ANALYTIC CONFIDENCE LEVEL CALCULATIONS USING THE LIKELIHOOD RATIO AND FOURIER TRANSFORM

H. Hu, J. Nielsen
Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Abstract
The interpretation of new particle search results involves a confidence level calculation on either the discovery hypothesis or the background-only ("null") hypothesis. A typical approach uses toy Monte Carlo experiments to build an expected experiment estimator distribution against which an observed experiment's estimator may be compared. In this note, a new approach is presented which calculates analytically the experiment estimator distribution via a Fourier transform, using the likelihood ratio as an ordering estimator. The analytic approach enjoys an enormous speed advantage over the toy Monte Carlo method, making it possible to quickly and precisely calculate confidence level results.

1. INTRODUCTION

A consistently recurring topic in experimental physics has been the interpretation and combination of results from searches for new particles. The fundamental task is to interpret the collected dataset in the context of two complementary hypotheses. The first hypothesis – the *null hypothesis* – is that the dataset is compatible with non-signal background production alone, and the second is that the dataset is compatible with the sum of signal and background production. In most cases, the search for new particles proceeds via several parallel searches for final states. The results from all of these subchannels are then combined to produce a final result.

All existing confidence level calculations follow the same general strategy [1, 2, 3]. A test statistic or *estimator* is constructed to quantify the "signal-ness" of a real or simulated experiment. Most calculation methods use an ensemble of toy Monte Carlo experiments to generate the estimator distribution against which the observed experiment's estimator is compared. This generation can be rather time-consuming when the number of toy Monte Carlo experiments is great (as it must be for high precision calculations) or if the number of signal and background events expected for each experiment is great (as it is for the case of searches optimized to use background subtraction).

In this note, we present an improved method for calculating confidence levels in the context of searches for new particles. Specifically, when the likelihood ratio is used as an estimator, the experiment estimator distribution may be calculated analytically with the Fourier transform. The most dramatic advantage of the analytic method over the toy Monte Carlo method is the increase in calculation speed.

2. LIKELIHOOD RATIO ESTIMATOR FOR SEARCHES

The likelihood ratio estimator is the ratio of the probabilities of observing an event under the two search hypotheses. The estimator for a single experiment is

$$E = C\frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}. \tag{1}$$

Here $\mathcal{L}_{s+b}$ is the probability density function for signal+background experiments and $\mathcal{L}_b$ is the probability density function for background-only experiments. Because the constant factor C appears in each event's estimator, it does not affect the ordering of the estimators. For clarity in this note, the constant is chosen to be e^s, where s is the expected number of signal events. In practice, not every event is equally signal-like. Each search may have one or more event variables that discriminate between signal-like and background-like events. For the general case, the probabilities $\mathcal{L}_{s+b}$ and $\mathcal{L}_b$ are functions of the observed events' measured variables.

As an example, consider a search using one discriminant variable m, the reconstructed Higgs mass. The signal and background have different probability density functions in m, defined as $f_s(m)$ and $f_b(m)$, respectively. (For searches with more than one discriminant variable, m would be replaced by a vector of discriminant variables $\overrightarrow{x}$.) It is then straightforward to calculate $\mathcal{L}_{s+b}$ and $\mathcal{L}_b$ for a single event, taking into account the event weighting coming from the discriminant variable:

$$E = e^s \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b} = e^s \frac{e^{-(s+b)}\left[sf_s(m) + bf_b(m)\right]}{e^{-b}\left[bf_b(m)\right]}. \quad (2)$$

The likelihood ratio estimator can be shown to maximize the discovery potential and exclusion potential of a search for new particles [3].

3. ENSEMBLE ESTIMATOR DISTRIBUTIONS VIA FAST FOURIER TRANSFORM (FFT)

One way to form an estimator for an ensemble of events is to generate a large number of toy Monte Carlo experiments, each experiment having a number of events generated from a Poisson distribution. Another way is to compute analytically the probability density function of the ensemble estimator given the probability density function of the event estimator. The discussion of this section pursues the latter approach.

The likelihood ratio estimator is a multiplicative estimator. This means the estimator for an ensemble of events is formed by multiplying the individual event estimators. Alternatively, the logarithms of the estimators may be summed. In the following derivation, $F = \ln E$, where E is the likelihood ratio estimator.

For an experiment with 0 events observed, the estimator is trivial:

$$E = e^s \frac{e^{-(s+b)}}{e^{-b}} = 1 \quad (3)$$

$$F = 0 \quad (4)$$

$$\rho_0(F) = \delta(F), \quad (5)$$

where $\rho_0(F)$ is the probability density function in F for experiments with 0 observed events.

For an experiment with exactly one event, the estimator is, again using the reconstructed Higgs mass m,

$$E = e^s \frac{e^{-(s+b)}\left[sf_s(m) + bf_b(m)\right]}{e^{-b}\left[bf_b(m)\right]}, \quad (6)$$

$$F = \ln \frac{sf_s(m) + bf_b(m)}{bf_b(m)}, \quad (7)$$

and the probability density function in F is defined as $\rho_1(F)$.

For an experiment with exactly two events, the estimators of the two events are multiplied to form an ensemble estimator. If the reconstructed Higgs masses of the two events are m_1 and m_2, then

$$E = \frac{\left[sf_s(m_1) + bf_b(m_1)\right]\left[sf_s(m_2) + bf_b(m_2)\right]}{\left[bf_b(m_1)\right]\left[bf_b(m_2)\right]} \quad (8)$$

$$F = \ln \frac{sf_s(m_1) + bf_b(m_1)}{bf_b(m_1)} + \ln \frac{sf_s(m_2) + bf_b(m_2)}{bf_b(m_2)}. \quad (9)$$

The probability density function for exactly two particles $\rho_2(F)$ is simply the convolution of $\rho_1(F)$ with itself:

$$\begin{aligned} \rho_2(F) &= \int\int \rho_1(F_1)\rho_1(F_2)\delta(F - F_1 - F_2)dF_1 dF_2 \qquad (10) \\ &= \rho_1(F) \otimes \rho_1(F). \qquad (11) \end{aligned}$$

The generalization to the case of n events is straightforward and encouraging:

$$\begin{aligned} E &= \prod_{i=1}^{n} \frac{sf_s(m_i) + bf_b(m_i)}{bf_b(m_i)} \qquad (12) \\ F &= \sum_{i=1}^{n} \ln \frac{sf_s(m_i) + bf_b(m_i)}{bf_b(m_i)} \qquad (13) \\ \rho_n(F) &= \int \ldots \int \prod_{i=1}^{n} \left[\rho_1(F_i)dF_i\right] \delta\left(F - \sum_{i=1}^{n} F_i\right) \qquad (14) \\ &= \underbrace{\rho_1(F) \otimes \ldots \otimes \rho_1(F)}_{n \text{ times}}. \qquad (15) \end{aligned}$$

Next, the convolution of $\rho_1(F)$ is rendered manageable by an application of the relationship between the convolution and the Fourier transform.

If $A(F) = B(F) \otimes C(F)$, then the Fourier transforms of A, B, and C satisfy

$$\overline{A(G)} = \overline{B(G)} \cdot \overline{C(G)}. \qquad (16)$$

This allows the convolution to be expressed as a simple power:

$$\overline{\rho_n(G)} = \left[\overline{\rho_1(G)}\right]^n. \qquad (17)$$

Note this equation holds even for $n = 0$, since $\overline{\rho_0(G)} = 1$. For any practical computation, the analytic Fourier transform may be approximated by a numerical Fast Fourier Transform (FFT).

How does this help to determine ρ_{s+b} and ρ_b? The probability density function for an ensemble estimator with s expected signal and b expected background events is

$$\rho_{s+b}(F) = \sum_{n=0}^{\infty} e^{-(s+b)} \frac{(s+b)^n}{n!} \rho_n(F), \qquad (18)$$

where n is the number of events observed in the experiment. Upon Fourier transformation, this becomes

$$\begin{aligned} \overline{\rho_{s+b}(G)} &= \sum_{n=0}^{\infty} e^{-(s+b)} \frac{(s+b)^n}{n!} \overline{\rho_n(G)} \qquad (19) \\ &= \sum_{n=0}^{\infty} e^{-(s+b)} \frac{(s+b)^n}{n!} \left[\overline{\rho_1(G)}\right]^n \qquad (20) \end{aligned}$$

$$\overline{\rho_{s+b}(G)} = e^{(s+b)\left[\overline{\rho_1(G)} - 1\right]}. \qquad (21)$$

The function $\rho_{s+b}(F)$ may then be recovered by using the inverse transform. In general, this relation, which holds for any multiplicative estimator, means that the probability density function for an arbitrary number of expected signal and background events may be calculated analytically once the probability density function of the estimator is known for a single event.

Two examples provide practical proof of the principle. For the first, assume a hypothetical estimator results in a probability density function of simple Gaussian form

$$\rho_1(F) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{22}$$

where $\sigma = 0.2$ and $\mu = 2.0$. For an expected $s+b = 20.0$, both the FFT method and the toy Monte Carlo method are used to evolute the event estimator probability density function to an experiment estimator probability density function. The agreement between the two methods (Fig. 1a) is striking. The higher

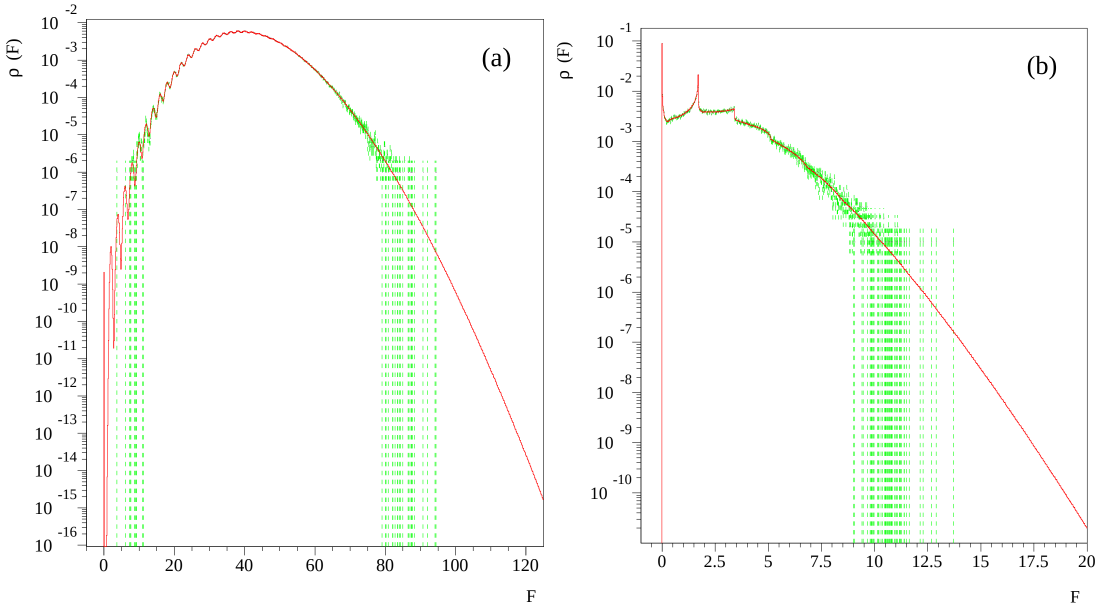

Fig. 1: The experiment estimator probability density functions for a Gaussian event estimator probability function (a) and for a typical non-Gaussian event estimator (b). The solid line is calculated with the FFT method, and the dashed line is calculated with the toy Monte Carlo method. Error bars associated with the Monte Carlo method are due to limited statistics.

precision of the FFT method is apparent, even when compared to 1 million toy Monte Carlo experiments. The periodic structure is due to the discontinuous Poisson distribution being convolved with a narrow event estimator probability function. In particular, the peak at $\ln E = 0$ corresponds to the probability that exactly zero events be observed ($e^{-(s+b)} = 2.1 \times 10^{-9}$). The precision of the toy Monte Carlo method is limited by the number of Monte Carlo experiments, while the precision of the FFT method is limited only by computer precision. For the second example, the probability density function of a typical non-Gaussian estimator is calculated for an experiment with $s = 5$ and $b = 3$ expected events (Fig. 1b). Again, the two methods agree well in regions where the toy Monte Carlo method is useful.

Finally, the obtained experiment estimator probability density function may be used to calculate confidence levels on the search hypotheses. For example, the final confidence coefficients c_{s+b} and c_b are simply integrals of the experiment estimator probability density function [4, 5].

4. DISCUSSION ON SYSTEMATIC UNCERTAINTIES

When the likelihood ratio estimator is used as a test statistic, the systematic uncertainty on the confidence level is due to the uncertainties on numbers of background events expected, the number of signal events expected, and the shapes of the discriminant variables. Since the shapes are nothing more than the density of signal and background events in the discriminant variable space, we focus only on the uncertainty due to uncertainties on background and signal numbers.

Consider one channel having k types of signal events and l types of background events. The number of each type of event is denoted by u_i, $(i = 1, 2, \ldots, k + l)$. Then the Fourier transform of the experiment estimator's density function is calculated using the previous results:

$$\overline{\rho(G)} = e^{\sum_{i=1}^{k+l} u_i \left[\overline{\rho_{1,i}(G)} - 1\right]} \tag{23}$$

where $\overline{\rho_{1,i}(G)}$ is the transformed density function for one event of the $i\mathrm{th}$ type. If the uncertainties follow a Gaussian distribution with a correlated error matrix

$$S_{ij} = \langle (u_i - \langle u_i \rangle) (u_j - \langle u_j \rangle) \rangle \tag{24}$$

between the $k + l$ types of events, then the systematic uncertainty on the experiment estimator's density function may be calculated analytically as

$$\begin{aligned} \overline{\rho_{\mathrm{sys}}(G)} &= \int \ldots \int e^{\sum_{i=1}^{k+l} u_i \left[\overline{\rho_{1,i}(G)} - 1\right]} \left(\frac{1}{\sqrt{2\pi}}\right)^{k+l} \frac{1}{\sqrt{|S|}} \\ & e^{\sum_{i=1}^{k+l} \sum_{j=1}^{k+l} -\frac{1}{2} (u_i - \langle u_i \rangle) S_{ij}^{-1} (u_j - \langle u_j \rangle)} \prod_i du_i \\ &= e^{\sum_{i=1}^{k+l} \langle u_i \rangle \left[\overline{\rho_{1,i}(G)} - 1\right] + \frac{1}{2} \sum_{i,j} \left[\overline{\rho_{1,i}(G)} - 1\right] S_{ij} \left[\overline{\rho_{1,j}(G)} - 1\right]} \end{aligned} \tag{25}$$

In general, the resolution function can be constructed by combining several Gaussian distributions, so the systematic uncertainty can be calculated analytically.

5. COMBINING RESULTS FROM SEVERAL SEARCHES

Given the multiplicative properties of the likelihood ratio estimator, the combination of several search channels proceeds intuitively. The estimator for any combination of events is simply the product of the individual event estimators. Consequently, construction of the estimator probability density function for the combination of channels parallels the construction of the estimator probability density function for the combination of events in a single channel. In particular, for a combination with N search channels:

$$\overline{\rho_{s+b}(G)} = \prod_{j=1}^{N} \overline{\rho_{s+b}^{j}(G)} \tag{26}$$

$$= e^{\sum_{j=1}^{N} (s_j + b_j) \left[\overline{\rho_1^j(G)} - 1\right]} \tag{27}$$

Due to the strictly multiplicative nature of the estimator, this combination method is internally consistent. No matter how subsets of the combinations are rearranged (*i.e.*, combining channels in different orders, combining different subsets of data runs), the result of the combination does not change.

6. CONCLUSION

A fast confidence level calculation with a multiplicative estimator makes possible studies that might have otherwise been too CPU-intensive with the toy MC method. These include studies of improvements in the event selections, of various working points, and of systematic errors. A precise calculation also makes possible rejection of null hypotheses at the level necessary for discovery.

Acknowledgements

The authors thank Haimo Zobernig for very useful discussions.

References

[1] P. Janot and F. Le Diberder, *Optimally combined confidence limits*, Nucl. Instrum. Methods **A411** (1998) 449.

[2] T. Junk, *Confidence Level Computation for Combining Searches with Small Statistics*, CERN-EP 99-041 and `hep-ex/9902006` (1999).

[3] A. L. Read, *Optimal Statistical Analysis of Search Results based on the Likelihood Ratio and its Application to the Search for the MSM Higgs Boson at* $\sqrt{s} = 161$ *and* 172 *GeV*, DELPHI note 97-158 PHYS 737 (1997).

[4] V. F. Obraztsov, Nucl. Instrum. Methods **A316** (1992) 388-390.

[5] S. Jin and P. McNamara, *The Signal Estimator Limit Setting Method*, `physics/9812030` (1998).

Discussion after talk of Jason Nielsen. Chairman: Roger Barlow.

M. Woodroofe

I hope I'm not making too many comments. The distribution of this sum up to the random Poisson is sometimes called a compound Poisson distribution. Is that a familiar term to you ?

J. Nielsen

Which one is this ?

M. Woodroofe

A compound Poisson distribution. It's the sum of a bunch of independent random variables where the number of terms in the sum has a Poisson distribution. Those arise, among other places, in the distribution of insurance claims, where the number of claims is a Poisson and the amount of the claim is a random variable. A lot of effort has gone into understanding the distribution of compound Poisson, much along the lines that you're talking about. You might want to connect what you've done to some of the earlier work.

Shan Jin

Can this method apply to the unified approach ?

J. Nielsen

Because it uses multiplicative estimator or additive estimator like the log of the estimator that I used, then it's not going to work if you are ever breaking up the pieces and renormalizing the estimator. As long as the estimator is a multiplicative estimator this will work, but if you are ever using, for example, the published unified approach, then I don't think it would work.