Statistical Treatment for the Razor Analysis

How to extract a limit on a multi-box counting experiment

C. Rogan, J. Lykken, M. Pierini, H. Prosper,
S. Sekmen, M. Spiropulu

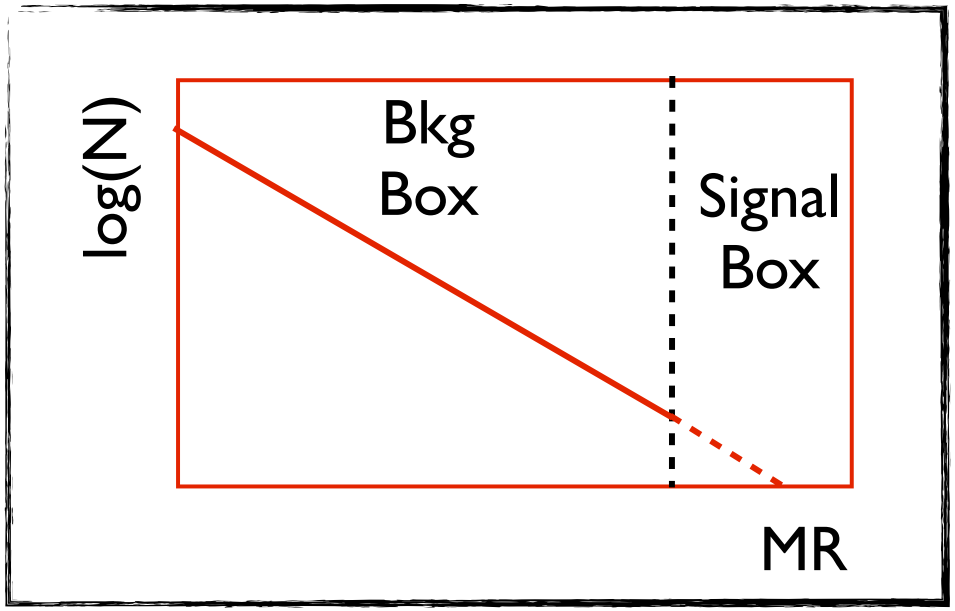
Outline

- Single Box example
- From One box to many boxes
- The full approach
- Comparison with other methods

The Single Box Case

- Let's take the simple case of an analysis with only ONE source of SM bkg
- We have one NP vs SM discriminating variable (MR). NP events cluster on the tail, while the low-MR region is NP free
- We know the shape of the MR distribution for our bkg.

- We see y events in the bkg in the Bkg Box
- We know the scaling factor b
- We can obtain an estimate $\bar{\mu}$ of the expected background μ as $\bar{\mu} = y/b$
- We see n events in the Signal Box, with an expected count of $s+\mu$
- We want to extract a PDF for s



Reference priors (I)

- In 1979 J. Bernardo established a procedure to associate to a given statistics the prior that maximizes the role of the likelihood in determining the posterior (ie. less importance to prior, more to the data)
- He started defining the distance between two functions using the Kullback-Leibler divergence of a function π from a function p

$$\boxed{D[\pi, p] \equiv \int p(\theta|x)\ \ln \frac{p(\theta|x)}{\pi(\theta)}\, d\theta}$$

- In our case p is the posterior and π is the prior. It quantifies the information gained by updated the a-priori knowledge with the result of the experiment. The “best” prior would then be the one that maximizes D[π, p]. But this would then depend on the result of the experiment (i.e. the prior is not established a-priori anymore)

Reference priors (II)

- Instead, Bernardo considers the expected distance, averaging over the possible observations from K repetitions of the same experiment

$$\boxed{I_K[\pi] \equiv \sum_{x_1=0}^{\infty} \cdots \sum_{x_K=0}^{\infty} m(x_{(K)})\, D[\pi, p(\theta|x_{(K)})],}$$

- In 1931 H. Jeffreys suggested that a non-informative prior should minimize the information to the posterior. This defines the Jeffreys prior as

$$\boxed{\pi(\theta) = \sqrt{\mathbb{E}\left[-\frac{d^2 \ln p(x|\theta)}{d\theta^2}\right]}}$$

- A known example is the prior $\pi(a) \sim 1/\mathrm{sqrt}(a)$ when a is Poisson mean. For all the problems of interest the reference prior reduces to the Jeffreys prior

Reference priors (III)

$y=\bar{\mu}b$ $\xrightarrow{\text{scaling factor } b \pm \delta_b}$ measure n

bkg estimate $\bar{\mu} \pm \delta\mu$

The likelihood: $p(n|\, s, \mu\,) = \dfrac{(\,s + \mu)^n}{n!}\, e^{-\,s\,-\mu}$

The prior on μ: $\pi(\mu\,|\,s) = \pi(\,\mu) = \dfrac{b(b\mu)^{y-1/2}\, e^{-b\mu}}{\Gamma(y+1/2)}$

where $\bar{\mu} = \dfrac{y+\frac{1}{2}}{b}, \quad \delta\mu = \dfrac{1}{\sqrt{y+\frac{1}{2}}}.$

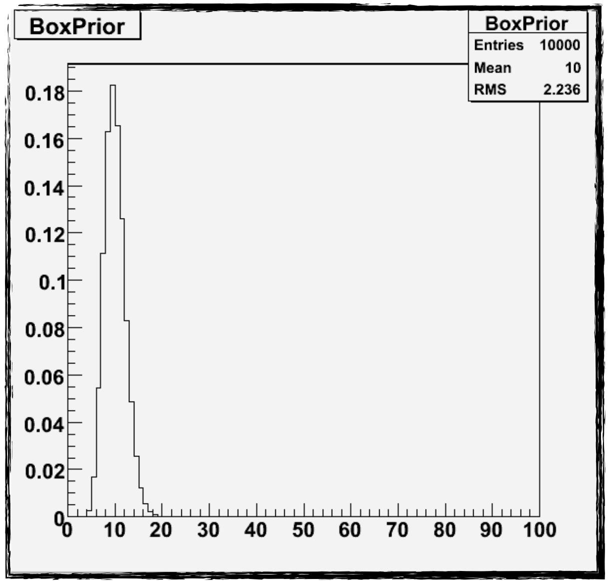
The prior on n: a Jeffreys prior ~1/sqrt(μ+s)

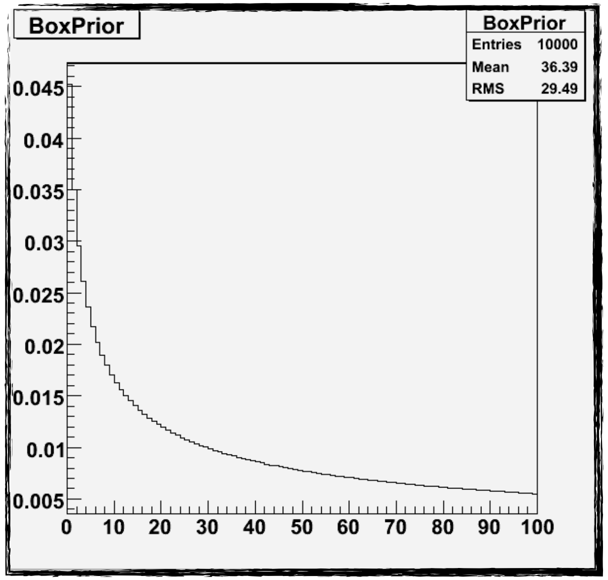
An example (I)

- We take

$$
\begin{array}{l}
- \ n = 13 \\
- \ \overline{\mu} = 10 \\
- \ b = 2.0 \pm 0.2
\end{array}
$$

- We start from the prior for μ

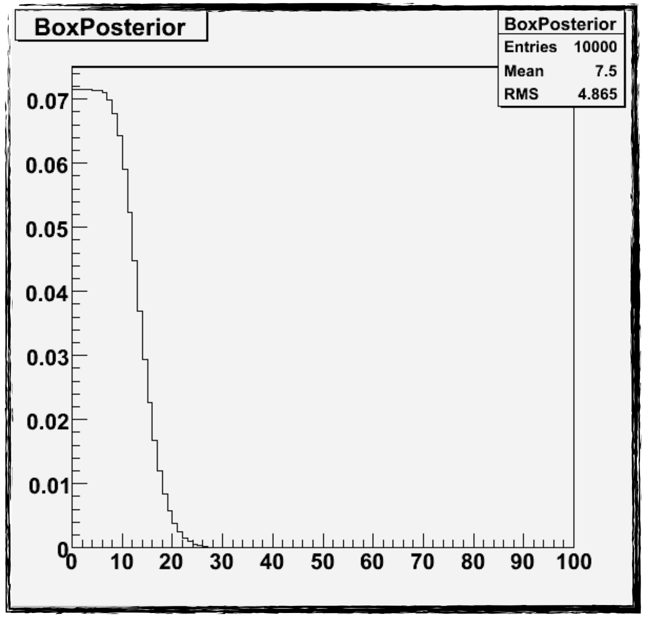


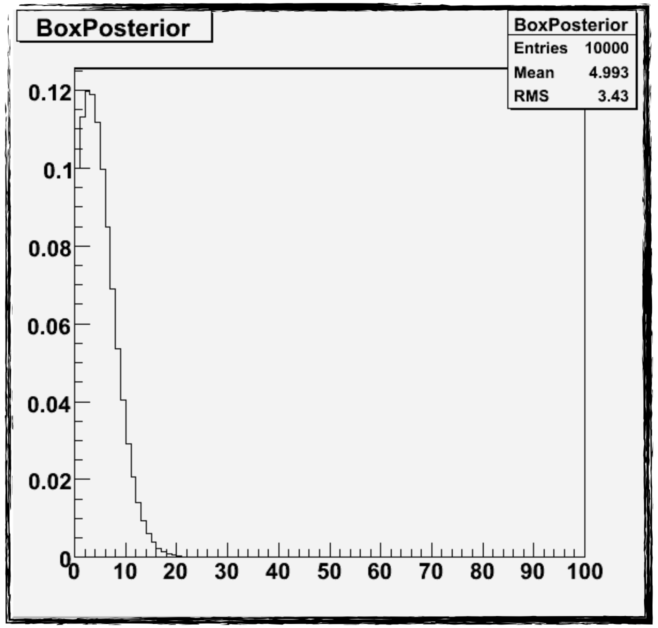


- For a given μ, the prior on s is $p(s|\mu) \sim 1/\mathrm{sqrt}(s+\mu)$

An example (II)

- We need to consider also the likelihood in the signal box



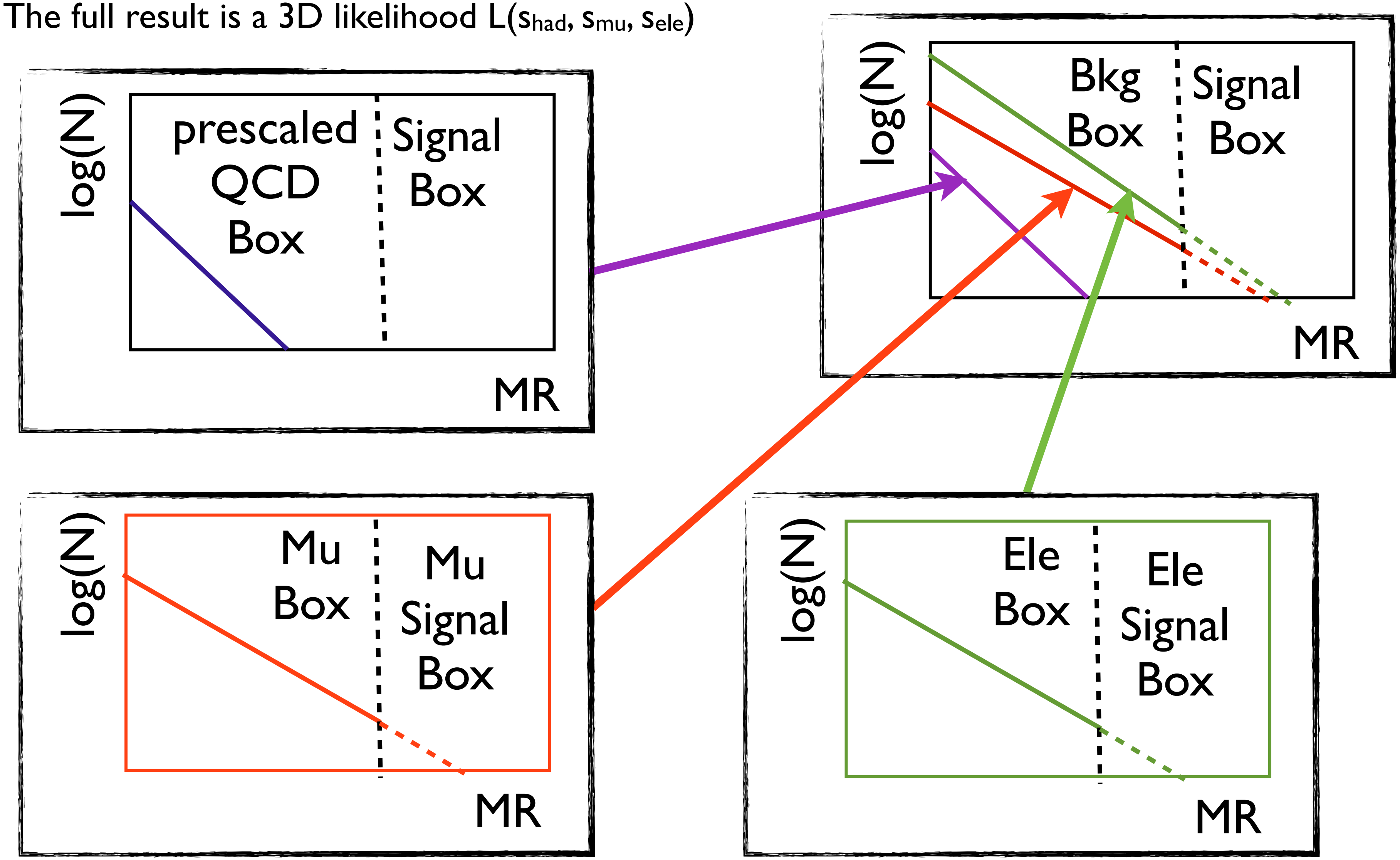


- The posterior on s is then obtained as the integral

$$p(s|n) \sim \int p(n|s,\mu)p(s|\mu)p(\mu)d\mu$$

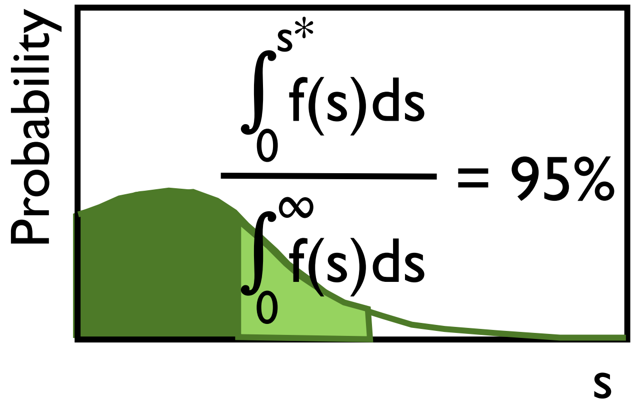
The Razors boxes

- The analysis tell us that QCD can be ignored (or take a bkg-only problem, with prediction $O(10^{-5})$)
- Had, Mu, and Ele control boxes as the previous example
- The counts are statistically independent but the parameters are related
- The full result is a 3D likelihood $L(s_{had}, s_{mu}, s_{ele})$

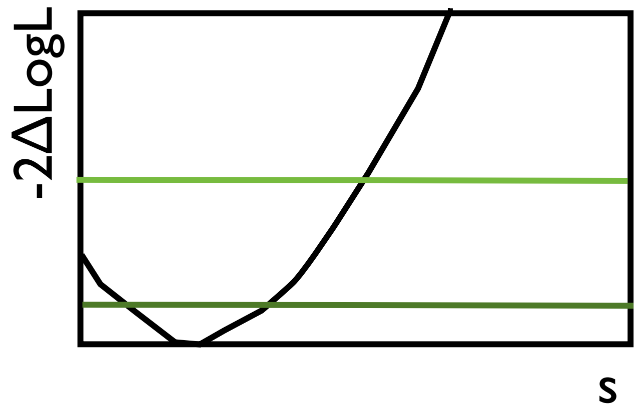


Other Approaches

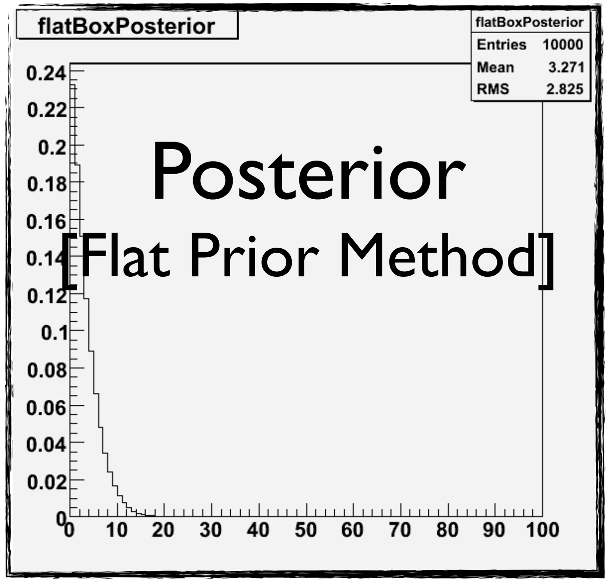
- Bayesian with Flat Prior (HEP "classic")
  - what people usually do
  - flat prior on s
  - Gaussian prior on μ (summing in quadrature stat and sys error)
  - not motivated beyond the fact that it is in the PDG

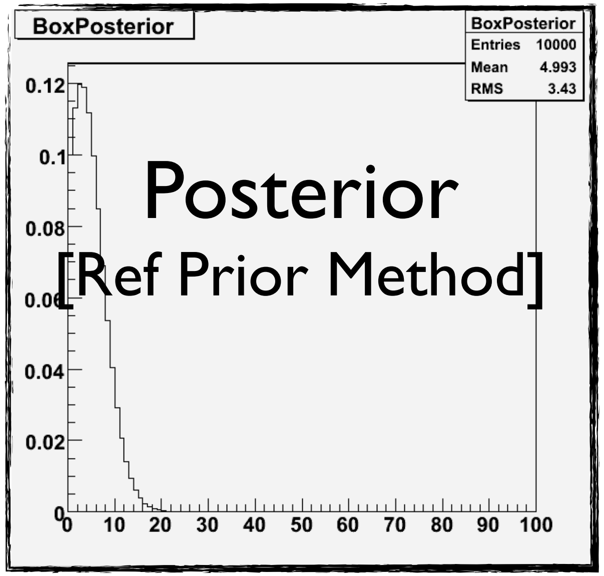


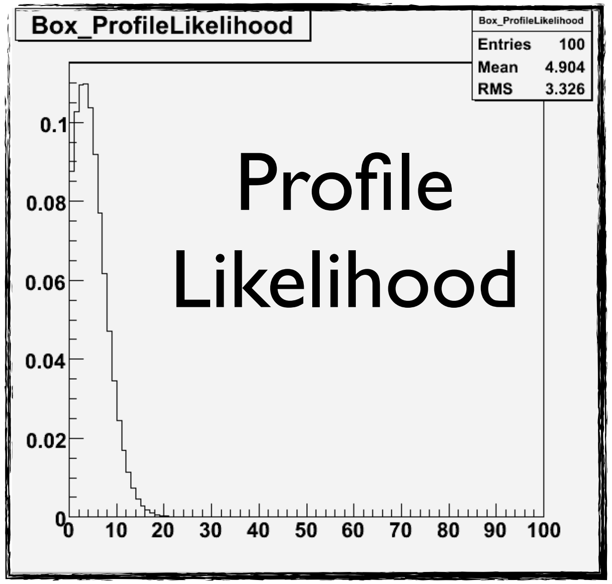
- Profile Likelihood
  - write the likelihood of s and m as L(s, μ)~Poisson(s+μ|n)P(bμ|y)
  - Get the likelihood L(s)= L(s, $\hat{\mu}$), where $\hat{\mu}$ is the value that maximizes L(s, μ) for that value of s
  - Use deltaLogLikelihood approach to do statistics inference (eg compute a limit)



Comparison of Results

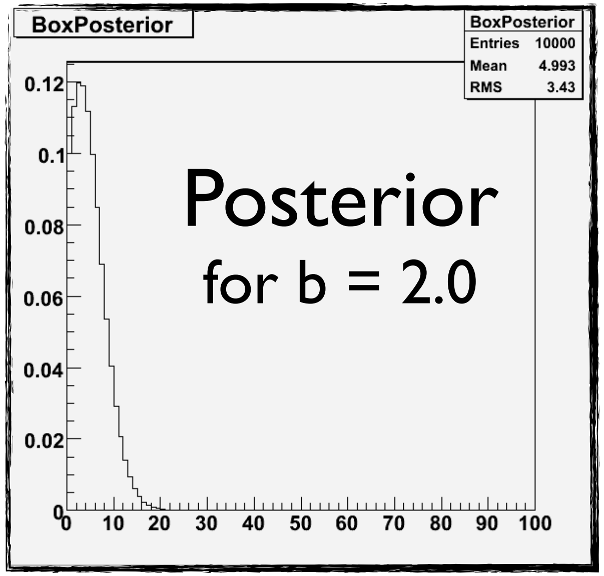




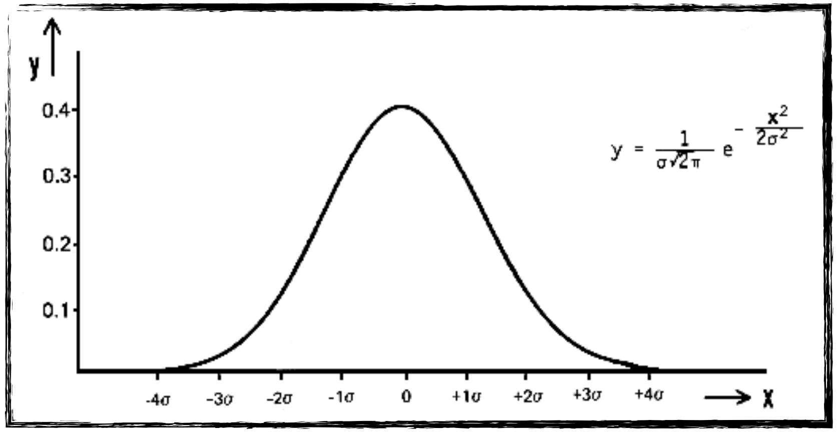


Method	95%	99%
Flat Prior	7.9	11.6
Ref Prior	10.5	14.0
Profile Lik	12.0	15.0

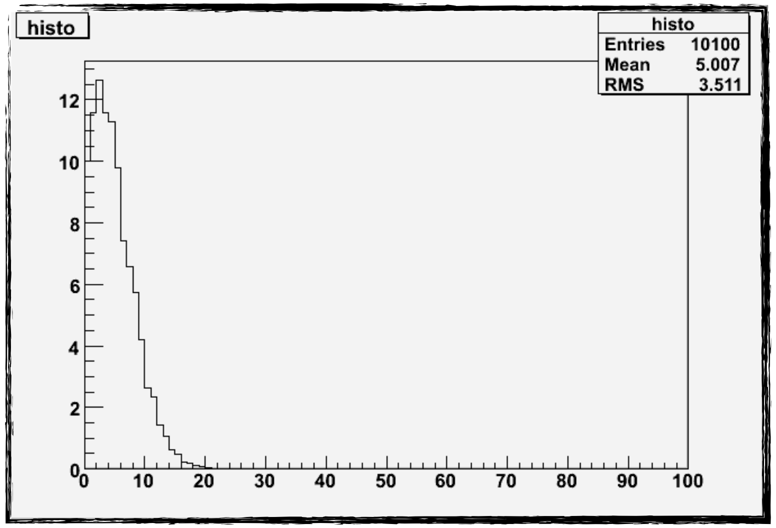
Including Systematics



- Write the posterior on s as a conditional probability on the parameter



- Multiply by the PDF of the parameter determining your systematic. In our case, use G(2.0, 0.2) for b and G(1.0,0.1) for s (Luminosity systematic)



- Multiply by the prior on the parameter and integrate

- The limit gets worse (10.6@95% Prob; 14.3@99% Prob)

Conclusions

- We intend to use a reference-prior based approach to set the exclusion limit
- Good properties in the context of both Bayesian and Frequentist treatment (see literature)
- It provides an answer close to the frequentist limit
- It provides a consistent framework to incorporate systematics
- Using ROOT-based C++ code. Can be distributed withe examples. Plan to add it to ROOTSTAT

Backup

Our Case

There are two complications in our case:

- we have a multi-box problem, meaning many s_i, many b_i, and many m_i
- The background in the "last" box is $b_N = \Sigma_{i<N} (\mu_i\ b_i/\mu_{i,N})$, $\mu_{i,N}$ being the scaling factor from the sideband of the box i to the signal box N (not the scale from the signal box n to its sideband)

What we know

- up to now the m parameters have been considered as pure numbers
- Writing $b_N = \Sigma_{i<N} (\mu_i\ b_i/\mu_{i,N})$ we are not adding ANY unknown

Solution

- We threat the problem of the n-1 disconnected boxes first
- We multiply the posterior by the likelihood

$$p(n|\, s_N,\mu_N) = \frac{(s_N+\mu_N)^n}{n!} e^{-s_N-\mu_N}$$

- We multiply the likelihood by the prior on the b parameters (a Jeffrey's prior for a Gaussian-distributed quantity) and their likelihoods (the measurements of b we have from control samples)
- We integrate all over the b parameters