

Mean of multiple Gaussian distributed measurements

- Suppose you have N Gaussian-distributed measurements x_i of some quantity μ , but they all have different uncertainties σ_i . 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 will come back to the very important χ^2

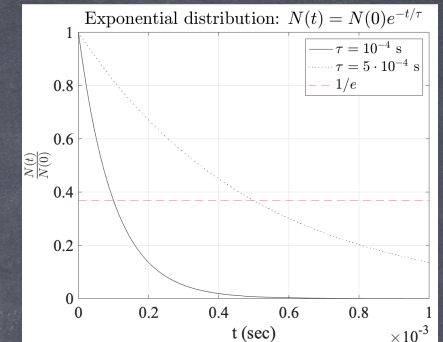
Mean of multiple Gaussian distributed measurements

- Suppose we make N independent measurements all with the same $\sigma_i \equiv \sigma$
- $\hat{\mu} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2} \pm \frac{1}{\sqrt{\sum_i 1 / \sigma_i^2}} \rightarrow \hat{\mu} = \frac{\sum_i x_i}{N} \pm \frac{\sigma}{\sqrt{N}}$
- 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? To overcome the $1/\sqrt{N}$ dependence we have to increase the dataset by a factor $10^2 = 100$.

Bonus: Alternate derivation of uncertainty in mean for single σ

- $\sigma_m^2 = E[(\bar{x} - \mu)^2]$
- $x = \mu + \sigma\delta$, where δ is a **unit normal distribution**, i.e., we express observables in terms of independent fluctuations around model parameters
- i.e., $E[\delta] = 0, E[\delta^2] = 1$
- $E[(\bar{x} - \mu)^2] = E\left[\left(\frac{1}{N}\sum x_i - \mu\right)^2\right] = E\left[\left(\frac{1}{N}\sum \sigma\delta_i\right)^2\right]$
- $= \frac{\sigma^2}{N^2}E\left[\sum \delta_i^2 + \sum_i \sum_{j \neq i} \delta_i \delta_j\right] = \frac{\sigma^2}{N^2}(N + 0) = \frac{\sigma^2}{N}$
- In other words, $\sigma_m = \frac{\sigma}{\sqrt{N}}$, the same as above

Exponential distribution



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

- $\frac{dN}{dt} = -\lambda N(t) \rightarrow N(t) = N_0 e^{-\lambda t}$

- Absorption of radiation

- $\frac{dI}{dx} = -\mu I(x) \rightarrow I(t) = I_0 e^{-\mu x}$

- What are λ and μ ?

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

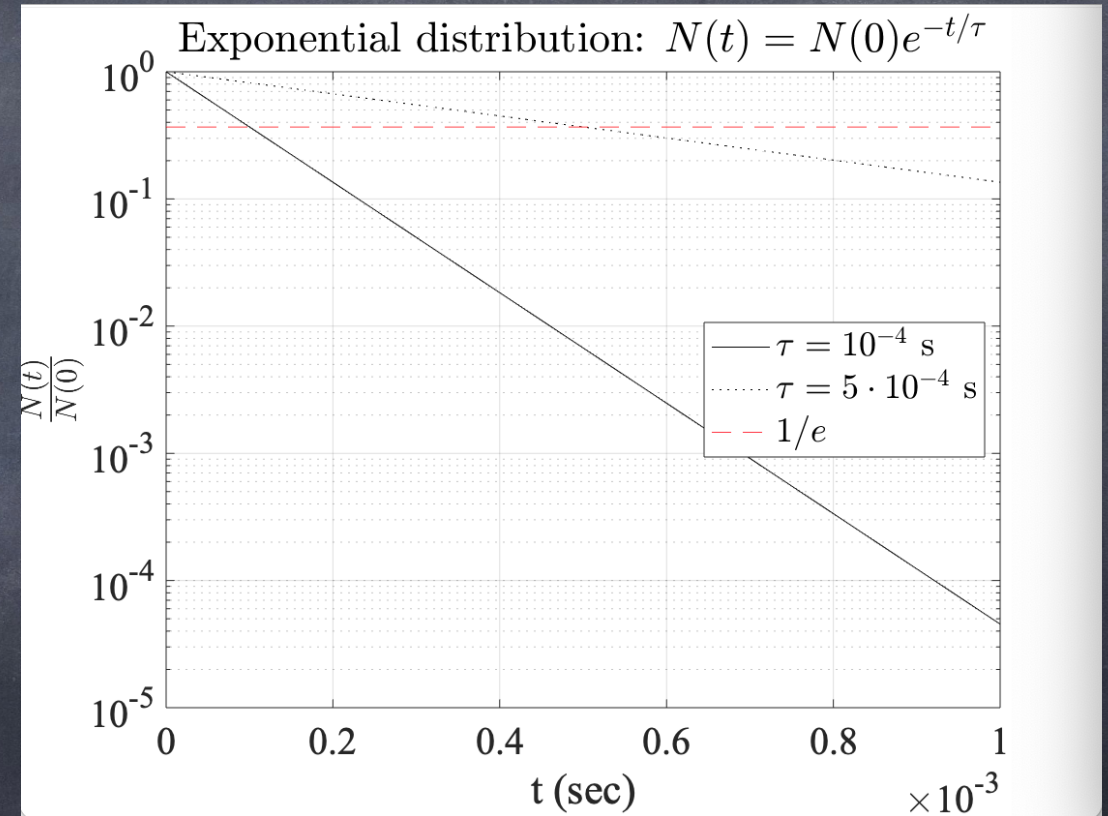
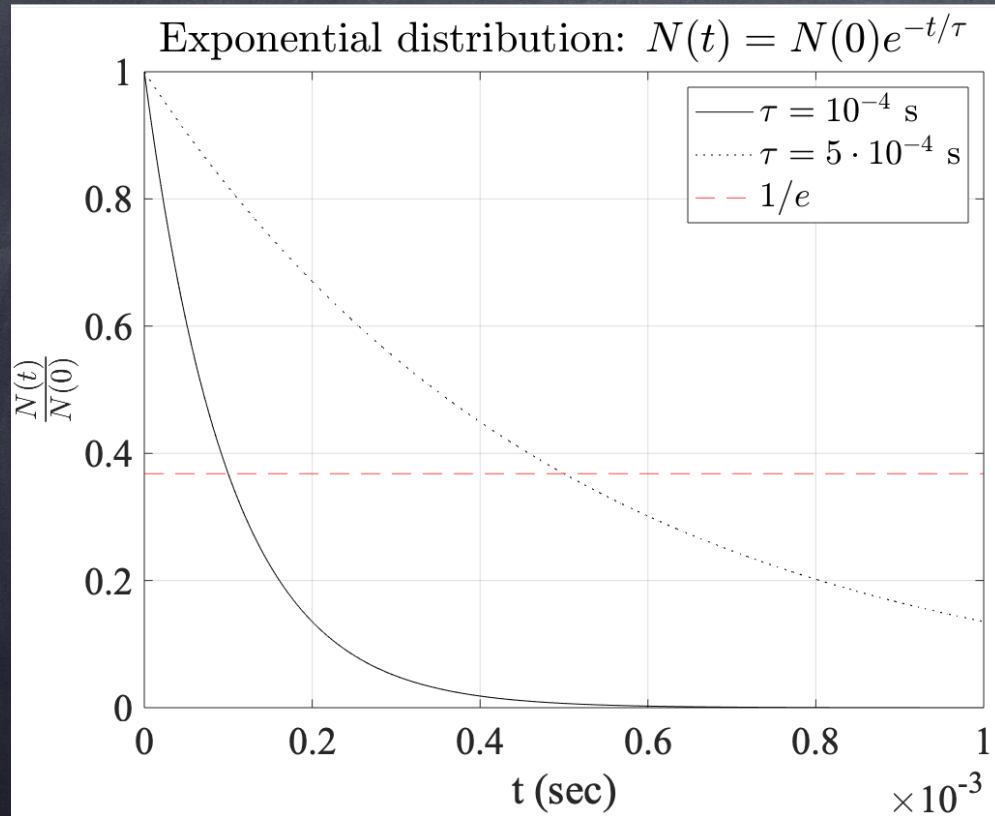
- λ is average decay rate per unit time or "decay constant"

- μ is average absorption per unit distance

Exponential distribution

- 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 = $\tau^2 = \frac{1}{\lambda^2}$

Exponential distribution



Breit-Wigner distribution

- In the formation of unstable nuclei and particles, the exponential distribution of decay times ($\propto e^{-\lambda t}$) results in a resonance phenomenon.
- If the lifetime is short enough, the resonance (distribution of invariant mass W) may be wide enough to be detectable:

$$f(W|M, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(Wc^2 - Mc^2)^2 + \Gamma^2/4}$$

- $\Gamma = \hbar\lambda = \frac{\hbar}{\tau}$ (In HEP it is standard practice to let $\hbar = c = 1$)
- Both $E[W]$ and $E[W^2]$ are ill-defined! $\Rightarrow$ The B-W can't be characterized by mean and variance, but rather by mode (peak position Mc^2) and FWHM (Γ)
- FWHM also useful for energy resolution, since there are often lower and upper tails larger than expected for a Gaussian

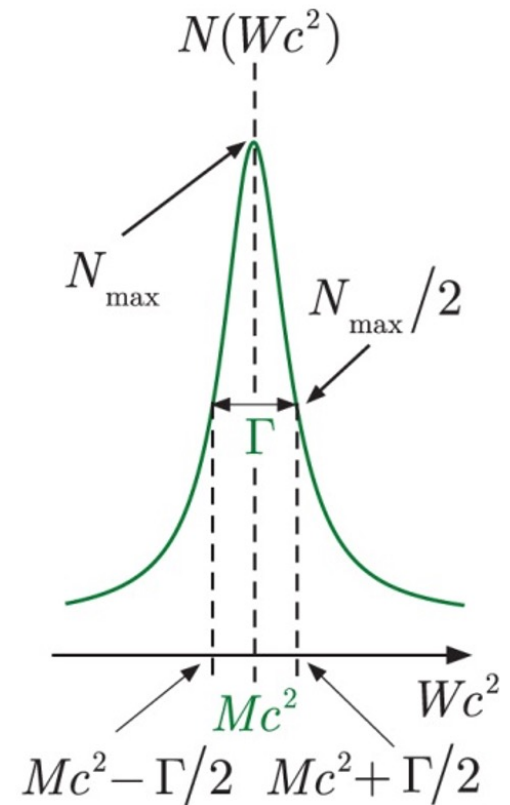


Figure 1.9 The Breit-Wigner formula (1.82).

Nuclear and Particle Physics: An Introduction, Third Edition. B.R. Martin and G. Shaw.
© 2019 John Wiley & Sons Ltd. Published 2019 by John Wiley & Sons Ltd.
Companion website: www.wiley.com/go/martin/nuclear3

FYI: Cauchy distribution

- In math and stats the **Cauchy distribution** is a special case of the **Breit-Wigner** (or Lorentzian) distribution
- Let $M \rightarrow 0, \Gamma \rightarrow 2, x \equiv W$: $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ is the Cauchy pdf

Summary of pdfs (so far)

	Gaussian	Exponential	Binomial	Poisson	Breit-Wigner
Observe	$f(x \mu, \sigma)$	$f(x \lambda)$	$f(n N, p)$	$f(n \mu)$	$f(x x_0, \Gamma)$
pdf	$\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\lambda e^{-\lambda x}$	$\frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$	$\frac{e^{-\mu} \mu^n}{n!}$	$\frac{1}{\pi} \frac{\Gamma/2}{(x-x_0)^2 + \Gamma^2/4}$
Mean	μ	$1/\lambda$	Np	μ	undefined
Variance	σ^2	$1/\lambda^2$	$Np(1-p)$	μ	undefined
Example	Position measurement	Decays of radioactive nuclei	Detector efficiency	Counting occurrences	Invariant mass

Part II

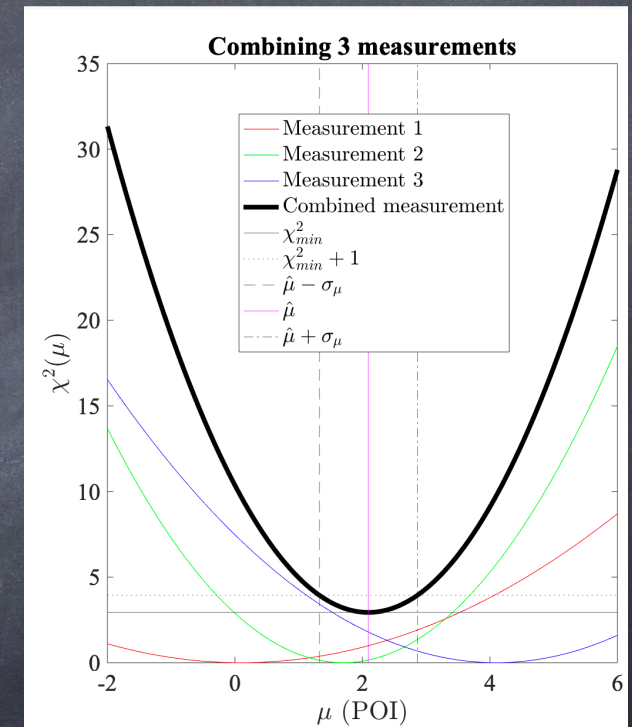
- χ^2 distribution
- Error propagation and covariance
- Rounding of experimental results
- Curve fitting
- I don't propose to go into details of numerical minimization techniques.

The χ^2 distribution

- Recall from mean of measurements with Gaussian pdf

$$-2\ln\mathcal{L} = \sum_i \frac{(x_i - \mu)^2}{\sigma_i^2} + \sum_i \ln(2\pi\sigma_i^2) \equiv \chi^2 + \sum_i \ln(2\pi\sigma_i^2)$$

- The second term, assuming we know the uncertainties σ_i , is constant
- In this problem we see that $\chi^2(\mu)$ is a parabola (how can I say this?) with a minimum at $\hat{\mu}$.
- The sum of two parabolas is also a parabola.
- What can we learn from the value of χ^2 at the minimum, i.e., what is its distribution?

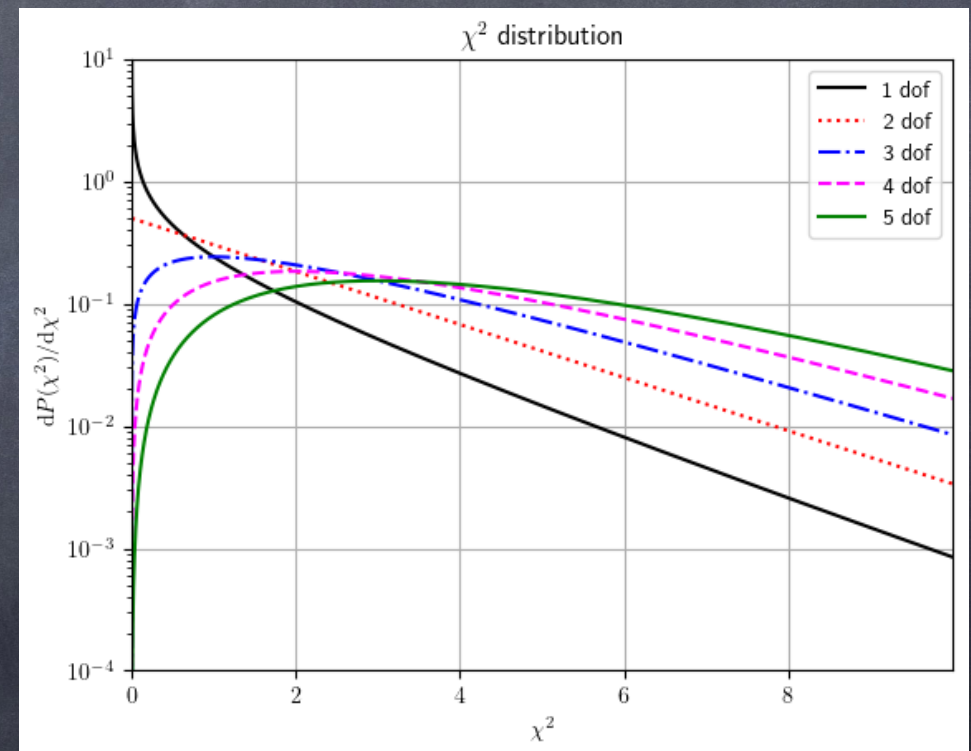


The χ_n^2 distribution

- We aren't going to go through the maths, but the probability density function (pdf) is

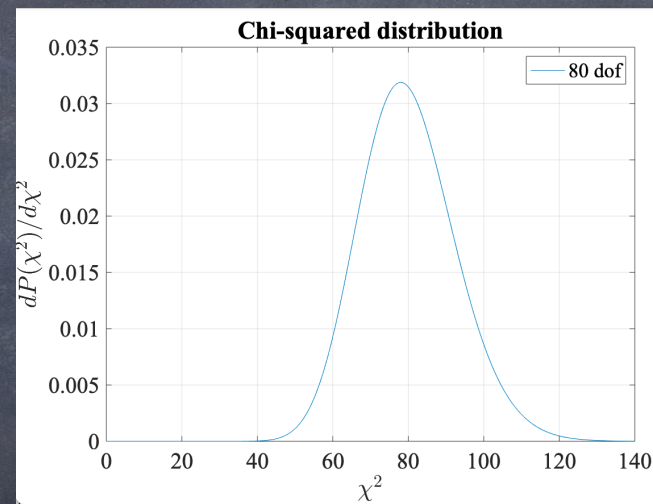
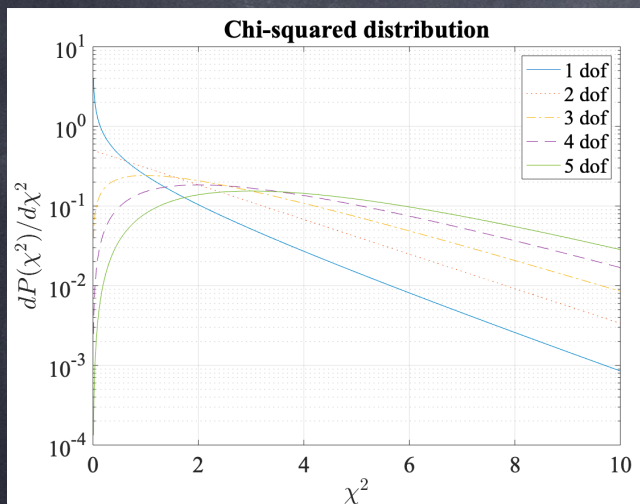
$$f(\chi^2, n) = \frac{dP(\chi^2, n)}{d\chi^2} = \frac{(\chi^2/2)^{n/2-1} e^{-\chi^2/2}}{2\Gamma(n/2)}$$

- n (i.e. n_{dof}) is the number of degrees of freedom.
- Typically $n_{dof} = n_{data} - n_{par}$
 - Later, when estimating variance of a sample using the measured mean, we will have $n_{dof} = N - 1$, the «missing» degree of freedom used by the mean
 - Later, we will do fits with several free parameters ($n_{par} > 1$)



The χ_n^2 distribution

$$\chi^2 = \sum_i \frac{(x_i - \mu)^2}{\sigma_i^2}$$



- What is χ^2 expressing?
Test if spread of results
consistent with model

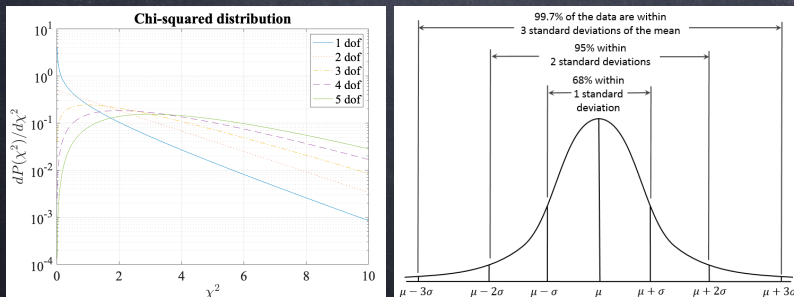
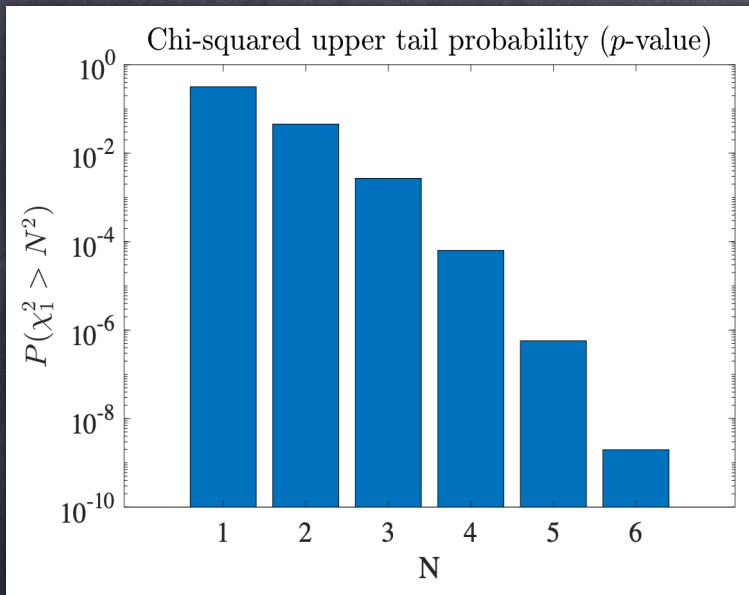
- For low n_{dof} (i.e. n) the distribution is skewed toward 0
- For large n_{dof} the distribution becomes more symmetric
- mean $E[\chi_n^2] = n$
- variance $E[(\chi_n^2 - n)^2] = 2n$
- Rule of thumb: χ_n^2/n much larger than 1 indicates an anomaly
- Interpretation of $\chi_n^2/n \ll 1$?
Also anomalous. Possibly overestimated uncertainties

The χ_n^2 distribution

Don't spend
time on this
Make a quiz

- Remember: mean $E[\chi_n^2] = n$, variance $E[(\chi_n^2 - n)^2] = 2n$
- Question: If we repeat a measurement with 100 dof a bunch of times, what average χ^2 do we expect? What standard deviation?
- Question: If we perform a single measurement with 100 dof and observe $\chi^2/dof = 1.28$, what might we conclude?
- Question: If we do a different measurement with 1 dof and observe $\chi^2/dof = 1.28$, what might we conclude?
- Have we learned something about the anomaly rule-of-thumb?!

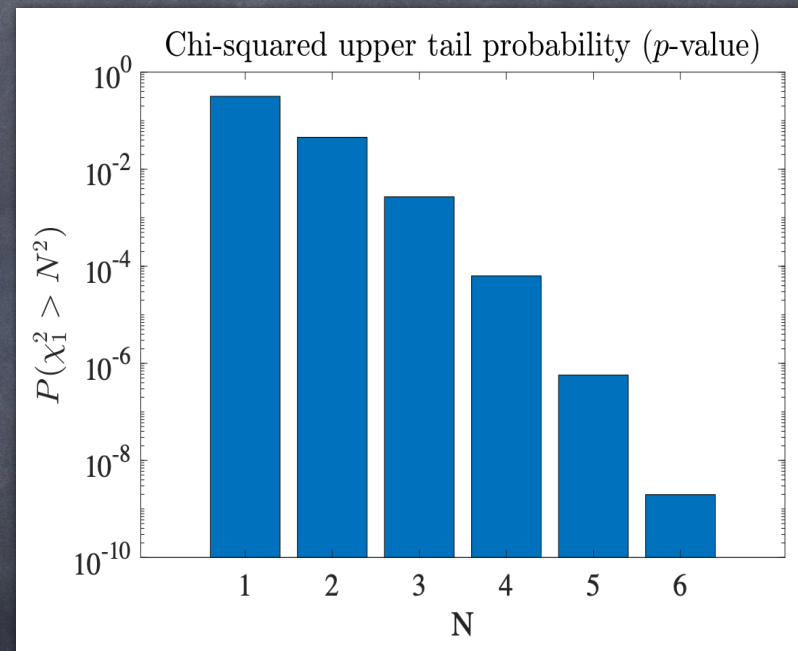
χ_n^2 tail probability: $P(\chi_n^2 > c)$



- p -value (which gets abused in some scientific fields and thus has gotten some bad press in recent years) tells something about how far out in the (improbable) tail of the distribution our data is.
- If the p -value is extreme enough we may make claims about «discovery» or «exclusion» of some model/theory/prediction, but **rigorous practice** includes showing that an alternate hypothesis is a better and reasonable description of the data.
- $$p = P(\chi_n^2 > c) = \int_c^\infty \frac{dP(\chi_n^2)}{d\chi_n^2} d\chi_n^2$$
- It's no accident that $P(\chi_1^2 > N^2) = 2P(x - \mu > N\sigma)$ for Gaussian distributed $G(x|\mu, \sigma)$!
- BONUS: In HEP we frequently use $\frac{1}{2}P(\chi_1^2 > N^2)$ for p -values, since **discoveries** are most often an excess of data rather than either excess or deficit.

χ_n^2 tail probability

$N\sigma$	$P(\chi_1^2 > N^2)$	$\frac{1}{2}P(\chi_1^2 > N^2)$
1	0.3172	0.1587
2	0.0455	0.0228
3	$2.70 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$
4	$6.33 \cdot 10^{-5}$	$3.17 \cdot 10^{-5}$
5	$5.73 \cdot 10^{-7}$	$2.87 \cdot 10^{-7}$



ROOT, MATLAB, Python

```
root [0] ROOT::Math::chisquared_cdf_c(25,1)  
(double) 5.7330314e-07
```

```
>>> from scipy.stats import chi2  
>>> chi2.sf(25,1)  
5.733031437583875e-07
```

```
>> chi2cdf(25,1,'upper')  
  
ans =  
  
5.7330e-07
```

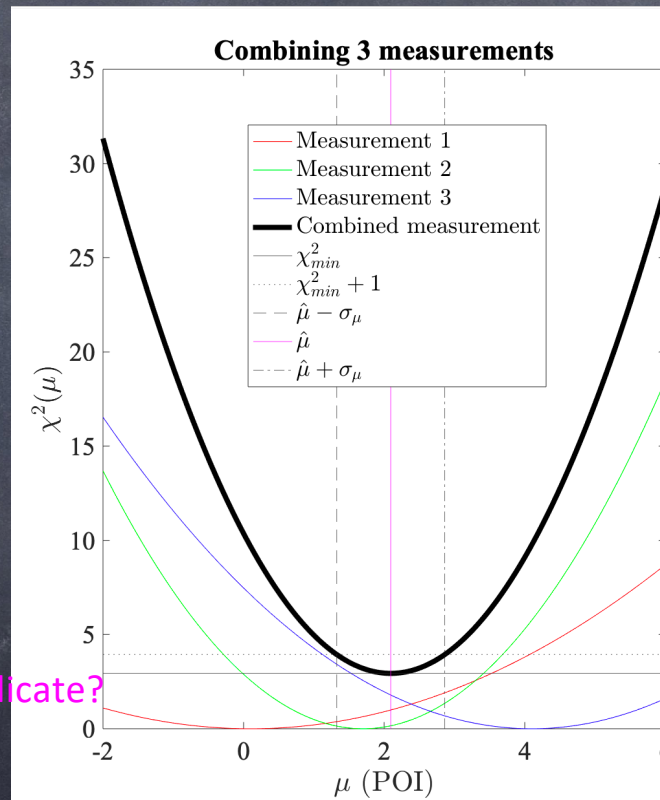
Combining 3 measurements $\chi^2 = \sum_i \frac{(x_i - \mu)^2}{\sigma_i^2}$

- Combine
 - 0.1 ± 2.0
 - 1.7 ± 1.0
 - 4.1 ± 1.5

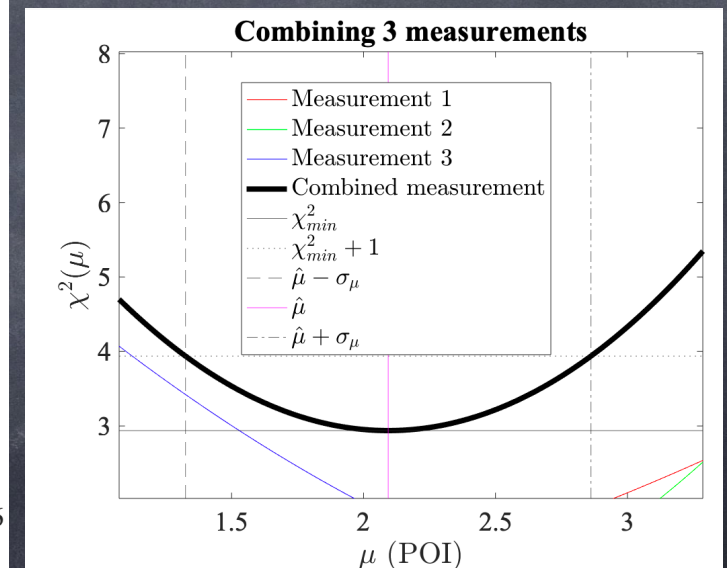
- Draw the χ^2 curves

- Calculate $\hat{\mu} \pm \sigma_\mu$ and compare to the combined χ^2

Question: What does the $\chi^2=2.94$ at $\hat{\mu}$ indicate?
 "Goodness of fit" $P(\chi^2_{3-1} > 2.94) = 0.23$
 so model probably consistent with data.



$$\frac{1}{\sigma_\mu^2} = \sum \frac{1}{\sigma_i^2}$$



P.S. This graphical approach is equivalent to previous mean w/error!

A. Read, U. Oslo

$\Delta\chi^2 = +N^2$ rule

- Typically, we are interested in $N = 1$, to find σ_μ
- How much does χ^2 increase at $\pm N\sigma_\mu$ from the minimum at $\hat{\mu}$ (for 1 dof)?
- $$\Delta\chi^2 \equiv \sum_i \frac{(x_i - (\hat{\mu} \pm N\sigma_\mu))^2}{\sigma_i^2} - \sum_i \frac{(x_i - \hat{\mu})^2}{\sigma_i^2} = \dots = N^2$$
- Alternative to $\frac{\partial^2 \chi^2}{\partial \mu^2} = \frac{2}{\sigma_\mu^2}$
- Also works well (MINOS procedure in MINUIT) for approximately parabolic $-2\ln\mathcal{L}(\mu)$, and often better than $\frac{\partial^2(-2\ln\mathcal{L})}{\partial \mu^2} = \frac{2}{\sigma_\mu^2}$

Propagation of errors

- Suppose we have a function f of several variables x_i , all of which have an uncertainty δx_i

- We are interested in the expectation value of the variance of f :

$$E(\text{var}(f)) = \langle (f - \bar{f})^2 \rangle = \langle (\delta f)^2 \rangle$$

- We approximate the deviations of f by $\delta f \approx \sum_i \frac{\partial f}{\partial x_i} \delta x_i$

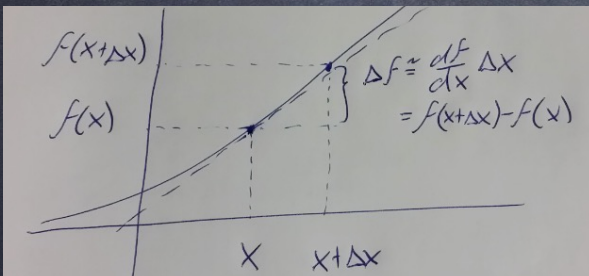
- Variances and covariances of the x_i :

$$\langle (\delta x_i)^2 \rangle \equiv \sigma_i^2 \text{ and } \langle \delta x_i \delta x_j \rangle \equiv \text{cov}(x_i, x_j)$$

- We square and take the average

$$E(\text{var}(f)) = \langle (\delta f)^2 \rangle \approx \sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + 2 \sum_i \sum_{j>i} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{covar}(x_i, x_j)$$

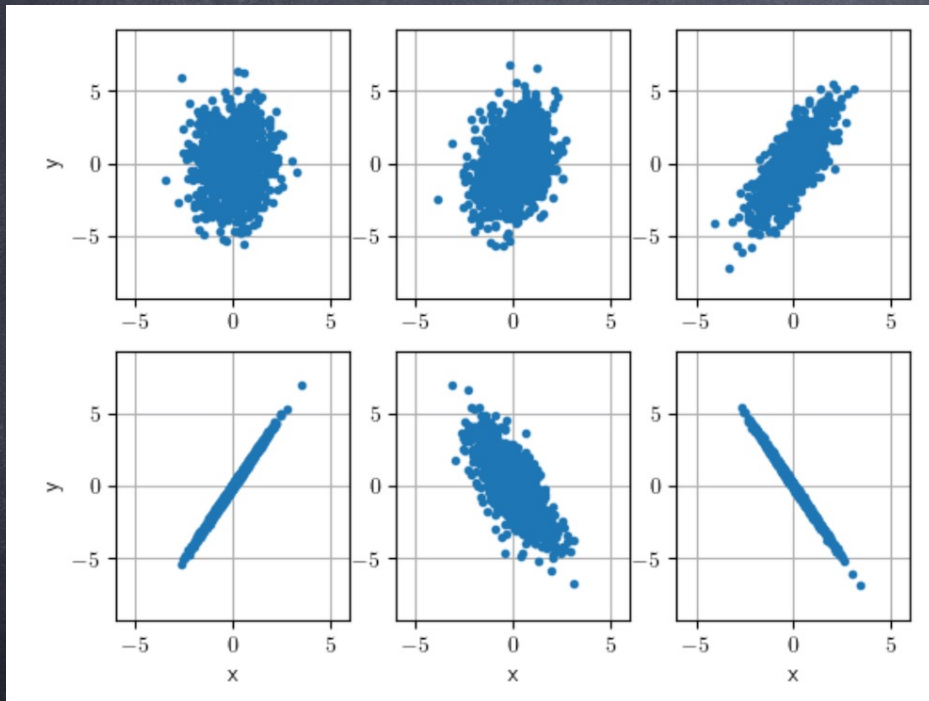
- Question: What pdf are we assuming for the x_i ? Gaussian distribution



Propagation of errors

- $var(f) = \langle (\delta f)^2 \rangle \approx \sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + 2 \sum_i \sum_{j>i} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} covar(x_i, x_j)$
- P.S. This is a bad approximation if we happen to hit a minimum of f !
- In many problems the covariance is zero (independent measurements of different quantities) and we have a relatively simple formula for propagating errors: $var(f) \approx \sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$
- P.S. $covar(x_1, x_2) \equiv \langle (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) \rangle$ has same form as $var(x_1) \equiv \langle (x_1 - \bar{x}_1)(x_1 - \bar{x}_1) \rangle$

Propagation of errors



Question: Roughly what are the (a) variances and (b) covariances here?

a) Variance in x is about 1 and in y about 2 (applies to all 5 plots).

b) Covariances are roughly zero, small positive, large positive, «100%» positive, large negative, «100%» negative.

Propagation of errors

- An example of a function of two variables, with no covariance between the measurements: $R = \frac{\rho^2 + h^2}{2h}$, $\rho \pm \sigma_\rho$, $h \pm \sigma_h$

- Apply $\sigma_f^2 \approx \sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$

- $\sigma_R^2 \approx (\rho/h)^2 \sigma_\rho^2 + \frac{1}{4}(1 - \rho^2/h^2)^2 \sigma_h^2$

Question: If $f = \frac{x}{y}$ and there is no covariance between x and y , what is $\left(\frac{\sigma_f}{f} \right)^2$? $\left(\frac{\sigma_f}{f} \right)^2 = \left(\frac{\sigma_x}{x} \right)^2 + \left(\frac{\sigma_y}{y} \right)^2$

Propagation of errors

- Combining experimental and/or theoretical uncertainty (systematic uncertainties) with statistical uncertainty can be thought of as adding terms with 0 contribution to the most probable value, each with uncertainty:

- $(\hat{\mu} \pm \sigma_{stat}) + (0 \pm \sigma_{expt}) + (0 \pm \sigma_{theory})$

- No reason for there to be covariances so this function is the addition of central values and the errors can be summed quadratically:

- $\hat{\mu} \pm \sigma_{tot}, \sigma_{tot}^2 = \sigma_{stat}^2 + \sigma_{expt}^2 + \sigma_{theory}^2$ («Pythagoros» in n orthogonal dimensions)

- Question: How do we combine multiple experimental sources of uncertainty that have no covariances? Sum the squared uncertainties!

Combining experimental and/or theoretical uncertainty (systematic uncertainties) with statistical uncertainty can be thought of as adding terms with 0 contribution to the most probable value, each with uncertainty:

$(\hat{\mu} \pm \sigma_{stat}) + (0 \pm \sigma_{expt}) + (0 \pm \sigma_{theory})$

No reason for there to be covariances so this function is the addition of central values and the errors can be summed quadratically:

$\hat{\mu} \pm \sigma_{tot}, \sigma_{tot}^2 = \sigma_{stat}^2 + \sigma_{expt}^2 + \sigma_{theory}^2$ «Pythagoros» in n orthogonal dimensions)

Question: How do we combine multiple experimental sources of uncertainty that have no covariances? Sum the squared uncertainties!

Rounding

- There is an art to rounding numbers. Several conventions exist.
- The PDG convention is widely used in HEP (see Sec. 5.3 of <https://pdg.lbl.gov/2016/reviews/rpp2016-rev-rpp-intro.pdf>)
- Results and errors should be rounded to compatible significant digits
 - Number of significant digits of the uncertainty should not exceed 2 (uncertainty of the uncertainty should not be overstated).
 - If the next digit is 0-4, then keep the last digit
 - If the next digit is 5-9, increase last digit by 1
 - 1.23456 can be rounded to 1, 1.2, 1.23, 1.235, 1.2346
- Examples: 23.4 ± 2.1 , 1.74 ± 0.04 , $(3.1415 \pm 0.0009) \times 10^{-43}$, 10400 ± 600

Rounding

- When you round, don't round 1 digit at a time!
- Leo's example: $2.346 \rightarrow 2.35 \rightarrow 2.4$, compared to $2.346 \rightarrow 2.3$
- Whatever your convention, use it consistently.

Rounding (PDG-style)

The basic rule states that if the three highest order digits of the error lie between 100 and 354, we round to two significant digits. If they lie between 355 and 949, we round to one significant digit. Finally, if they lie between 950 and 999, we round up to 1000 and keep two significant digits. In all cases, the central value is given with a precision that matches that of the error. So, for example, the result (coming from an average) 0.827 ± 0.119 would appear as 0.83 ± 0.12 , while 0.827 ± 0.367 would turn into 0.8 ± 0.4 .

$100 - 354 \rightarrow 2 \text{ digits}$

$355 - 949 \rightarrow 1 \text{ digit}$

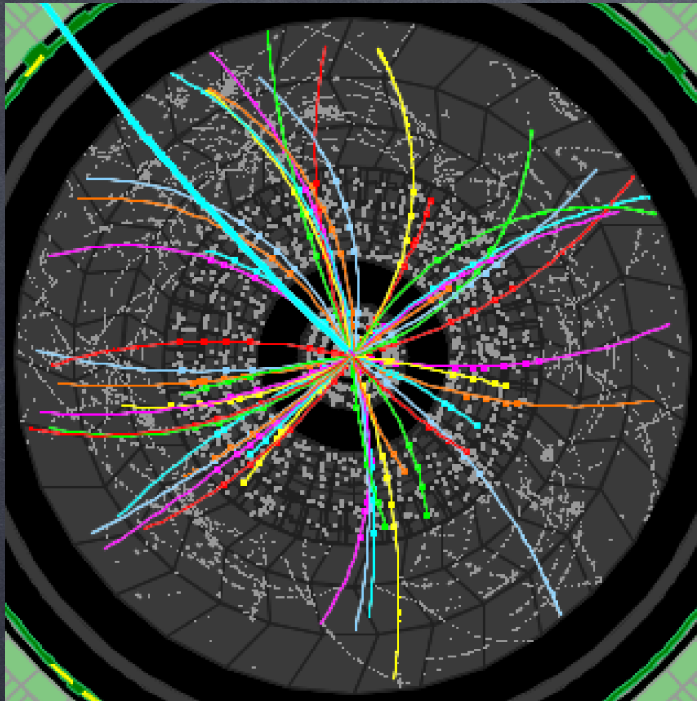
$950 - 999 \rightarrow 10(00)$

$112.3462 \pm 0.967 \rightarrow ? \quad 112.3 \pm 1.0$

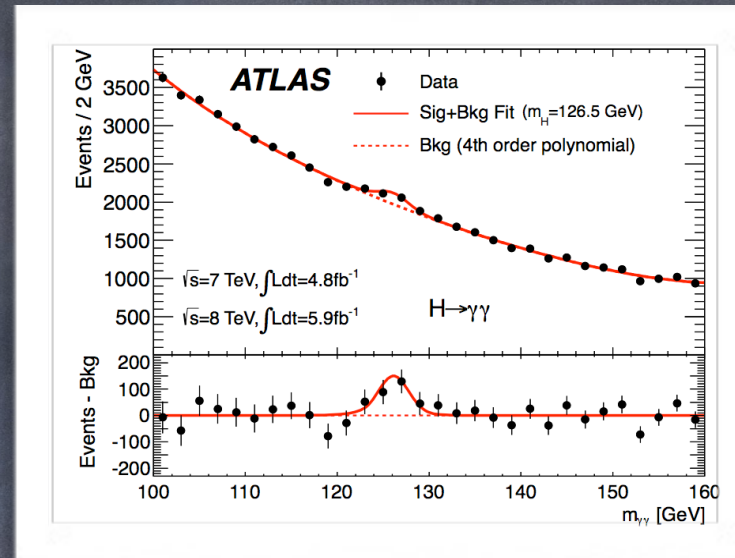
$(0.827 \pm 0.055) \times 10^{-12} \rightarrow ?$

$(0.83 \pm 0.06) \times 10^{-12}$

Curve-fitting (examples)



- Following particles through an experiment (charged particles in a magnetic field)



- Extracting information about a signal from a fit of data to a parametric statistical model

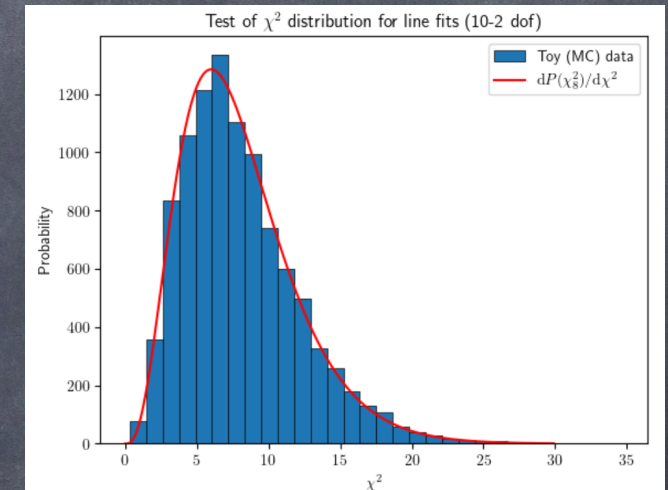
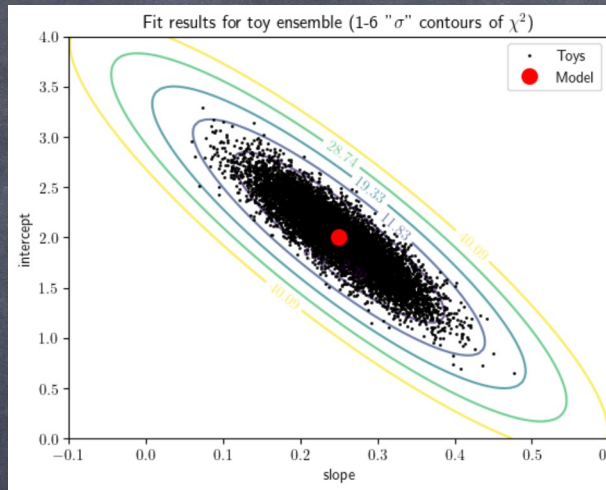
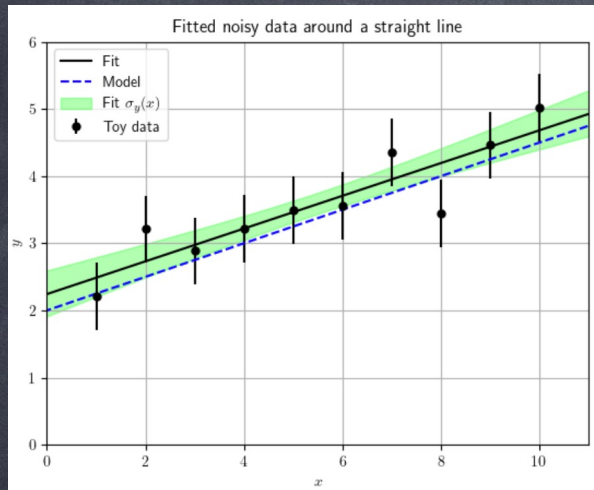
Curve-fitting basics

- $\chi^2 = \sum_{i=1}^n \left(\frac{y_i - f(x_i, \mu)}{\sigma_i} \right)^2$ and minimize: $\frac{\partial \chi^2}{\partial \mu} = 0$, i.e., solve $\frac{\partial \chi^2}{\partial \mu_j} = 0$ for all j
- Leo p. 104 calls this «least-squares», but traditional least-squares is done with $\sigma_i = 1$, i.e., no prior knowledge of uncertainties
- Can be solved analytically (matrix algebra) for functions that are linear in the parameters, e.g. $f(x_i, \mu) = \sum_{k=1}^m \mu_k g_k(x_i)$
- Occasionally it is possible to make the problem linear by an appropriate transformation, e.g., $N = N_0 e^{-\lambda t} \rightarrow \ln N = \ln N_0 - \lambda t$
- Numerical methods used in general (e.g. MINUIT package in ROOT, or `scipy.optimize.curve_fit` in Python)

Curve-fitting

- Leo (p. 105) demonstrates the analytic solution to the straight-line fit.
- An alternate (but a bit more complicated) path to the same results for the covariance matrix is by propagating the uncertainties on the fit parameters! For the details see (in Norwegian)
<https://www.uio.no/studier/emner/matnat/fys/FYS2150/v19/kursmateriell/usikkerhet/fys2150-linjetilpasning-2019.pdf>

Curve-fitting (line example)



fit to 1 «toy» data set

an ensemble of «toys»

Let's first take a practical approach and use the numerical curve-fitting (χ^2 -minimization etc) of our favorite tool.