Statistical Methods for Experimental Particle Physics

Theory and Lots of Examples

Thomas R. Junk
Fermilab

TRIUMF Summer Institute
July 20 - 31, 2009

Day 1: Introduction
Binomial, Poisson,
and Gaussian Distributions
Making Measurements

Useful Reading Material

Particle Data Group reviews on Probability and Statistics. http://pdg.lbl.gov

Frederick James, “Statistical Methods in Experimental Physics”, 2nd edition, World Scientific, 2006

Louis Lyons, “Statistics for Nuclear and Particle Physicists” Cambridge U. Press, 1989

Glen Cowan, “Statistical Data Analysis” Oxford Science Publishing, 1998

Roger Barlow, “Statistics, A guide to the Use of Statistical Methods in the Physical Sciences”, (Manchester Physics Series) 2008.

Bob Cousins, “Why Isn’t Every Physicist a Bayesian” Am. J. Phys **63**, 398 (1995).

http://www.physics.ox.ac.uk/phystat05/
http://www-conf.slac.stanford.edu/phystat2003/
http://conferences.fnal.gov/cl2k/

So Why Study Probability and Statistics?

- It’s how to be a scientist! The basic steps:
  - Ask a question about Nature
  - Formulate hypotheses to test
  - Design an Experiment
  - Build the Experiment
  - Run the Experiment : Count collision events
  - Make statements about the hypotheses
  - Publish

Can be contentious!

In practice, these steps are often done in many different orders, sometimes in parallel, often repeatedly.

- Detector and accelerator upgrades to do better physics
- New topics become interesting with more data
- New ideas from theorists and other experiments

Figures of Merit

Our jobs as scientists are to

- **Measure quantities as precisely as we can**
  Figure of merit: the **uncertainty** on the measurement

- **Discover new particles and phenomena**
  Figure of merit: the **significance** of evidence or observation -- try to be first!
  Related: the **limit** on a new process

To be counterbalanced by:

- Integrity: All sources of systematic uncertainty must be included in the interpretation.
- Large collaborations and peer review help to identify and assess systematic uncertainty

Figures of Merit

Our jobs as scientists are to

- **Measure quantities as precisely as we can**
  Figure of merit: the **expected** uncertainty on the measurement

- **Discover new particles and phenomena**
  Figure of merit: the **expected** significance of evidence or observation -- try to be first!
  Related: the **expected** limit on a new process

To be counterbalanced by:

- Integrity: All sources of systematic uncertainty must be included in the interpretation.
- Large collaborations and peer review help to identify and assess systematic uncertainty

Expected Sensitivity is used in Experiment and Analysis Design

Probability and Statistics

Statistics is largely the inverse problem of Probability

Probability: Know parameters of the theory → Predict distributions of possible experiment outcomes

Statistics: Know the outcome of an experiment → Extract information about the parameters and/or the theory

Probability is the easier of the two -- solid mathematical arguments can be made.

Statistics is what we need as scientists. Much work done in the 20th century by statisticians.

Experimental particle physicists rediscovered much of that work in the last two decades.

In HEP we often have complex issues because we know so much about our data and need to incorporate all of what we know

Some Probability Distributions useful in HEP

Binomial:

Given a repeated set of *N* trials, each of which has probability *p* of “success” and 1 - *p* of “failure”, what is the distribution of the number of successes if the *N* trials are repeated over and over?

$$\mathrm{Binom}(k\,|\,N,p)=\binom{N}{k}p^k(1-p)^{N-k},\quad \sigma(k)=\sqrt{\mathrm{Var}(k)}=\sqrt{Np(1-p)}$$

k is the number of “success” trials

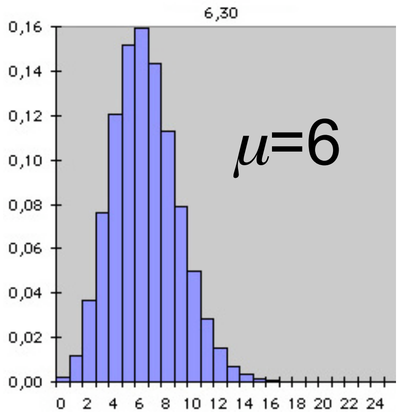
Example: events passing a selection cut, with a fixed total *N*

Some Probability Distributions useful in HEP

Poisson:

Limit of Binomial when $N \rightarrow \infty$ and $p \rightarrow 0$ with $Np = \mu$ finite

$$\mathrm{Poiss}(k \mid \mu) = \frac{e^{-\mu}\mu^{k}}{k!} \qquad \sigma(k) = \sqrt{\mu}$$



Normalized to unit area in two different senses

$$\sum_{k=0}^{\infty} \mathrm{Poiss}(k \mid \mu) = 1, \quad \forall \mu$$

$$\int_{0}^{\infty} \mathrm{Poiss}(k \mid \mu)\, d\mu = 1 \quad \forall k$$

All counting results in HEP are assumed to be Poisson distributed

Binomial is formally more correct since the number of bunch crossings and particles per bunch are finite -- but very large).

Composition of Poisson and Binomial Distributions

Example: Efficiency of a cut, say lepton p_T in leptonic *W* decay events at the Tevatron

Total number of *W* bosons: *N* -- Poisson distributed with mean μ

The number passing the lepton p_T cut: *k*

Repeat the experiment many times. ***Condition*** on *N* (that is, insist *N* is the same and discard all other trials with differnet *N*. Or just stop taking data).

p(*k*) = Binom(*k*|*N*,ε) where ε is the efficiency of the cut

Composition of Poisson and Binomial Distributions

But the number of *W* events passing the cut is just another counting experiment -- it must be Poisson distributed.

But that means no longer conditioning on *N* ($\mu = \sigma L$)

$$\mathrm{Poiss}(k \mid \varepsilon\sigma L) = \sum_{N=0}^{\infty} \mathrm{Binom}(k \mid N,\varepsilon)\mathrm{Poiss}(N \mid \sigma L)$$

A more general rule: The law of conditional probability

P(A and B) = P(A|B)P(B) = P(B|A)P(A) more on this one later

And in general, $$P(A) = \sum_{B} P(A \mid B)P(B)$$

Joint Probability of Two Poisson Distributed Numbers

Example -- two bins of a histogram
Or -- Monday's data and Tuesday's data

$$\mathrm{Poiss}(x \mid \mu) \times \mathrm{Poiss}(y \mid \nu) = \mathrm{Poiss}(x + y \mid \mu + \nu) \times \mathrm{Binom}\left(x \mid x + y, \frac{\mu}{\mu + \nu} \right)$$

The sum of two Poisson-distributed numbers is Poisson-distributed with the sum of the means

$$\sum_{k=0}^{n} \mathrm{Poiss}(k \mid \mu)\mathrm{Poiss}(n - k \mid \nu) = \mathrm{Poiss}(n \mid \mu + \nu)$$

Application: You can rebin a histogram and the contents of each bin will still be Poisson distributed (just with different means)

Question: How about the difference of Poisson-distributed variables?

Application to a test of Poisson Ratios

Our composition formula from the previous page:

$$\mathrm{Poiss}(x \mid \mu) \times \mathrm{Poiss}(y \mid \nu) = \mathrm{Poiss}(x + y \mid \mu + \nu) \times \mathrm{Binom}\left(x \mid x + y, \frac{\mu}{\mu + \nu} \right)$$

Say you have n_s in the “signal” region of a search, and n_c in a “control” region -- example: peak and sidebands

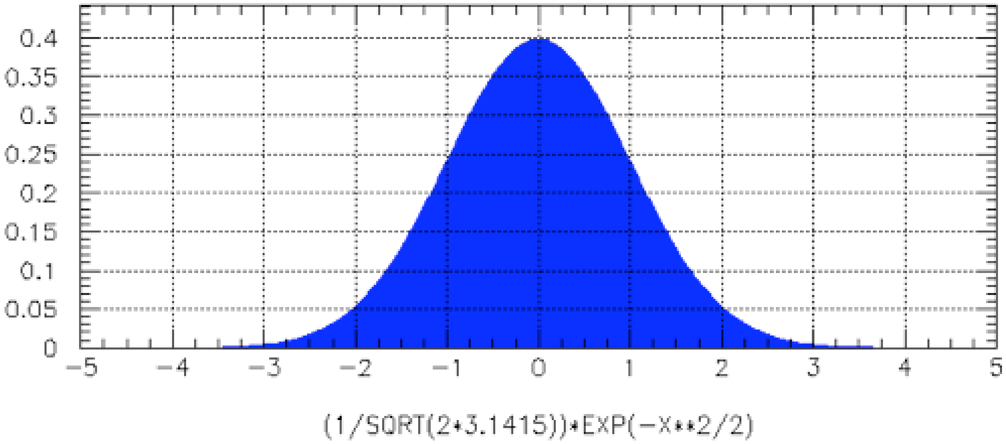
n_s is distributed as Poiss(s+b)
n_c is distributed as Poiss(τb)

Suppose we want to test H_0: s=0. Then $n_s/(n_s+n_c)$ is a Binomial variable that measures $1/(1+\tau)$

Another Probability Distribution useful in HEP

Gaussian:

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



It's a parabola on a log scale.

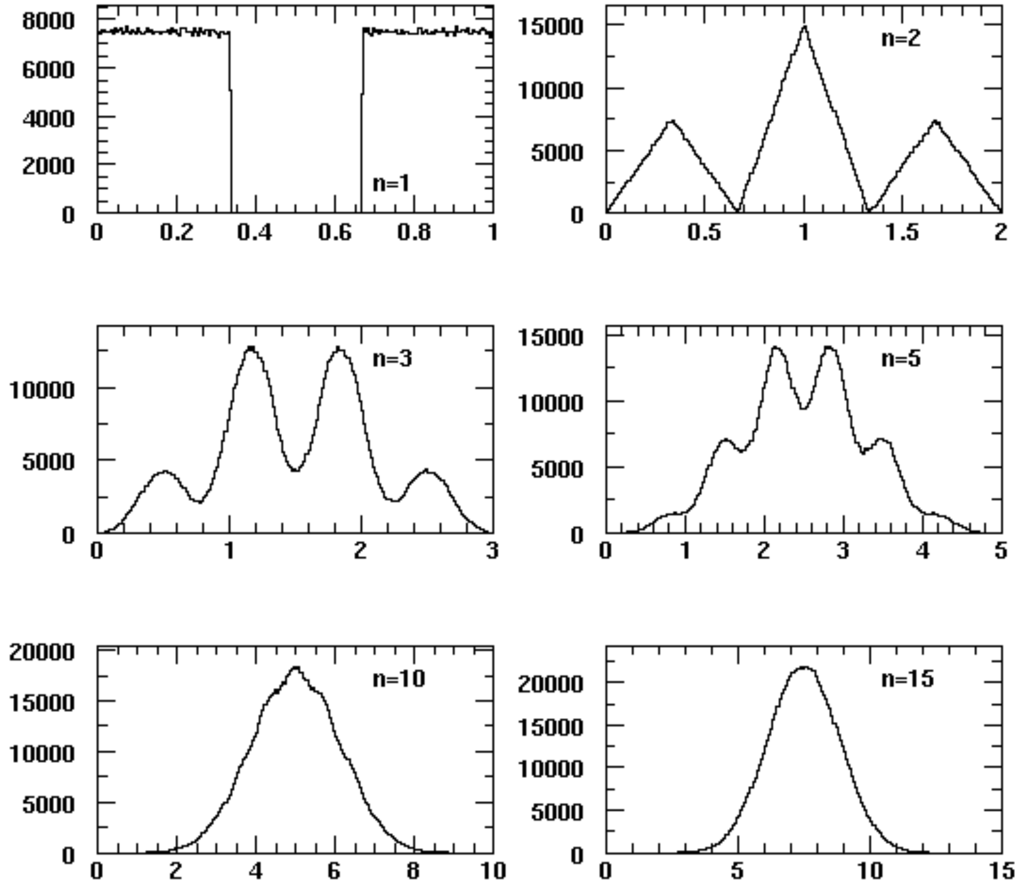
Sum of Two Independent Gaussian Distributed Numbers is Gaussian with the sum of the means and the sum in quadrature of the widths

$$\mathrm{Gauss}\left(z,\mu+\nu,\sqrt{\sigma_x^2+\sigma_y^2}\right) = \int_{-\infty}^{\infty} \mathrm{Gauss}(x,\mu,\sigma_x)\mathrm{Gauss}(z-x,\nu,\sigma_y)dx$$

A difference of independent Gaussian-distributed numbers is also Gaussian distributed (widths still *add* in quadrature)

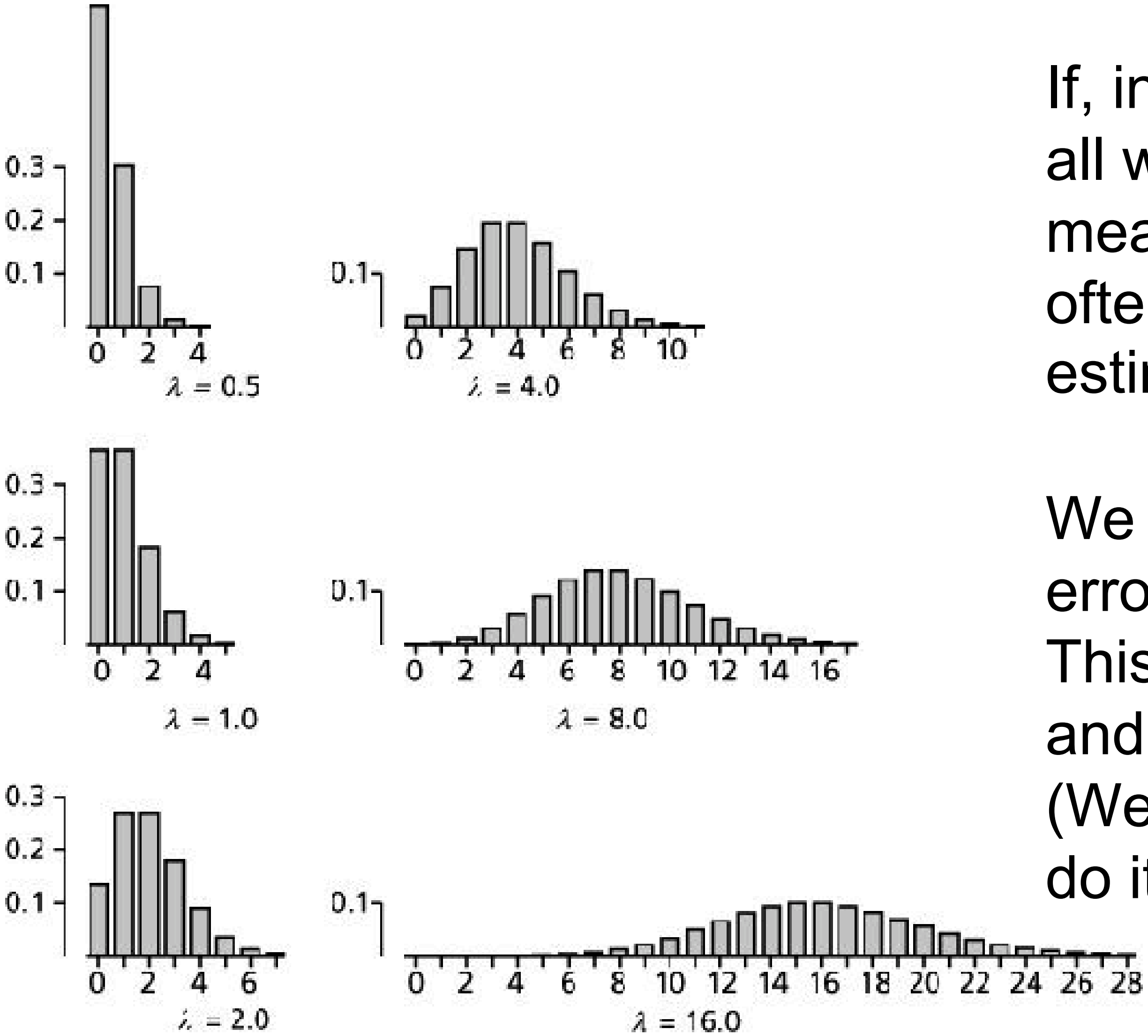
The Central Limit Theorem

The sum of many small, uncorrelated random numbers is asymptotically Gaussian distributed -- and gets more so as you add more random numbers in. Independent of the distributions of the random numbers (as long as they stay small).



Poisson for large μ is Approximately Gaussian of width

$$\sigma = \sqrt{\mu}$$



If, in an experiment all we have is a measurement n, we often use that to estimate μ.

We then draw $\sqrt{n}$ error bars on the data. This is just a *convention*, and can be misleading. (We still recommend you do it, however)

Why Put Error Bars on the Data?

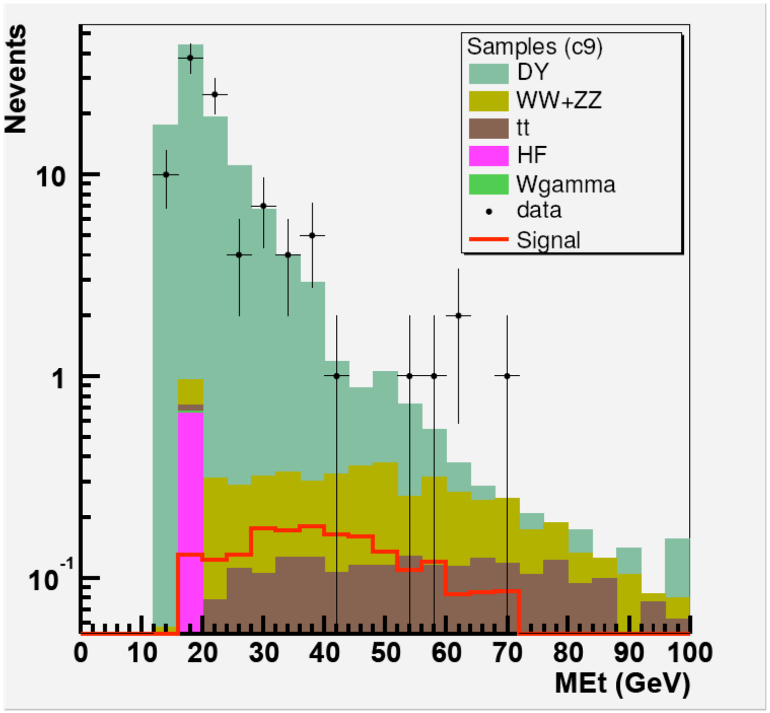
- To identify the data to people who are used to seeing it this way

- To give people an idea of how many data counts are in a bin when they are scaled (esp. on a logarithmic plot).

- So you don't have to explain yourself when you do something different (better)

But: $\sqrt{n} \neq \sqrt{\mu}$

The true value of μ is usually unknown

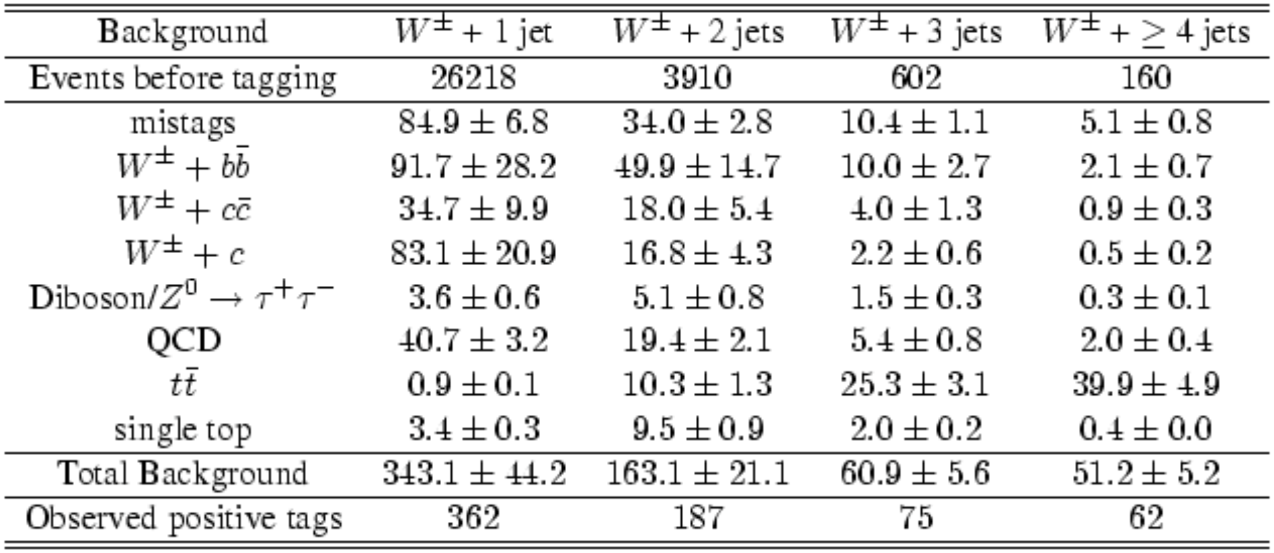


Aside: Errors on the Data? (ans: no)

Standard to make histograms with no errors: MC model points with error bars $n_{\rm obs} \pm \sqrt{n_{\rm obs}}$

But we are not uncertain of $n_{\rm obs}$! We are only uncertain about how to interpret our observations; we know how to count.

Correct presentation of data and predictions



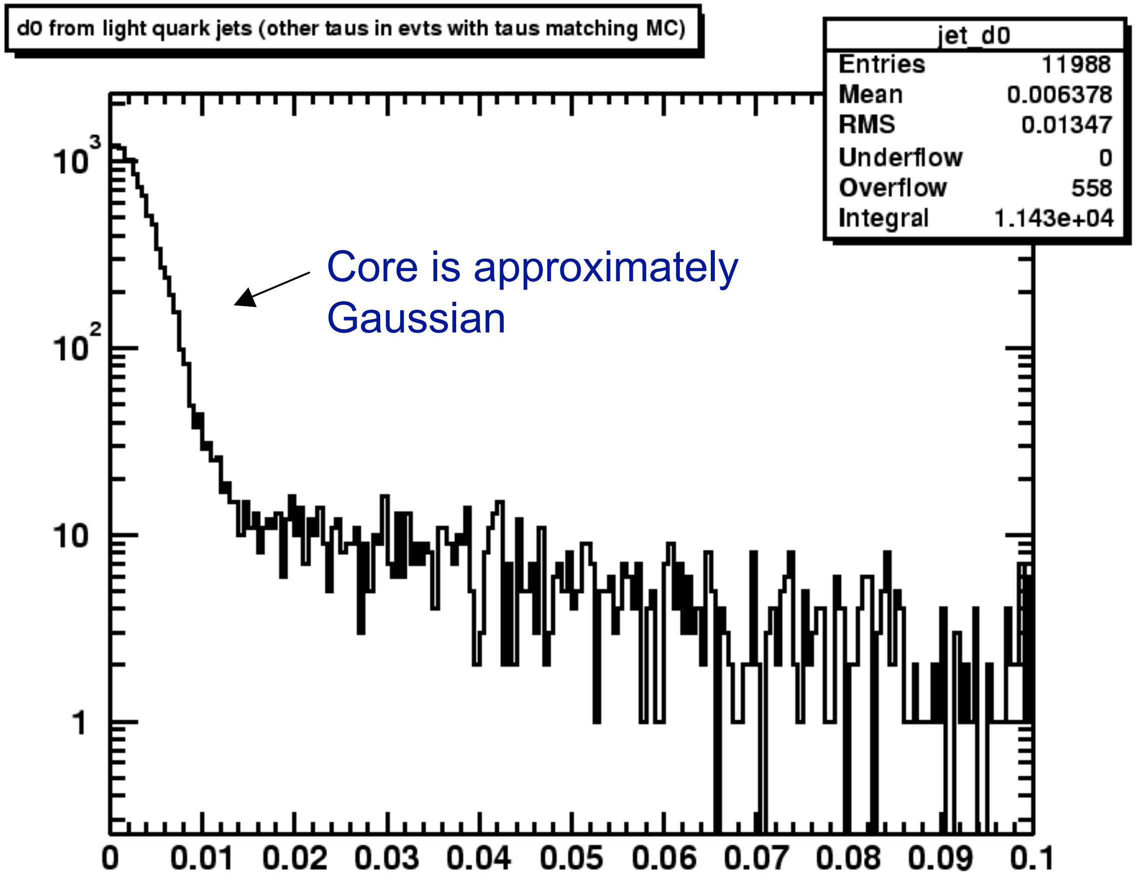
Background	$W^\pm + 1$ jet	$W^\pm + 2$ jets	$W^\pm + 3$ jets	$W^\pm + \geq 4$ jets
Events before tagging	26218	3910	602	160
mistags	84.9 ± 6.8	34.0 ± 2.8	10.4 ± 1.1	5.1 ± 0.8
$W^\pm + b\bar{b}$	91.7 ± 28.2	49.9 ± 14.7	10.0 ± 2.7	2.1 ± 0.7
$W^\pm + c\bar{c}$	34.7 ± 9.9	18.0 ± 5.4	4.0 ± 1.3	0.9 ± 0.3
$W^\pm + c$	83.1 ± 20.9	16.8 ± 4.3	2.2 ± 0.6	0.5 ± 0.2
Diboson/$Z^0 \to \tau^+\tau^-$	3.6 ± 0.6	5.1 ± 0.8	1.5 ± 0.3	0.3 ± 0.1
QCD	40.7 ± 3.2	19.4 ± 2.1	5.4 ± 0.8	2.0 ± 0.4
$t\bar{t}$	0.9 ± 0.1	10.3 ± 1.3	25.3 ± 3.1	39.9 ± 4.9
single top	3.4 ± 0.3	9.5 ± 0.9	2.0 ± 0.2	0.4 ± 0.0
Total Background	343.1 ± 44.2	163.1 ± 21.1	60.9 ± 5.6	51.2 ± 5.2
Observed positive tags	362	187	75	62

Table 10: The number of observed positive tagged events and the background summary for an integrated luminosity of 161.6 pb^{-1} for CEM and CMUP and 305.2 pb^{-1} for CMX.

Not all Distributions are Gaussian

Track impact parameter distribution for example

Multiple scattering -- core: Gaussian; rare large scatters; heavy flavor, nuclear interactions, decays (taus in this example)



Core is approximately Gaussian

"All models are false. Some models are useful."

Different Meanings of the Idea "Statistical Uncertainty"

- Repeating the experiment, how much would we expect the answer to fluctuate?

  -- approximate, Gaussian

- What interval contains 68% of our belief in the parameter(s)?

  Bayesian credibility intervals

- What construction method yields intervals containing the true value 68% of the time?

  Frequentist confidence intervals

In the limit that all distributions are symmetric Gaussians, these look like each other. We will be more precise later.

Why Uncertainties Add in Quadrature

Common situation -- a prediction is a sum of uncertain components, or a measured parameter is a sum of data with a random error, and an uncertain prediction

“statistical” “systematic”

e.g., Cross-Section = (Data-Background)/(A*ε*Luminosity) where Background, Acceptance and Luminosity are obtained somehow from other measurements and models.

Probability distribution of a sum of Gaussian distributed random numbers is Gaussian with a sum of means and a sum of variances.

$$\int_{-\infty}^{\infty} g(x,\mu_1,\sigma_1)g(x'-x,\mu_2,\sigma_2)dx = g(x',\mu_1+\mu_2,\sqrt{\sigma_1^2+\sigma_2^2})$$

Convolution assumes variables are independent.

Statistical Uncertainty on an Average of Random Numbers Drawn from the Same Gaussian Distribution

N measurements, $x_{\mathrm{i}} \pm \sigma$ are to be averaged

$$a = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The uncertainty on the sum is $\sqrt{N\sigma^2}$

so the uncertainty on the average is $\sigma_a = \frac{\sigma}{\sqrt{N}}$

You can look up the uncertainty on the width σ
in the PDG if you measure that with the RMS of N measurements.

Uncertainties That Don't Add in Quadrature

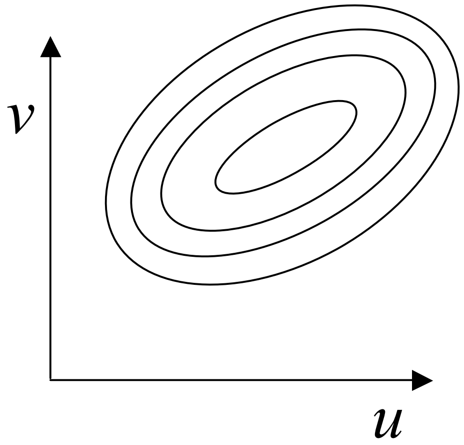
Some may be correlated! (or partially correlated). Doubling a random variable with a Gaussian distribution doubles its width instead of multiplying by $\sqrt{2}$

Example: The same luminosity uncertainty affects background prediction for many different background sources in a sum. The luminosity uncertainties all add linearly. Other uncertainties (like MC statistics) may add in quadrature or linearly.

Strategy: Make a list of independent sources of uncertainty -- these each may enter your analysis more than once. Treat each error source as independent, not each way they enter the analysis. Parameters describing the sources of uncertainty are called nuisance parameters (distinguish from parameter of interest)

Propagation of Uncertainties

Covariance: $\sigma_{uv}^2 = \langle (u-\bar{u})(v-\bar{v}) \rangle$



If

$$x = au + bv$$

then

$$\sigma_x^2 = a^2\sigma_u^2 + b^2\sigma_v^2 + 2ab\sigma_{uv}^2$$

This can even vanish! (anticorrelation)

In general, if $x = f(u,v)$

$$\sigma_x^2 = \left(\frac{\partial x}{\partial u}\right)^2 \sigma_u^2 + \left(\frac{\partial x}{\partial v}\right)^2 \sigma_v^2 + 2\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right)\sigma_{uv}^2$$

Relative and Absolute Uncertainties

If $x = auv$

then $\sigma_x^2 = a^2 v^2 \sigma_u^2 + a^2 u^2 \sigma_v^2 + 2a^2 uv\sigma_{uv}^2$

or, more easily memorized:

$$\frac{\sigma_x}{x} = \sqrt{\frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + 2\frac{\sigma_{uv}^2}{uv}}$$

“relative errors add in quadrature” for multiplicative uncertainties (but watch out for correlations!)

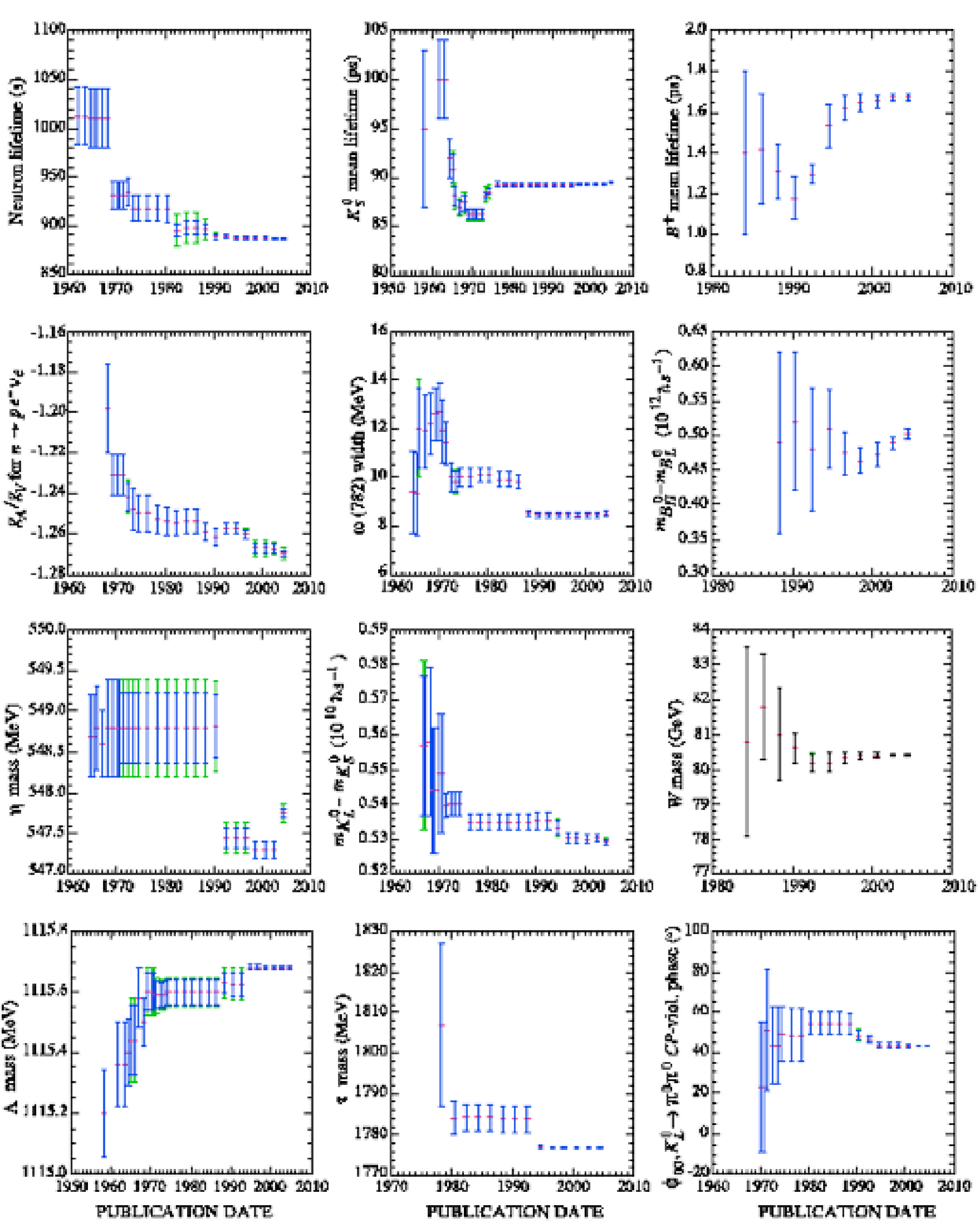
The same formula holds for division (!) but with a minus sign in the correlation term.

How Uncertainties get Used

- Measurements are inputs to other measurements -- to compute uncertainty on final answer need to know uncertainty on parts.
- Measurements are averaged or otherwise combined -- weights are given by uncertainties
- Analyses need to be optimized -- shoot for the lowest uncertainty
- Collaboration picks to publish one of several competing analyses -- decide based on sensitivity
- Laboratories/Funding agencies need to know how long to run an experiment or even whether to run.

Statistical uncertainty: scales with data. Systematic uncertanty often does too, but many components stay constant -- limits to sensitivity.

Examples from the front of the PDG



χ^2 and Goodness of Fit

For n independent Gaussian-distributed random numbers, the probability of an outcome (for known σ_i and μ_i) is given by

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} g(x_i, \mu_i, \sigma_i)$$

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(x_i-\mu_i)^2/2\sigma_i^2}$$

If we are interested in fitting a distribution (we have a model for the μ_i in each bin with some fit parameters) we can maximize *p* or equivalently minimize

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} = -2\ln p + c$$

σ_i includes stat. and syst. errors

For fixed μ_i this χ^2 has *n* degrees of freedom (DOF)

Counting Degrees of Freedom

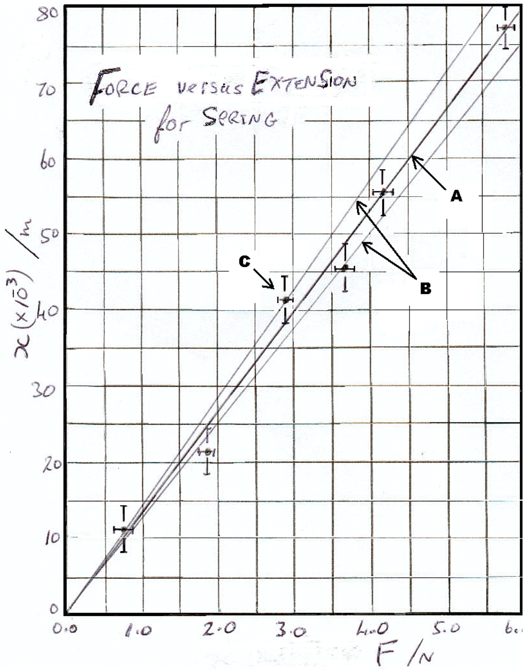
$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$ has *n* DOF for fixed μ_i and σ_i

If the μ_i are predicted by a model with free parameters (e.g. a straight line), and χ^2 is minimized over all values of the free parameters, then

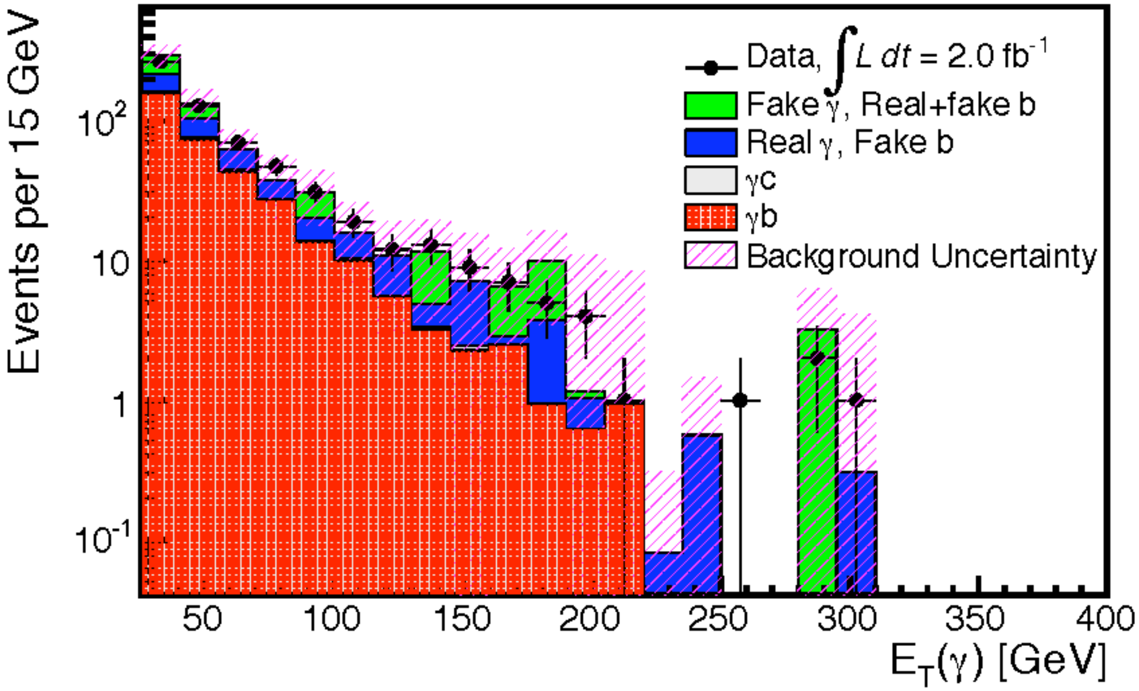
DOF = n - #free parameters in fit.

Example: Straight-line least-squares fit:
DOF = npoints - 2 (slope and intercept float)

With one constraint: intercept = 0,
6 data points, DOF = ?



MC Statistics and “Broken” Bins



NDOF=?

- Limit calculators cannot tell if the background expectation is really zero or just a downward MC fluctuation.
- Real background estimations are sums of predictions with very different weights in each MC event (or data event)
- Rebinning or just collecting the last few bins together often helps.

- Advice: Make your own visible underflow and overflow bins (do not rely on ROOT’s underflow/overflow bins -- they are usually not plotted. Limit calculators should ignore ROOT’s u/o bins).

χ^2 and Goodness of Fit

- Gaussian-distributed random numbers cluster around μ_i -- 68% within 1σ. 5% outside of 2σ. Very few outside 3 sigma.

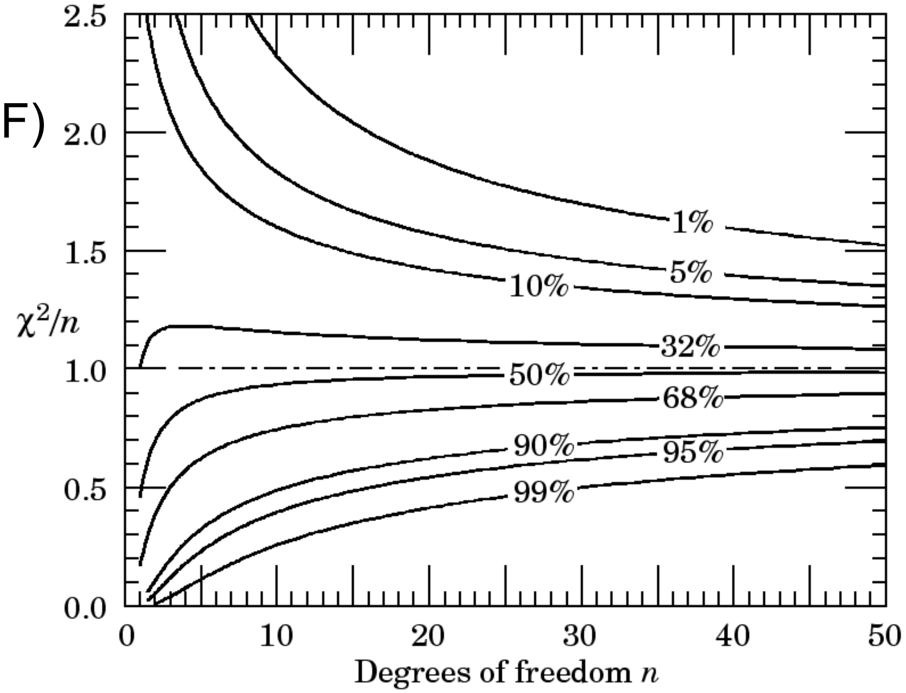
  Average contribution to χ^2 per DOF is 1. χ^2/DOF converges to 1 for large *n*

TMath::Prob(Double_t Chisquare,Int_t NDOF)

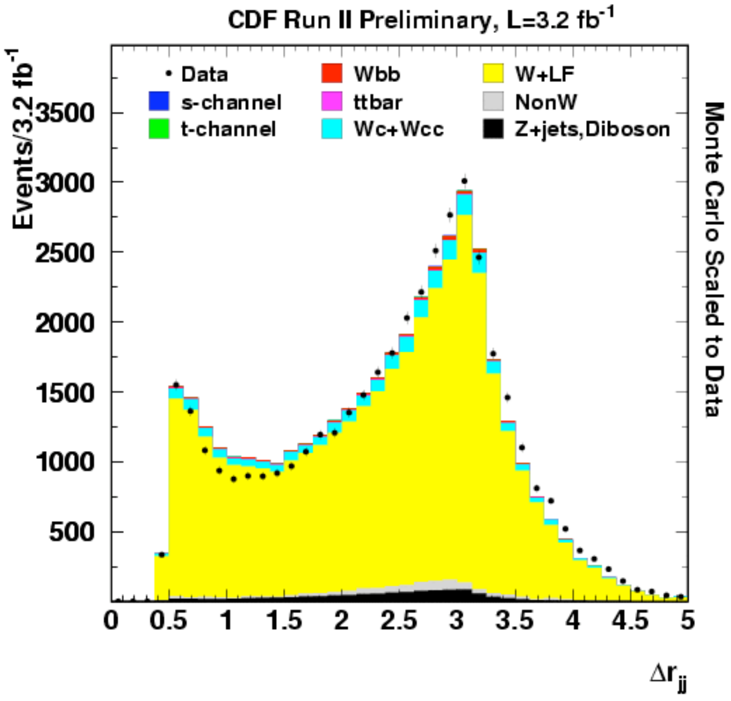
Gives the chance of seeing the value of Chisquared or bigger given NDOF.

This is a **p-value** (more on these later)

CERNLIB routine: PROB.

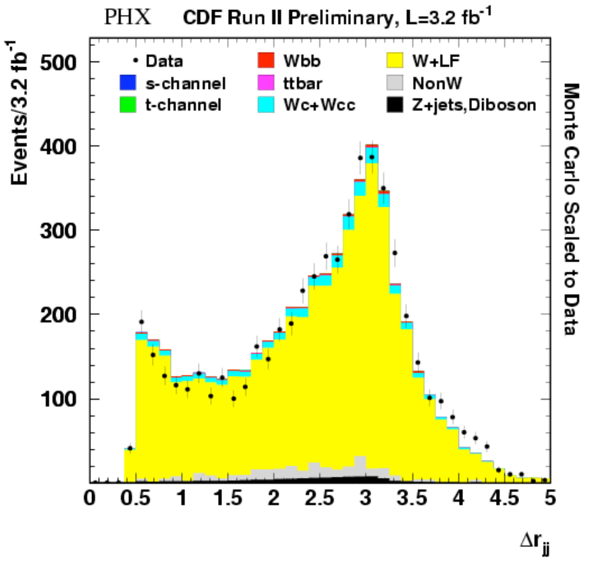


Chisquared Tests for Large Data Samples

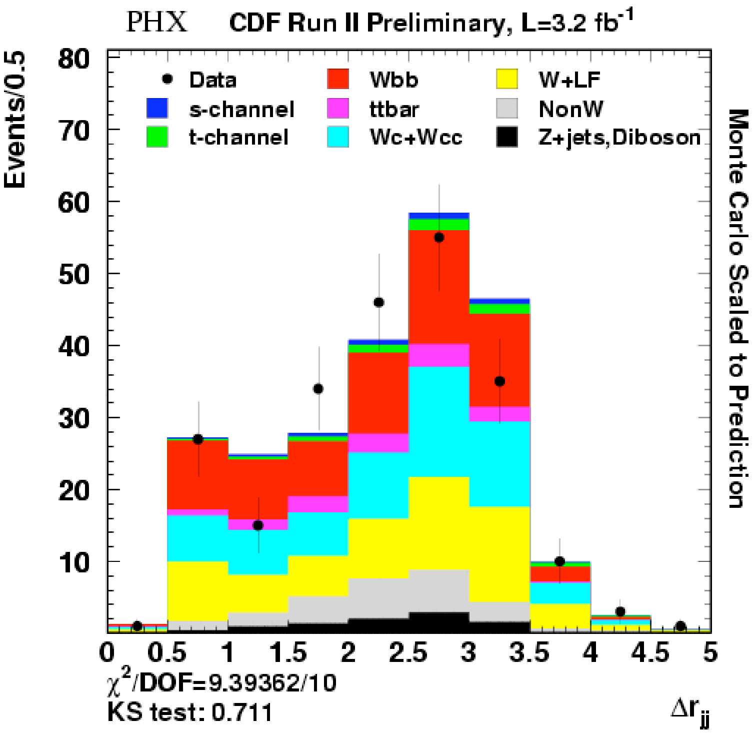


A large value of χ^2/DOF -- p-value is microscopic. We are very very sure that our model is slightly wrong. With a smaller data sample, this model would look fine (even though it is still wrong.

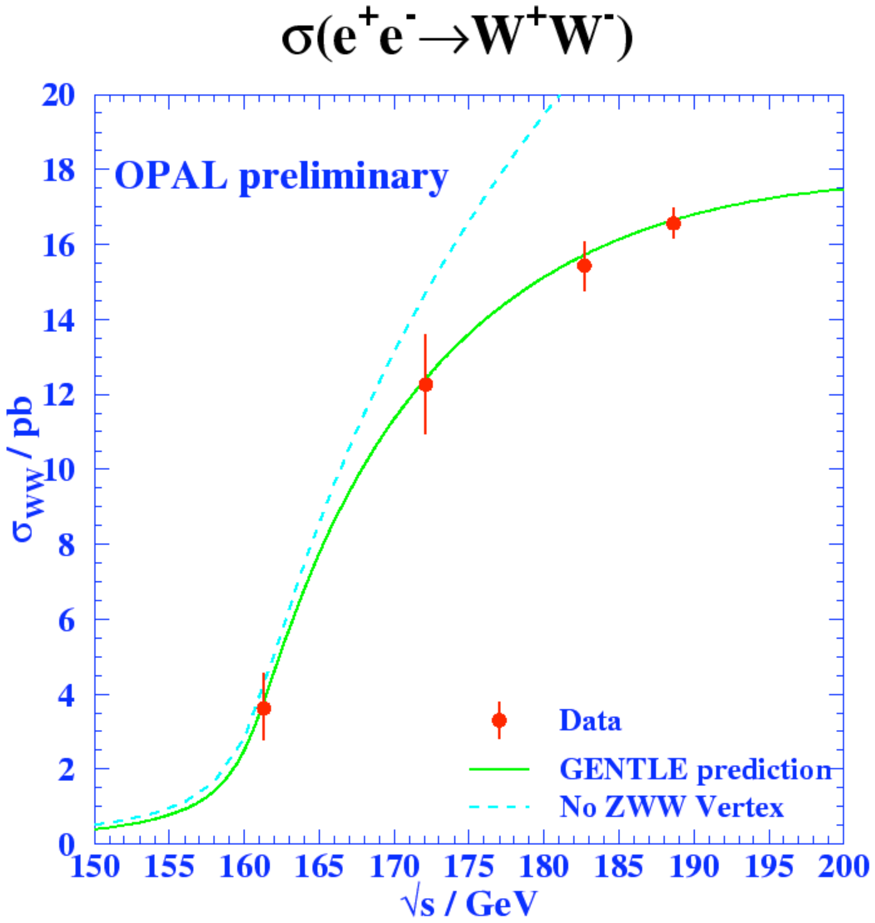
χ^2 depends on choice of binning.



Smaller data samples: harder to discern mismodeling.



χ^2 Can Sometimes be so Good as to be Suspicious



It should happen sometimes! But it is a red flag to go searching for correlated errors or overestimated errors

no free parameters in model

(happy ending: further data points increased χ^2)

Including Correlated Uncertainties in χ^2

Example with

- Two measurements $a_1 \pm u_1 \pm c_1$ and $a_2 \pm u_2 \pm c_2$ of one parameter x
- Uncorrelated errors u_1 and u_2
- Correlated errors c_1 and c_2 (same source)

$$\chi^2(x) = \sum_{i,j=1,2} (x - a_i)\mathcal{C}^{-1}_{ij}(x - a_j)$$

$$\mathcal{C} = \begin{pmatrix} u_1^2 + c_1^2 & c_1 c_2 \\ c_1 c_2 & u_2^2 + c_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

If there are several sources of correlated error c_i^p then the off-diagonal terms become $\sum_p c_1^p c_2^p$

Combining Precision Measurements with **BLUE**

$$\chi^2(x) = \sum_{i,j=1,2} (x-a_i)\mathcal{C}_{ij}^{-1}(x-a_j)$$

Procedure: Find the value of *x* which minimizes χ^2

This is a maximum likelihood fit with symmetric, Gaussian uncertainties.

Equivalent to a weighted average:

$$x_{best} = \sum_i w_i a_i \qquad \text{with} \qquad \sum_i w_i = 1$$

1 standard-deviation error from $\chi^2(x_{best}\pm\sigma_0)$-$\chi^2(x_{best})$=1

Can be extended to many measurements of the same parameter *x*.

More General Likelihood Fits

$$L = P(\mathrm{data} \mid \vec{\nu}, \vec{\theta})$$

$\vec{\nu}$: "Parameters of Interest" mass, cross-section, b.r.
$\vec{\theta}$: "Nuisance Parameters" Luminosity, acceptance, detector resolution.

Strategy -- find the values of $\vec{\theta}$ and $\vec{\nu}$ which maximize *L*

Uncertainty on parameters: Find the contours in $\vec{\nu}$ such that

ln(*L*) = ln(L_{max}) - $s^2/2$, to quote *s*-standard-deviantion intervals. Maximize L over $\vec{\theta}$ separately for each value of $\vec{\nu}$. Buzzword: "*Profiling*"

More General Likelihood Fits

Advantages:

- "Approximately unbiased"
- Usually close to optimal
- Invariant under transformation of parameters. Fit for a mass or mass^2 doesn't matter.

Unbinned likelihood fits are quite popular. Just need $L = P(\text{data} \mid \vec{\theta}, \vec{\nu})$

Warnings:

- Need to estimate what the bias is, if any.
- Monte Carlo Pseudoexperiment approach: generate lots of random fake data samples with known true values of the parameters sought, fit them, and see if the averages differ from the inputs.
- More subtle -- the uncertainties could be biased.

  -- run pseudoexperiments and histogram the "pulls" (fit-input)/error -- should get a Gaussian centered on zero with unit width, or there's bias.
- Handling of systematic uncertainties on nuisance parameters by maximization can give misleadingly small uncertainties -- need to study L for other values than just the maximum (L can be bimodal)

Example of a problem:

Using Observed Uncertainties in Combinations Instead of Expected Uncertainties

Simple case: 100% efficiency. Count events in several subsets of the data. Measure *K* times $n_i \pm \sqrt{n_i}$ each with the same integrated luminosity.

Total: $N_{tot} = \sum_{i=1}^{K} n_i$

Weighted average: (from **BLUE**)

Best average: $n_{avg} = N_{tot}/K$

$$n_{avg} = \frac{\sum_{i=1}^{K} n_i/\sigma_i^2}{\sum_{i=1}^{K} 1/\sigma_i^2} = \frac{\sum_{i=1}^{K} n_i/n_i}{\sum_{i=1}^{K} 1/n_i} = \frac{K}{\sum_{i=1}^{K} 1/n_i}$$

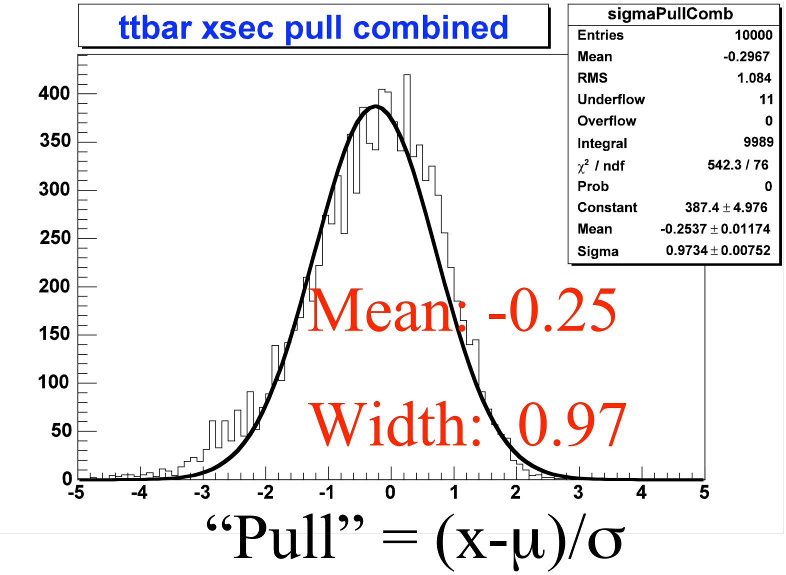
crazy behavior (especially if one of the n_i=0)

What Went Wrong?

-- low measurements have smaller uncertainties than larger measurements.

True uncertainty is the scatter in the measurements for a fixed set of true parameters

Solution: Use the expected error $\sqrt{\mu}$ for the true value of the parameter after averaging -- need to iterate!



But: Sometimes the "observed" uncertainty carries some real information! Statisticians prefer reporting "observed" uncertainties as lucky data can be more informative than unlucky data.

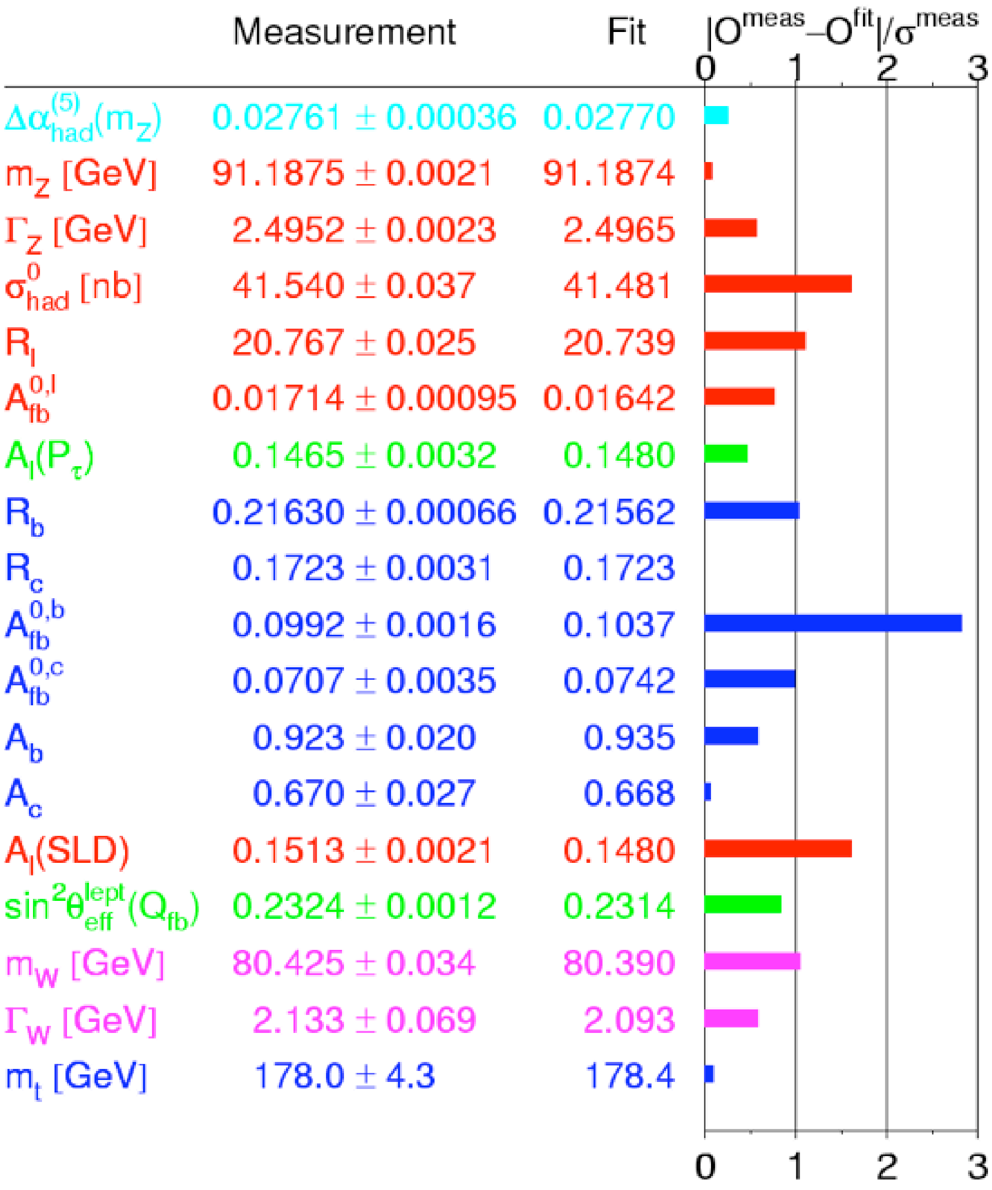
Example: Measuring M_Z from one event -- leptonic decay is better than hadronic decay.

A Prominent Example of Pulls -- Global Electroweak Fit

χ^2/DOF = 18.5/13

probability = 13.8%

Didn’t expect a 3σ result in 18 measurements, but then again, the total χ^2 is okay



	Measurement	Fit
$\Delta\alpha^{(5)}_{had}(m_Z)$	0.02761 ± 0.00036	0.02770
m_Z [GeV]	91.1875 ± 0.0021	91.1874
Γ_Z [GeV]	2.4952 ± 0.0023	2.4965
σ^0_{had} [nb]	41.540 ± 0.037	41.481
R_l	20.767 ± 0.025	20.739
$A^{0,l}_{fb}$	0.01714 ± 0.00095	0.01642
$A_l(P_\tau)$	0.1465 ± 0.0032	0.1480
R_b	0.21630 ± 0.00066	0.21562
R_c	0.1723 ± 0.0031	0.1723
$A^{0,b}_{fb}$	0.0992 ± 0.0016	0.1037
$A^{0,c}_{fb}$	0.0707 ± 0.0035	0.0742
A_b	0.923 ± 0.020	0.935
A_c	0.670 ± 0.027	0.668
A_l(SLD)	0.1513 ± 0.0021	0.1480
$\sin^2\theta^{lept}_{eff}(Q_{fb})$	0.2324 ± 0.0012	0.2314
m_W [GeV]	80.425 ± 0.034	80.390
Γ_W [GeV]	2.133 ± 0.069	2.093
m_t [GeV]	178.0 ± 4.3	178.4

Bounded Physical Region

What happens if you get a best-fit value you know can't possibly be the true?

Examples:

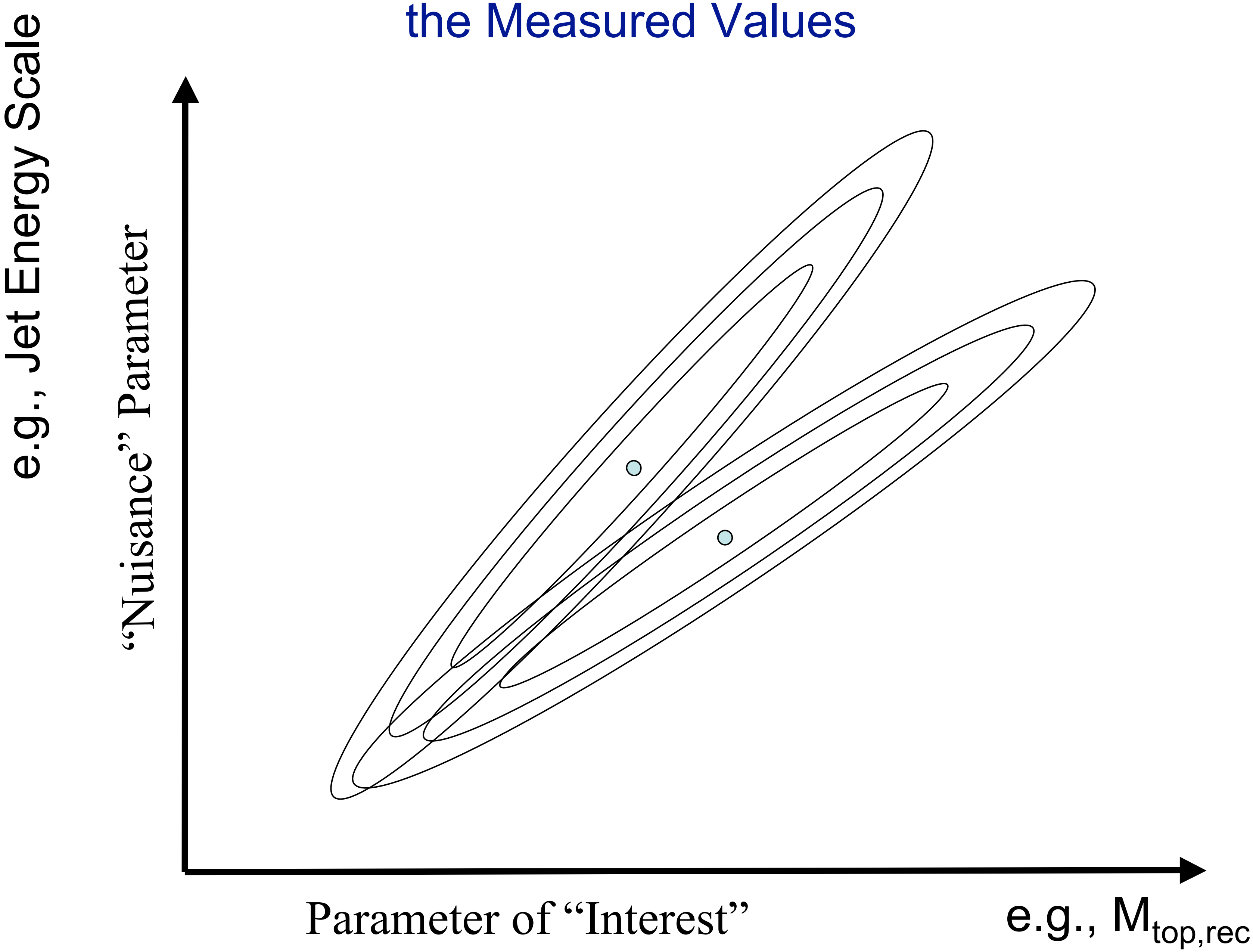
Cross Section for a signal < 0
m^2(new particle) < 0
sinθ < -1 or > +1

These measurements are important! You should report them without adjustment.
(but also some other things too)

An average of many measurements without these would be biased.
Example: Suppose the true cross section for a new process is zero.
Averaging in only positive or zero measurements will give a positive answer.

Later discussion: confidence intervals and limits -- take bounded physical regions into account. But they aren't good for averages, or any other kinds of combinations.

Odd Situation: BLUE Average of Two Measurements not Between the Measured Values


e.g., Jet Energy Scale
"Nuisance" Parameter
Parameter of "Interest"
e.g., $M_{top,rec}$