Subtracting and Fitting Histograms using Profile Likelihood

<u>Andre Nepomuceno</u>
Fernando Marroquim

Federal University of Rio de Janeiro

Outline

- *Likelihood and Profile Likelihood*
- *The profile χ^2 function for Poisson signal and background*
- *Subtracting Histograms and setting Limits*
- *Fitting signal data with profile χ^2 function*
- *Summary*

Motivation

- *Many interesting signals expected at LHC will be strongly contaminated by large backgrounds.*
- *These signals will be difficult to detect during the first years of LHC operation due to the initial low luminosity.*
- *Techniques on how to extract signal information with very low statistics will be useful in such situations.*

Likelihood and Profile Likelihood I

Let us assume a counting experiment where signal and background are completely *independent* and both obey Poisson distributions, with *k* data events and *m* background events previously estimated by MC. The likelihood is

$$L(s, b; k, m, \tau) \propto (s+b)^k e^{-(s+b)} (\tau b)^m e^{-\tau b}$$

where *s* and *b* are related to signal and background distributions, and

$$\tau = \frac{\mathscr{L}_{MC}}{\mathscr{L}_{\exp}}$$

$\mathscr{L}_{MC}$ **and** $\mathscr{L}_{\exp}$ **are the MC and experimental luminosities, respectively.** $\tau > 0$**, but when one has limited computer resources,** $0 < \tau < 1$**. Our knowledge about the background increases when** $\tau \to \infty$.

Likelihood and Profile Likelihood I

In order to obtain a b independent likelihood, one can find the maximum likelihood estimator of the background

$$\frac{\partial}{\partial b} \log L(s, b; k, m)\Big|_{b=\hat{b}} = 0$$

$$\hat{b} = \max\left(0, \frac{k+m-(1+\tau)s+\Delta(s)}{2(1+\tau)}\right)$$

$$\Delta = \sqrt{\left[k+m-(1+\tau)s\right]^2+4m(1+\tau)s}$$

$\Rightarrow$ $\hat{b} \geq 0$ due to physical constraints.

Likelihood and Profile Likelihood II

Replacing b by $\hat{b}$ in $L(s, b; k, m, \tau)$, one obtains the profile likelihood (see Rolke *et al*, NIM A458, 2001)

$$L_P(s, k, m, \tau) \;\alpha\; (s + \hat{b})^k e^{-(s+\hat{b})} (\tau \hat{b})^m e^{-\tau \hat{b}}$$

From the equation above, we can obtain the most probable value of s

$$\hat{s} = max\left(0, k - \frac{m}{\tau}\right)$$

$\Rightarrow \hat{s} \geq 0$ due to physical constraints.

The profile χ^2 Function I

We want to construct an approximate χ^2 function using the profile likelihood L_P. The maximum profile likelihood ratio is

$$\lambda_P = \frac{L_P(s, k, m, \tau)}{L_P(\hat{s}, k, m, \tau)}$$

According to maximum likelihood ratio theorem

$$\chi^2_P \sim -2 \log \lambda_P$$

$$\chi^2_P = 2\left\{(s-\hat{s})+(\tau+1)\left(\hat{b}(s)-\hat{b}(\hat{s})\right)+k \ln\left(\frac{\hat{s}+\hat{b}(\hat{s})}{s+\hat{b}(s)}\right)+m \ln\left(\frac{\hat{b}(\hat{s})}{\hat{b}(s)}\right)\right\}$$

which is the profile χ^2 function.

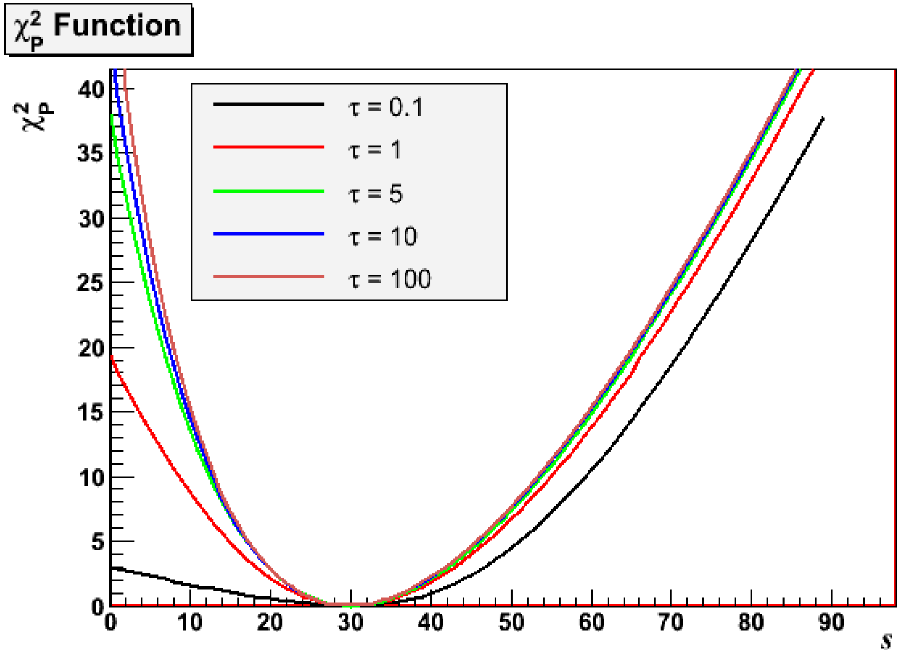
The profile χ^2 Function II

When $\hat{b} = 0$ and $m = 0$, one has

$$\chi^2_P = 2\left[s - k + k\left(\ln(k) - \ln(s)\right)\right]$$

that is the expression obtained by Cousins *et al* (see PDG 2006, pg 302), when there is no background.

The profile χ^2 Function III



In each curve $m/\tau = 10$.

Notice that the main variation of the profile χ^2 with respect to τ occurs in the region $0 < s < 30$.

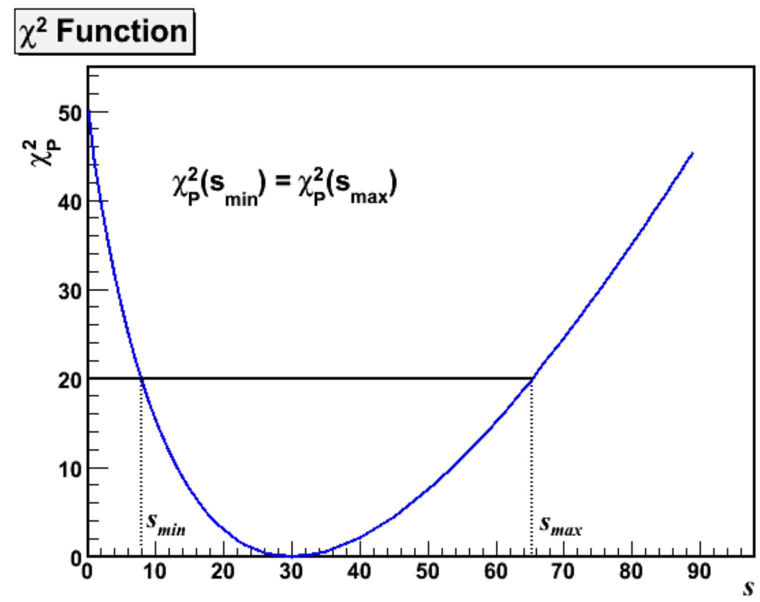
No large difference between $\tau = 5$ and $\tau = 100$

This means that $\tau > 5$ does not increase significantly our knowledge about the background. The signal statistical fluctuations dominates the behaviour of the profile χ^2 function. It will be a waist of cpu time to generate background events with $\tau > 5$.

The profile χ^2 Function IV

The profile χ^2 function is useful to obtain signal limits, by solving the equation system

$$\begin{cases} \displaystyle\int_{s_{min}}^{s_{max}} f_P(s;k,m,\tau)\,ds = 1-\alpha \\ \chi^2_P(s_{min}) = \chi^2_P(s_{max}) \\ 0 \le s_{min} \le s_{max} \end{cases}$$



Where f_P is the normalized probability density of s obtained from $L_P(s;k,m,\tau)$, and α is related to the desired confidence level.

Subtracting and Fitting Histograms

There are two ways of extracting signal information:

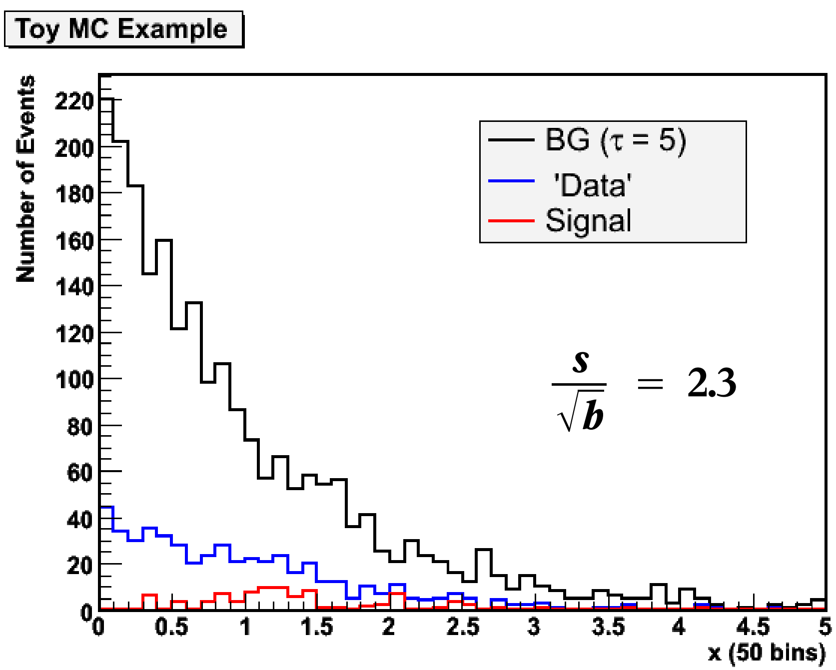
- **one is subtracting two histograms:**
  data histogram - background histogram
  In this case we do not know the signal shape.

- **the other one is using the profile χ^2 function to fit a function of knowing shape.**

Subtracting Histograms and setting Limits I

A MC Example

- **500 Toy MC data events generated**
- **Background: *B ~Exp(-x)* (450 events)**
- **Signal: *S ~ Gauss*(1.2, 0.2) (50 events)**
- **BG: 2250 events corresponding to $\tau = 5$**



- **One wants to extract the red histogram information from the blue one.**
- **The red histogram shows the signal content in the "data"**
- **Taking $\tau > 5$ would be waist of computer time, since this does not increase substantially our knowledge about the background.**

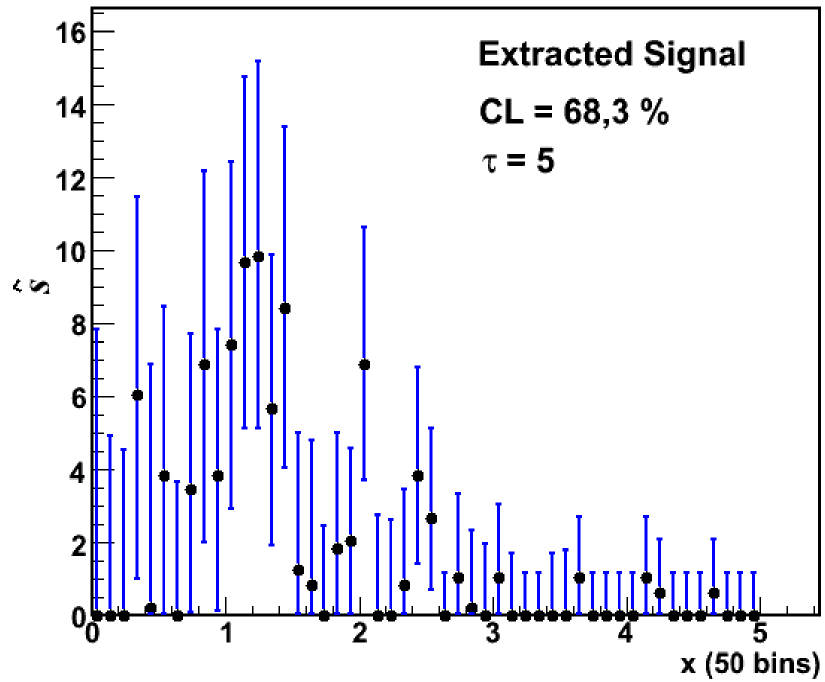
Subtracting Histograms and setting Limits II

The signal estimated for each bin and its error are given by

$$\hat{s}_i = max\left(0, k_i - \frac{m_i}{\tau}\right)$$

$$\begin{cases} \displaystyle\int_{s_{min}}^{s_{max}} f_P(s_i; k_i, m_i, \tau)\, ds = 1-\alpha \\ \chi^2_{P_i}(s_{min}) = \chi^2_{P_i}(s_{max}) \\ 0 \le s_{min} \le s_{max} \end{cases}$$

CL = 68 , 3%



No bins with negative values due the to the constraint $s \geq 0$.

Subtracting Histograms and setting Limits III

The "subtraction quality" can be obtained approximately by doing

$$\chi^2_P/ndof = \frac{1}{N}\sum_{i=1}^{N}\chi^2_{P_i}$$

where *N* is the number of bins. If profile χ^2_P/ndof goes to zero, this means no signal.

Another way is calculating the *p*-value for this case, which is given by (Cowan, *Statistical Data Analysis*, pag 62)

$$P = \int_{\chi^2_P}^{\infty} g(z;n_d)\,dz \qquad \text{where} \qquad g(z;n_d) = \frac{z^{\frac{n}{2}-1}e^{\frac{-z}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

Subtracting Histograms and setting Limits IV

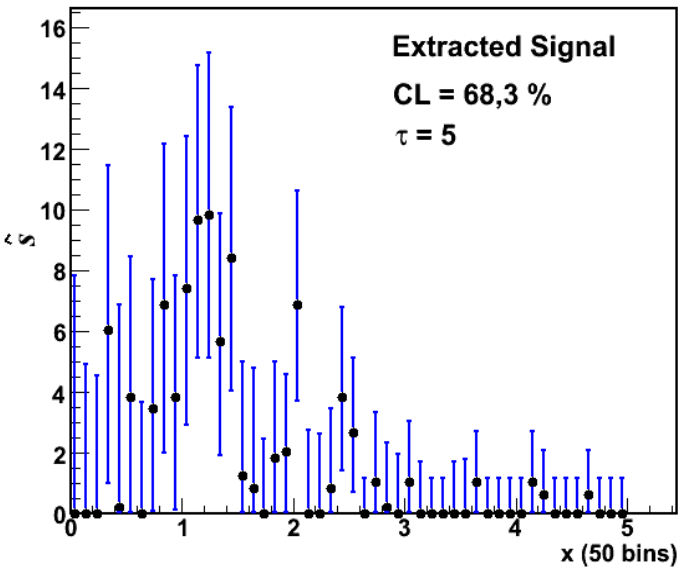
Taking into account just the bins between x = 0.85 and x = 2.5, we have, under the hypothesis that there is no signal (*s* = 0)

$$P = \int_{\chi_P^2}^{\infty} g(z;n_d=18)\,dz = 0.022$$

or, taking just the bins between x = 0.85 and x = 2.0 (*N* = 13 bins)

$$P = \int_{\chi_P^2}^{\infty} g(z;n_d=13)\,dz = 0.009$$

One has a probability 0,022 (0,009) of the data and background events be compatible. This is a strong signal evidence.



Fitting Histograms I

One can use the profile χ^2 to fit histograms, when the signal shape is known.

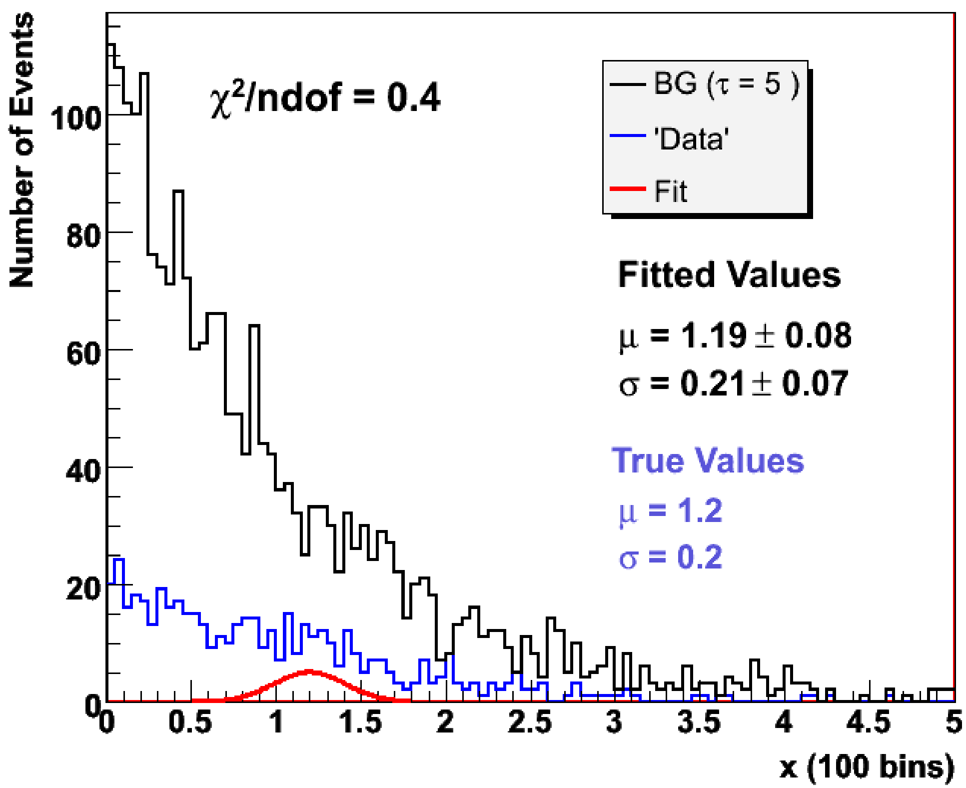
The χ^2_P is given by the sum of all $\chi^2_{P_i}$ correspondenting to N bins

$$\chi^2_P = \sum_{i=1}^{N} \chi^2_{P_i}$$

- The s_i in the ith bin is replaced by the function $f(x_i, \vec{\theta})$ to be fitted , where x_i is the function argument at the ith bin, and $\vec{\theta}$ is the vector of parameters to be fitted.
- It is not necessary to know the background shape.
- Less parameters to be fitted, no parameters related to the background.
- Even the bins with contents equal to zero contribute the profile χ^2 function

Fitting Histograms II

- In this example, one fits a Gaussian to the same MC data that was showed before.
- In this plot, the MC data are distributed in a histogram of 100 bins
- Good agreement between fitted and true values



The quality of the fit can be done using the usual χ^2_P and *p*-values procedures.

Summary

- The profile χ^2 function is useful to extract signal information even of unknown shape.
- The number of background events is estimated by MC. No improvement by taking $\tau > 5$.
- It is not necessary to fit the background distribution thus it is not necessary also to know the background shape.
- The signal parameters errors incorporated automatically the background fluctuations due to our knowledge about it represented by the τ values.
- It is not necessary to exclude bins with content equal zero as in the Least Square Method. This is important for example in long tail data.
- It is only necessary to fit the signal function parameters $\Rightarrow$ reduce the number of parameters to be fitted.

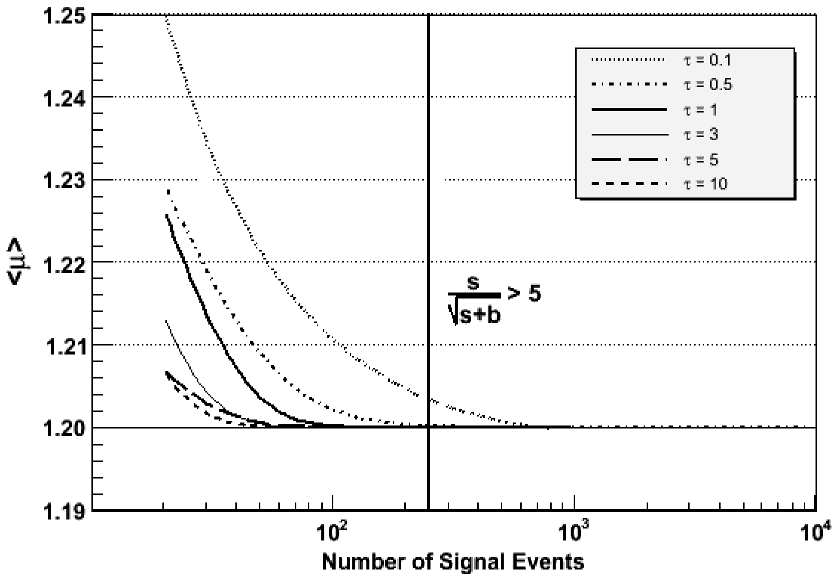
Backup Slides I

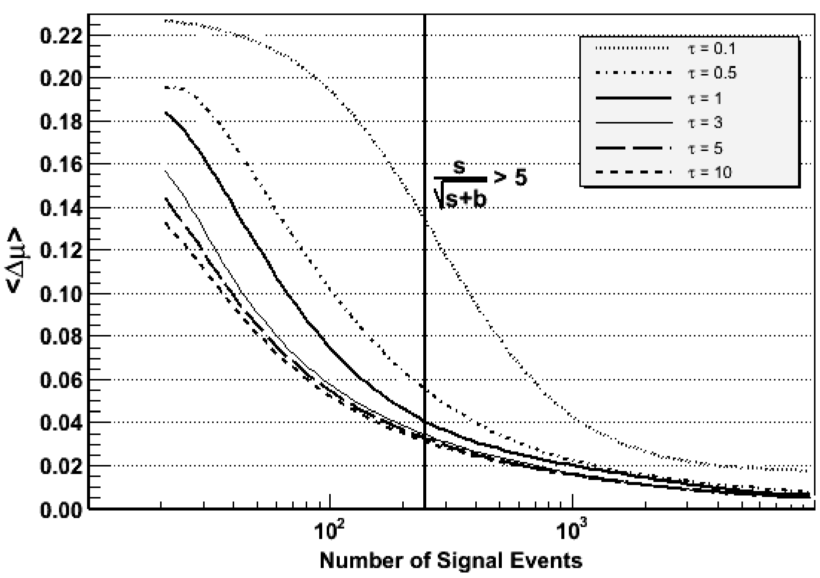
Systematic Study

- Toy Monte Carlo: generated number of signal events from 20 to 10 000 in a histogram of 100 bins.
- For each number of entries, were generated 10 000 sets of points and fitted each of them.
- The ratio between signal and total events was kept constant equal to 0.1
- The study was performed for different values of τ and different signal and background distributions.
- Comparisons with LSM were done and we obtained faster convergence and no fluctuations around the true values.

Backup Slides II

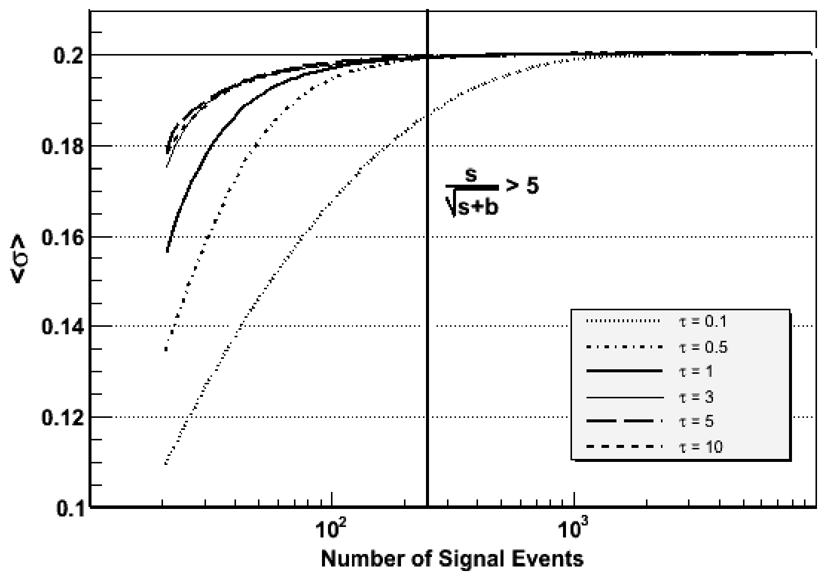
Some results for exponential background and Gaussian signal
Fitting the Gaussian mean and its fluctuation.

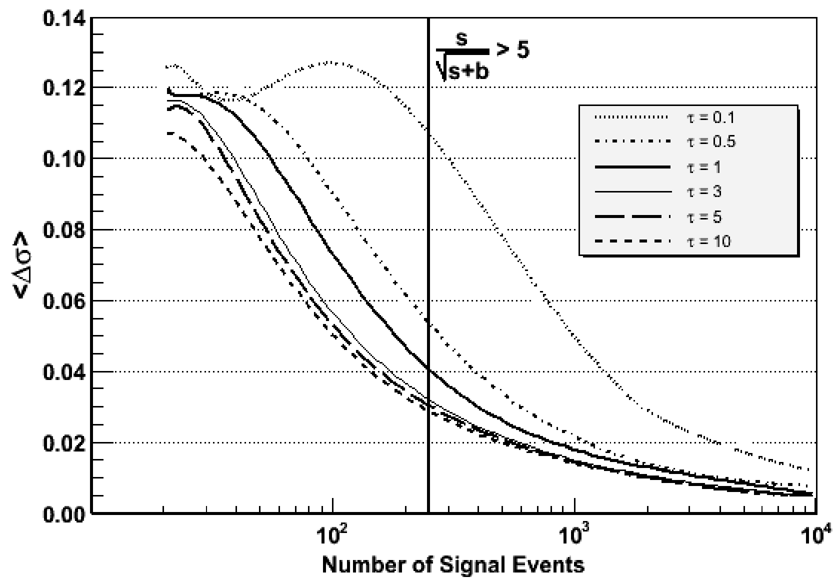




Backup Slides III

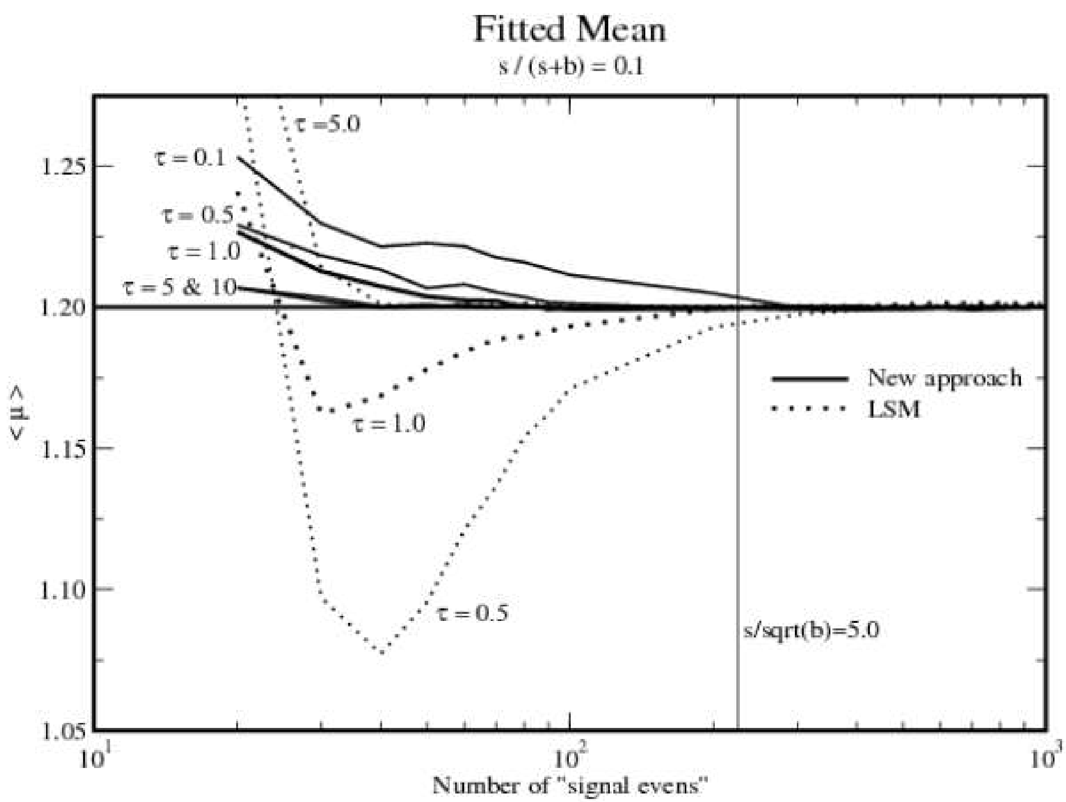
Some results for exponential background and Gaussian signal
Fitting the Gaussian width and its fluctuation.





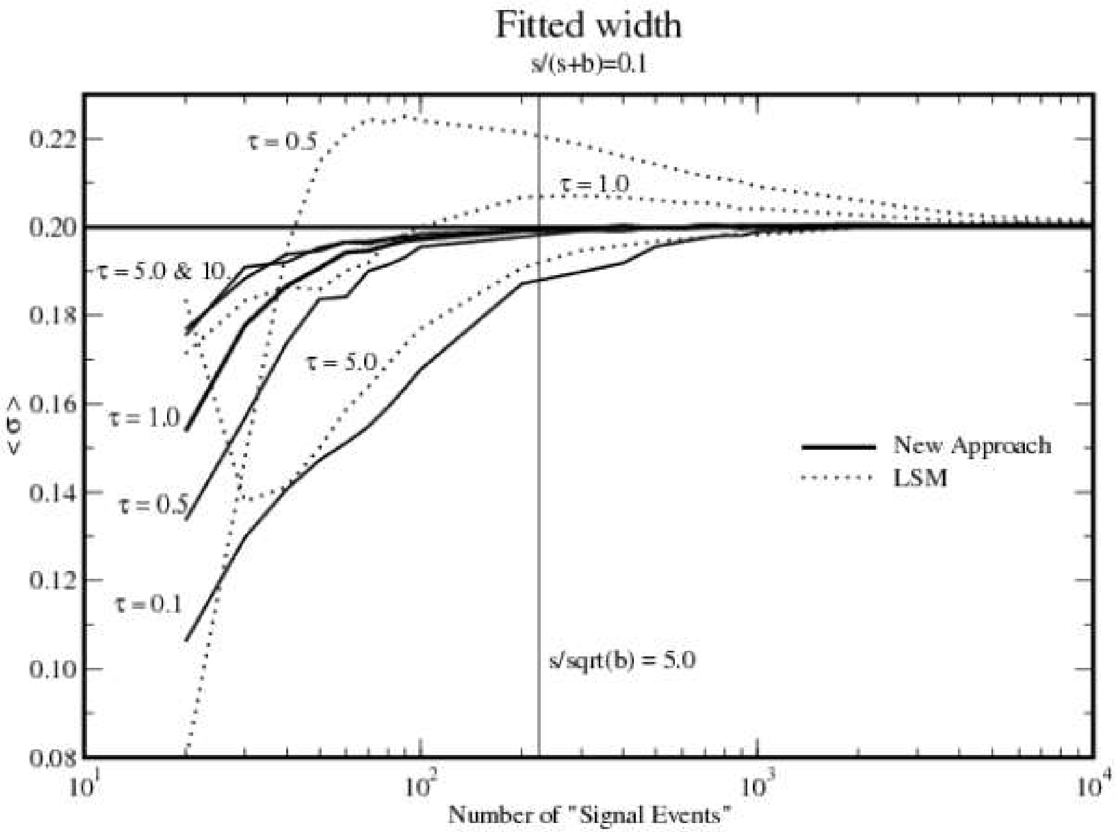
Backup Slides IV

Comparison with LSM
Fitting the Gaussian mean



Backup Slides V

Comparison with LSM
Fitting Gaussian width



Backup Slides VI

For more details on profile likelihood, see S. A. Murphy et al, *Journal of American Statistical Association*, Jun 2000; 95, 450.