Bayesian Upper Limits for Poisson Processes

Luc Demortier

– CDF Joint Physics Meeting –

Friday, February 4, 2005

Outline of talk:

- Critique of the CDF 7117 study;
- Objective versus subjective priors;
- Robustness analysis;
- Choice of prior for cross section measurements;
- The correlated prior method;
- Frequentist properties of the correlated prior method;
- Software and documentation;
- Summary.

Critique of the CDF 7117 Study

CDF note 7117 describes a setting with two experiments:

1. a "subsidiary" experiment to measure the signal acceptance ϵ,
2. a "primary" experiment to measure the signal rate s.

Both measurements have Poisson likelihoods and non-informative, flat priors for their respective parameters, ϵ and s.

The advantage of this setting is that it allows one to calculate the frequentist properties of an objective Bayesian method.

The disadvantage is that it is not very realistic: usually one does not have the luxury of a single measurement of the acceptance for a process of unknown rate.

In a more realistic setting, acceptance information comes from several sources: Monte Carlo studies, comparisons between data and Monte Carlo, theoretical ideas, etc. This information is then "assembled" into a probability distribution that can serve as acceptance prior for the rate measurement.

$\Rightarrow$ How does one "assemble" a prior distribution?

Objective versus Subjective Priors (1/2)

Depending on how much information is available about a parameter before the measurement, there are two approaches for choosing a prior:

1. Subjective approach
   Consider the construction of a prior for the acceptance in a cross section measurement:

$$
\left\{\begin{array}{l}\text{Auxiliary measurements}\\ \text{Monte Carlo studies}\\ \text{Theoretical ideas}\\ \text{Informed guesses}\\ \ldots\end{array}\right.
\begin{array}{l}
\Rightarrow \left\{\begin{array}{l}\text{“Central Value”}\\ \text{“Uncertainty”}\\ \ldots\end{array}\right. \Rightarrow \\[2em]
\Rightarrow \left\{\begin{array}{l}\text{Gaussian}\\ \text{gamma}\\ \text{log-normal}\\ \ldots\end{array}\right. \Rightarrow
\end{array}
\left.\vphantom{\begin{array}{l}a\\a\\a\\a\\a\\a\\a\\a\\a\end{array}}\right\}
\boxed{\begin{array}{c}\textit{Prior}\\ \textit{Density}\end{array}}
$$

   There clearly is a lot of subjectivity involved in the above choices. . .

   In HEP, nuisance parameters such as energy scale, tracking efficiency, background normalization, etc., are typically assigned subjective priors.

Objective versus Subjective Priors (2/2)

2. Objective approach

   When there is no prior information available about a parameter, or one wishes to pretend that this is the case, then one is led to the concept of "ignorance" priors, also called "noninformative," "reference," "objective," or "non-subjective."

   The form of these priors is determined by a formal rule, e.g.:

   - Insufficient reason
   - Invariance
   - Maximal entropy
   - Coverage matching
   - Maximal "missing" information
   - etc.

   Objective priors are often improper (infinite normalization), which can cause various kinds of difficulties with the posterior.

Robustness Analysis

It is well known that subjective Bayesian analysis is the only coherent mode of behavior.

Unfortunately, one cannot make arbitrarily fine discriminations in judgments about probabilities. Therefore, subjective priors are by nature imprecise and one needs to check how robust one's inferences are against reasonable changes in the prior(s). For example:

- If the default prior for a positive parameter is a truncated Gaussian, try a gamma or a log-normal, or a linear combination of these;
- For an asymmetric prior, see what happens when the estimated "central value" of the parameter is used as the median or mode of the prior distribution, instead of its mean.

Without checking for robustness, one could be seriously misled as to the accuracy of the conclusion.

If the range of answers is too large, the question of interest may not be settled without more data or more prior information. This is only realistic.

Choice of Prior for Cross Section Measurements (1/3)

Suppose we measure the cross section σ for a Poisson process. The likelihood function is:

$$\mathcal{L}(\sigma, A \,|\, n) \;=\; \frac{(\sigma A L + b L)^n}{n!}\, e^{-\sigma A L - b L},$$

where A is the acceptance, L the integrated luminosity, b the expected background rate, and n the observed number of events. For simplicity, assume that L and b are very well known, whereas there is an uncertainty ΔA on the acceptance. (Note: in Joel's talk, $s \longleftrightarrow \sigma$, $\epsilon \longleftrightarrow AL$, and $b \longleftrightarrow bL$.)

In this setting, we have one parameter of interest (σ) and one nuisance parameter (A). It would seem that a reasonable prior for these parameters would be flat in σ ("insufficient reason"), and truncated Gaussian in A (if ΔA is not too large).

Unfortunately, with this choice of prior, posterior upper limits diverge to infinity. The reason for this is the flat improper prior for σ. Can we solve this problem by introducing a "cutoff" in the cross section prior?

Choice of Prior for Cross Section Measurements (2/3)

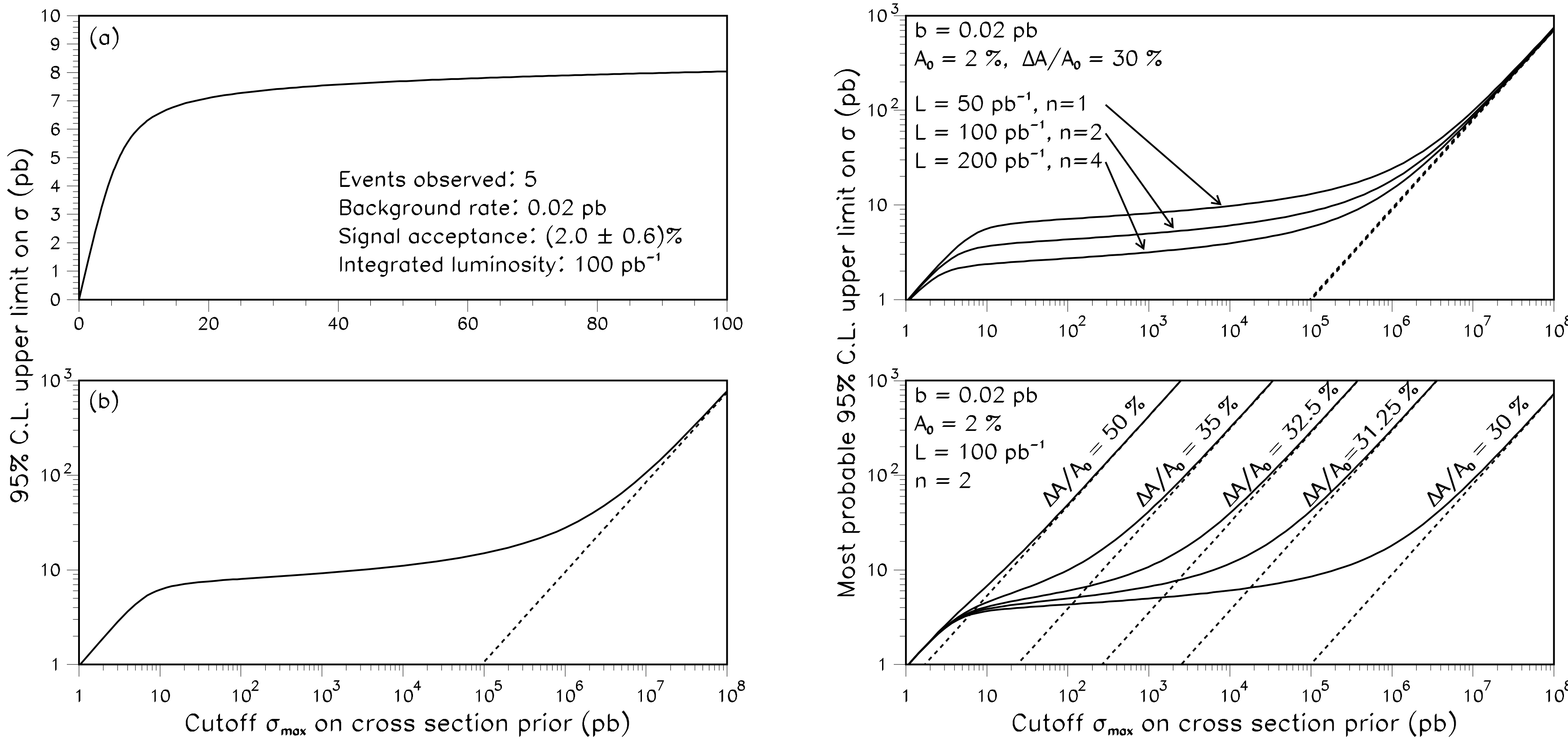

Figure 1: Bayesian upper limits at the 95% credibility level on a hypothetical cross section σ, as a function of the cutoff $\sigma_{\max}$ on the flat prior for σ.

Choice of Prior for Cross Section Measurements (3/3)

Truncating the flat cross section prior yields finite upper limits; however:

- The resulting upper limits may be very sensitive to the acceptance uncertainty;
- There is no obvious choice for the cutoff on the cross section prior that is general enough to be useful for everybody all of the time.

As shown by Joel Heinrich in his talk, using a gamma acceptance prior instead of a truncated Gaussian one yields finite upper limits, even with a flat improper cross section prior. This solution is not satisfactory however, because:

- It means that one has to give up on a fair robustness analysis.
  Joel showed that for a 10% acceptance uncertainty, the gamma and truncated Gaussian are quite similar. Therefore, the upper limit should not be sensitive to which one is chosen as prior, but it is.
- It leads to incoherence.
  With a flat improper prior for the cross section and a gamma prior $\pi(A)$ for the acceptance, the marginal *posterior* for the acceptance is proportional to $\pi(A)/A$, regardless of the integrated luminosity. In other words, the acceptance information gets updated even if the experiment is not done.

The Correlated Prior Method (1/2)

Since the real source of the problem is the improper cross section prior, this is where we must look for a solution. First, note that the likelihood only depends on the *product* of the cross section σ and acceptance A. So, some information in the data gets used to update σ, and some gets "wasted" on updating A.

The question is then: can we choose a prior that maximizes the transfer of information from the data to the cross section parameter, or equivalently, that minimizes the updating of the uninteresting acceptance parameter?

The answer is yes: instead of a prior that factorizes as usual in terms of σ and A, choose one that factorizes in terms of the expected number of signal events $\mu \equiv \sigma AL$ and A:

$$\pi(\sigma, A) \;=\; \left[\pi(\mu, A)\,\frac{\partial(\mu, A)}{\partial(\sigma, A)}\right]_{\mu=\sigma AL} \;=\; \pi(\mu)\Big|_{\mu=\sigma AL}\,\pi(A)\,AL.$$

Note that for this prior, the conditional prior density of the cross section given the acceptance depends on the acceptance: $\pi(\sigma \,|\, A) = \pi(\mu)|_{\mu=\sigma AL}\, LA$, hence there is a prior correlation between σ and A.

The Correlated Prior Method (2/2)

Properties of the correlated prior:

1. It is an "objective" prior, since it is derived from the formal requirement to minimize the updating of the prior information about the nuisance parameter. As is the case with many objective priors, this requirement only partially determines the form of the prior.
2. For the single-channel problem, it yields a marginal acceptance posterior that is exactly equal to the marginal acceptance prior.
3. Because of its more efficient use of the information in the data, it yields finite upper limits on the cross section, even with a flat improper prior on the expected number of signal events μ and a truncated Gaussian prior on the acceptance. A robustness analysis can be done.
4. The posterior obtained by this method can also be derived as a multiplicative convolution of two independent posteriors: one from a measurement of the number of signal events, and one from a measurement of the acceptance.
5. This method was used by the ZEUS collaboration in their searches for sleptons and squarks.

Frequentist Properties of the Correlated Prior Method (1/2)

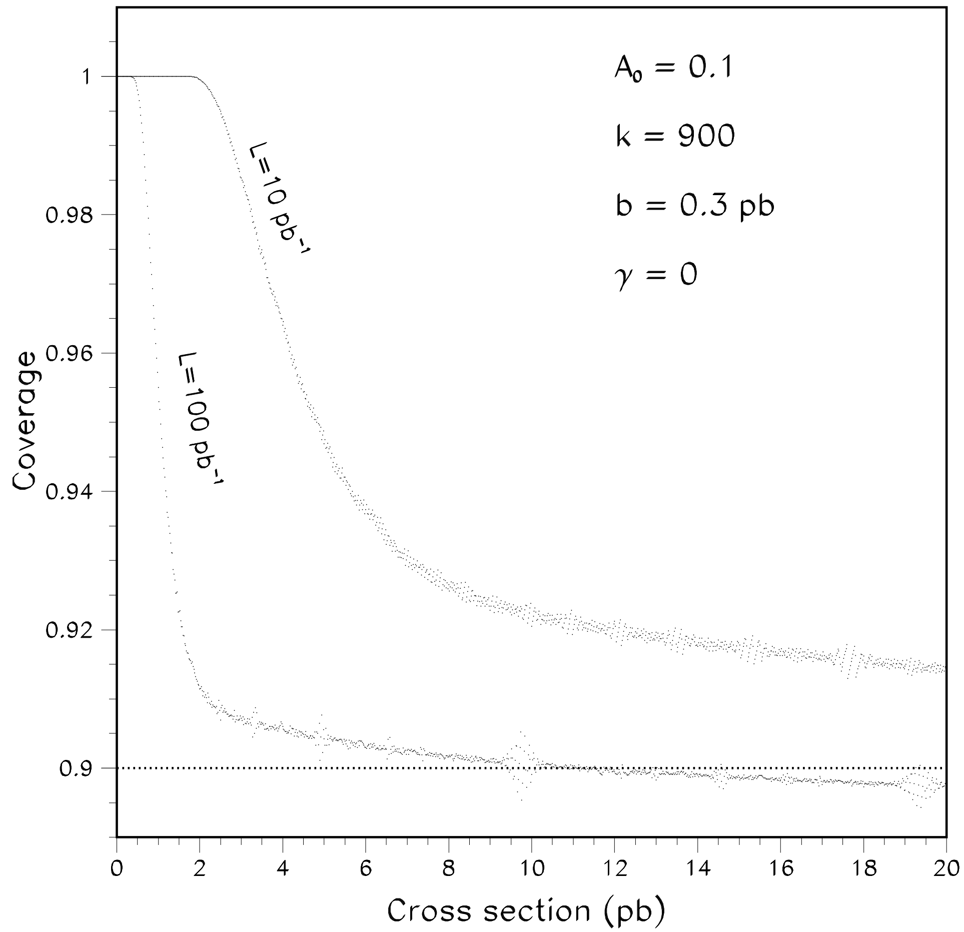

Figure 2: Coverage of 90% credibility level upper limits as a function of true cross section, for the correlated prior method.

Frequentist Properties of the Correlated Prior Method (2/2)

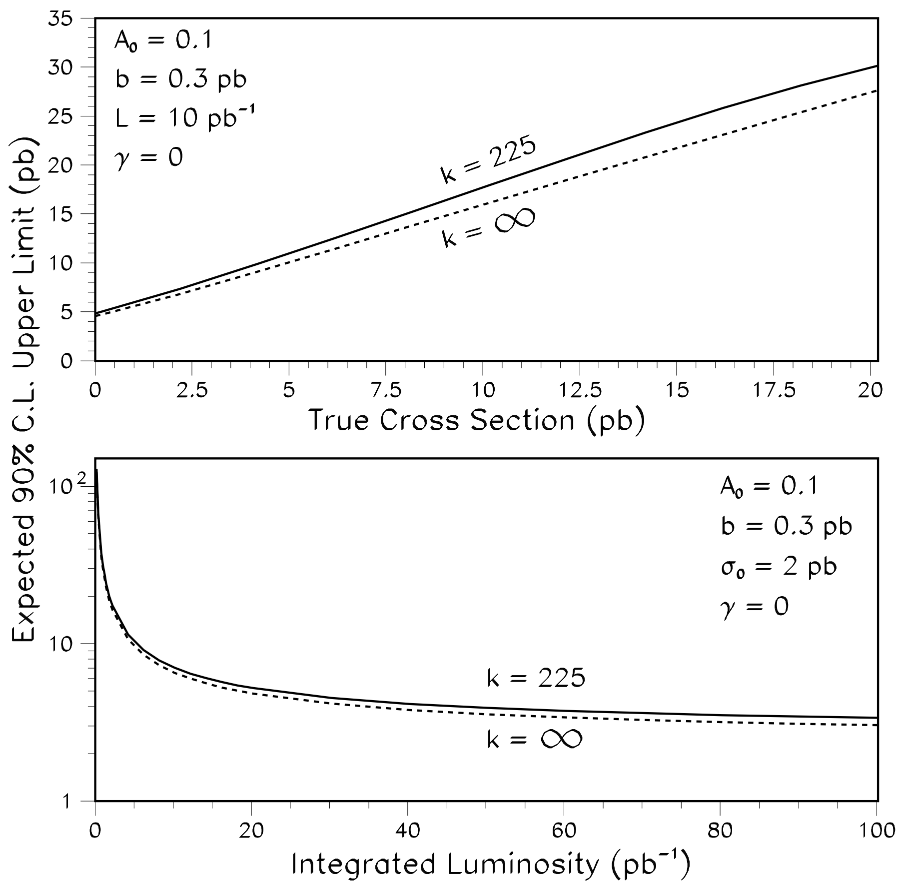

Figure 3: Expected 90% credibility level upper limits as a function of true cross section (top) and integrated luminosity (bottom), for the correlated prior method.

Software and Documentation

- The correlated prior method is documented in CDF note 5928.
- There exists an example program that combines the CDF Run I limits on Higgs production (7 channels, 22 nuisance parameters), and uses importance sampling to do the integrations.
- We will provide a program that calculates upper limits in the simplest case (one channel, a few nuisance parameters).

Summary

- Robustness analysis is a crucial component of any Bayesian calculation.
- The correlated prior method is optimal in the sense that it maximizes the transfer of information from the data to the parameter of interest.
- Because of this optimality, upper limits do not diverge and one can do a sensible robustness analysis.