Techniques for setting limits: frequentist approach

Fabrice Couderc
fabrice.couderc@cea.fr

Overview

- Introduction
- Minimizing systematic effects using profile likelihood techniques
- A word on CL_s
- Conclusions

dapnia SPP
cea
saclay

What do we want to know ?

- We observe $\boldsymbol{x}$, realization of the random variable X:
  - Is X due to "background only" (hypothesis $\mathcal{H}_b$) ?
  - Is X due to or "signal + background" (hypothesis $\mathcal{H}_{s+b}$) ?
- Likelihood: $LH(\mathcal{H}\,;\boldsymbol{x}) \propto P(X=\boldsymbol{x} \mid \mathcal{H})$. [requires: $P(X|\mathcal{H})$]
- Law of likelihood: We have evidence that $\boldsymbol{x}$ supports $\mathcal{H}_A$ over $\mathcal{H}_B$ if $LH(\mathcal{H}_A\,;\boldsymbol{x}) > LH(\mathcal{H}_B\,;\boldsymbol{x})$. The strength of that evidence is measured by the ratio **Q**:

$$\mathbf{Q} = LH(\mathcal{H}_A\,;\boldsymbol{x})/LH(\mathcal{H}_B\,;\boldsymbol{x})$$

Q > **1** supports $\mathcal{H}_A$ over $\mathcal{H}_B$

Q < **1** supports $\mathcal{H}_B$ over $\mathcal{H}_A$

- Use this **ratio Q** to answer the questions:
  - Are data compatible with background ? To which extent ?
  - (If yes) Which signal models are compatible with data ?
  - (If yes) Which signal models are **not** supported by data ?

Program

- We have: a sensitive test Q = $LH(\mathcal{H}_{s+b}\ ;\mathbf{x})/LH(\mathcal{H}_b\ ;\mathbf{x})$ and its realization in actual data: q_{obs}

- To quantify the strength of the evidence, whether $q_{obs}>1$ (signal like) or $q_{obs}<1$ (background like), we miss two probabilities :
  - $P_b\ (Q > q_{obs}) = 1 - CL_b$: probability that an experiment with background only would fluctuate as "high" as in actual data.
  - $P_{s+b}(Q < q_{obs}) = CL_{s+b}$: probability that an experiment with signal +background would fluctuate as "low" as in actual data.

- The "Frequentist" solution: two sets of MC experiments
  - MC experiments following $\mathcal{H}_{s+b} \Rightarrow P_{s+b}(Q < q)$
  - MC experiments following $\mathcal{H}_b \Rightarrow P_b\ (Q < q)$

Some likelihood of interests

Know only the expected number of events (good old days…)

Expect μ under hypothesis $\mathcal{H}$

Observe: N

pure counting

$$LH(N;\mu)=e^{-\mu}\frac{\mu^{N}}{N!}$$

Know shapes + expected number of events (usual one)

Shape follows the (normalized) continuous distribution with ∫f =1 (f = pdf of X)

"unbinned"

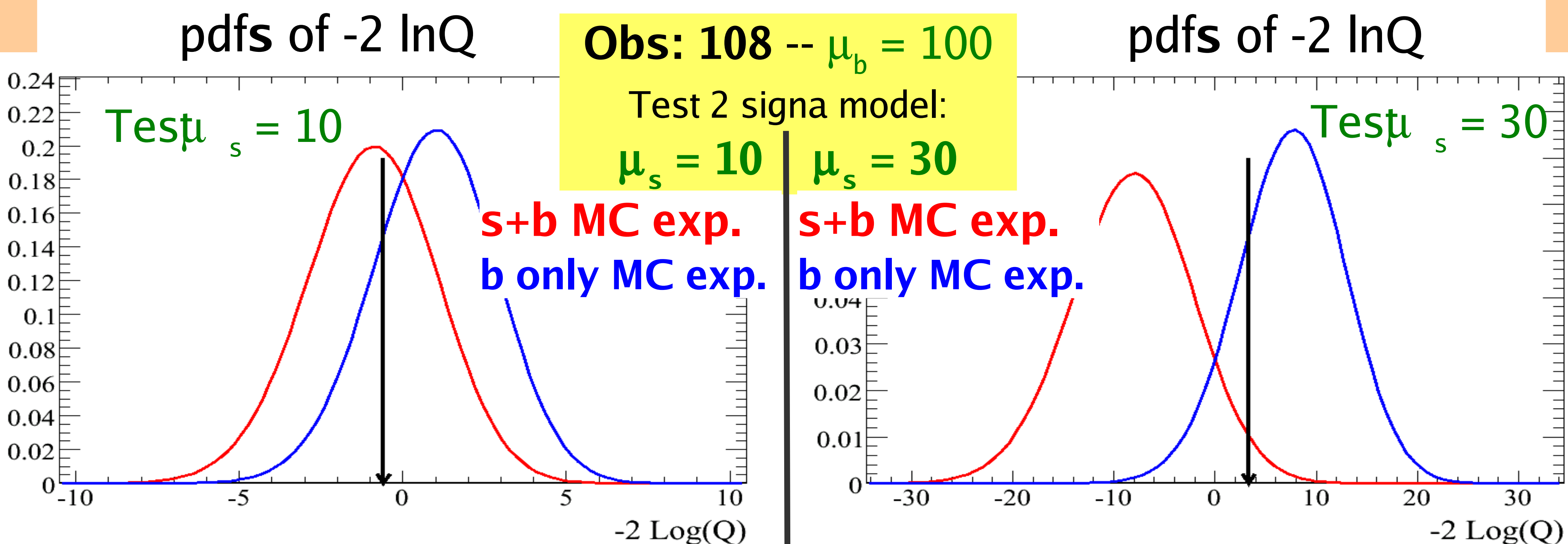
$$LH(x=(x_1,\ldots,x_N);\mu,f)=e^{-\mu}\frac{\mu^{N}}{N!}\prod_{i=1}^{N}f(x_i)$$

"binned"version

$$LH(N=(N_1,\ldots,N_{nbin});\mu=(\mu_1,\ldots,\mu_2))=\prod_{ib=1}^{nbin}e^{-\mu_{ib}}\frac{\mu_{ib}^{N_{ib}}}{N_{ib}!}$$

First example

- Technically, one uses -2 lnQ:
  - Technically more convenient (no more exponentials, $\prod \rightarrow \sum$)
  - -2 ln LH follows approximately a χ^2 distribution
- Compute -2 lnQ prob. density function (pdf) for "s+b"exp., "b only"exp.:
  - Compare them to data: CL_{s+b}[obs] , 1-CL_b[obs]
  - Get expected sensitivities (using q_{median}): $<CL_{s+b}>_b$, $<1-CL_b>_{s+b}$
- Example, counting experiment (using corresponding LH)



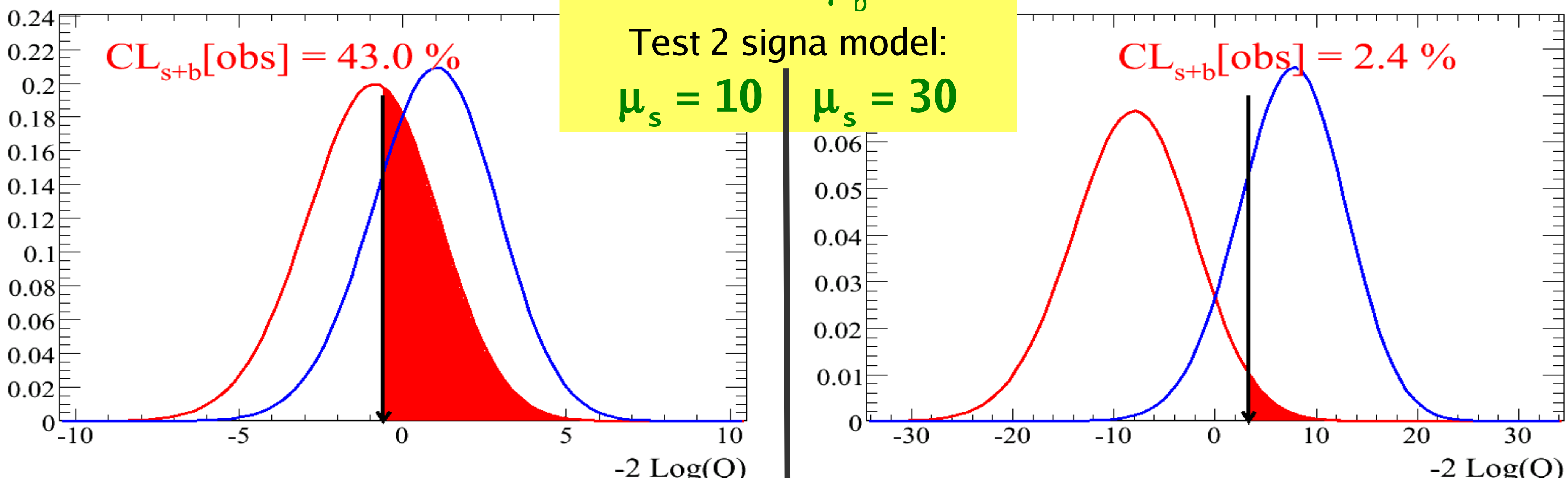
First example

- Technically, one uses -2 lnQ:
  - Technically more convenient (no more exponentials, $\prod \to \sum$)
  - -2 ln LH follows approximately a χ^2 distribution
- Compute -2 lnQ prob. density function (pdf) for "s+b" exp., "b only" exp.:
  - Compare them to data: CL_{s+b}[obs] , 1-CL_b[obs]
  - Get expected sensitivities (using q_{median}): $\langle CL_{s+b}\rangle_b$, $\langle 1\text{-}CL_b\rangle_{s+b}$
- Example, counting experiment (using corresponding LH)

Obs: 108 -- μ_b = 100

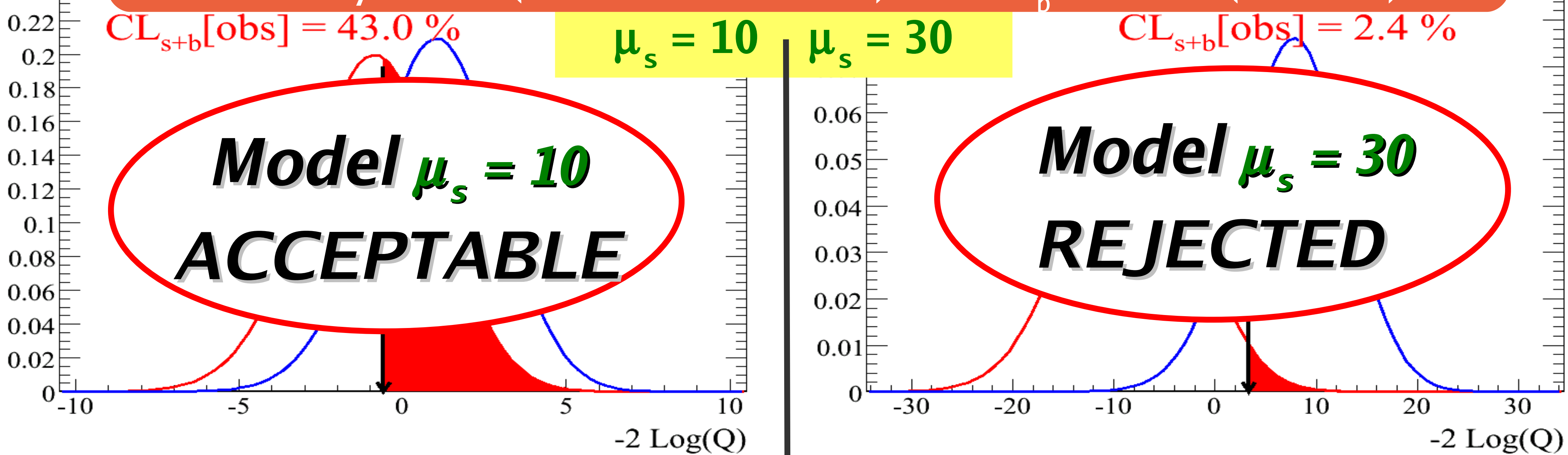
Test 2 signa model:

μ_s = 10 | μ_s = 30



First example

- Technically, one uses -2 lnQ:
  - Technically more convenient (no more exponentials, $\prod \rightarrow \sum$)
  - -2 ln LH follows approximately a χ^2 distribution
- Compute -2 lnQ prob. density function (pdf) for "s+b" exp., "b only" exp.:
  - Compare them to data: CL_{s+b}[obs] , 1-CL_b[obs]
  - Get expected sensitivities (using q_{median}): $<CL_{s+b}>_b$, $<1-CL_b>_{s+b}$
- Standard numbers:
  - The hypothesis s+b is rejected @ 95 (90)% CL : CL_{s+b} < 5 (10)%
  - Discovery @ 5σ (Observation @3σ) : 1-CL_b < 6 10^{-7} (2.7 10^{-3})



What about systematics ?

- Usually, background and signal models not perfectly known. They depend on some uncertain parameters that we do not really care about: nuisance parameters.
- Including systematics in the limit: For each MC experiments, nuisance parameters are randomly moved by $\pm 1\ \sigma$ at the generation. -2lnQ is then computed using the mean of nuisance parameters.
- Example:
  - Background level known by theory up to 5 %:

$$b = 100 \pm 5$$

  - For each MC exp, generate a new value of μ_b according to the law $\mathcal{N}(100,\mathbf{5})$

Profile Likelihoods

- When systematics are large, data can help! One can use the shape of a variable to constraint some nuisance parameters in a region free of signal!
- Include in the likelihood the information from theorists:

$$LH(x_1,..,x_N,\underline{b};\mu_s,\mu_b,f)=e^{-(\mu_s+\mu_b)}\frac{\mu_s+\mu_b^{\ N}}{N!}\prod_{i=1}^{N} f_{s+b}(x_i) \times e^{-\frac{1}{2}\frac{(b-\mu_b)^2}{\sigma_b^2}}$$

usual statistical part

Constraint on μ_b

This requires some distribution !

Profile Likelihoods

- When systematics are large, data can help! One can use the shape of a variable to constraint some nuisance parameters in a region free of signal!
- Include in the likelihood the information from theorists:

$$LH\left(x_1,..,x_N,\underline{b};\mu_s,\mu_b,f\right)=\underbrace{e^{-(\mu_s+\mu_b)}\frac{\mu_s+\mu_b{}^N}{N!}\prod_{i=1}^N f_{s+b}\left(x_i\right)}_{\text{usual statistical part}} \times \underbrace{e^{-\frac{1}{2}\frac{(b-\mu_b)^2}{\sigma_b^2}}}_{\text{Constraint on } \mu_b}$$

usual statistical part

Constraint on μ_b

This requires some distribution !

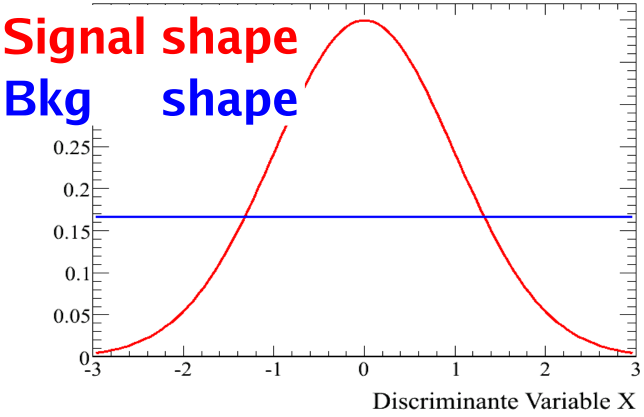
- Maximizing LH vs μ_b we obtained the

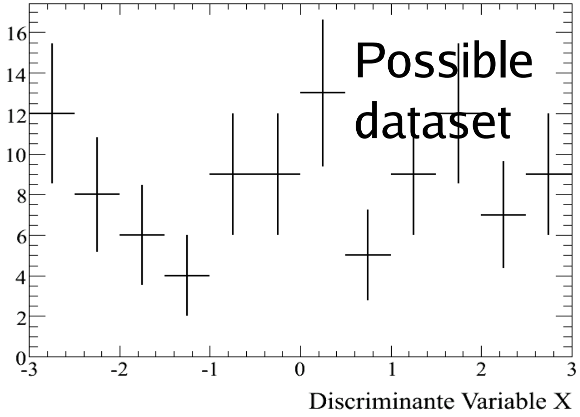
  "Profile Likelihood": $LH_P(\mu_s,f) = \max\_\mu_b\ LH(\mu_s,\mu_b,f)$

- Use LH_P to construct Q : $Q_P = LH_P(\mu_s,f) / LH_P(0,f)$
- For each MC experiments, LH is maximized (therefore this maximization depends on the hypothesis s or s+b!).

A more sophisticated example

- Search a gaussian on top of a flat background



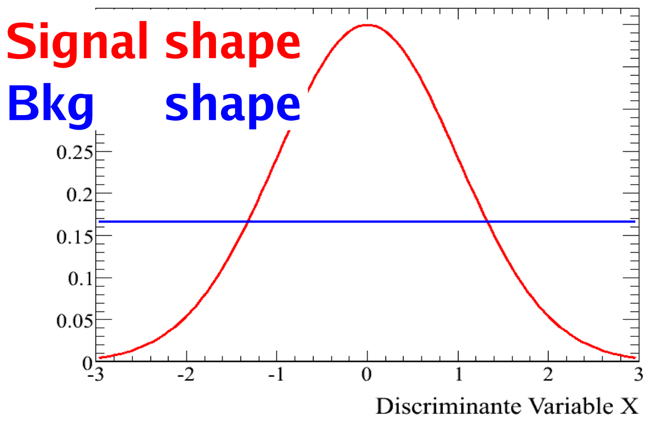


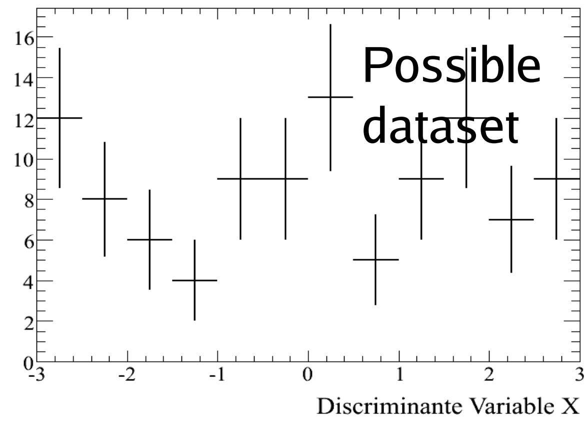
$$b = 100 \pm 5$$

- Compute the expected limit (under bkg only assumption):

A more sophisticated example

- Search a gaussian on top of a flat background





Case 1:
"not too large"syst

$$b = 100 \pm 5$$

- Compute the expected limit (under bkg only assumption):

Counting experiment : $\mu_s < 19$ @95 %CL

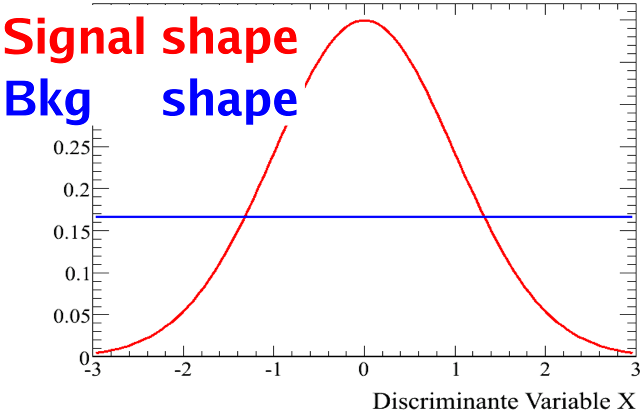
Counting experiment AND profiling : $\mu_s < 19$ @95 %CL — Profile does not improve, can not use sidebands to constrain

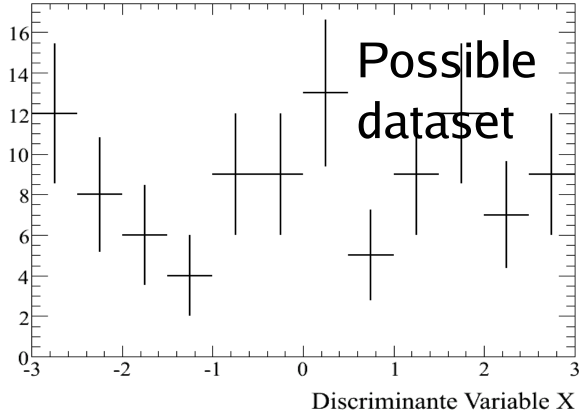
Using standard LH : $\mu_s < 15$ @95 %CL — The shapes help (better is the resolution, better is the result)

Using the shapes in LH AND profiling : $\mu_s < 15$ @95 %CL — Profiling does not help

A more sophisticated example

- Search a gaussian on top of a flat background





Case 2:
"quite large"syst

$$b = 100 \pm 20$$

- Compute the expected limit (under bkg only assumption):

Counting experiment : $\mu_s < 38$ @95 %CL

Counting experiment AND profiling : $\mu_s < 38$ @95 %CL

Using standard LH : $\mu_s < 25$ @95 %CL

Using the shapes in LH AND profiling : $\mu_s < 19$ @95 %CL

Background is more constrained by sidebands than by theory.
-> It really helps (gain 25%)!

Thoughts about profile Likelihood

- This technique begins to be used (at TeVatron at least). It looks promising though it needs to be used with care:
  - It needs confidence in the bkg shape: constrain signal region from sidebands
  - Bkg shape might be fluctuated according to a systematic (requires an alternate shape and a way to inter(extra)polate them, not so easy!)
  - It needs a prior on the syst distribution (gaussian is not always a good solution: ex theoretical cross sections). Prior is used 2 times: when generating MC exp and when minimizing -2lnQ.
- On the other way, it can constrained some external parameters (efficiencies, bkg cross sections, lumi...)

Should I build a tool myself ?

- Several tools exist, all dealing with the "binned" likelihood case (ie, inputs = binned histograms):
  - Trolke (ROOT built-in): profile LH, but only counting possible
  - mcLimit, TLimit (ROOT built-in) : no profiling. Be careful with expected sensitivity when syst are large
  - mclimit_csm www.hep.uiuc.edu/home/trj/cdfstats/mclimit_csm1/index.html (quite slow)
  - Some other codes used in Dzero: (no web page, you can ask for codes)
- In special cases, it might be easier to built a dedicated tool:
  - Your preferred likelihood might not be the usual poissonian one, unbinned LH, analytic shapes... (RooFit might help for generating MC exp.)
  - Discovery 1-CL_b: very large number of MC experiments needed, difficult in practice, some technique exist: FFT (Fast Fourier Transform), weights implemented in mclimitxx but these weights are biased when large syst.
  - shapes comparisons...
  - Profiling LH, sometimes LH minimization is analytic (cpu time gain)!
- Warning: personal code need special care. Lot of debugging, correct treatment of systematic errors.

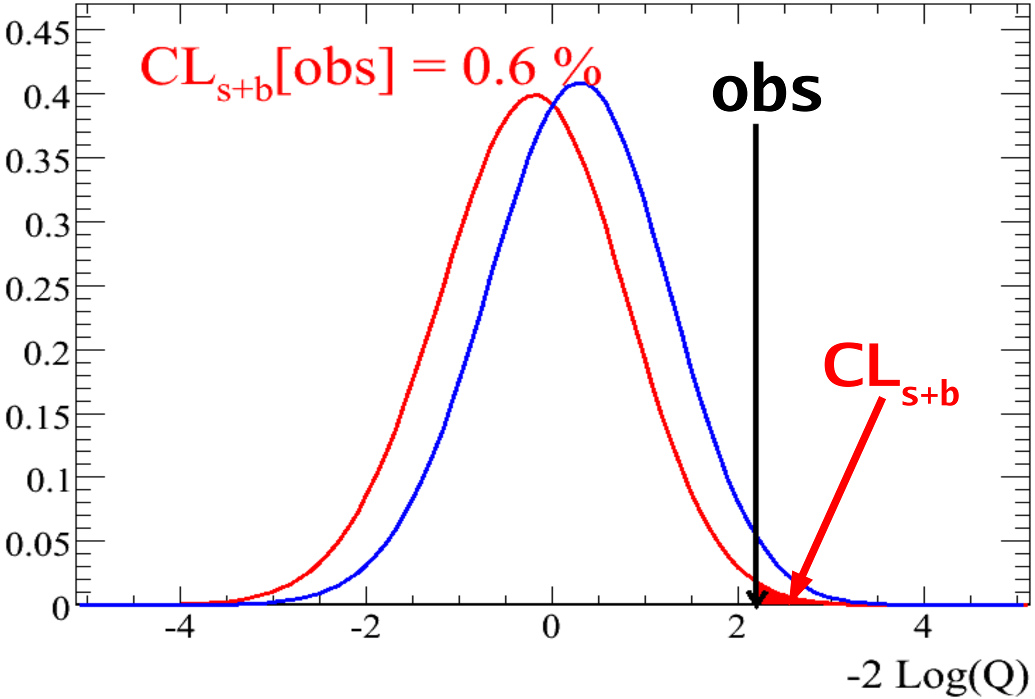
A word on CL_s

- Pure frequentist : use $CL_{s+b} < 5$ % to exclude @ 95 %CL
- Modified frequentist : Form $CL_s = CL_{s+b} / CL_b$ and use this for exclusion.
  - Limit @ 95 %CL $\Leftrightarrow$ $CL_s < 5$ % ($CL_b < 1 \Rightarrow CL_{s+b} < CL_s$)
  - Properties : conservative, not too aggressive when large downward background fluctuation.

By construction:

$CL_{s+b} < CL_b$

Even the extreme vanishing signal case can be excluded in case of low bkg fluctuation (case of all $1\text{-}CL_b > 95$ %)



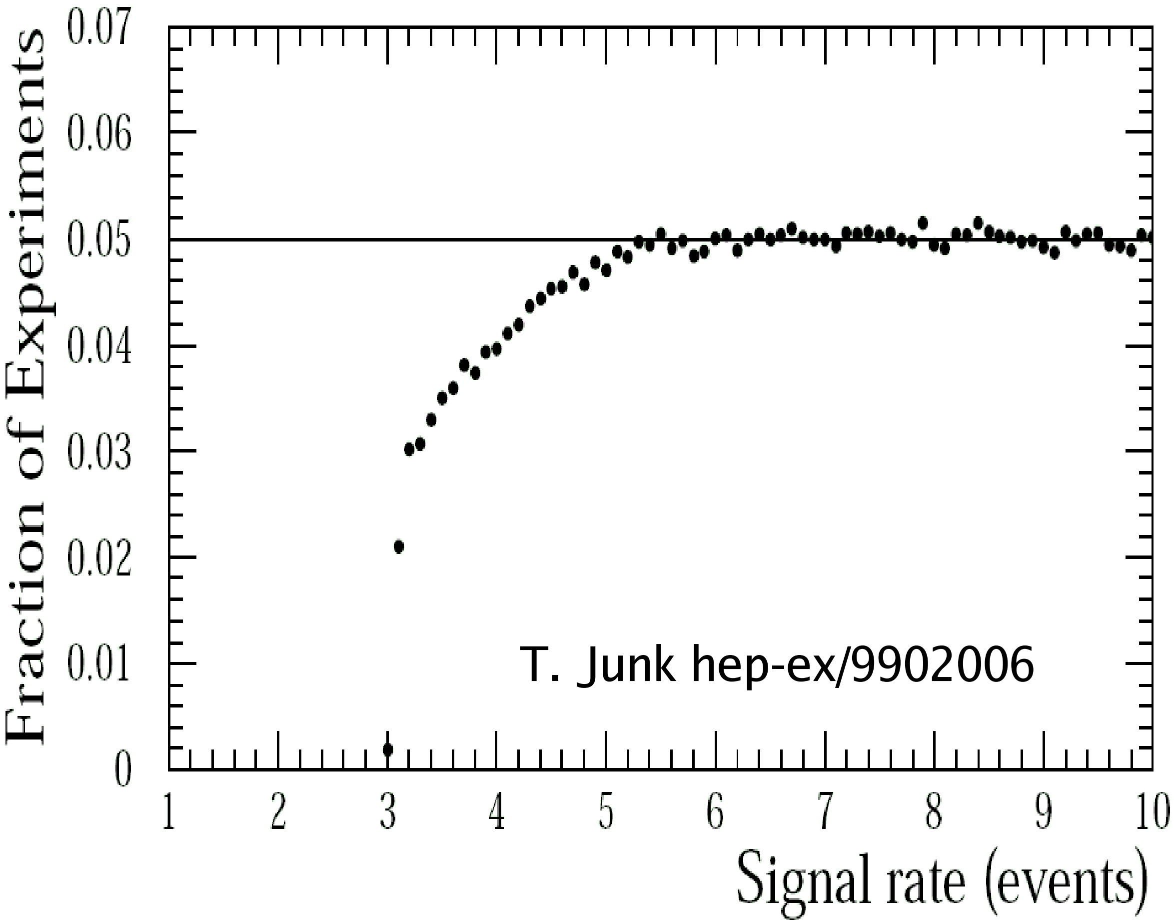
Conclusions

- Described only one approach in this talk. There are other ones Bayesian techniques, Feldman-Cousins...
- Very robust and simple approach, no need for any assumptions: no prior, no need to use the large N limit to get P(Q). Very easy to combine several channels: add LogLH!
- Profile Likelihood techniques, though quite old, are now more and more used. They allow to constrain nuisance parameters (efficiencies, bkg cross sections...) but assumes those bkg shapes are known.
- Further readings

  Confidence Level, confidence interval, limits: R. Barlow
  http://www-group.slac.stanford.edu/sluo/Lectures/Stat2006_Lectures.html
  CLs : http://www.iop.org/EJ/abstract/0954-3899/28/10/313
  mcLimit weighting : http://root.cern.ch/root/doc/TomJunk.pdf
  mclimit_csm: http://www.hep.uiuc.edu/home/trj/cdfstats/mclimit_csm1/index.html
  Profile LH : http://lanl.arxiv.org/abs/physics/0403059
  FFT : physics/9906010

A word on CLs

- Setting a limit using CL_s is not frequentist. Look at the following false exclusion rates (percentage of time a signal is excluded when it is really in data).



TLimit patch

ROOT Tlimit patch

1- CL_b obs.

CL_{s+b} obs. (kFALSE)

CL_{s+b} obs. (kTRUE)

CL_{s+b} exp. for b only

personnal code
(green curve is from TLimit)