8.882 LHC Physics
Experimental Methods and Measurements

Likelihoods and Selections
[Lecture 20, April 22, 2009]

Organization

Project 2 ...

- Matthew handed in, not yet corrected

Project 3

- looks like people have no particular issues
- recitation in Friday was rather quiet

Conference Schedule

- Tuesday May 19 at 12:00 Kolker Room

Final Conference Project

LHC Physics: “Experimental Methods and Measurements”

Plenary Session (12:00–13:30, May 19, Kolker Room)

- Welcome and LHC Overview (C.Paus)
- Search for Standard Model Higgs Boson: Overview (?)
- Search for Higgs in $H \rightarrow ZZ^*$ (Mattew Chan)
- Search for Higgs in $H \rightarrow WW^*$ (?)
- Search for Higgs in $qqH \rightarrow qqWW^*$ (?)

Physics
Colloquium Series

‘09
Spring

The Physics Colloquium Series

Thursday, April 23 at ***4:15 pm*** *in room* ***10-250***

Alain Aspect

Institut d'Optique, Palaiseau, France

"Wave particle duality for a single photon: from Einstein's LichtQuanten to Wheeler's Delayed Choice Experiment"

For a full listing of this semester’s colloquia,
please visit our website at **web.mit.edu/physics**

Lecture Outline

Likelihoods and Selections

- likelihoods and fits
  - statistical uncertainties
  - full likelihood for lifetimes
- checking whether it makes sense
  - goodness of fits
  - projections

Sophisticate Selections

- likelihoods
- neural networks

Maximum Likelihood Estimator

Taylor expansion around minimum, p_{fit}

$$\log L(p) = logL|_{p_{\mathrm{fit}}} + \frac{1}{2}\frac{\partial^2 \log L}{\partial p^2}|_{p_{\mathrm{fit}}}(p - p_{\mathrm{fit}})^2$$

$$L(p) = L|_{p_{\mathrm{fit}}} \exp\left(\frac{1}{2}\frac{\partial^2 \log L}{\partial p^2}|_{p_{\mathrm{fit}}}(p - p_{\mathrm{fit}})^2\right)$$

Consider this as a PDF for true value of parameter p

- PDF is a Gaussian with mean value p_{fit}
- variance is given as

$$\sigma^2 = V(p) = -\frac{1}{\frac{\partial^2 \log L}{\partial p^2}|_{p_{\mathrm{fit}}}}$$

Maximum Likelihood Estimator

Again Taylor expansion around minimum, p_{fit}, but using definition of the variance σ

$$\log L(p) = logL|_{p_{\text{fit}}} - \frac{1}{2}\frac{(p - p_{\text{fit}})^2}{\sigma^2}$$

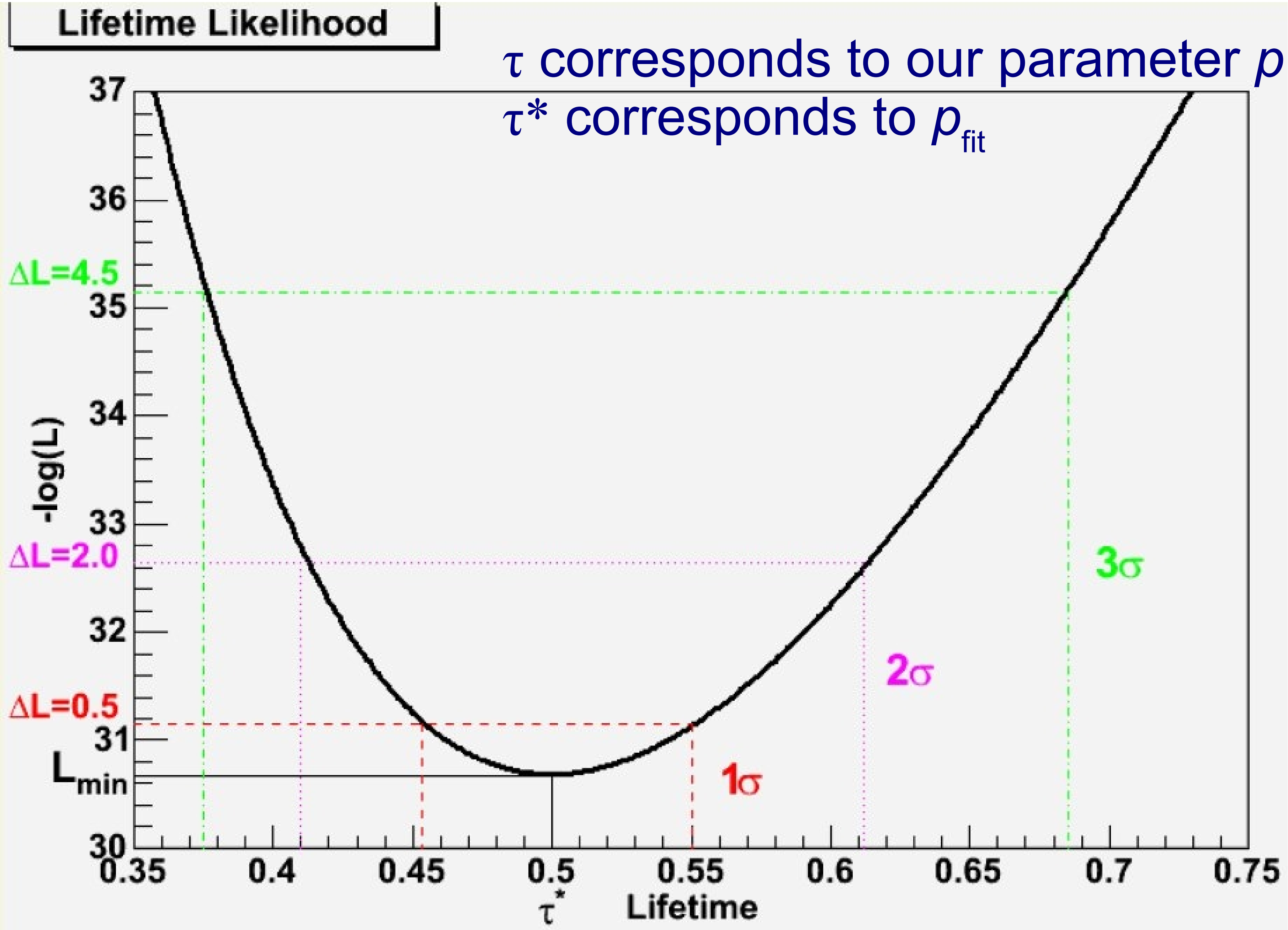
Values of the likelihood for 1, 2, n σ (standard deviations) from the central value are

$$\log L(p_{\text{fit}} \pm 1\sigma) = logL|_{p_{\text{fit}}} - \frac{1}{2}$$

$$\log L(p_{\text{fit}} \pm 2\sigma) = logL|_{p_{\text{fit}}} - 2$$

$$\log L(p_{\text{fit}} \pm n\sigma) = logL|_{p_{\text{fit}}} - \frac{1}{2}n^2$$

Picture of Uncertainties



τ corresponds to our parameter *p*

τ* corresponds to p_{fit}

Correspondence to χ2

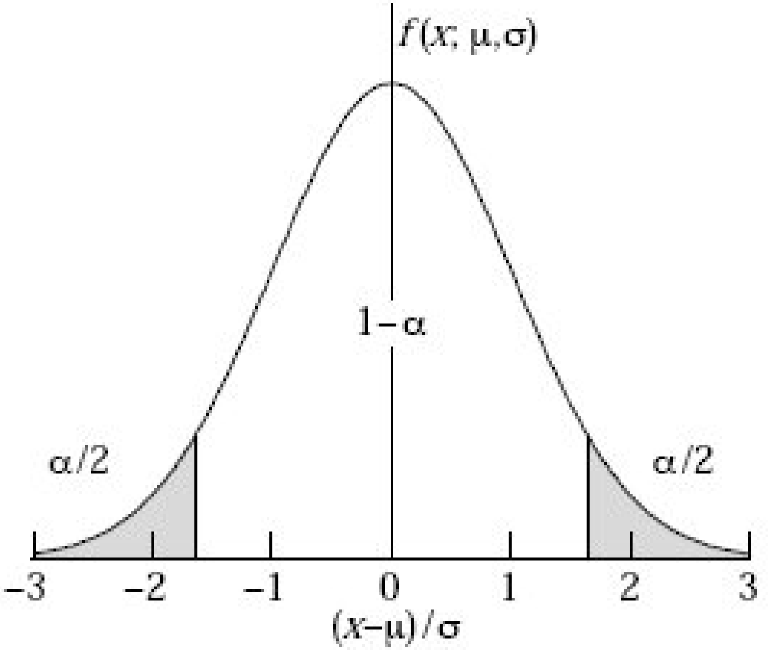
For Gaussian PDF we know

$$1-\alpha = \frac{1}{\sqrt{2\pi\sigma}} \int_{\mu-\delta}^{\mu+\delta} dx \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\sigma}\right)$$

$$.\chi^2 = \sum_i \frac{(f_i - x_i)^2}{\sigma_i^2}$$ follows Gaussian

One standard deviation is interval which includes 68%

- change in minimum χ^2 by $1^2 = 1$
- two standard deviations correspond to $\Delta\chi^2 = 2^2 = 4$ (95%)
- or n standard deviations correspond to $\Delta\chi^2 = n^2$



Correspondence to χ^2

Confidence level intervals for Gaussian *n* sigma

- in root: α = 1.0 – TMath::Erf(1.0**n*/sqrt(2))
- or: *P* = TMath::Erf(1.0**n*/sqrt(2))
- probability for 5 standard deviations is astonishingly small well, it should be

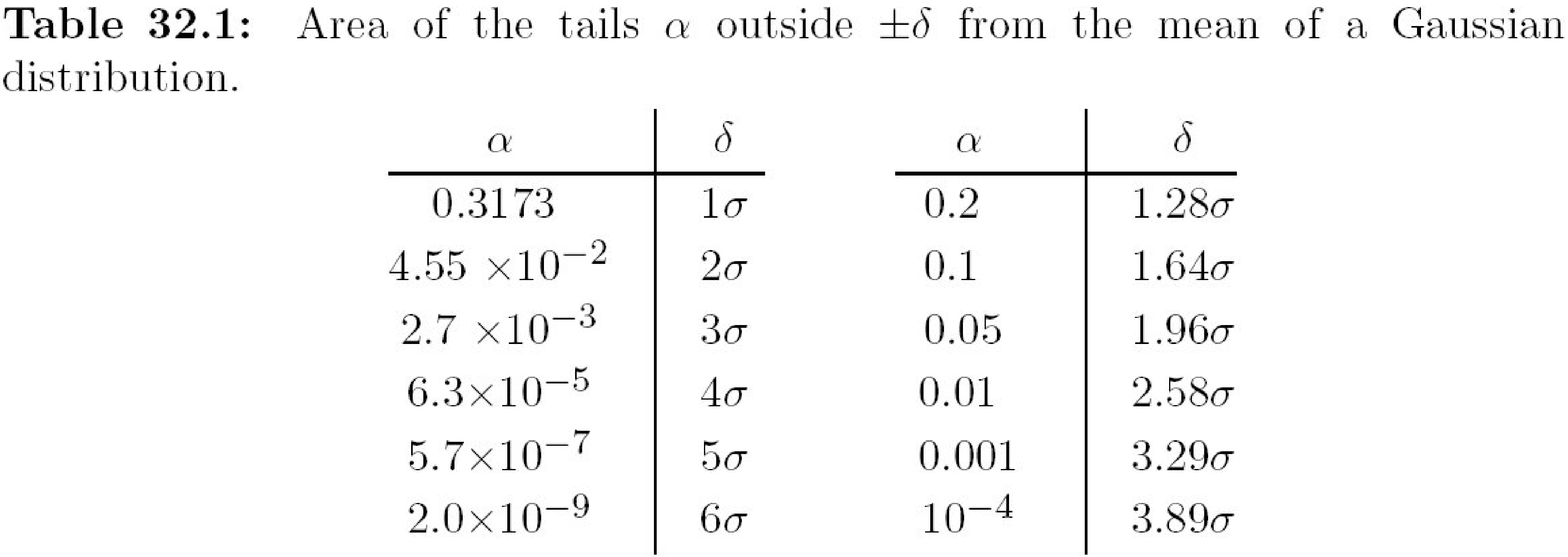


Table 32.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
4.55×10^{-2}	2σ	0.1	1.64σ
2.7×10^{-3}	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ

Analytical Estimate of Variance

Lifetime likelihood and variance:

$$-\log L = N \log \tau + \frac{1}{\tau} \sum_{i=1}^{N} t_i \qquad \sigma^2 = V(\tau) = -\frac{1}{\frac{\partial^2 \log L}{\partial p^2}\Big|_{\tau_{\text{fit}}}}$$

$$\frac{\partial(-logL)}{\partial \tau} = 0 = \frac{N}{\tau_{\text{fit}}} - \frac{1}{\tau_{\text{fit}}^2} \sum_{i=1}^{N} t_i$$

$$\frac{\partial^2(-logL)}{\partial \tau^2} = -\frac{N}{\tau_{\text{fit}}^2} + \frac{2}{\tau_{\text{fit}}^3} \sum_{i=1}^{N} t_i = \frac{N}{\tau_{\text{fit}}^2}$$

$$\rightarrow \sigma^2 = -\frac{1}{\frac{\partial^2 \log L}{\partial p^2}\Big|_{\tau_{\text{fit}}}} = \frac{\tau_{\text{fit}}^2}{N}$$

Full Likelihood for our Analysis

So far Likelihood for 1 measurement type as input, t_i

- but we are using more measurement types ....
- mass, (uncertainty of mass,) uncertainty of proper time

How to account for these additional dimensions?

- very simple for likelihood: multiply PDFs (should be independent): $P(t,m) = P(t)\,P(m)$
- also treat signal and background separately

$$P(t, m|\tau, f_{\rm bg}, m_B, a, b) = f_{\rm sig}\left(\frac{1}{\tau}\exp(-t/\tau)\right) G(m|m_B, \sigma_B) +$$

$$(1 - f_{\rm sig})\delta(t)\,(am + b)$$

signal: has a lifetime

Gaussian mass distribution

background: no lifetime, t = 0

flat mass distribution

Full Likelihood for our Analysis

Including detector imperfections

- proper time has uncertainty attached to it
- smears out the signal exponential distribution as well as the δ distribution of the background

$$P(t, m|\tau, f_{\text{bg}}, m_B, a, b) = f_{\text{sig}} \left(\frac{1}{\tau} \exp(-t'/\tau) \otimes G(t'|0, \sigma) \right) G(m|m_B, \sigma_B) + (1 - f_{\text{sig}}) G(t|0, \sigma) \, (am + b)$$

In principle *P(Δt)* and *P(Δm)* to be included

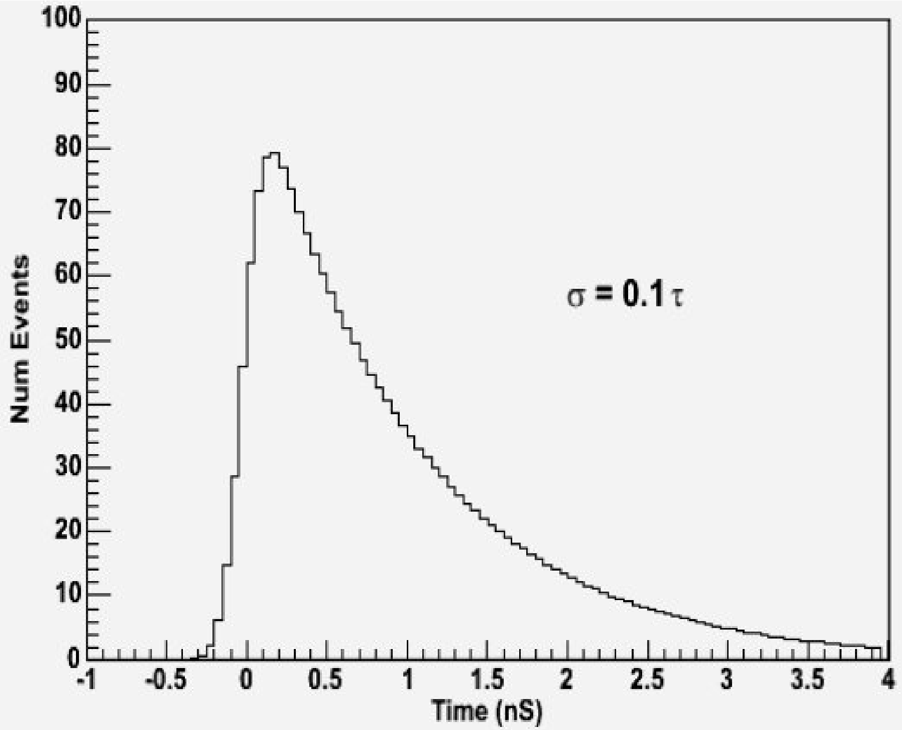
- if not we implicitly assume them to be the same
- mass is fine but proper time is not: looks different for signal and background
- add $P_{\text{sig}}(\Delta t)$ and $P_{\text{bg}}(\Delta t)$ factors to the two components, need a template for this (not needed for your assignment)

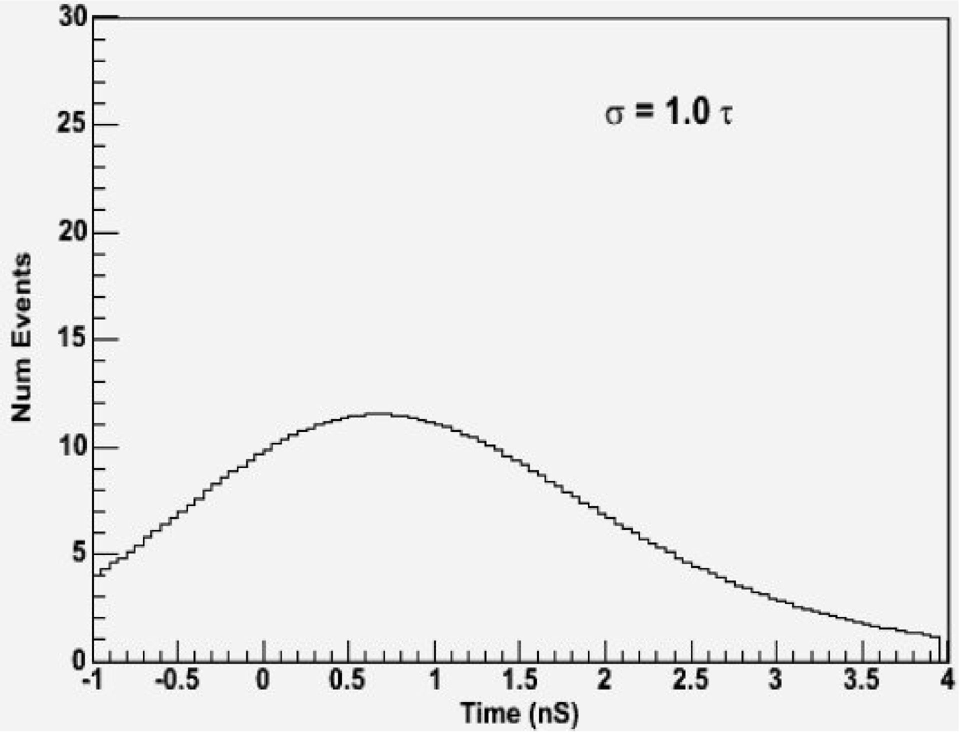
Full Likelihood for our Analysis

Including detector imperfections

$$P(t|\tau) = \frac{1}{\sqrt{2\pi}\sigma\tau}\int_0^\infty dt'\,\exp(-t/\tau)\,\exp\left(-\frac{\sigma^2}{2\tau^2}\right)\exp\left(-\frac{(t'-(t-\sigma^2/\tau)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\tau}\exp(-t/\tau)\,\exp(-\frac{\sigma_2}{2\tau^2})\frac{1}{\pi}\mathrm{erfc}\left(-\frac{1}{\sqrt{2}}\left(\frac{t}{\sigma}-\frac{\sigma}{\tau}\right)\right)$$



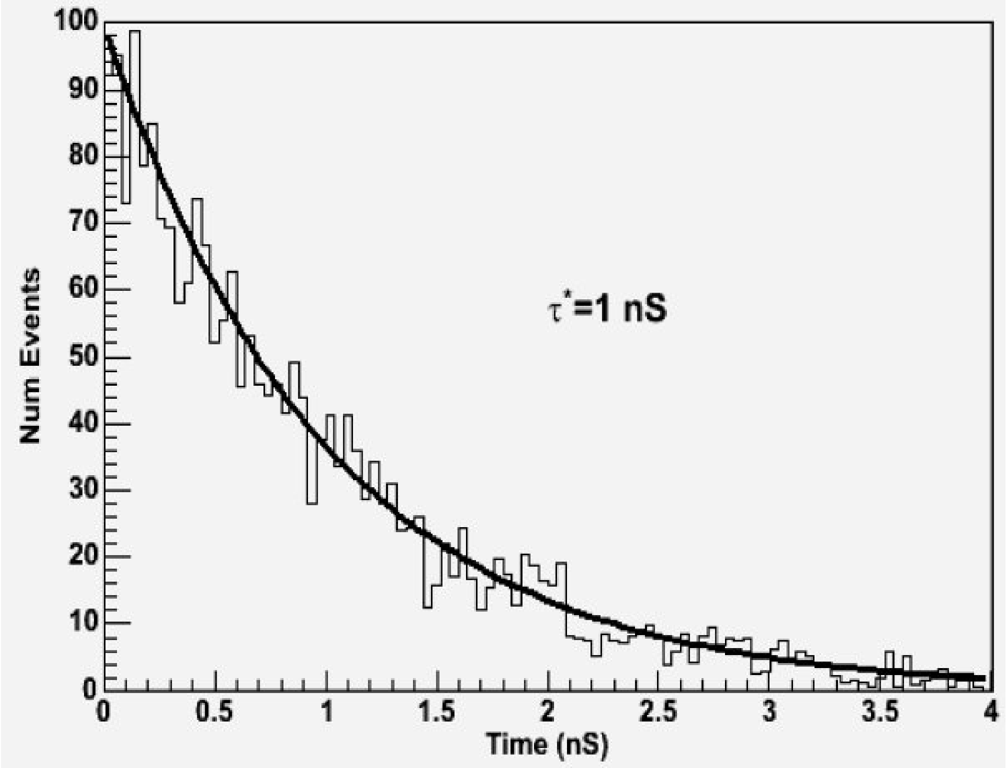


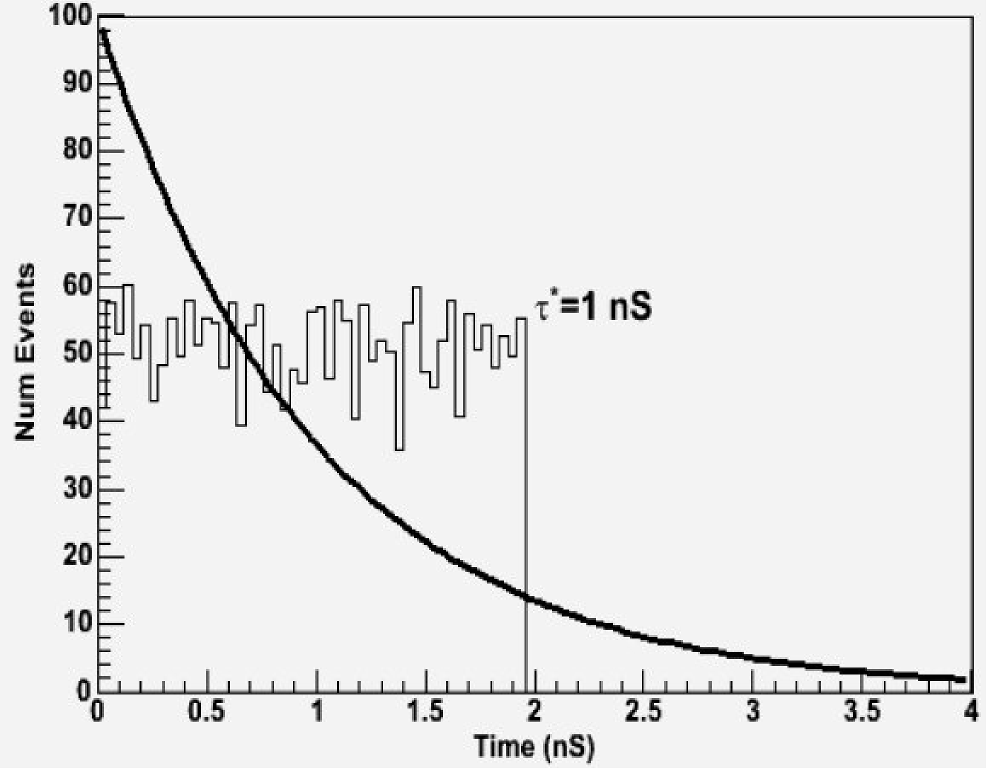
Goodness of Fit

Least square fit tells explicitly probability of fit

- *P* = TMath::Prob(Chi2Min, nDoF)
- comes out very close to zero or one? something is wrong!

Minimum likelihood does not work like that





Average lifetime of samples the same → max. likelihood the same

Goodness of Fit

How to get a goodness of fit from max. likelihood?

- general answer: statisticians are still writing papers about it! → no unique and fully accepted answer

Let's take a physicist's approach

- need a chi2 like quantity for all observables
- make sure that they have reasonable probabilities
- need: histograms with data and theory curve
  - data looks simple, we got that but need to find binning so we have enough events to apply Gaussian statistics
  - not easy: each event has potentially different theory curve!
  - sum up full theory curve for all events
- take $\chi 2$ value to determine probability for the picture
- nDoF= (number of bins – number of parameters in picture)

Testing for Biases

Likelihood fits often are complex and very difficult to implement correctly

- test for fitting bias is absolutely essential
- in some cases biases cannot be completely avoided

How to safe yourself from trouble?

- toy Monte Carlo is the answer
  - implement a toy in which you generate data exactly according to your implemented model
  - generating a large number of toy experiments should give you on average exactly the correct answer (the lifetime you put in)
- uncertainties should be as expected from statistics
- this means, the pulls $((p-p_{\mathrm{fit},i})/\Delta p_{\mathrm{fit},i})^2$ are Gaussian with
  - mean equals 0, within uncertainties
  - width equals 1, within uncertainties

Testing for Biases

Example (updated: fitCTauBuJpsiK.C) provided at

- ~paus/8.882/614/MixFit/scripts/fitCTauBuJpsiK.C

Sequence to perform pulls

- perform the fit and store the results to re-initialize your fit with them when you start (iMode = iModeFit)
- edit the initialization values by hand:

```
gF->AddParameter(new Parameter("buCTau"    ,+0.047412,  0.001, 0.100,0.001));
```

initial value

- now generate data for about 100 experiments, setting the number of events equal to the events in your input file (iMode = iModePulls), check iModePulls in the script... most of the parameters are self explanatory

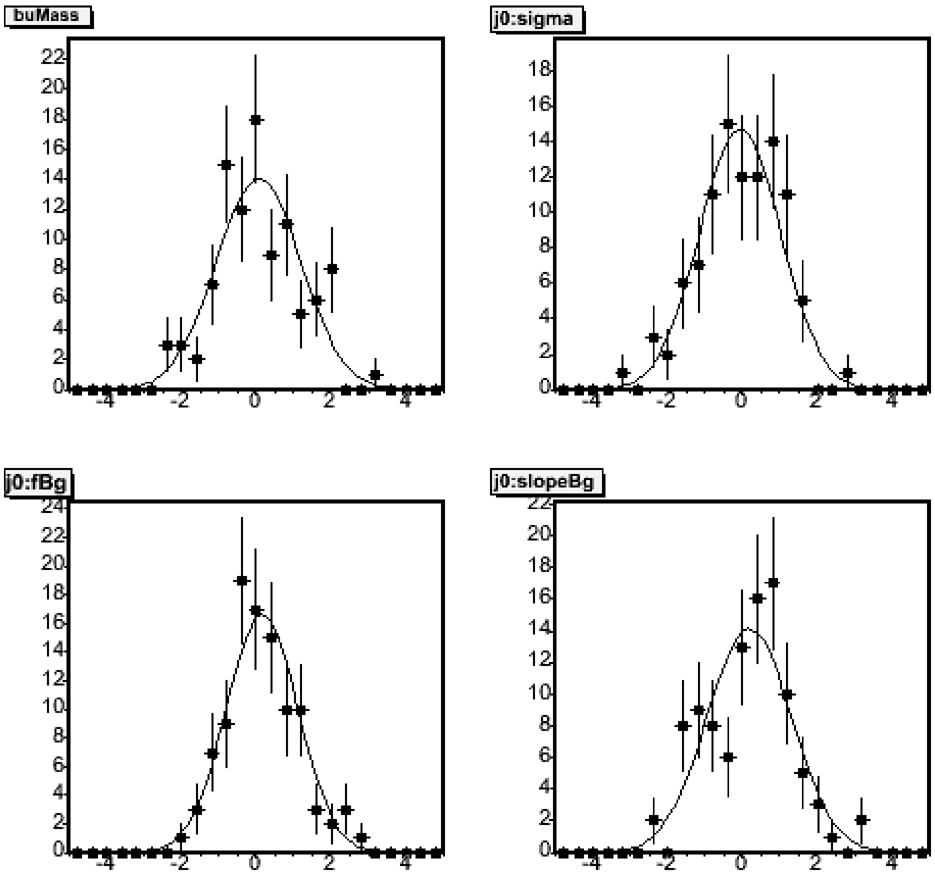
Testing for Biases

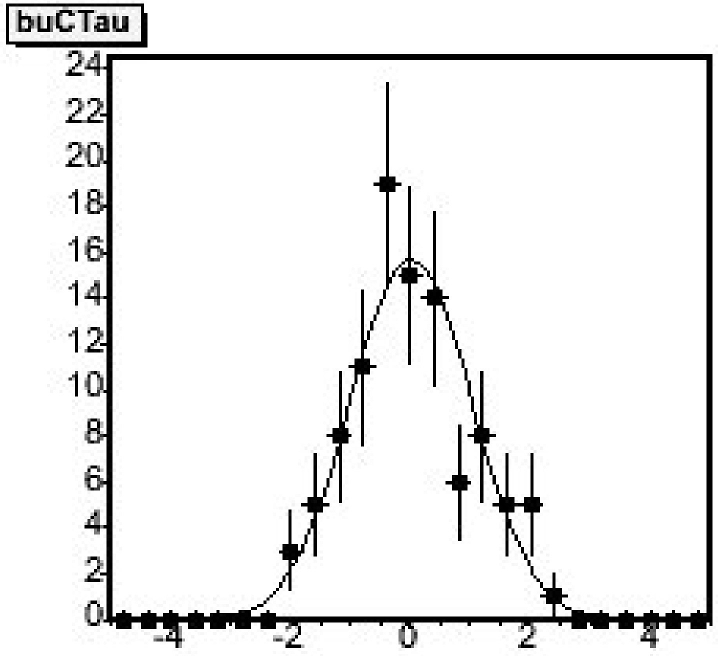
Sample plots of biases for limited set of parameters

- to exclude parameter from fit or pull just set the initial step size to zero

```
gF->AddParameter(new Parameter("buCTau"    ,+0.047412,  0.001, 0.100,0.001));
```

initial step size



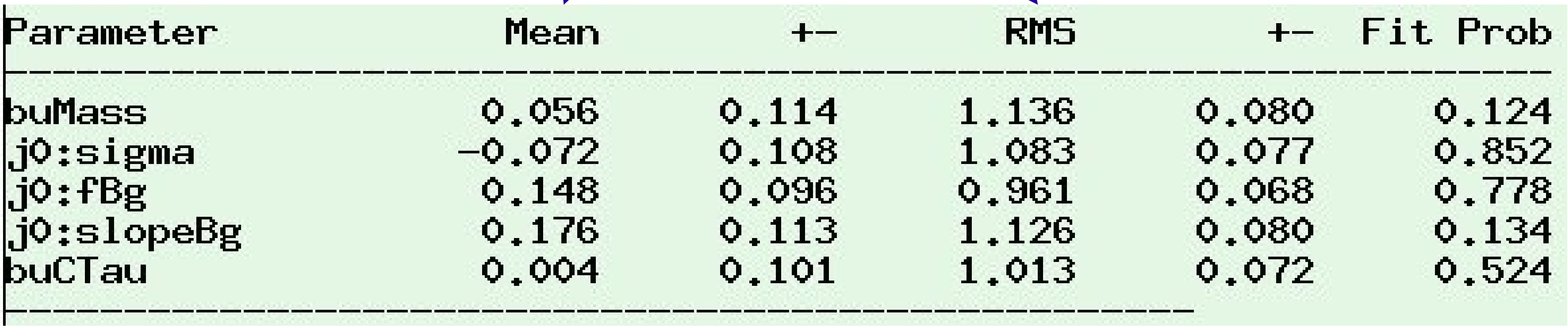


Testing for Biases

Parameter biases in numbers

- have to be consistent with the pictures....
- well, they are consistent with a unit Gaussian

$$G(x|\mu = 0, \sigma = 1)$$



```
Parameter                   Mean          +-          RMS          +-   Fit Prob
--------------------------------------------------------------------------------
buMass                     0.056       0.114        1.136       0.080      0.124
j0:sigma                  -0.072       0.108        1.083       0.077      0.852
j0:fBg                     0.148       0.096        0.961       0.068      0.778
j0:slopeBg                 0.176       0.113        1.126       0.080      0.134
buCTau                     0.004       0.101        1.013       0.072      0.524
-------------------------------------------------------------
```

Optimizing Your Selection

Step 1: determine cut variables

- make sure they do not bias *ct* distribution
- choose independent variable: *ex.* χ^2 and prob. not useful
- suggestions: p_T, χ^2, limited *z* range, ...

Step 2: determine quality criterion

- ultimately the full uncertainty of a lifetime fit
- less spectacular: statistical uncertainty of lifetime fit
- just: $n_S/\sqrt{n_S + n_B}$

Step 3: prepare grid for variables

- p_T in 100 MeV steps?

Step 4: find best selection values, quality maximal

- step through grid and find cuts for which quality is best

Optimizing Your Selection

For the quality criterion: $n_S/\sqrt{n_S + n_B}$

- determine relevant events in tight mass window around peak: up to 3 standard deviations, maybe less
- number of background events from sideband extrapolation
- number of signal events from the Monte Carlo
  - using data will cause biases towards statistical fluctuations
  - make sure to apply the same mass window and fitting form to guarantee identical procedure

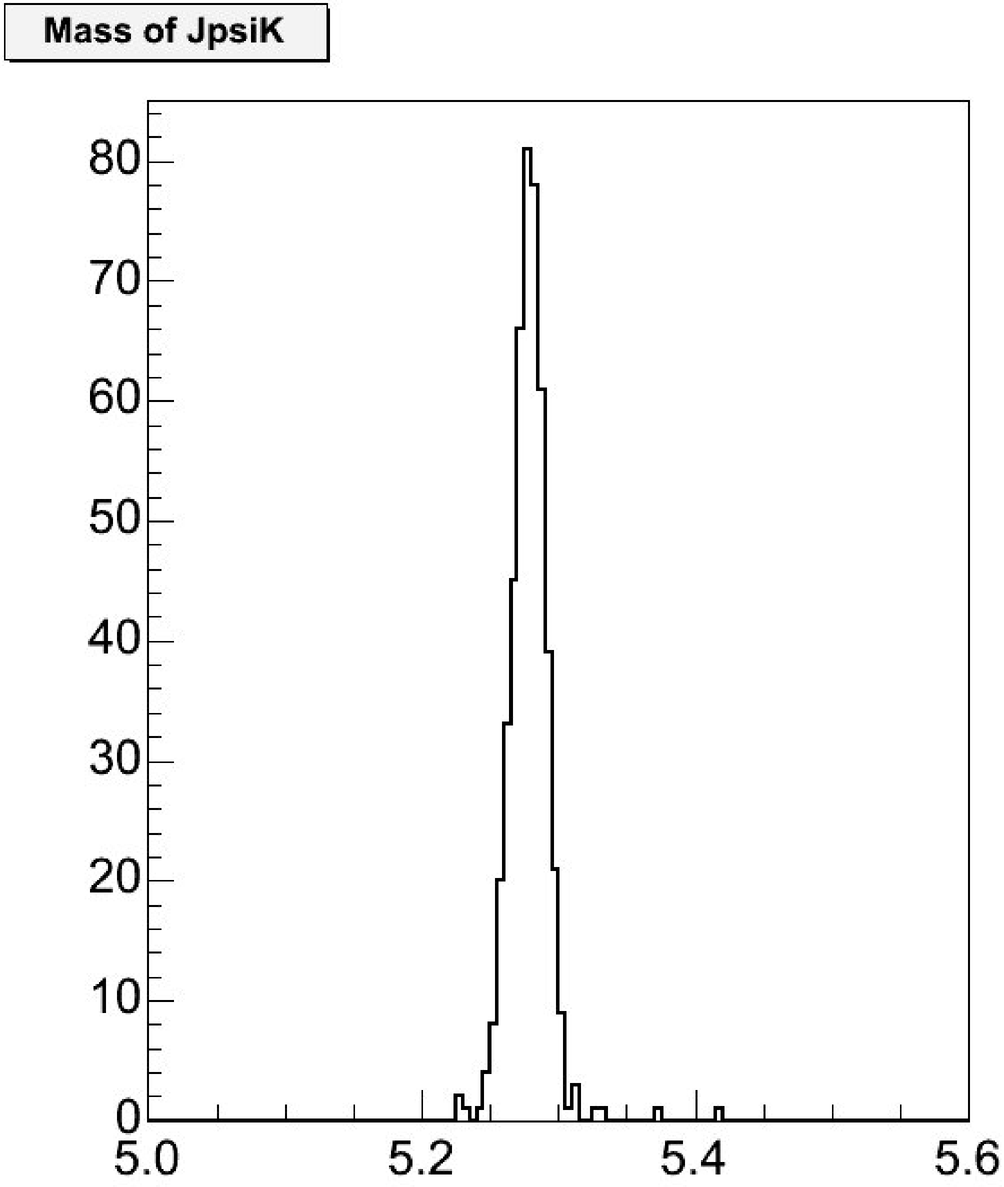
In general

- lifetime measurements independent of instantaneous luminosity (second order effects, different background?)
- if data and Monte Carlo disagree it is not a disaster as for cross section, you might just not get the optimal point

Optimizing Your Selection

Produced fresh MC

- runs very quickly, minimal changes to analysis script
- book:
  - “skims/bjps-71/bujk__0001-00”
- dataset:
  - “B-JpsiK”
- check:
  - runJpsiK.C
  - in: ~paus/8.882/614/Ana/scripts/
- switch: isMc selects between data and MC



Conclusion

Likelihood fits and all that

- uncertainties from likelihood are simple:
  - minimize [-2 log(*L*)] and treat it like χ2 (this is what TMinuit does)
  - correspondingly all intervals
- goodness of fit is an unsolved problem
  - ball is in the users court
  - perform test: toy studies to remove biases, projections to compare relevant variables

Selection and its optimization

- avoid bias on *ct* or think about it carefully
- optimization: use MC to calculate signal events, data fro background

Hand-in of project 3 (lifetime) during next week!

Next Lectures

Neural networks and data driven techniques

Higgs Searches and Other Essentials

- guest lecturer being identified fo rthe Higgs lecture?!
- overview over the High p_T physics and searches in particular