Likelihood Method II: Further Expansion

T.J. Haines
LANL, April 3, 2000

This note is a follow-on to my previous memo on the likelihood method. First, I will review the method established in the previous note. I will then discuss some of the formalism that shows for the first time that the previous method is/was: 1) not normalized properly, 2) equivalent to the so-called "binned maximum likelihood" method, 3) will maximize at $\mu_s = N_o - \mu_b$, sort-of what one would expect, though this maximum occurs at all values of μ_s (including $\mu_s < 0$), and 4) the method of McCullough does similarly. I will then finish this note by including background estimation error in the formalism, one of the main points of this work since its inception. I finish by showing the usual estimate for a signal strength is systematically wrong, underestimating signal strength by a factor of typically 10% – 50%.

Introduction & Jargon

Let me here again define a few symbols from the previous note

μ_s	True level of signal in the GRB region
μ_b	True background flux in the GRB region
N_b	Number of background events in region
N_s	Number of signal events in region
N_o	Total number of events observed in the region

With these definitions, it is clear that $N_o = N_s + N_b$ is our constraint; all of the above numbers are by assumption (or by prior choice in a Baysian's view) positive. And remember that $P(N|\mu)$ is a properly normalized Poisson distribution.

With the above definitions, the previous note developed an appropriate likelihood test, which is reviewed again here. We want to maximize the ratio of the probability that our observation is due to a signal of true "strength" μ_s and true background μ_b compared to the probability that our observation is only due to background ($\mu_s = 0$). The only unknown parameter that we are concerned about (so far) is the signal level; we will maximize the likelihood with respect to this variable, so we must write down the probability of having a particular μ_s given our observation of N_o. The probability given in the previous note that we want is, for a particular unknown N_s,

$$\mathrm{Prob}(N_s,\mu_s) = \mathrm{Prob}(N_s|\mu_s) * \mathrm{Prob}(N_b|\mu_b) = \mathrm{Prob}(N_s|\mu_s) * \mathrm{Prob}(N_o - N_s|\mu_b)$$

where we have invoked our constraint that $N_b + N_s = N_b$. Now, we do not know the actual number of signal or background events that we observed, we only know the total number that we observed. So, we must include all possibilities for the actual number of signal (and thus background) events that composed our observation. Said differently, our observation may have resulted from observing $N_s = 0$, $N_b = N_o$ events, or $N_s = 1$, $N_b = N_o - 1$ events, or $N_s = 2$, $N_b = N_o - 2$ events, or … etc. Thus the total probability is

$$P(\mu_s) = \sum_{N_b=0}^{N_o} P(N_b \mid \mu_b) * P(N_o - N_b \mid \mu_s)$$

The likelihood ratio can then be written as

$$L(\mu_s) = \frac{P(\mu_s)}{P(\mu=0\ )} \quad = \quad \frac{\sum_{N_b=0}^{N_o} P(N_b \mid \mu_b) * P(N_o - N_b \mid \mu_s)}{P(N_0 \mid \mu_b)} = e^{-\mu_s} \left(\frac{\mu_s}{\mu_b}\right)^{N_o} \sum_{N_b=0}^{N_o} \left(\frac{\mu_s}{\mu_b}\right)^{N_b} \frac{1}{N_b!(N_o - N_b)!} \quad eq.1^*$$

which is (essentially) eq. 1 from the previous note.

(Im)Proper Normalization

Upon inspection it should become obvious that the numerator in eq. 1 is not properly normalized. That is to say, the numerator should represent a properly normalized probability—it should have unit probability when all possibilities are considered for the relevant variable. Thus, it should have unit area when integrated over μ_s. It does not. Instead

$$\int_0^\infty P(\mu_s) d(\mu_s) \quad = \quad \sum_{N_b=0}^{N_o} P(N_b \mid \mu_b) * \int_0^\infty d(\mu_s) P(N_o - N_b \mid \mu_s) = \sum_{N_b=0}^{N_o} P(N_b \mid \mu_b)$$

so I must divide eq. 1 by this expression. The denominator, when summed over all possible values of the "variable" N_0, is properly normalized.

Equivalence to the "Binned Likelihood" Method

Now eq. 1[*] becomes, after proper normalization,

$$L(\mu_s) = \frac{P(\mu_s)}{P(\mu=0\ )} \quad = \quad \frac{\dfrac{\sum_{N_b=0}^{N_o} P(N_b \mid \mu_b) * P(N_o - N_b \mid \mu_s)}{\sum_{N_b=0}^{N_o} P(N_b \mid \mu_b)}}{P(N_0 \mid \mu_b)} \qquad eq.1$$

where the probabilities are now normalized. Consider the numerator for a moment, and allow me to re-arrange some terms as follows:

$$\sum_{N_b=0}^{N_o} P(N_b \mid \mu_b) * P(N_o - N_b \mid \mu_s) = \sum_{N_b=0}^{N_o} \frac{1}{N_b!} e^{-\mu_b} \mu_b{}^{N_b} \frac{1}{(N_o - N_b)!} e^{-\mu_s} \mu_s{}^{(N_o - N_b)}$$

$$= \frac{1}{N_o!} e^{-(\mu_{+\mu_s})b} \sum_{N_b=0}^{N_o} \frac{N_o!}{N_b!(N_o - N_b)!} \mu_b{}^{N_b} \mu_s{}^{(N_o - N_b)} = \frac{1}{N_o!} e^{-(\mu_{+\mu_s})b} (\mu_b + \mu_s)^{N_o} = P(N_o \mid (\mu_b + \mu_s))$$

The so-called "binned likelihood" [as described in e.g. "Milagrito's Point Spread Function and Mrk 501(Revision 1)",10-2-99 by Stefan Westerhoff] method considers this very same Poisson probability, except that μ_s (and in principle μ_b) is taken as a function of distance from a candidate source position, given by a single parameter (the source strength) and a known/God-given point spread function. The difference simply is that this method uses only one (large) bin. I will continue with the original formalism for a while longer, as I want to generalize the method to include error in the background estimation.

Analytic Derivation of Maximum Likelihood

Using the above result, eq. 1 now becomes

$$L(\mu_s) = \frac{P(\mu_s)}{P(\mu = 0\ )} = \frac{P(N_0 \mid \mu_b + \mu_s)}{P(N_0 \mid \mu_b) * \sum_{N_b=0}^{N_o} P(N_b \mid \mu_b)}$$

so that finding the maximum of this function wrt. μ_s is equivalent to finding the maximum of the numerator. It is straightforward to show that $\partial P(N|\mu)/\partial\mu = 0$ when $\mu = N$ (and similarly straightforward to show this is a maximum), for any value of N or μ; for N=0, the probability becomes $e^{-\mu}$, which obviously maximizes at $\mu=0$. For $N \neq 0$, the first derivative is zero at $\mu = N$, and the second derivative is negative, thus it is a maximum; you probably knew that already, didn't you?

So, we find the maximum likelihood at $N_0 = \mu_s + \mu_b$. Of course, this means that the likelihood can maximize in the "un-physical" region $\mu_s = N_0 - \mu_b < 0$ if $\mu_b > N_0$. This remains a weakness of the method in my opinion.

Imperfect Background Estimation

So far I have assumed that the level of the expected background in the bin was perfectly well determined. Whether in the search for nucleon decay, gamma-ray point sources, or in many other situations, the expected background level has an error associated with it; it is here that this method deviates from the "binned likelihood" method, and it is for this reason that I have kept the more cumbersone expression for the likelihood.

In those applications in which I am interested, the error in the background estimate is determined (at least in part) by counting statistics of the background itself; there may also be an important contribution from a systematic measurement error. I will not deal with that situation here, except to say that if the distribution of the error is known (i.e. a Gaussian of some width), then the method I now give is easily generalized.

Glancing back up at eq. 1, you can readily see that we are interested in the probability that N_0 events was observed given that we expected μ_b background level and μ_s signal level. This probability is the odds that our observation may have resulted from observing $N_s = 0$, $N_b = N_o$ events, or $N_s = 1$, $N_b = N_o - 1$ events, or $N_s = 2$, $N_b = N_o - 2$ events, or … etc. To derive eq. 1, we assumed that these probabilities are Poisson distributed. Here we change that assumption somewhat. Note that this formalism follows directly from work documented in my Ph.D. thesis, and expanded upon in the "Source Techniques" paper (D.E. Alexandreas *et al.*, NIM **A328**, 570(1992)).

In the applications mentioned above, the error in the background estimate is due in part simply to counting statistics of the background rate. That is the situation I will deal with here in this context; it is obvious how to generalize this to include other errors (though it is not usually obvious how to quantify their probability distribution functions). Usually the background is determined (either through Monte Carlo simulations, or randomization of data) from a data set that has an exposure $1/\alpha$ times larger than the data. Before going forward, let me first (re-)define some variables:

α	Known exposure of the signal region relative to the background region
μ_s	Unknown true level of signal in the region of interest
μ_b'	Unknown true background flux in the background region
μ_b^*	Estimated background rate in signal region; $\mu_b^* = \alpha N_b^*$
N_b^*	Number of background events in background region
N_s	Number of signal events in region; as always, $N_s + N_b = N_0$
N_o	Total number of events observed in the signal region

and we will let $P'(N_b|\mu_b^*)$ be the to-be-determined probability of some number N_b events given the estimated background level μ_b^*. The normalization of this probability should be such that

$$\sum_{Nb=0}^{\infty} P'(N_b \mid \mu_b^*) = 1$$

which we will worry about once we derive P'; you can see I still assume the number of signal events fluctuates via a Poisson distribution about the true signal level. The probability that we are genuinely interested in, will be

$$P(\mu_s) = \sum_{N_b=0}^{N_o} P'(N_b \mid \mu_b^*) P(N_o - N_b \mid \mu_s)$$

Now, by assumption, the number of background events in the background region fluctuates via Poisson statistics about some true mean value μ_b'. The probability of N_b events given some background estimate αN_b^* is then the probability of observing N_b and αN_b^* given a true background μ_b', or observing N_b and αN_b^* given a true background

μ_b'+ δ, or observing N_b and αN_b^* given a true background μ_b'+ 2δ, or ….; this statement can be shown formally from Bayes' theorem (with flat priors). Turned into mathematical symbols and not worrying about the above normalization for the moment, this becomes:

$$P'(N_b \mid \mu_b{}^* = \alpha N_b{}^*) = \int_0^\infty P(N_b \mid \alpha\mu_b) * P(N_b{}^* \mid \mu_b{}')d(\mu_b{}') = \frac{\alpha^{N_b}}{(1+\alpha)^{(N_b+N^*{}_b+1)}} \frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!}$$

where the symbols have their usual definitions.

First I can hear people saying "this just doesn't look right—its not a Poisson!" That one expects a Poisson is the mistake. We have calculated an "average" probability (the first term above), if you will, folded over the probability distribution of background fluctuations (the second term). As these fluctuations become increasingly small, the above probability should go to a Poisson. This can be demonstrated by the following.

The second Poisson becomes narrower in the limit $N_b^*\to\infty$, $\alpha\to 0$ such that $\mu_b^* = \alpha N_b^*$ remains a constant. In this limit it is not too difficult to show that

$$\frac{\alpha^{N_b}}{(1+\alpha)^{(N_b+N^*{}_b+1)}} \to \alpha^{N_b} e^{-N_b};\frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!} \to \frac{N_b{}^{*N_b}}{N_b!}; so \quad that$$

$$\frac{\alpha^{N_b}}{(1+\alpha)^{(N_b+N^*{}_b+1)}} \frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!} \to P(N_b \mid \mu_b{}^*)$$

as one would like—we end up with a Poisson distribution that we expect in the right limit.

Now we must deal with the normalization of P'. We need to require

$$\sum_{\mathrm{Nb}=0}^{\infty} \mathrm{P}'(\mathrm{N_b} \mid \mu_b{}^*) = 1$$

so I will introduce a normalization constant A into the expression for P'. We then have

$$\frac{A}{(1+\alpha)^{N_b{}^*+1}} \sum_{\mathrm{Nb}=0}^{\infty} \frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!}\left(\frac{\alpha}{1+\alpha}\right)^{N_b} = 1$$

I have yet to be able to figure out a closed form for the summation. However, for $N_b^*{>>}N_0$, I can approximate (similar to above)

$$\frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!} \approx \frac{N_b{}^{*N_b}}{N_b!}$$

so that the sum becomes approximately $e^{\mu b^*/(1+\alpha)}$ so that in this approximation I can write

$$P'(N_b \mid \mu_b{}^* = \alpha N_b{}^*) \approx \quad e^{-\mu_b{}^*/(1+\alpha)} \frac{(N_b + N_b{}^*)!}{N_b!N_b{}^*!}\left(\frac{\alpha}{1+\alpha}\right)^{N_b}$$

is the simplest (approximate) form I can write the properly normalized probability I am interested in.

Let me then collect the two probabilities together to write in this approximation

$$P(\mu_s) = \sum_{N_b=0}^{N_o} \mathrm{P'}(\mathrm{N_b} \mid \mu_b *) P(N_o - N_b \mid \mu_s) \approx e^{-\mu_b */(1+\alpha)} e^{-\mu_s} \sum_{N_b=0}^{N_0} \frac{(N_b + N_b *)!}{N_b! N_b *!} \left(\frac{\alpha}{1+\alpha} \right)^{N_b} \frac{\mu_s^{N_0 - N_b}}{(N_0 - N_b)!}$$

$$\approx P(N_0 \mid (\mu_s + \frac{\mu_b}{(1+\alpha)}))$$

where the last step used the same approximation as above and involved some re-arrangement of terms.

Note that this will maximize at the usual place for a Poisson, $N_0 = \mu_s + \mu_b/(1+\alpha)$. What does this mean? It means that the "usual" estimate for the signal strength, $\mu_s = N_0 - \mu_b$ is systematically wrong, an underestimate of the signal strength by a factor of $1/(1+\alpha)$ which is typically 10% – 50%.