



**Fermi National Accelerator Laboratory**

**FERMILAB-TM-2104**

## **A Recipe for the Construction of Confidence Limits**

I. Bertram et al.

*Fermi National Accelerator Laboratory  
P.O. Box 500, Batavia, Illinois 60510*

April 2000

## **Disclaimer**

*This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.*

## **Distribution**

*Approved for public release; further dissemination unlimited.*

## **Copyright Notification**

*This manuscript has been authored by Universities Research Association, Inc. under contract No. DE-AC02-76CHO3000 with the U.S. Department of Energy. The United States Government and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government Purposes.*

## A Recipe for the Construction of Confidence Limits

I. Bertram, G. Landsberg, J. Linnemann, R. Partridge M. Paterno, H.B. Prosper

(DØ Collaboration)

Fermi National Accelerator Laboratory, Batavia, Illinois 60510

(7 January 1998)

### Abstract

This is the description of the recipe for the construction of confidence limits recommended by the “Search Limits Committee” for the use of the DØ collaboration.

### I. INTRODUCTION

In this note, we present the recipe recommended by the “Search Limits Committee” for the construction of confidence intervals. In another note, currently in preparation, we will present the rationale for this recipe, a critique of the current literature on this topic, and several examples of the use of the method. This note is intended to fill the need of the collaboration to have a reference available until the more complete note is finished.

Section II introduces the notation we use in this note, and Section III contains the suggested recipe.

### II. NOTATION

For discrete  $A$ , we use the symbol  $P(A|B)$  to represent the probability of proposition  $A$ , given that proposition  $B$  is true. When dealing with a *continuous* parameter  $x$ , we define a

differential probability density  $\rho(x|B)$ , such that, given  $B$ , the probability of the continuous variable being between  $x$  and  $x + dx$  is given by  $P(x|B) = \rho(x|B) dx$ .

In the probabilities below,  $I$  represents any relevant prior information in the problem, and its presence emphasizes the fact that *all* probabilities are contingent upon other specified information, and that it is improper to combine probabilities contingent upon different prior information.

### III. THE RECIPE

We recommend the use of the Bayesian technique for the setting of confidence intervals, and present the following recipe for the construction of such limits.

1. *Define the model.* For the case of new particle searches,<sup>1</sup> there is only one accepted model: the expected number of events  $\mu$  is related to the signal cross section  $\sigma$ , the signal efficiency  $\epsilon$ , the integrated luminosity  $\mathcal{L}$ , and the expected background  $b$ , as:

$$\mu = b + \mathcal{L}\epsilon\sigma . \tag{3.1}$$

In different analyses, one may replace  $b$  by some more complicated expression, containing perhaps a sum of many terms, each of which is a product of background acceptances, background cross sections, branching fractions, and the integrated luminosities.

2. *Determine the likelihood function for the data, given the model.* In the case of counting experiments, one conventionally accepted likelihood function is the Poisson distribution with expectation value (mean)  $\mu$ . Given some expectation value  $\mu$ , the probability of observing  $k$  events is

---

<sup>1</sup>We limit our discussion to the case of counting experiments, and leave the discussion of spectrum analysis to another paper.

$$P(k|\mu, I) = \frac{e^{-\mu} \mu^k}{k!}, \quad (3.2)$$

where  $I$  reflects all the information used to build  $\mu$ , as well as the assumption about the relevance of the Poisson distribution for describing the results. With the above model, the likelihood function is simply

$$P(k|\sigma, \mathcal{L}, \epsilon, b, I) = \frac{e^{-(b+\mathcal{L}\epsilon\sigma)} (b + \mathcal{L}\epsilon\sigma)^k}{k!}. \quad (3.3)$$

It is worth emphasizing that the above is the probability for observing  $k$  events, given some  $\sigma$ ,  $\mathcal{L}$ ,  $\epsilon$ , and  $b$ , and not the probability that  $\sigma$ ,  $\mathcal{L}$ ,  $\epsilon$ , or  $b$ , are right.

3. *Assign the prior probabilities for all parameters.* One must next use the available information (such as the knowledge of the integrated luminosity, within some bound of uncertainties) to assign prior probabilities to each of the parameters in the problem. In general, the prior can contain correlations between the parameters. In many cases, the correlations can be neglected; if so, then the prior is factorizable into a product of independent priors. In most cases, the prior can be factored into  $P(\sigma|I)$  and  $P(\theta|I)$ , where  $\theta$  represents all the parameters *except*  $\sigma$ .

Often (but not necessarily always!), when there is an estimate of both the value and uncertainty in a given quantity, the appropriate distribution for a prior is a Gaussian. In general, the principle of maximizing the Shannon Entropy [1] can be imposed to determine an appropriate prior.

Because the theory of “uninformative priors” is the least well-developed portion of Bayesian statistics, the choice of prior for the cross section for signal poses a problem. We suggest using a flat prior probability density of finite range:

$$\rho(\sigma|I) = \begin{cases} 1/\sigma_{\max} & \text{if } 0 \leq \sigma \leq \sigma_{\max} \\ 0 & \text{otherwise,} \end{cases} \quad (3.4)$$

where  $\sigma_{\max}$  is chosen sufficiently large that the likelihood function for  $\sigma > \sigma_{\max}$  is negligible. We also suggest that the dependence of the result on the nature of the

prior be investigated to assess the sensitivity to the choice. The prior of most interest is the so-called *Jeffreys prior*:

$$\rho(\sigma|I) = \lim_{\substack{\alpha \rightarrow 0 \\ \beta \rightarrow \infty}} \begin{cases} 0 & \text{if } \sigma < \alpha \\ 1/[\sigma \ln(\beta/\alpha)] & \text{if } \alpha \leq \sigma \leq \beta \\ 0 & \text{if } \sigma > \beta, \end{cases} \quad (3.5)$$

where the limits are to be taken only at the end of the calculation, and not at some intermediate step.

Unless the lower limit  $\alpha$  for the Jeffreys prior has some finite positive value, the posterior probability density for  $\sigma$  will contain divergent integrals when there is background ( $b > 0$ ). This will also be the case for  $k = 0$ , when there is no background. Confidence limits will therefore not be able to be determined in such situations. Nevertheless, even for these cases, the integrals for the means and variances in  $\sigma$  converge, and can still be calculated.

It should be recognized that the divergence of the posterior density for  $\sigma$  based on the Jeffreys prior is such that smaller values of  $\alpha$  lead to *more* stringent limits on cross sections relative to those given by flat priors, and, in this sense, the flat prior of Eq. (3.4) is more “conservative”.

#### 4. Apply Bayes’ Theorem to find the posterior probability.

Bayes’ Theorem relates the pre-data knowledge of the parameters (the prior probabilities) to the post-data knowledge of the parameters (the posterior probabilities), with the linkage being the likelihood function and a normalization constant. In general, we can write:

$$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|C)}. \quad (3.6)$$

The denominator in Eq. (3.6) is determined by the normalization condition

$$\sum_{\text{all } A} P(A|BC) = 1. \quad (3.7)$$

When applying these ideas to probabilities over ranges of continuous variables, it should be recognized that Eq. (3.6) is specified in terms of probabilities (at discrete values), and not probability densities. Especially when normalizations are ambiguous, this may require great care. Under such circumstances, it is always better to evaluate the ratio in Eq. (3.6) over some specified range of the continuous parameters, rather than take the ratio of the individual multiplicative terms. This will avoid “paradoxes” that can arise from an incorrect approach to the limit when the sum in Eq. (3.7) becomes an integral.

To apply Bayes’ Theorem to our problem, we identify the following relationships:

- $A$  reflects the probability of the cross section being between  $\sigma$  and  $\sigma + d\sigma$ , the integrated luminosity between  $\mathcal{L}$  and  $\mathcal{L} + d\mathcal{L}$ , the signal efficiency between  $\epsilon$  and  $\epsilon + d\epsilon$ , and the background count between  $b$  and  $b + db$ .
- $B$  reflects the  $k$  events observed in the data.
- $C$  contains all relevant prior knowledge. This includes the descriptions of the knowledge of the parameters  $\sigma$ ,  $\mathcal{L}$ ,  $\epsilon$ , and  $b$ , as well as the assumptions that went into the entire model.

Thus, Bayes’ Theorem for our problem becomes:

$$P(\sigma, \mathcal{L}, \epsilon, b|k, I) \propto \frac{e^{-(b+\mathcal{L}\epsilon\sigma)}(b + \mathcal{L}\epsilon\sigma)^k}{k!} P(\sigma|I)P(\mathcal{L}, \epsilon, b|I), \quad (3.8)$$

where the constant of proportionality is determined by the condition

$$\int_0^\infty d\sigma \int_0^\infty d\mathcal{L} \int_0^1 d\epsilon \int_0^\infty db \rho(\sigma, \mathcal{L}, \epsilon, b|k, I) = 1. \quad (3.9)$$

5. *Remove nuisance variables.* Because the interest is in  $\sigma$ , and not in  $\mathcal{L}$ ,  $\epsilon$ , or  $b$ , the last three ( $\mathcal{L}$ ,  $\epsilon$ , and  $b$ ) are often termed *nuisance parameters*. To remove any dependence on the nuisance variables, Eq. (3.8) is integrated over them. The result is the posterior distribution for  $\sigma$ :

$$\rho(\sigma|k, I) = \int_0^\infty d\mathcal{L} \int_0^1 d\epsilon \int_0^\infty db \rho(\sigma, \mathcal{L}, \epsilon, b|k, I) . \quad (3.10)$$

6. *Use the resulting posterior probability distribution to calculate quantities of interest.*

The full information of the Bayesian analysis is contained in the posterior density function for  $\sigma$ . However, for some problems, less information is sufficient. For example, in a search for some new phenomenon, the relevant issue is usually the upper limit on the cross section ( $\sigma_{\text{UL}}$ ), calculated by integrating Eq. (3.10). This upper limit, specified at some level of confidence of  $100 \times \beta\%$ , is defined by:

$$\beta = \int_0^{\sigma_{\text{UL}}} d\sigma \rho(\sigma|k, I) . \quad (3.11)$$

When the posterior distribution for  $\sigma$  is peaked significantly away from zero, it may be more appropriate to quote the mean and the variance of the distribution — and to claim discovery!

## REFERENCES

- [1] E.T. Jaynes, *IEEE Trans. on Systems Sci. and Cybernetics* **SSC-4**, 227 (1968);  
*Proc. of IEEE*, **70**, 939 (1982).