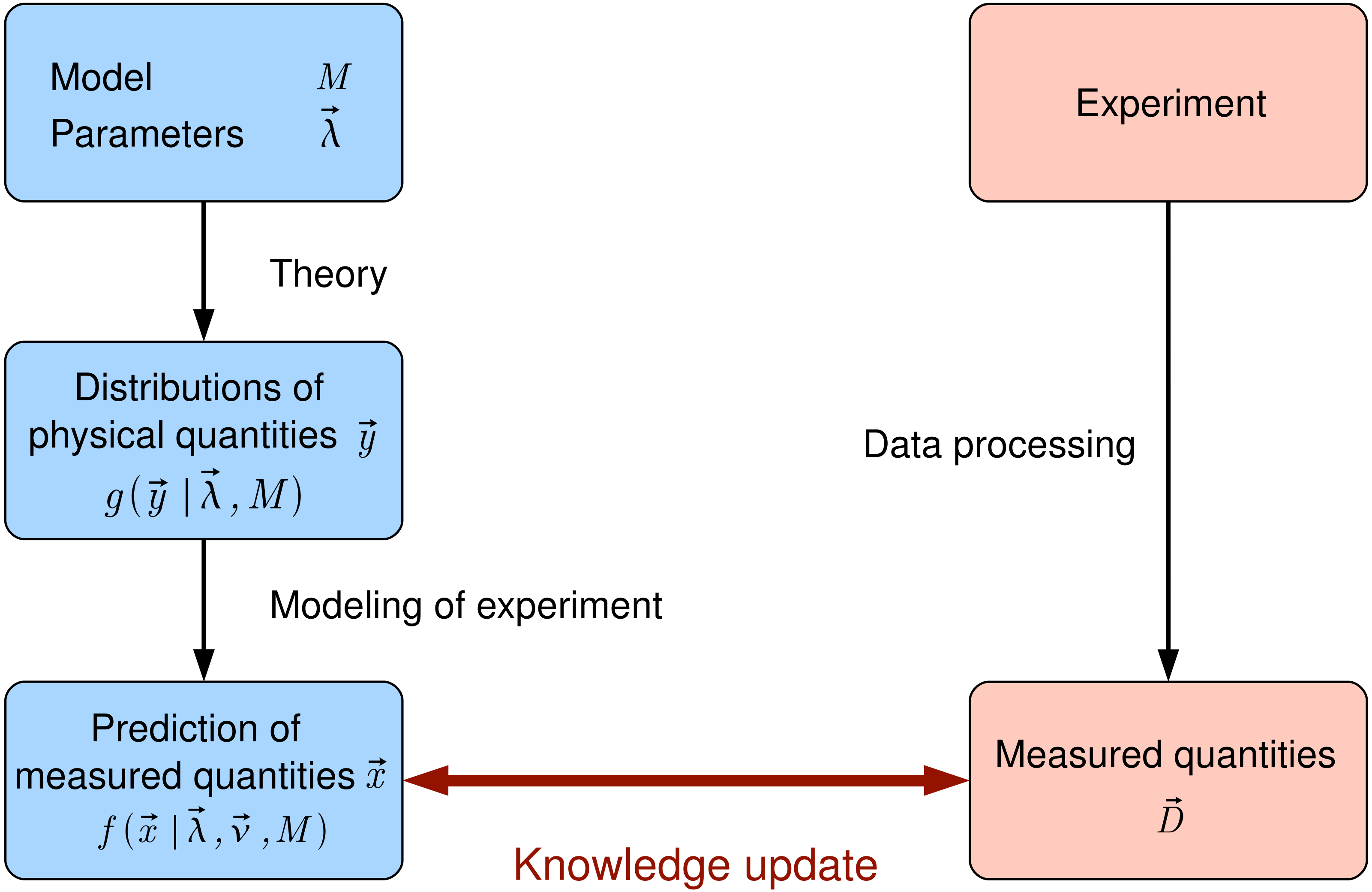
IMPRS Block Course on Probability and Data Analysis

Day 2

Morning – lecture on learning from data, Bayes Theorem, parameter estimation, calculating probability intervals and limits

Afternoon – analyzing charged current deep inelastic scattering data on your computer to find limits on the cross section for a putative right-handed W boson coupling to electrons and neutrinos.

How we learn



Model M
Parameters $\vec{\lambda}$
Theory
Distributions of physical quantities $\vec{y}$
$g(\vec{y}\,|\,\vec{\lambda}, M)$
Modeling of experiment
Prediction of measured quantities $\vec{x}$
$f(\vec{x}\,|\,\vec{\lambda}, \vec{\nu}, M)$
Experiment
Data processing
Measured quantities
$\vec{D}$
Knowledge update

How we Learn

We learn by comparing measured data with distributions for predicted results assuming a theory, parameters, and a modeling of the experimental process.

What we typically want to know:

• Is the theory reasonable ? I.e., is the observed data a likely result from this theory (+ experiment)

• If we have more than one potential explanation, then we want to be able to quantify which theory is more likely to be correct given the observations

• Assuming we have a reasonable theory, we want to estimate the most probable values of the parameters, and their uncertainties. This includes setting limits (>< some value at XX% probability).

Radioactive Decay

As an example, we will consider measuring the decay rate for a radioactive isotope, in the presence of background. We assume the lifetime is long compared to the time needed for the measurement.

We take two measurements, one with the source absent, to measure the background rate, and once with the source present.

Data Set	Source in/out	Run Time	Events
1	Out	100	100
2	In	100	110

What can we say about the decay rate for our isotope ?

$$N = N_0 e^{-t/\tau} \qquad \frac{dN}{dt} = -\frac{N}{\tau}$$

Formulation

The expected distribution (density) of the data assuming a model M and parameters $\vec{\lambda}$ is written as $f(\vec{x}|\vec{\lambda}, M)$ where $\vec{x}$ is a possible realization of the data. There are different possible definitions of this function.

We require that

$$f(\vec{x}|\vec{\lambda}, M) \geq 0 \qquad \int f(\vec{x}|\vec{\lambda}, M) d\vec{x} = 1$$

although as we will see the normalization condition is not really needed.

Formulation

The modeling of the experiment will typically add other (nuisance) parameters. E.g., there are often uncertainties, such as, e.g., the energy scale of the experiment. Different assumptions on these lead to different predictions for the data. Can have

$$f(\vec{x}|\vec{\lambda}, \vec{\nu}, M)$$

where $\vec{\nu}$ represents our nuisance parameters.

Example: for our decay example, we could choose:

$M \rightarrow R = R_S + R_B$ Total rate is sum of the signal rate + background rate

$\vec{\lambda} \rightarrow R_S (\approx N_0/\tau)$

$\vec{\nu} \rightarrow R_B$

The observed number of events in a time window is assumed to follow a Poisson distribution with expectation RT (T is time of observation). R_S and R_B are assumed to be constant.

Formulation

For the model, we have $0 \leq P(M) \leq 1$. For a fully Bayesian analysis, we require

$$\sum_i P(M_i) = 1$$

For the parameters, assuming a model, we have:

$$\begin{aligned} P(\vec{\lambda}|M_i) &\geq 0 \\ \int P(\vec{\lambda}|M_i) d\vec{\lambda} &= 1 \end{aligned}$$

The joint probability distribution is $P(\vec{\lambda}, M) = P(\vec{\lambda}|M)P(M)$

and $\sum_i P(M_i) \int P(\vec{\lambda}|M_i) d\vec{\lambda} = 1$

Learning Rule

$$P_{i+1}(\vec{\lambda}, M|\vec{D}) \propto f(\vec{x} = \vec{D}|\vec{\lambda}, M)P_i(\vec{\lambda}, M)$$

where the index represents a 'state-of-knowledge'

We have to satisfy our normalization condition, so

$$P_{i+1}(\vec{\lambda}, M|\vec{D}) = \frac{f(\vec{x} = \vec{D}|\vec{\lambda}, M)P_i(\vec{\lambda}, M)}{\sum_M \int f(\vec{x} = \vec{D}|\vec{\lambda}, M)P_i(\vec{\lambda}, M)d\vec{\lambda}}$$

We usually write $P_i = P_0$. This is our 'prior' information before performing the measurement (e.g., $R_S \geq 0$ in our radioactive decay example.)

Learning Rule

$$P(\vec{\lambda}, M|\vec{D}) = \frac{f(\vec{x}=\vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}, M)}{\sum_M \int f(\vec{x}=\vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}, M)d\vec{\lambda}}$$

The denominator is the probability to get the data summing over all possible models and all possible values of the parameters.

$$P(\vec{D}) = \sum_M \int f(\vec{x}=\vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}, M)d\vec{\lambda}$$

If the function $f(\vec{x}=\vec{D}|\vec{\lambda}, M)$ is the probability to get the data with this model and parameter set, then we get

$$P(\vec{\lambda}, M|\vec{D}) = \frac{P(\vec{D}|\vec{\lambda}, M)P(\vec{\lambda}, M)}{P(\vec{D})}$$

Bayes Equation

Bayes-Laplace Equation

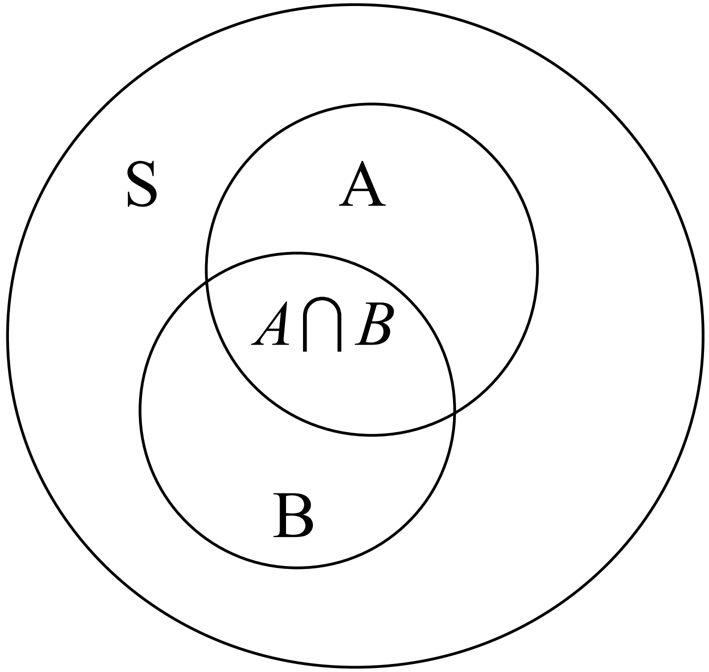
Here is the standard derivation:

$$P(A,B) = P(A|B)P(B)$$

$$P(A,B) = P(B|A)P(A)$$

So

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$



Clear for logic propositions and well-defined S,A,B.

In our case, B=model+parameters, A=data

Radioactive Decay

Let's try it out on our example:

1. First try to get an estimate of the background rate from the first data set. Assume we don't know very much. How do we represent this initial lack of knowledge ? Pick a simple form:

$$P_0(R_B) = \text{constant}$$

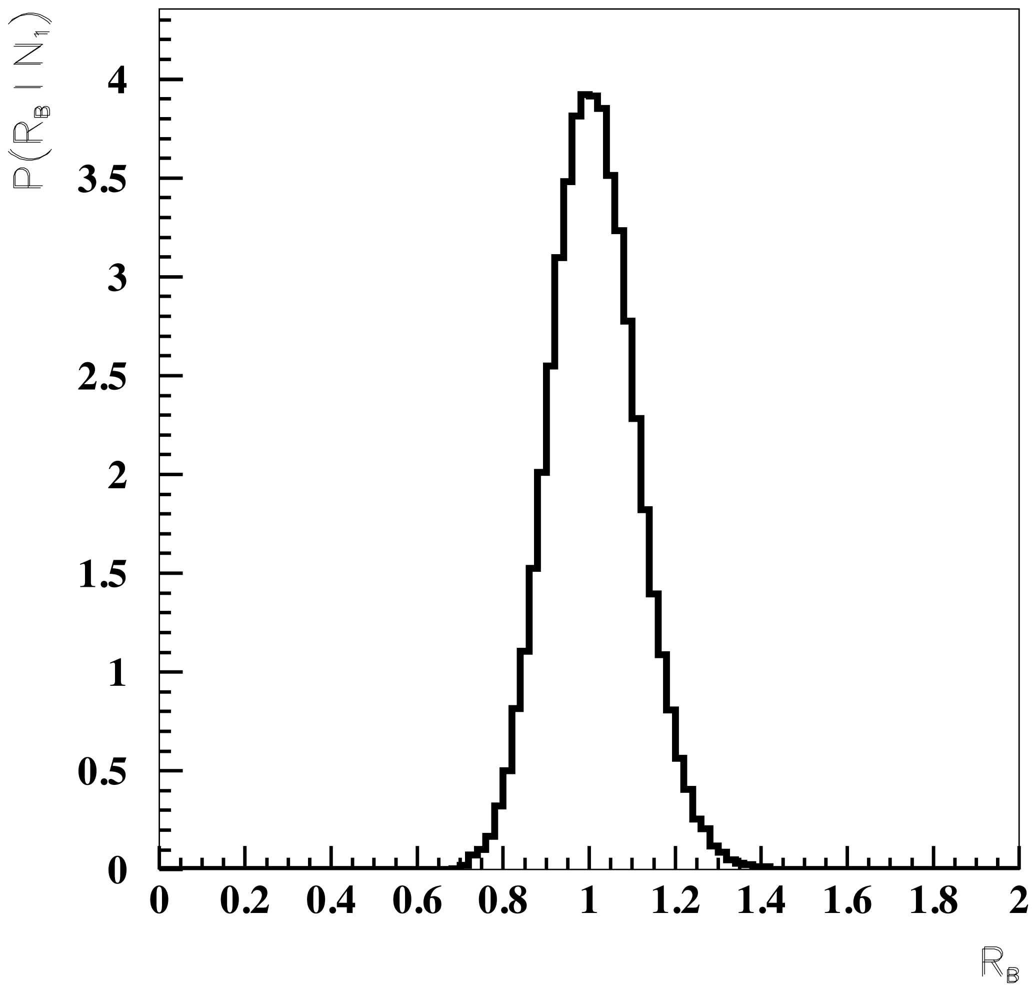
In this type of experiment, the number of counts in a time window follows a **Poisson Distribution**:

$$P(N_1|R_B) = \frac{e^{-n_B} N_1^{n_B}}{N_1!} \qquad n_B = R_B \cdot T$$

$$P(R_B|N_1) \propto P(N_1|R_B)$$

Radioactive Decay

Data Set	Source in/out	Run Time	Events
1	Out	100	100



Peak is at 1, as expected, and the Gaussian width is about 0.1. I.e., we know the background rate with about 10% certainty (100 events measured)

Radioactive Decay

2. Now try to get an estimate of the signal (and better estimate of the background). We can either analyze both data sets simultaneously or take what we learned from the first data set as prior and just analyze the second (same results). Choose

$$P_0(R_B, R_S) = P_0(R_B)P_0(R_S) = \text{constant}, R_S > 0, R_B > 0$$

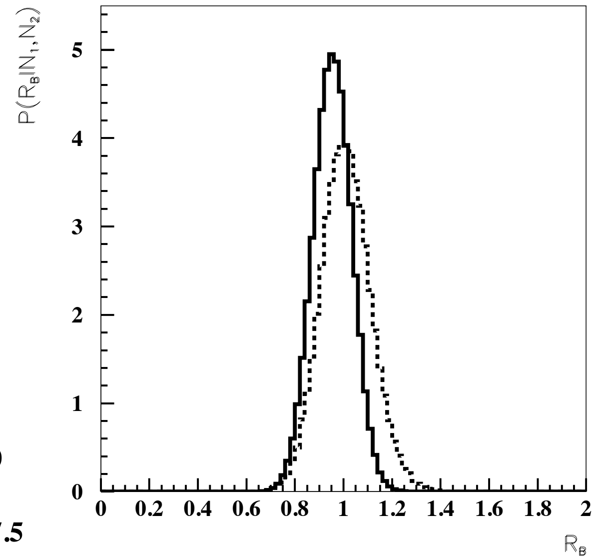
Analyze both data sets simultaneously

$$\begin{aligned} P(N_1, N_2|R_B, R_S) &= \frac{e^{-n_B} N_1^{n_B}}{N_1!} \frac{e^{-n} N_2^{n}}{N_2!} \\ n_B &= R_B \cdot T \\ n &= (R_B + R_S)T \end{aligned}$$

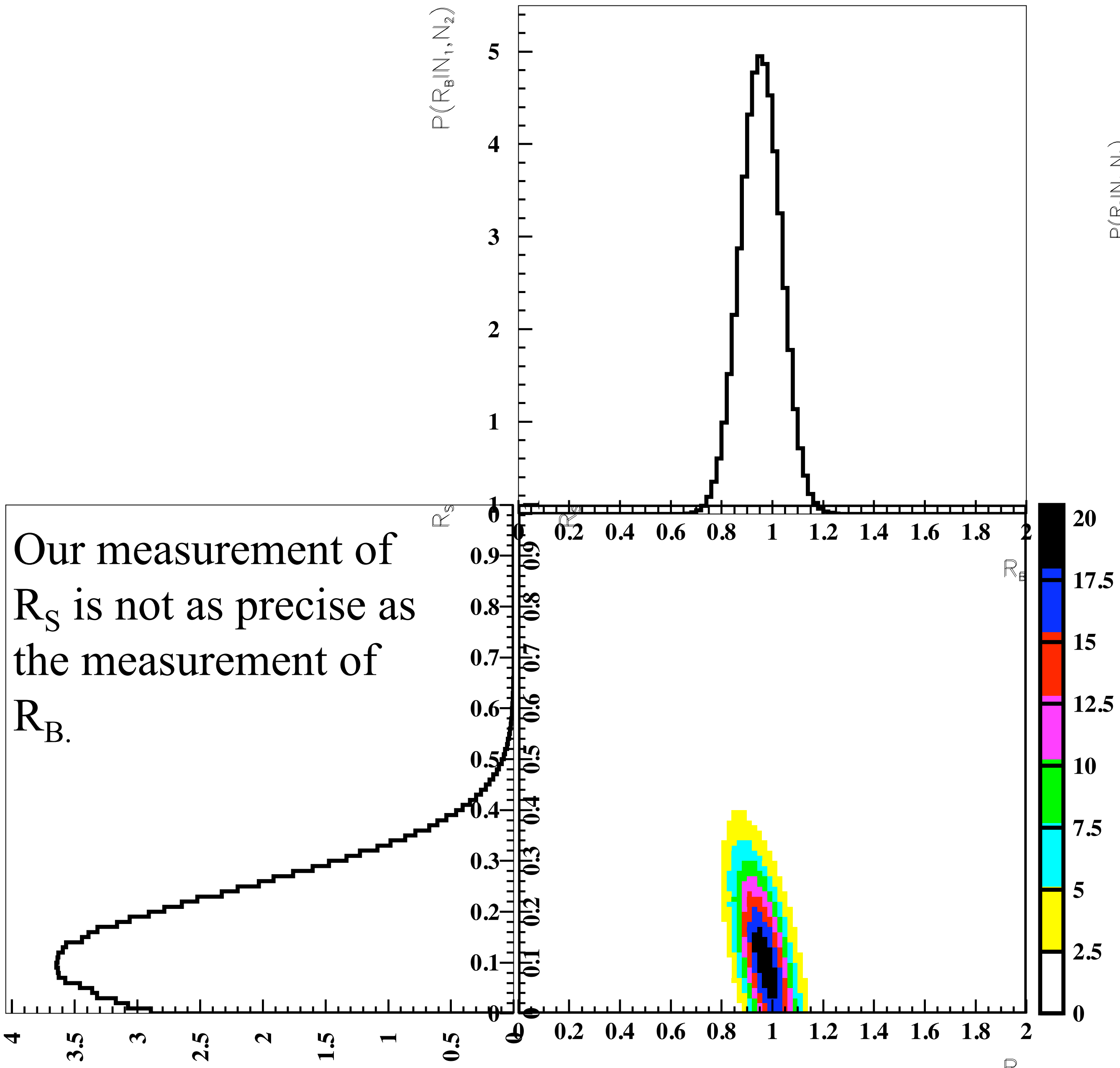
$$P(R_B, R_S|N_1, N_2) \propto P(N_1, N_2|R_B, R_S)$$

Radioactive Decay

The measurement of R_B has improved.



Our measurement of R_S is not as precise as the measurement of R_B.



Data Set	Source in/out	Run Time	Events
1	Out	100	100
2	In	100	110

Parameter Estimation

The posterior pdf gives the full probability distribution for all parameters, including all correlations – no approximations. If interested in subset of parameters, then marginalize. E.g., for one parameter:

$$P(\lambda_i|\vec{D},M) = \int P(\vec{\lambda}|\vec{D},M)d\vec{\lambda}_{\mathrm{J}\neq i}$$

Can calculate what you need from the posterior pdf. E.g.,

Mode $\max\limits^{\lambda_i} \{P(\lambda_i|D,M)\}$

+ probability intervals, …

Mean of λ_i $<\lambda_i>= \int P(\lambda_i|\vec{D},M)\lambda_i d\lambda_i$

Median $\int_{\lambda_{min}}^{\lambda_{med}} P(\lambda_i|\vec{D},M)d\lambda_i = 0.5$

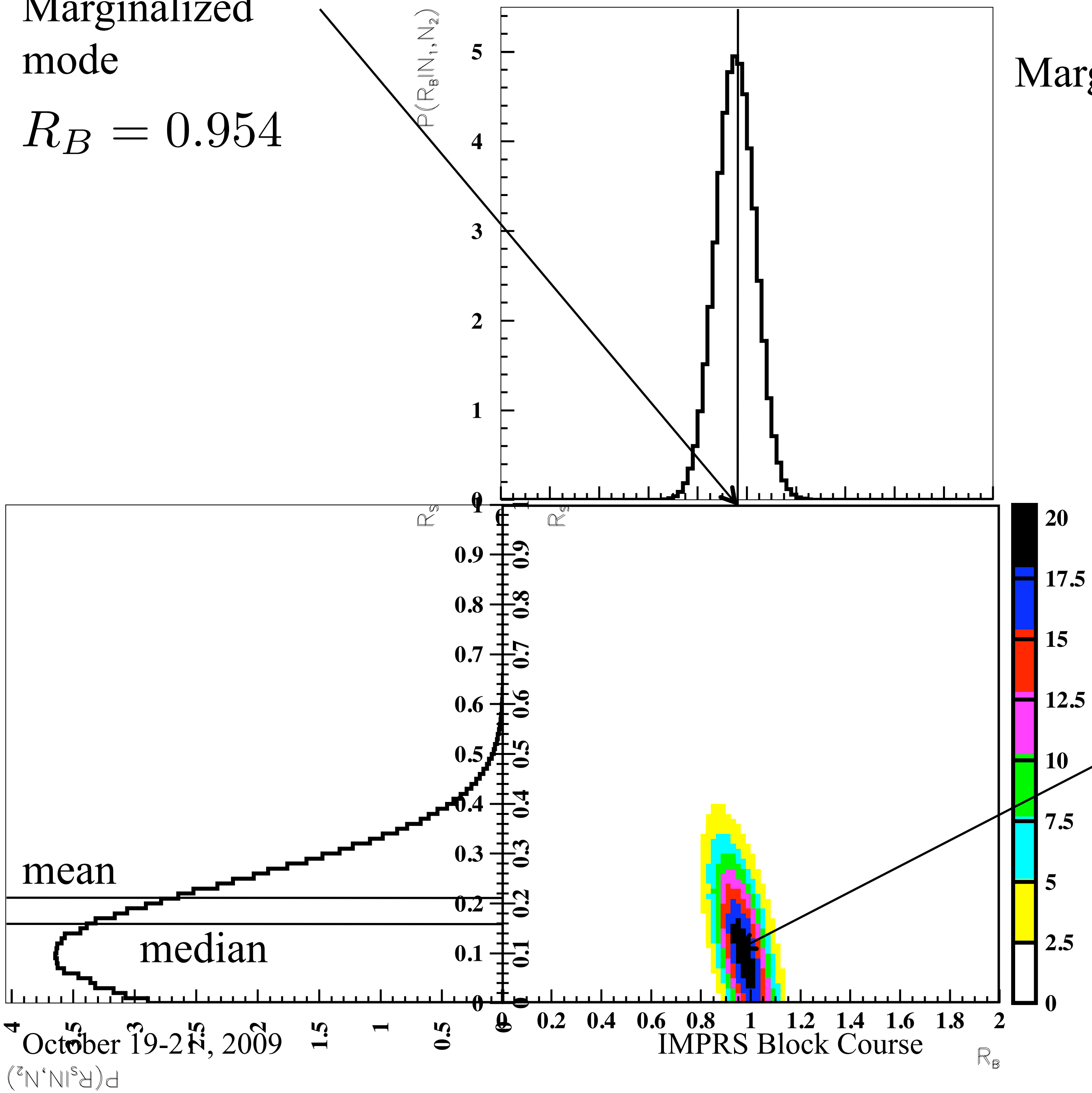
Can also perform uncertainty propagation w/o approximations

Radioactive Decay

Marginalized mode

$R_B = 0.954$

Marginalized distribution



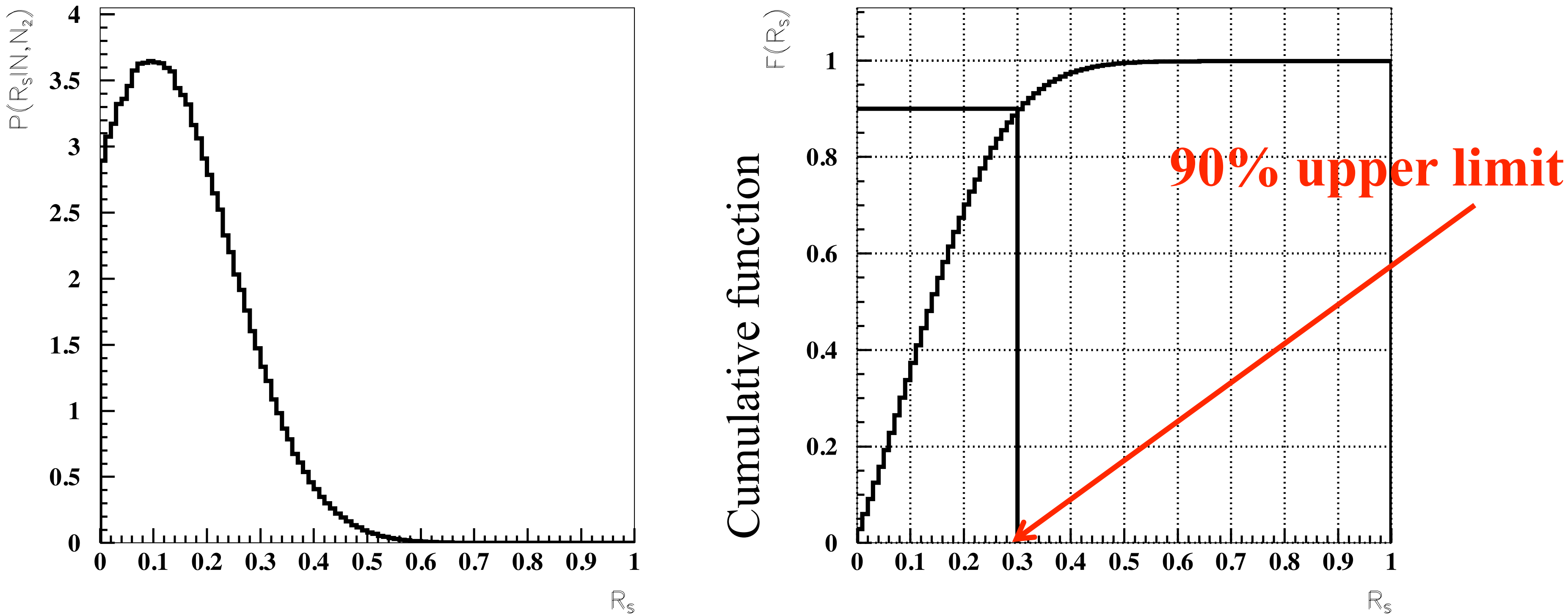
Global mode

$$R_B = 0.99999$$
$$R_S = 0.10011$$

Setting Limits

Setting limits is easy – just integrate the posterior pdf. E.g., 90% upper limit:

$$0.9 = \int_{\lambda_{min}}^{\lambda_{upper}} P(\lambda_i|\vec{D}, M)d\lambda_i$$



Or calculate contours in higher dimensional spaces

Binomial Distribution

We look now examples where our model uses Binomial distributions.

Suppose we perform N trials and have r successes. What is the 'true' probability of a success, p ? We implicitly assume that this probability is constant, and therefore the frequency distribution is a binomial. We want to know the parameter p of the binomial distribution. From Bayes' Theorem:

$$f(p \mid r,N) = \frac{f(r \mid p,N)\, f(p)}{\int_0^1 f(r \mid p,N)\, f(p)\, dp} = \frac{\frac{N!}{(N-r)!\,r!} p^r (1-p)^{N-r} f(p)}{\int_0^1 \frac{N!}{(N-r)!\,r!} p^r (1-p)^{N-r} f(p)\, dp}$$

Binomial – cont.

If we assume that f(p) is a constant

$$f(p\,|\,r,N)=\frac{f(r\,|\,p,N)f(p)}{\int_0^1 f(r\,|\,p,N)f(p)dp}=\frac{\frac{N!}{(N-r)!r!}p^r(1-p)^{N-r}}{\int_0^1 \frac{N!}{(N-r)!r!}p^r(1-p)^{N-r}dp}$$

The integral is technically a β function, and for integer r,N reduces to

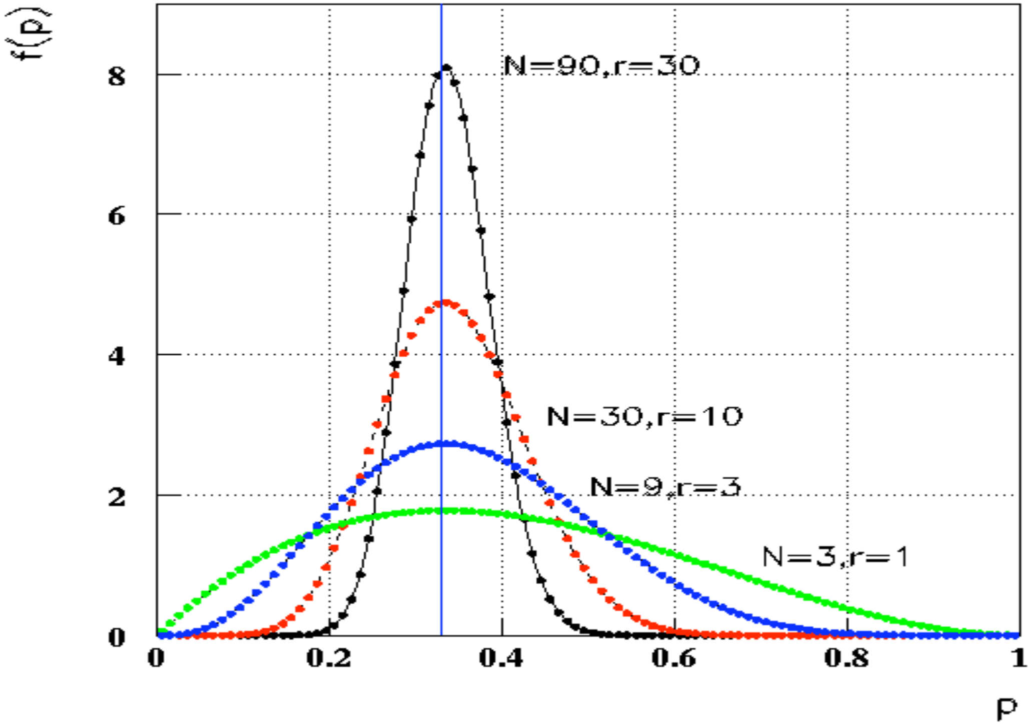
$$\int_0^1 p^x(1-p)^{n-x}dp=\frac{x!(n-x)!}{(n+1)!}$$

so

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

Note maximum at p=r/N

Binomial - cont.



The expectation value and variance are:

$$< p >= \int_0^1 \frac{(N+1)!}{r!(N-r)!} p^{r+1} (1-p)^{N-r} dp = \frac{(N+1)!}{r!(N-r)!} \frac{(r+1)!(N-r)!}{(N+2)!} = \frac{r+1}{N+2}$$

$$\sigma^2 = \frac{(r+1)(N-r+1)}{(N+3)(N+2)^2} =< p > (1- < p >) \frac{1}{N+3}$$

Binomial - cont.

Comments:

- The expectation value is not r/N, the 'observed' value of p, but rather r+1/N+2. As $N \rightarrow \infty$, this yields the expected answer. Note however that f(p|N,r) does peak at r/N. This should be kept in mind. The form r+1/N+2 gives reasonable values for r=0 and r=N (even for N=0)

- The variance also assumes the 'expected' form as $N \rightarrow \infty$.

- The formula can be used to learn, I.e, f(p) can be recalculated based on existing measurements rather than evaluating everything in one step. The same results are obtained. E.g.,

Binomial – cont.

Consider two special cases: r=0, r=N

$$\boxed{r = N}: \quad f(N \mid N, p) = \binom{N}{N} p^N (1-p)^{N-N} = p^N$$

$$f(p \mid r = N, N) = \frac{p^N}{\int_0^1 p^N dp} = (N+1) p^N$$

where we assumed f(p) = constant

For setting limits on parameters, we need the cumulative distribution function

$$F(p \mid r = N, N) = \int_0^p (N+1) p'^N dp' = p^{N+1}$$

Binomial - cont.

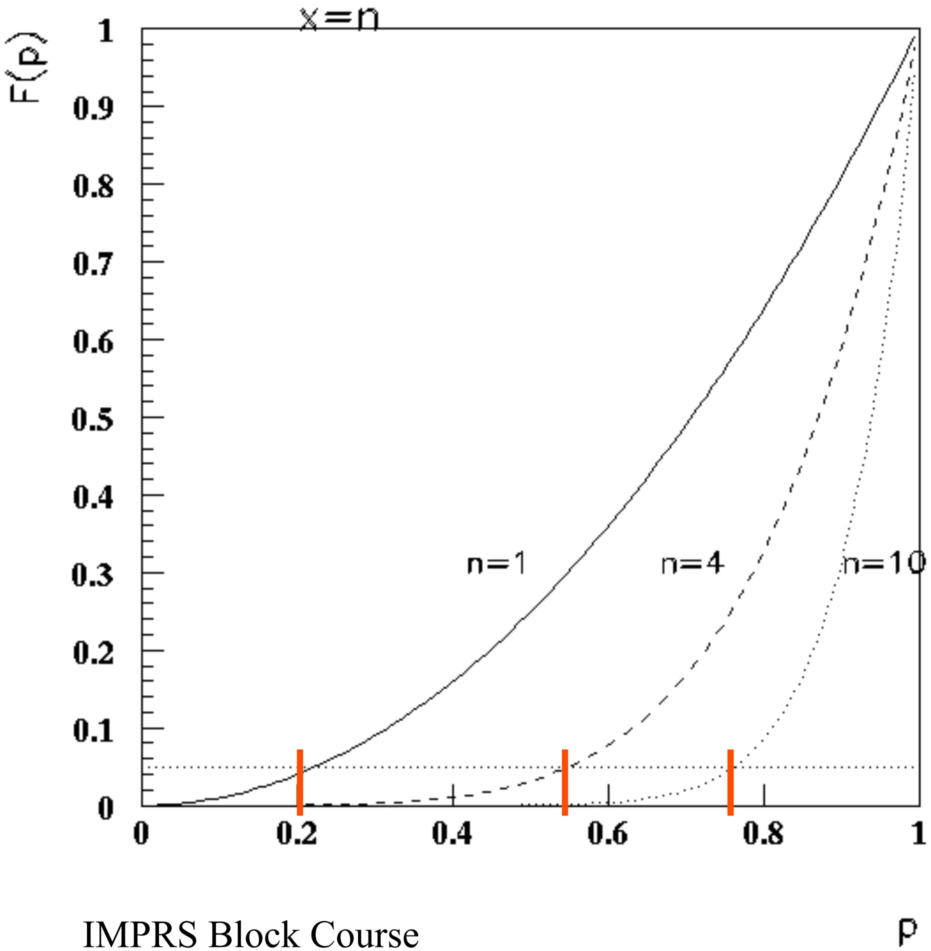
For a 95% lower limit : $F(p_0 \mid r=N,N)=0.05,\;$ so $\;p_0 = \sqrt[N+1]{0.05}$

95% probability that $p > p_0$

N=1, p_0=0.22

N=4, p_0=0.55

N=10, p_0=0.76



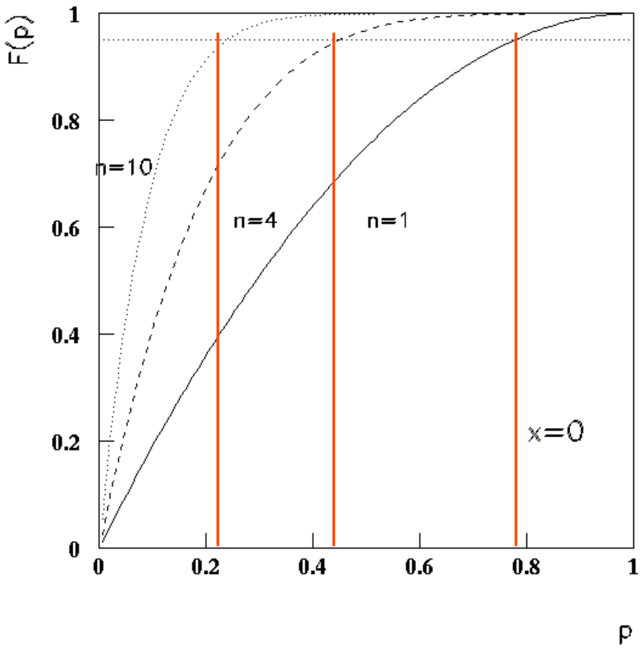
Binomial – cont.

$$r = 0: \quad f(0 \mid N, p) = \binom{N}{0} p^0 (1-p)^{N-0} = (1-p)^N$$

$$f(p \mid r=0, N) = \frac{(1-p)^N}{\int_0^1 (1-p)^N \, dp} = (N+1)(1-p)^N$$ where we assumed f(p) = constant

and $$F(p \mid r=0, N) = \int_0^p (N+1)(1-p')^N \, dp' = 1 - (1-p)^{N+1}$$

For a 95% upper limit : $F(p_0 \mid r=0, N) = 0.95$, so $p_0 = 1 - \sqrt[N+1]{0.05}$



95% probability that $p < p_0$

Data Analysis-Poisson Distribution

Typical examples – counting experiments such as source activity, failure rates, cross sections,… A very common application is the discovery process: a certain number of events were observed, and you want to know the probability that you have a discovery.

$$f(\nu \mid x) = \frac{f(x \mid \nu) f(\nu)}{\int_0^\infty f(x \mid \nu) f(\nu) d\nu} = \frac{\frac{\nu^x e^{-\nu}}{x!} f(\nu)}{\int_0^\infty \frac{\nu^x e^{-\nu}}{x!} f(\nu) d\nu}$$

This is our master formula. Result will depend on choice of prior, *f(ν)*.

Poisson - cont.

If we assume a flat prior,

$$f(\nu \mid x) = \frac{\frac{\nu^x e^{-\nu}}{x!} f(\nu)}{\int_0^\infty \frac{\nu^x e^{-\nu}}{x!} f(\nu) d\nu} = \frac{\frac{\nu^x e^{-\nu}}{x!}}{\int_0^\infty \frac{\nu^x e^{-\nu}}{x!} d\nu}$$

$$\int_0^\infty \frac{\nu^x e^{-\nu}}{x!} d\nu = \frac{1}{x!} \int_0^\infty \nu^x e^{-\nu} d\nu = 1$$

so

$$f(\nu \mid x) = \frac{\nu^x e^{-\nu}}{x!}$$

Note: peak at ν=x

Poisson - cont.

The cumulative distribution function:

$$F(\nu \mid x) = \int_0^{\nu} \frac{\nu'^{x} e^{-\nu'}}{x!} d\nu' = \frac{1}{x!}\left[-\nu'^{x} e^{-\nu'} \Big|_0^{\nu} + x\int_0^{\nu} \nu'^{x-1} e^{-\nu'} d\nu' \right]$$

$$F(\nu \mid x) = 1 - e^{-\nu} \sum_{n=0}^{x} \frac{\nu^n}{n!}$$

The expectation value and standard deviation:

$$<\nu> = \int_0^{\infty} \frac{\nu^x e^{-\nu}}{x!} \nu \, d\nu = \frac{(x+1)!}{x!} = (x+1)$$!!

$$\sigma^2 = \int_0^{\infty} \frac{\nu^x e^{-\nu}}{x!} \nu^2 \, d\nu - <\nu>^2 = \frac{(x+2)!}{x!} - (x+1)^2 = (x+1)$$!!

Poisson - cont.

Note: *x=0* $<\nu>=1$ *???*

From prior, expect

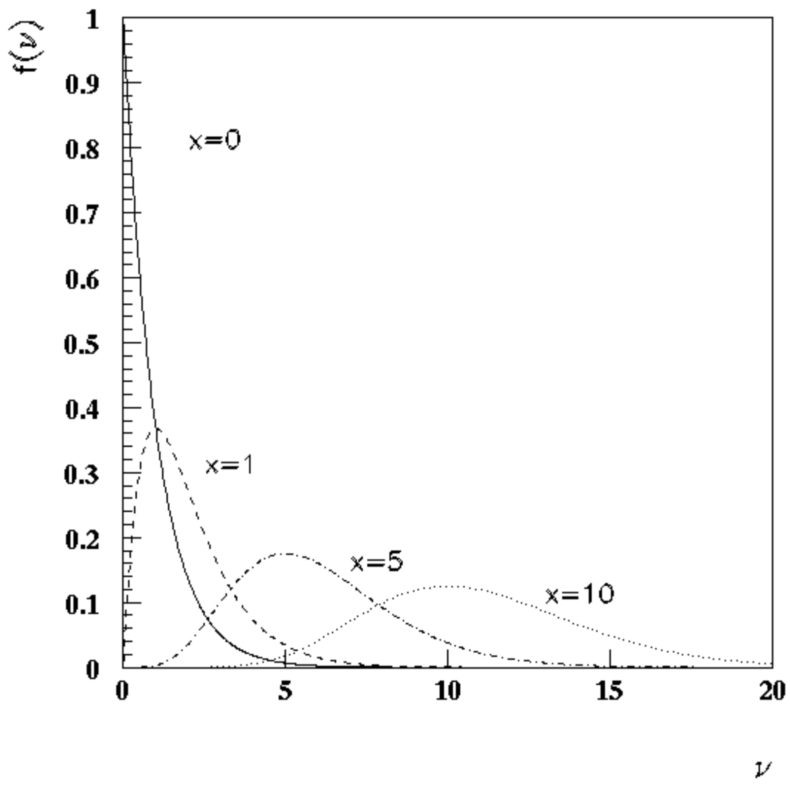
$$<\nu>=\int_0^\infty \nu f(\nu)d\nu=\lim_{\nu'\to\infty}\frac{1}{\nu'}\int_0^{\nu'}\nu d\nu=\lim_{\nu'\to\infty}\frac{\nu'}{2}=\infty$$

What happened ?

x=0 is a measurement !

Poisson – cont.

Some example *f(ν)*



Comments:

• As with the binomial, the expectation value is ≠ measured value, but the peak of the probability distribution (maximum likelihood) gives the 'correct' value

• The variance is also larger than the naïve expectation.

Poisson – cont.

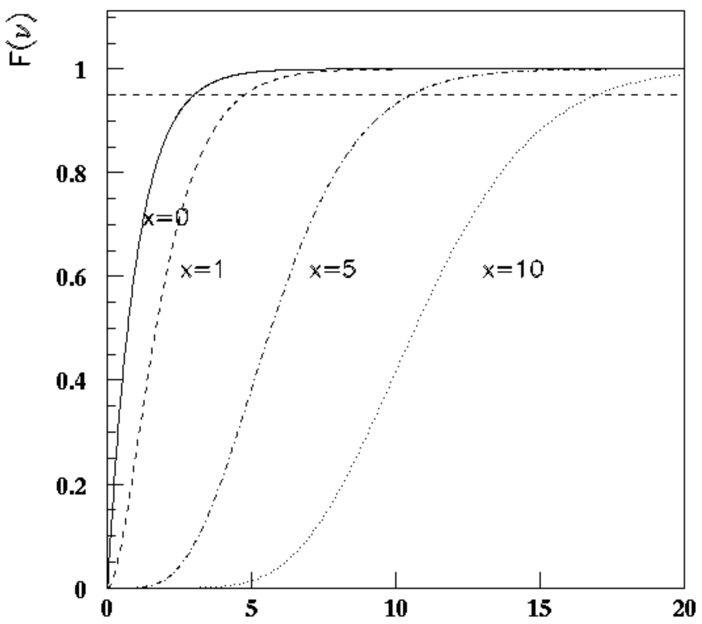
Some examples. First, no background, measure zero counts.

$$f(\nu \mid x=0) = \frac{\nu^0 e^{-\nu}}{0!} = e^{-\nu} \quad \text{with the flat prior assumption}$$

$$F(\nu \mid x=0) = 1 - e^{-\nu}$$

For a 95% upper limit, set $F(\nu \mid 0) = 0.95 = 1 - e^{-\nu}$

which gives $\nu < 3$ (very common result)



Note that there is nothing magic about the number 3 - it is a coincidence of choosing the 95% probability cutoff

Poisson – cont.

What if we cannot take a flat prior ? (e.g., we have previous information, such as non-observation in previous experiments or theoretical bias)

Suppose we can model the prior belief as $f(\nu) = \frac{1}{10} e^{-\nu/10}$

Now Bayes tells us $f(\nu \mid x=0) = \frac{f(0 \mid \nu) f(\nu)}{\int_0^\infty f(0 \mid \nu) f(\nu) d\nu} = \frac{e^{-\nu} \frac{1}{10} e^{-\nu/10}}{\int_0^\infty \frac{1}{10} e^{-11\nu/10} d\nu} = \frac{11}{10} e^{-11\nu/10}$

$$< \nu > = \int_0^\infty \frac{11}{10} e^{-11\nu/10} \nu d\nu = 0.91$$

$P(\nu \le 2.7) = 95\%$, i.e., $\nu \le 2.7$ with 95% probability

Peaked prior gives smaller upper limit. Lesson: choice of prior important

Poisson – cont.

And now suppose we have background:

Imagine that we observe a certain number of events, x, but there are two different possible sources - the signal we are looking for, and background events which mimic the signal but are uninteresting.

Notation: λ is the expected number of background events

σ_B is the uncertainty on the background (typically assumed to be Gaussian distributed). We will consider first the case where $\sigma_B \ll \lambda$ and can be neglected.

As discussed previously, the expected number of events again follows a Poisson distribution.

$$\mu = \lambda + \nu, \quad p(x \mid \mu) = \frac{e^{-\mu}\mu^x}{x!}$$

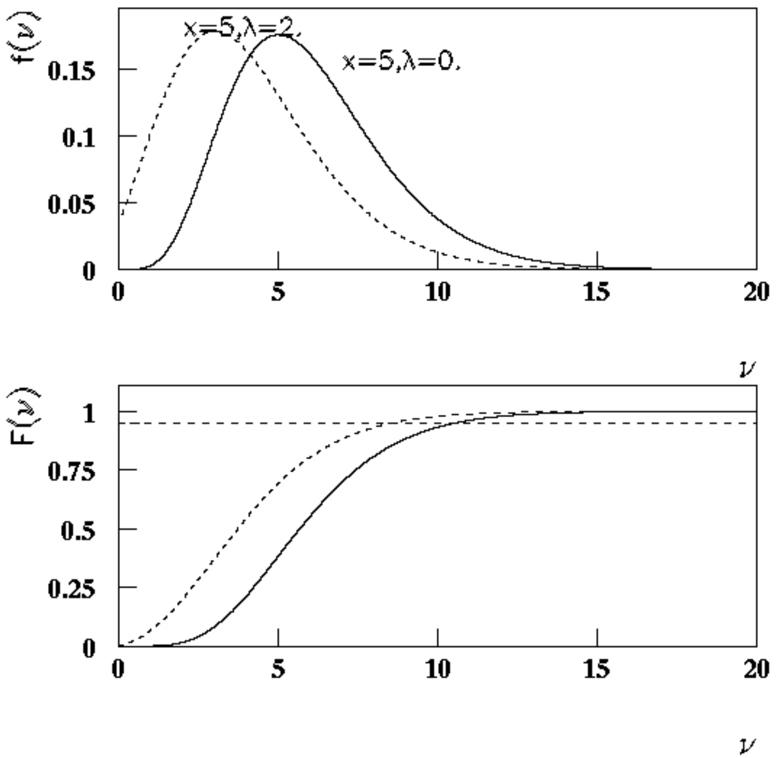
$$f(\nu \mid x, \lambda) = \frac{\left(e^{-(\lambda+\nu)}(\lambda+\nu)^x / x! \right) f(\nu)}{\int_0^\infty \left(e^{-(\lambda+\nu)}(\lambda+\nu)^x / x! \right) f(\nu) d\nu}$$

Poisson – cont.

If we again assume a flat f(ν) and integrate by parts.

$$f(\nu \mid x, \lambda) = \frac{e^{-\nu}(\lambda + \nu)^x}{x! \sum_{n=0}^{x} \frac{\lambda^n}{n!}}$$

$$F(\nu \mid x, \lambda) = 1 - \frac{e^{-\nu} \sum_{n=0}^{x} \frac{(\lambda + \nu)^n}{n!}}{\sum_{n=0}^{x} \frac{\lambda^n}{n!}}$$



Comments:

- The previous formula for no background is recovered when λ=0
- For x=0, f(ν|x, λ)=$e^{-\nu}$. It does not matter how much background you have, you get the same probability distribution for the signal. Source of much confusion & discussion.

Poisson – cont.

Now assume we have a non-negligible uncertainty on the background. We can no longer ignore the prior probability for the background. Call this $g(\lambda)$. We assume that the background is uncorrelated with the signal (independent prior probabilities). Then

$$f(\lambda,\nu) = f(\nu)g(\lambda)$$

Following Bayes'

$$f(\nu,\lambda \mid x) = \frac{f(x \mid \lambda,\nu) f(\lambda,\nu)}{\iint f(x \mid \lambda,\nu) f(\lambda,\nu) d\nu d\lambda} = \frac{f(x \mid \lambda,\nu) f(\nu) g(\lambda)}{\iint f(x \mid \lambda,\nu) f(\nu) g(\lambda) d\nu d\lambda}$$

And we marginalize wrt ν

$$f(\nu \mid x) = \frac{\int f(x \mid \lambda,\nu) f(\nu) g(\lambda) d\lambda}{\iint f(x \mid \lambda,\nu) f(\nu) g(\lambda) d\lambda d\nu}$$

Applying Poisson statistics

$$f(\nu \mid x,\lambda) = \frac{\int \left(e^{-(\lambda+\nu)} (\lambda+\nu)^x \big/ x! \right) f(\nu) g(\lambda) d\lambda}{\iint \left(e^{-(\lambda+\nu)} (\lambda+\nu)^x \big/ x! \right) f(\nu) g(\lambda) d\lambda d\nu}$$

Typically has to be solved numerically

Gaussian Uncertainties

We focus now on data with Gaussian probability distributions. I.e., we expect our results to follow a Gaussian distribution around the 'true' value.

The Gaussian assumption is usually valid for samples with large statistics. We also use it to parameterize the systematic uncertainties using the CLT as a vague justification.

We often do not know the true form of the distribution (measured-true), and the assumption of a Gaussian form is a default.

Bayes and Gaussian Measurements

The pdf for the mean given a single measurement, and assuming the width is known, is

$$p(\mu \mid x) = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]p(\mu)}{\int \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]p(\mu)d\mu}$$

Where x is the measured value

Now we make the usual starting assumption about the prior probability – it is a constant.

$$\int \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]d\mu = 1 \quad \text{since the integral is symmetric in } x,\mu$$

so

$$p(\mu \mid x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Bayes & Gaussians – cont.

The probability distribution for μ is a Gaussian distribution about x

• The mode is at $\mu = x$

• The standard deviation is $\sigma_\mu = \sigma$ (the width of the uncertainty assigned to measurement)

• The probability intervals for μ are

Probability Level	Interval
68.3%	$x \pm \sigma$
90.0%	$x \pm 1.65\sigma$
95.0%	$x \pm 1.96\sigma$
99.0%	$x \pm 2.58\sigma$
99.7%	$x \pm 3\sigma$

Bayes & Gaussians – cont.

Suppose now we make another measurement of the same quantity, and have a new estimate for the mean (switch to subscripts). Assume again that the resolution is known, although it could be different than before.

$$p_2(\mu \mid \vec{x}_2) = \frac{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]p_1(\mu)}{\int \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]p_1(\mu)d\mu}$$

We use the result from the first set of measurements for the prior.

$$p_1(\mu) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right].$$

The denominator is

$$\int \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]d\mu =$$

$$\int \frac{1}{2\pi\sigma_2\sigma_1}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]d\mu$$

Bayes & Gaussians - cont.

and the numerator is

$$\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]=$$

$$\frac{1}{2\pi\sigma_2\sigma_1}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]$$

First rewrite the numerator

$$\frac{1}{2\pi\sigma_2\sigma_1}\exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]=$$

$$\frac{1}{2\pi\sigma_2\sigma_1}\exp\left[-\frac{x_2^2}{2\sigma_2^2}-\frac{x_1^2}{2\sigma_1^2}+\mu\left(\frac{x_2}{\sigma_2^2}+\frac{x_1}{\sigma_1^2}\right)-\mu^2\left(\frac{1}{2\sigma_2^2}+\frac{1}{2\sigma_1^2}\right)\right]$$

Bayes & Gaussian – cont.

define

$$x_A = \left. \left(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \right) \middle/ \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \right.$$

and

$$\frac{1}{\sigma_A^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Then the numerator becomes

$$\frac{1}{2\pi\sigma_2\sigma_1} \exp\left[-\frac{x_2^2}{2\sigma_2^2} - \frac{x_1^2}{2\sigma_1^2} + \mu\frac{x_A}{\sigma_A^2} - \frac{\mu^2}{2\sigma_A^2} \right]$$

For the denominator, we integrate this over all μ

$$\int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_2\sigma_1} \exp\left[-\frac{x_2^2}{2\sigma_2^2} - \frac{x_1^2}{2\sigma_1^2} + \mu\frac{x_A}{\sigma_A^2} - \frac{\mu^2}{2\sigma_A^2} \right] d\mu =$$

$$\frac{1}{2\pi\sigma_2\sigma_1} \exp\left[-\frac{x_2^2}{2\sigma_2^2} - \frac{x_1^2}{2\sigma_1^2} \right] \int_{-\infty}^{+\infty} \exp\left[\mu\frac{x_A}{\sigma_A^2} - \frac{\mu^2}{2\sigma_A^2} \right] d\mu$$

Bayes & Gaussian – cont.

For the remaining integral, we 'complete the square'

$$-\mu\frac{x_A}{\sigma_A^2}+\frac{\mu^2}{2\sigma_A^2}=(A\mu+B)^2+C$$

$$A^2=\frac{1}{2\sigma_A^2} \qquad 2AB=-\frac{x_A}{\sigma_A^2} \qquad \text{so} \quad B=-\frac{x_A}{\sqrt{2}\sigma_A} \quad \text{and} \quad C=-B^2=-\frac{x_A^2}{2\sigma_A^2}$$

$$\int_{-\infty}^{+\infty}\exp\left[\mu\frac{x_A}{\sigma_A^2}-\frac{\mu^2}{2\sigma_A^2}\right]d\mu=\exp[-C]\int_{-\infty}^{+\infty}\exp\left[-(A\mu+B)^2\right]d\mu=\frac{\sqrt{\pi}}{A}\exp[-C]$$

$$\int_{-\infty}^{+\infty}\exp\left[\mu\frac{x_A}{\sigma_A^2}-\frac{\mu^2}{2\sigma_A^2}\right]d\mu=\sqrt{2\pi}\sigma_A\exp\left[\frac{x_A^2}{2\sigma_A^2}\right]$$

Gaussian – cont.

Putting it all together

$$P(\mu|x_1,x_2) = \frac{1}{\sqrt{2\pi}\sigma_A} e^{-\frac{(x_A-\mu)^2}{2\sigma_A^2}}$$

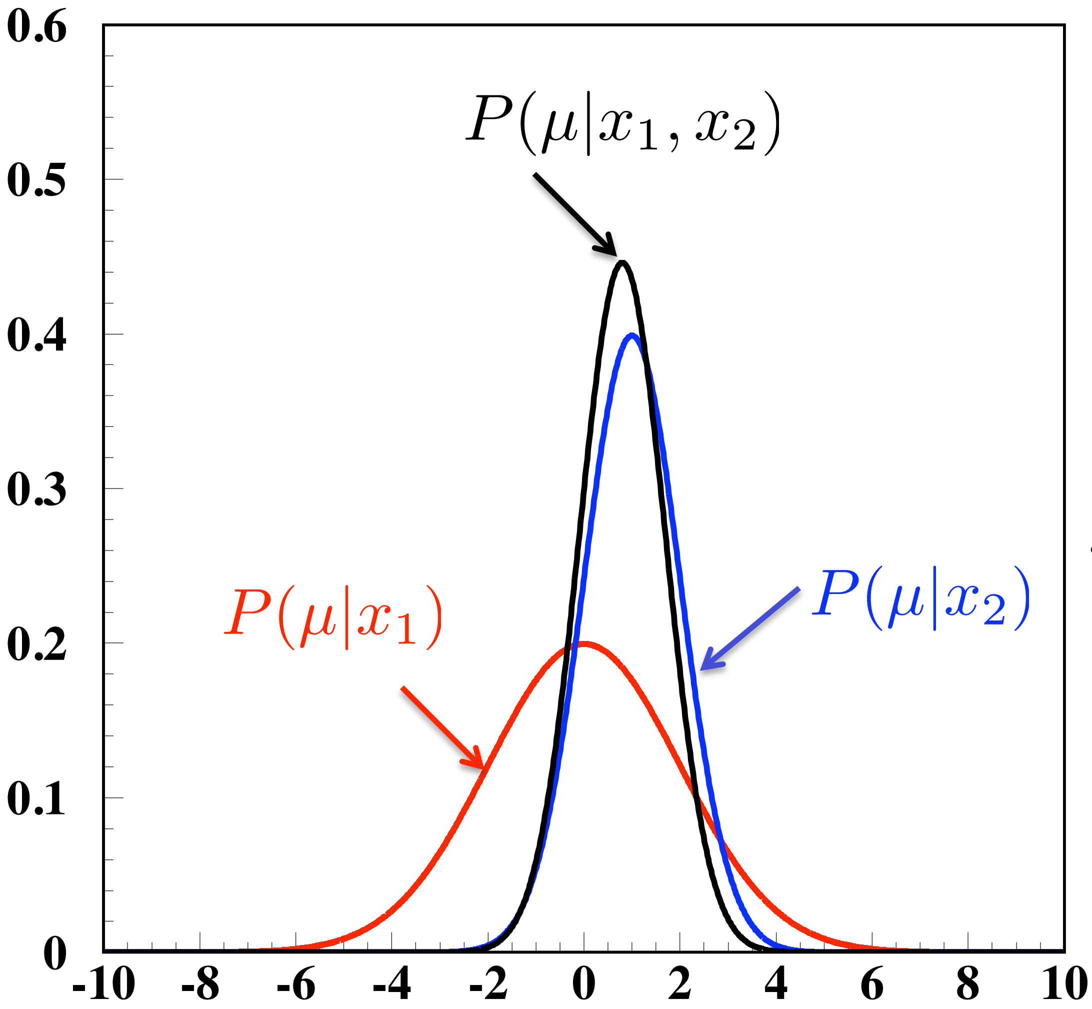
$$x_A = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} \qquad \frac{1}{\sigma_A^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Or in general

$$x_A = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2} \qquad \frac{1}{\sigma_A^2} = \sum_i \frac{1}{\sigma_i^2}$$

Very simple summation rule for results with Gaussian uncertainties.

Two Gaussian Measurements



$x_1 = 0 \;\; \sigma_1 = 2.$

$x_2 = 1 \;\; \sigma_2 = 1.$

$x_{1+2} = 0.8 \;\; \sigma_{1+2} = \sqrt{0.8}$

Example

Detector responses are usually modeled as Gaussians. E.g., the distribution of the measured energy is

$$p(E|E_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-E_0)^2}{2\sigma^2}}$$

where $$\sigma = \sqrt{a^2 \cdot E_0 + b^2 \cdot E_0^2}$$

This formula comes from assuming that there are two sources of fluctuations possible and they both can be assumed to be Gaussian distributed. The first term is from counting, i.e., Poisson but with high statistics. The second is a systematic calibration uncertainty of different cells in the calorimeter. a,b are parameters describing the performance.

Example

Suppose we now make a measurement of the energy. What can we say about the 'true' value ? If we assume a flat prior, we get

$$P(E_0|E) = P(E|E_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E_0-E)^2}{2\sigma^2}}$$

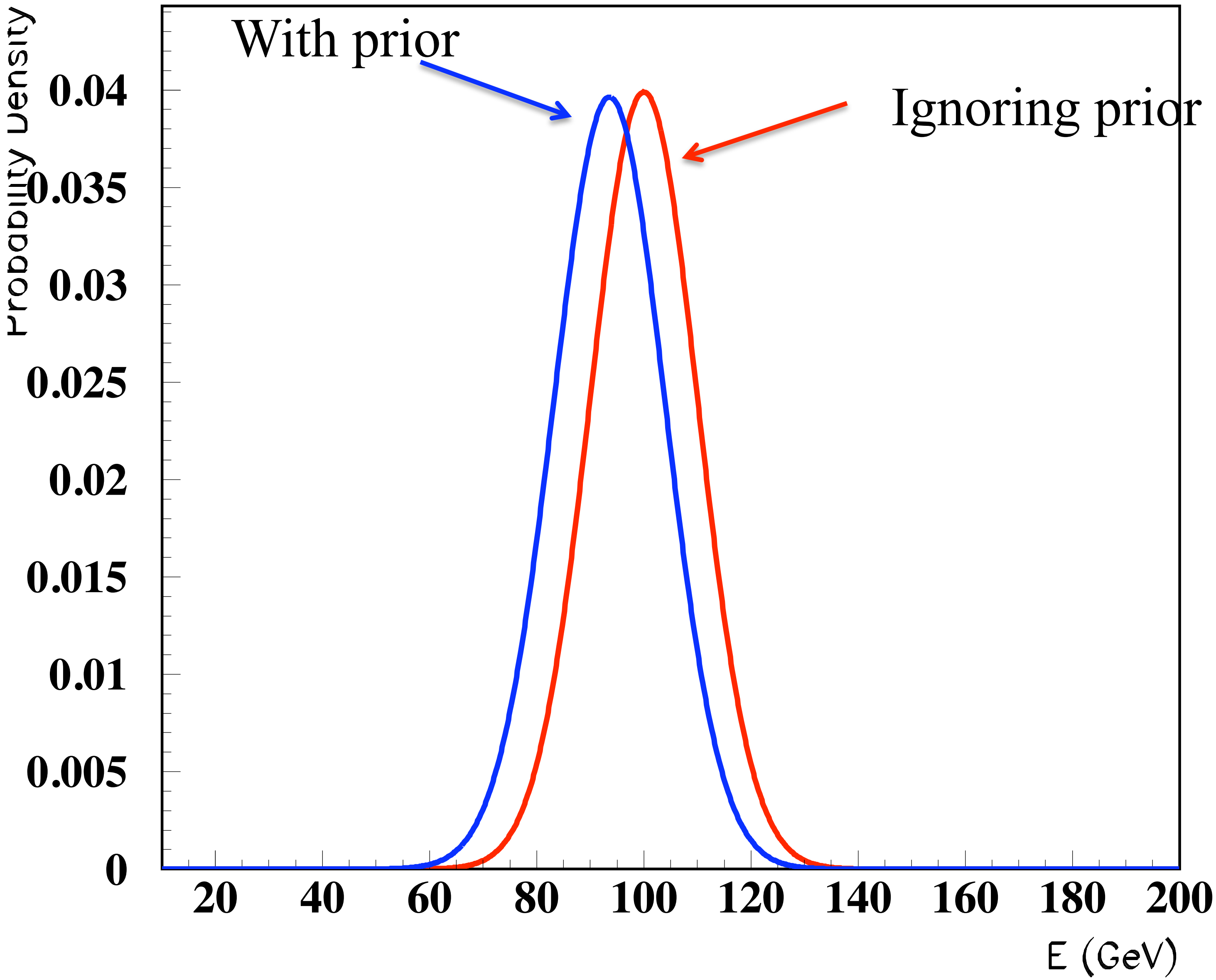
The probability distribution of the true value is a Gaussian centered on the measured value. However, energy distributions often have a steep distribution. Suppose the starting distribution was

$$P_0(E_0) \propto E_0^{-6}$$

then

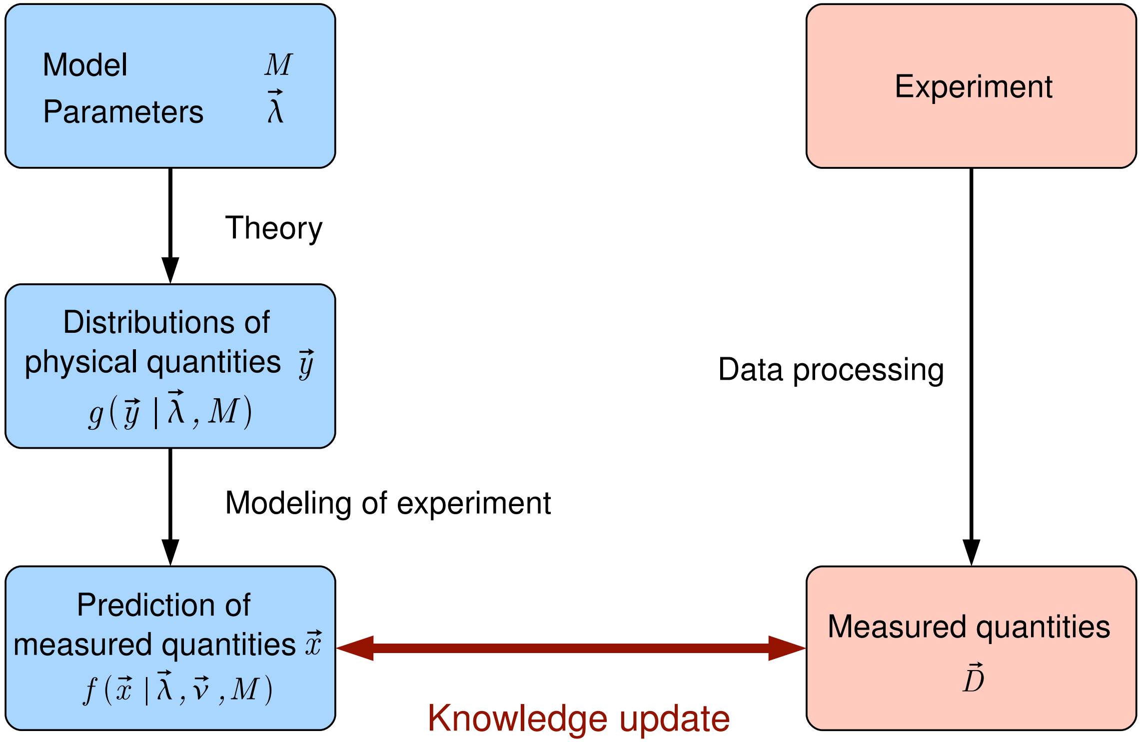
$$P(E_0|E) \propto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E_0-E)^2}{2\sigma^2}} E_0^{-6}$$

one measurement of the energy, resolution 10 GeV, measured 100 GeV



Power for Energy Spectrum

Suppose what we are trying to extract is the power of the underlying energy distribution. How would we proceed ?



In this case, assume $g(E_0|\lambda, M) \propto E_0^{-\lambda}$

Power example

We assume the measured values are related to the true:

$$P(E|E_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E_0-E)^2}{2\sigma^2}}$$

Now apply the 'law of total probability'

$$P(E|\lambda) = \int P(E|E_0)P(E_0|\lambda)dE_0$$

And Bayes' equation yields $P(\lambda|E) \propto \prod_i P(E_i|\lambda)P_0(\lambda)$

$$P(\lambda|E) \propto \left[\prod_i \int \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(E_i-E_{0,i})^2}{2\sigma_i^2}} E_{0,i}^{-\lambda} dE_{0,i}\right] P_0(\lambda)$$

Power example

$$P(\lambda|E) \propto \left[\prod_i \int \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(E_i - E_{0,i})^2}{2\sigma_i^2}} E_{0,i}^{-\lambda} dE_{0,i}\right] P_0(\lambda)$$

Need numerical approach.

1. Either integrate numerically many many time during parameter scan

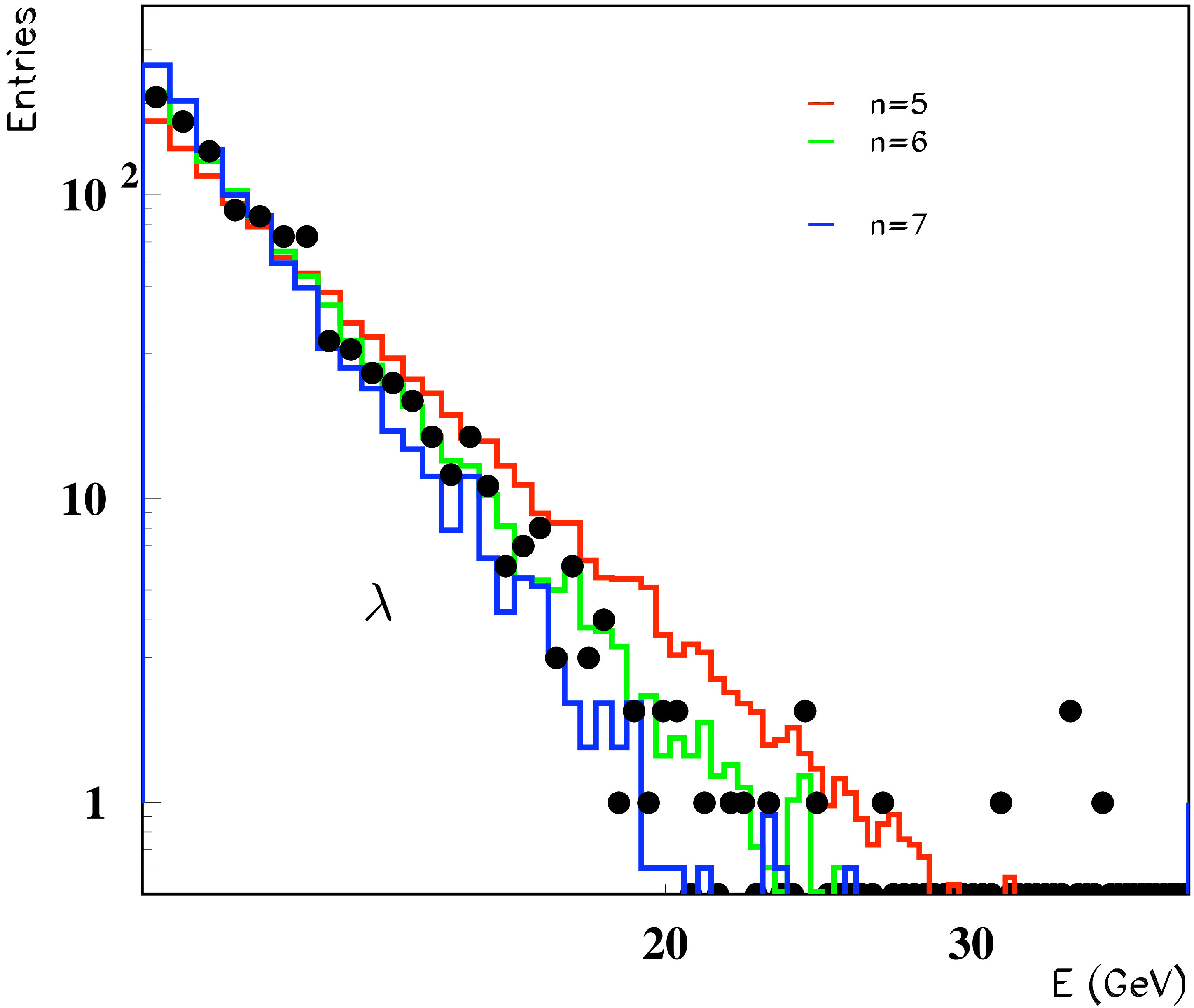
2. Make a histogram of expected number of entries in measured energy bins from your event simulation, then reweight the distribution for different values of λ and see how the agreement between expected and measured varies (Poisson statistics). Note that this does not use the equation above – in this case

$$P(E|\lambda) = \prod_{i=1}^{Nbins} \frac{e^{-\nu_i}\nu_i^{n_i}}{n_i!}$$

n_i Number of events in energy bin *i*

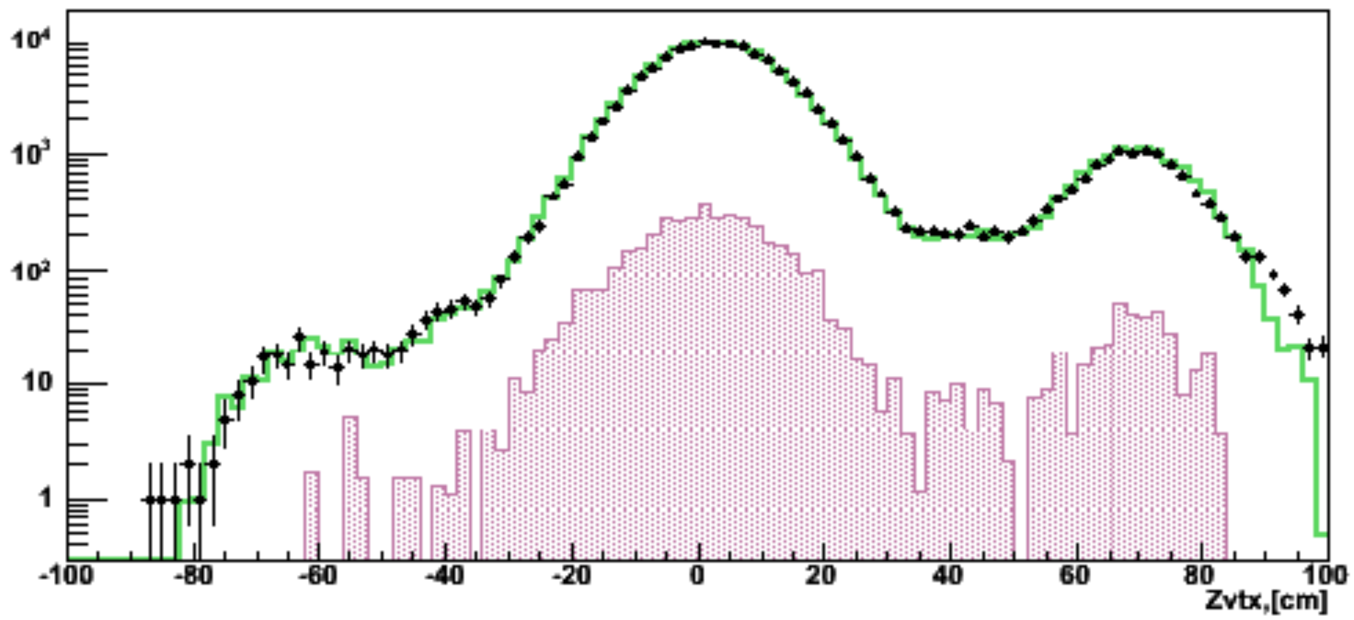
$\nu_i = \nu_i(\lambda)$ Expectation based on λ

Power example



HERA Vertex Example

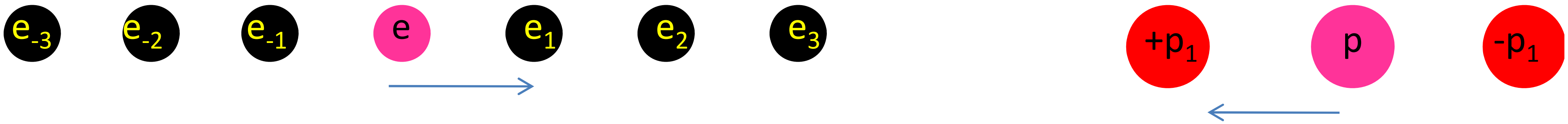
Here is a real-life example: the event vertex distribution for events recorded in the ZEUS detector. This is what we observe:



Presence of 'satellites', asymmetric distribution

HERA Vertex Example

Source of satellites:



r.f = 500 MHz => λ ~ 60 cm

r.f = 208 MHz => λ ~ 144.2 cm

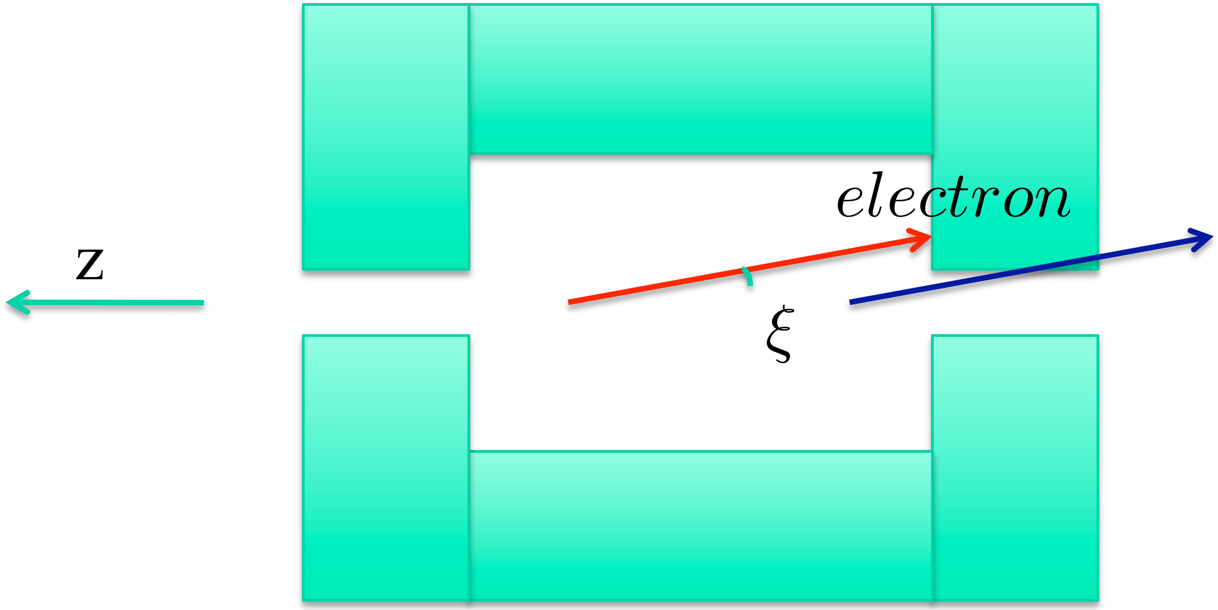
Possible interactions:

Interaction	type of vertex	relative position from main peak
ep	nominal	0 cm
e_1p	electron positive satellite	30. cm
e_2p	electron positive satellite	60. cm
e_3p	electron positive satellite	90. cm
$e_{-1}p$	electron negative satellite	-30. cm
$e_{-2}p$	electron negative satellite	-60. cm
$e_{-3}p$	electron negative satellite	-90. cm
ep_1	proton positive satellite	72.1 cm
ep_{-1}	proton negative satellite	-72.1 cm

HERA Vertex Example

Source of asymmetry:

$$electron + proton \rightarrow electron + X$$



At small angles $d\sigma/d\xi^2 \propto \frac{1}{\xi^4}$. At larger Z, acceptance for smaller angles.

HERA Vertex Example

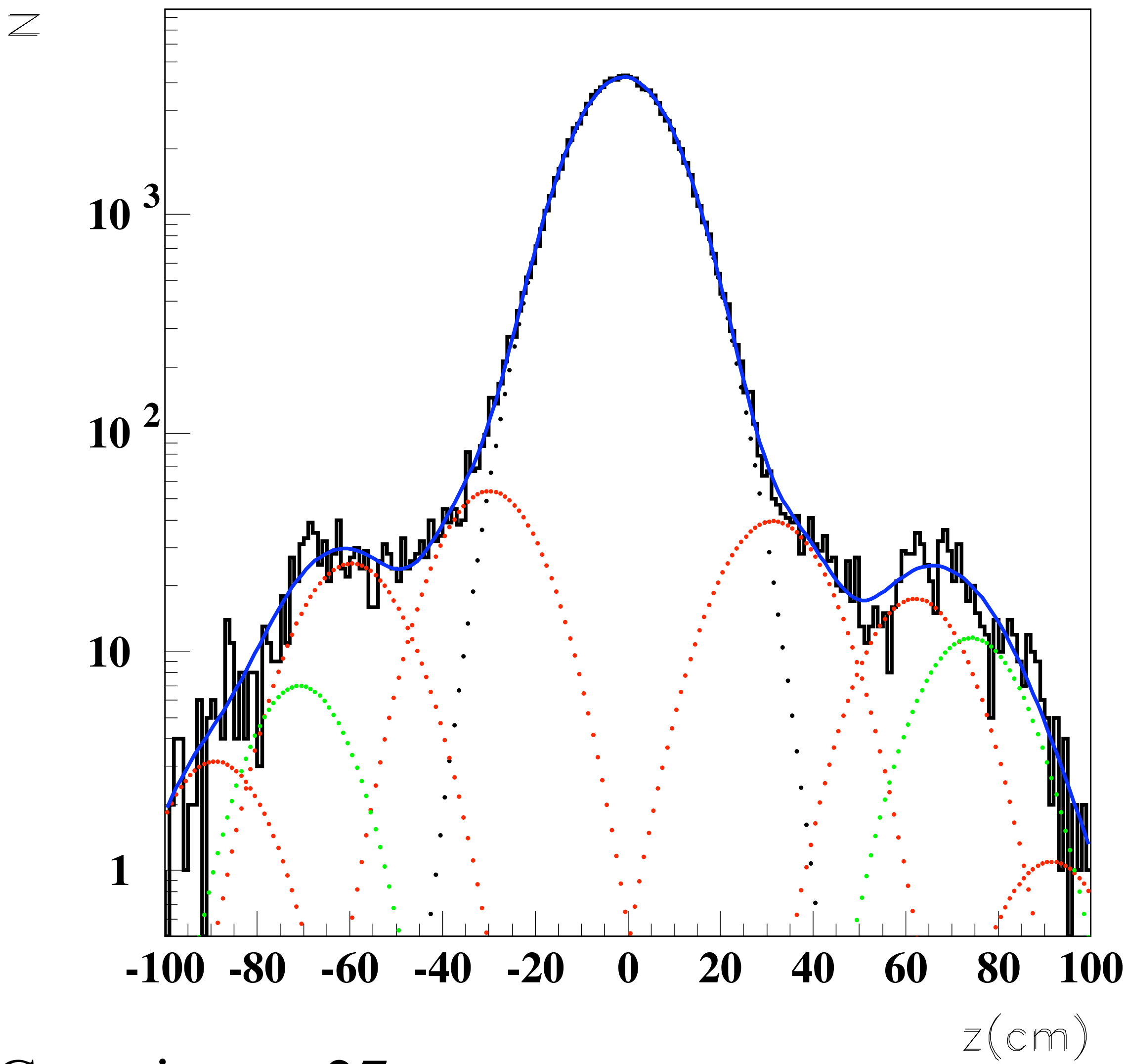
Special selection of events where event vertex bias largely removed (angular cuts)

Produce model with 9 Gaussians for underlying vertex distribution

Use the Monte Carlo simulation to tell us the relationship between true vertex and the measured vertex

Bin the data, and use Poisson statistics to compared the predictions with observations in bins of the reconstructed vertex

HERA Vertex Example



BAT fit, 9 Gaussians – 27 parameters