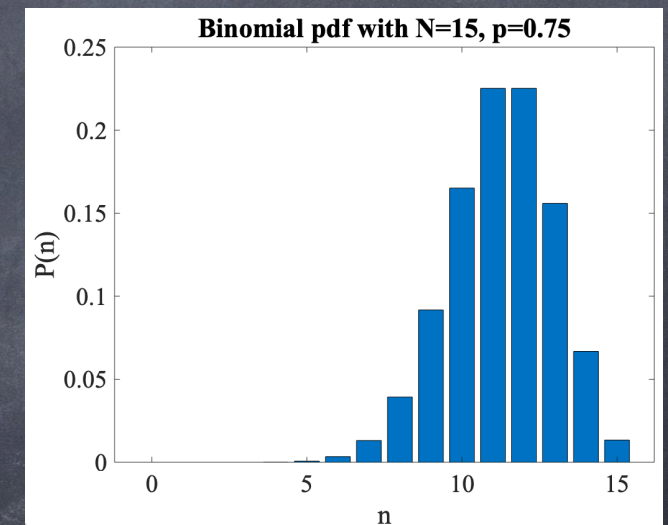
$$f(n|N, p) = P(n|N, p) = \frac{N!}{n! (N - n)!} p^n (1 - p)^{N-n}$$

- Combinatorics $\times$ Probability of n T $\times$ Probability of $(N-n)$ F
- Question: What should $\sum_{n=0}^N P(n|N, p)$ be? 1!!
- 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 , $E(n)$, given $N = 15, p = 0.75$?
- $E[n] = \sum_{n=0}^N nP(n|N, p) = Np = 11.25$
- Note that the expectation value doesn't have to be an integer!
- So if we observed 12, this appears to be a likely outcome,
 - but we want to be a bit more quantitative...



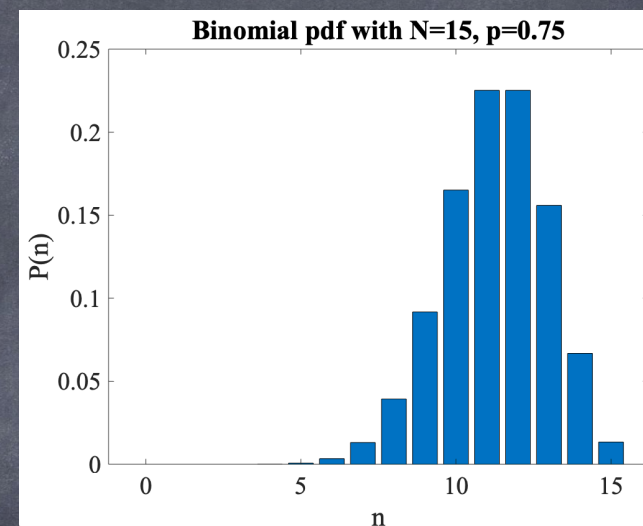
Binomial distribution

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

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

- In our measurement we don't know p (e.g., ϵ), we are trying to measure it!
- We use the observed number to infer $\hat{p} = \frac{n}{N}$.
- We use the variance to estimate the uncertainty σ_p on p :

$$\sigma_p = \frac{\sigma_n}{N} = \sqrt{p(1 - p)/N}$$



Binomial distribution

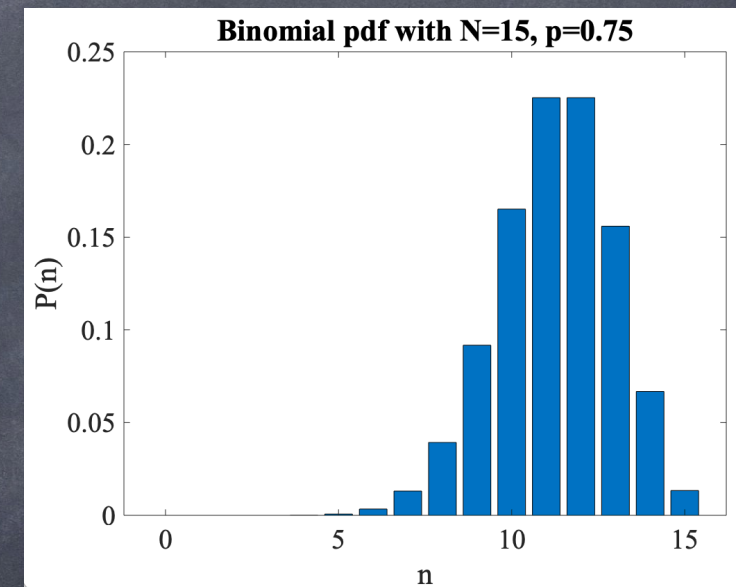
- Since $\hat{\epsilon} = \frac{n}{N}$, and N is constant, we have

$$\sigma_{\epsilon} = \frac{\sigma_n}{N} = \sqrt{\epsilon(1-\epsilon)/N}$$

- So, coming back to our example ($n = 12, N = 15$), 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 $p > 0.95$?! Since the binomial distribution is symmetric with respect to the two classes, $p > 0.95$ is as much of a challenge as $p < 0.05$.

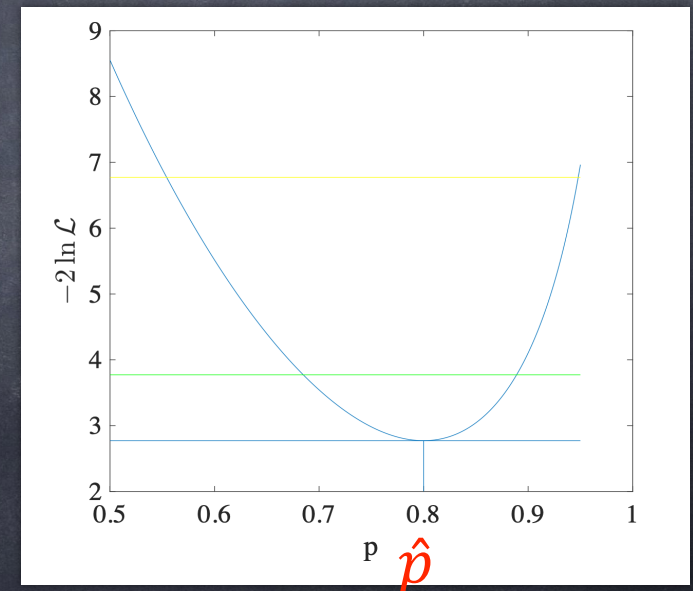


Maximizing binomial likelihood

- 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? Given the number of trials N and the success rate p , $\mathcal{L}$ gives the probability to observe n counts.
- So, we are going to find the $\hat{p}$ that maximizes $\mathcal{L}$ given our data N and n

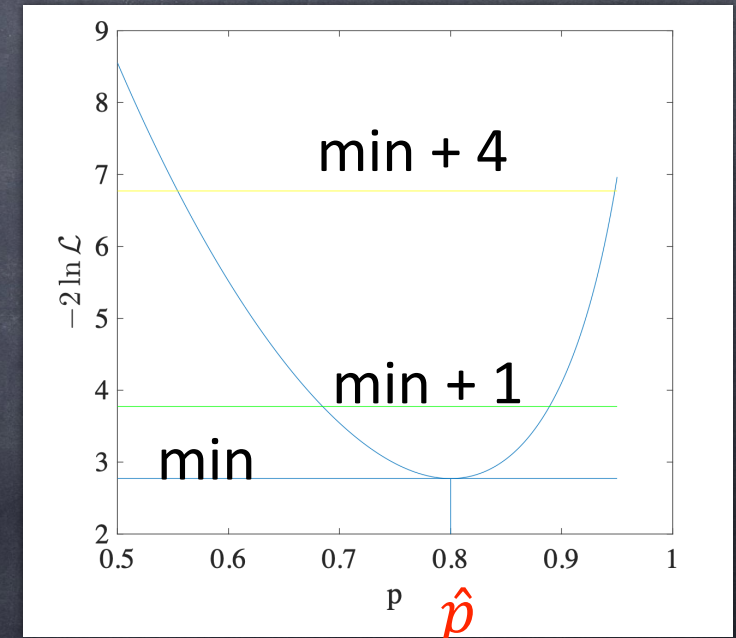
Maximizing binomial likelihood

- $\mathcal{L}(n|p, N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$
- $-2\ln\mathcal{L}(n|p, N) = -2 \left[\ln \left(\frac{N!}{n!(N-n)!} \right) + n\ln p + (N-n)\ln(1-p) \right]$
- 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 and our intuition



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? The likelihood profile hits «walls» at $p=0$ and $p=1$ – the profile on the side of the nearest wall has to be steeper.**



Maximizing Gaussian likelihood

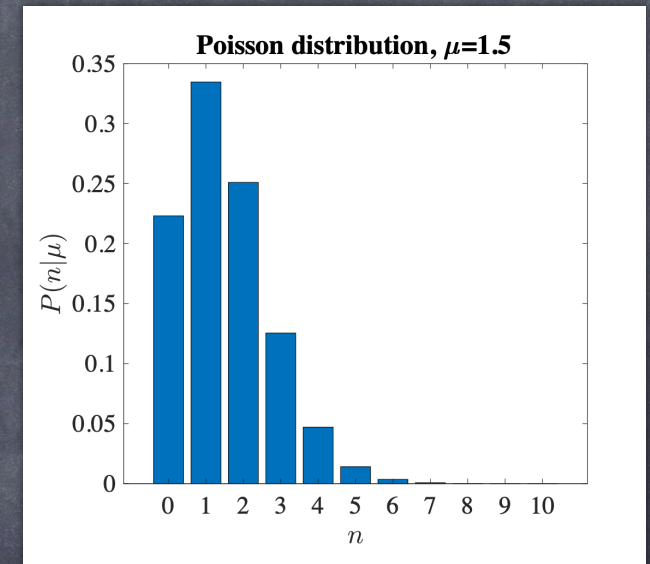
- $\mathcal{L}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $-2\ln\mathcal{L} = \frac{(x-\mu)^2}{\sigma^2} + \ln 2\pi\sigma^2$
- Assume that we know σ
- $-2 \frac{d\ln\mathcal{L}}{d\mu} = 0 = -2 \frac{(x-\hat{\mu})}{\sigma^2} \Rightarrow \hat{\mu} = x$
- It's straightforward 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 log-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), including the number of entries in a histogram bin
 - e.g. how many radioactive nuclear decays in a certain time
 - e.g. number of Higgs boson candidates in an LHC data set
- BONUS:
 - Can be derived as a limiting case of the binomial distribution for N , i.e., $Np = \text{constant}$ while $N \rightarrow \infty$.
 - 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

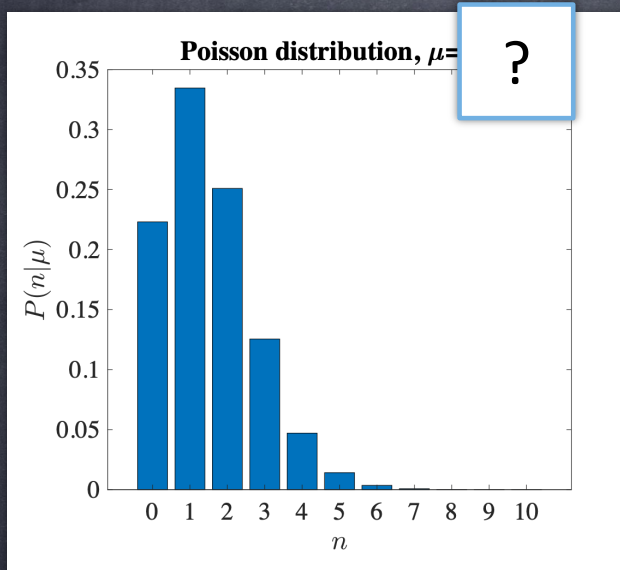
- pdf $P(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$, and $\sum_{n=0}^{\infty} P(n|\mu) = 1$
- mean $E[n] = \sum_{n=0}^{\infty} nP(n|\mu) = \mu$
- variance $E[(n - \mu)^2] = \sum_{n=0}^{\infty} (n - \mu)^2 P(n|\mu) = \mu$
- We (again) use expected variance as estimate for uncertainty (squared)
- Note that, unlike for the binomial distribution, n goes to ∞ (but in practice the probabilities over a few times μ are tiny)



Maximizing Poisson likelihood

- $\mathcal{L}(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$, $\ln\mathcal{L} = -\mu + n\ln\mu - \ln n!$
- $-2 \frac{d\ln\mathcal{L}}{d\mu} = 0 = -2 \left(-1 + \frac{n}{\hat{\mu}} \right) \Rightarrow \hat{\mu} = n$
- $-\frac{d^2\ln\mathcal{L}}{d\mu^2} = \frac{n}{\hat{\mu}^2} = \frac{1}{\hat{\mu}} = \frac{1}{n} = \frac{1}{\sigma^2}$, i.e. $\hat{\sigma}^2 = n$
- Consistent with the variance calculation!
- Question: If we observe 25 counts of some process, what is our estimate of μ , with uncertainty? $\hat{\mu} = 25 \pm 5$

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? $\mu = \frac{150}{100} = 1.5$

Poisson process in pieces

- Say we want to compare a single measurement of μ with N measurements of μ/N
 - e.g., a 10-minute measurement compared to 10 1-minute measurements
- 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!