Likelihood Analysis Performance and Tradeoffs

Patrick Nolan (Stanford)

SSC-LAT Software Workshop

July 15-16, 2003, SLAC

The Problem

The data will consist of a cloud of distinct photons with various energies and directions, all modified by the instrument response functions. We want to think in terms of point and diffuse gamma-ray sources.

We seem to have a consensus that many of the tools we use to translate from data to models will use likelihood. Bayesians and frequentists can agree on this foundation.

There is a concern that the calculation of likelihood functions may be onerous if we do it properly.

How to do the calculation right

The likelihood function to be maximized looks like this:

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln M(\boldsymbol{\lambda}_i) - N_{\mathrm{pred}},$$

where

$$N_{\mathrm{pred}} = \int_{\boldsymbol{\Lambda}} M(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$$

Here λ is shorthand for all the measurable characteristics of a photon: energy, direction, time, conversion plane, etc. *M* is the predicted density of events in this space. N_{pred} is the total number of predicted events. The index *i* represents the individual detected photons. The Region of Interest (Λ) is determined by the chosen data cuts.

Things get complicated because *M* can be hard to calculate and the N_{pred} integral is multidimensional. For a single point source, M can be decomposed like this:

$$M(E', \mathbf{p}', t, k) = \int_0^\infty D_k(E'; E, \mathbf{p} \cdot \mathbf{z}(t)) P_k(|\mathbf{p}' - \mathbf{p}|; E, \mathbf{p} \cdot \mathbf{z}(t)) A_k(E, \mathbf{p} \cdot \mathbf{z}(t))\, s(E)\, dE$$

Where E' and p' are the measured energy and direction, k is a place-holder for the internal event type variables, E and p are the true energy and direction, z is the instrument pointing vector, D is the energy dispersion function, P is the point spread function, and A is the effective area.

How to do it right (cont.)

The functions *A*, P and D will be tabulated. Each is a function of 4 or 5 variables, so storage won't be trivial. Producing this *M*, integrating over its 4 variables, and summing over all photons, will require a 6-dimensional integral/sum.

Things get complicated when we consider the zenith cutoff. Photons must be rejected if the come from a direction near the horizon. The Earth is a very bright gamma ray source. The horizon moves with respect to the field of view, so the limits of all the integrals change in an awkward way.

We like to think of a quantity called exposure which describes how deeply a spot on the sky has been examined. Maps of exposure can be made and used in the likelihood calculation. It looks like this:

$$\begin{aligned}\mathcal{E}_m(\mathbf{p}) &= \int dE' \int d\mathbf{p}' \int dt\, h(t) \sum_k H(\mathbf{u}(t)\cdot\mathbf{p}' - \mu_k(E')) \qquad (1)\\ &\quad \times A_k(E_m, \theta(\mathbf{p},t), \phi(\mathbf{p},t))\, P_k(\mathbf{p}'; E_m, \theta(\mathbf{p},t), \phi(\mathbf{p},t))\, D_k(E'; E_m, \theta(\mathbf{p},t), \phi(\mathbf{p},t))\\ &= \int dE' \int d\mathbf{p}' \int dt\, h(t) \sum_k H(\mathbf{u}(t)\cdot\mathbf{p}' - \mu_k(E'))\, R_k(E', \mathbf{p}'; E_m, \theta(\mathbf{p},t), \phi(\mathbf{p},t))\end{aligned}$$

It's a 5-dimensional integral for each point on the sky.

Possible Approximations

In the N_{pred} term we could replace the point spread function with a delta function. This is equivalent to using the true direction rather than the observed direction in the zenith cuts and ignoring any net loss or gain of events.

Similarly we can replace the energy dispersion with a delta function in N_{pred}, which is like using the true energy in zenith cuts.

We can also ignore the energy dispersion in the individual-photon terms, assuming that the off-diagonal terms don't contribute much to the final result. With wide energy bins, this might be justified. A quantitative study needs to be done.

Rather than tabulating the true forms of the response functions A, P, and D, we could use simple analytical approximations. For instance a gaussian point spread function whose width is a power-law function of energy.

The current situation

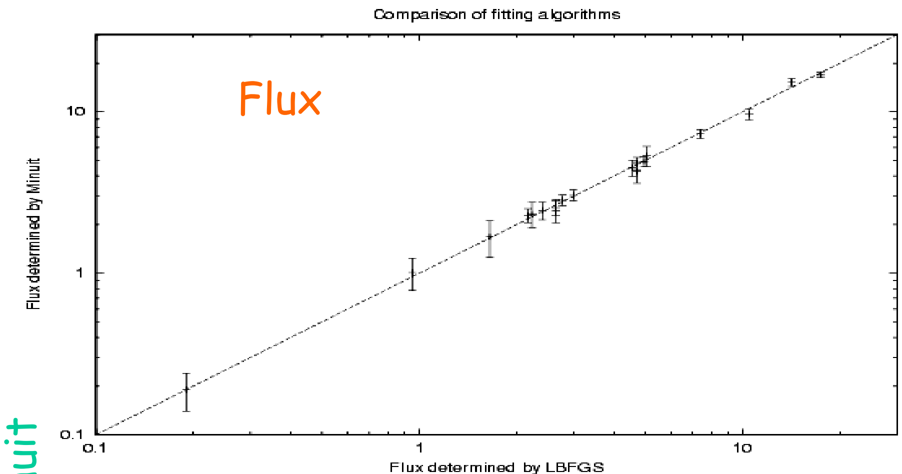
Jim Chiang has done a lot of work to implement the likelihood calculation. I added some optimization methods. Currently there are no zenith angle cuts, the PSF is the sum of two gaussians, and energy dispersion is ignored.

Here's a small dose of reality. I created a set of 10,000 photons based on a list of point sources in a 30 degree circle including the Cygnus region produced by Seth Digel. This includes 19 EGRET sources and 115 fainter ones added at random. This many events corresponds to about 4.7 days' worth of data. I fitted the flux and spectral index of the 19 EGRET sources, holding their positions fixed, pretending that the other sources are background. I used both the familiar Minuit algorithm and the LBFGS quasi-Newton method. Both converged to satisfactory answers in about 1000 seconds of CPU time on a 1.4 GHz Pentium III.

If I had allowed the positions to move, I expect a considerable increase in the time required, as well as possible problems with convergence. Getting rid of some of the approximations will probably increase the load by a factor of several. This isn't really interactive computing.

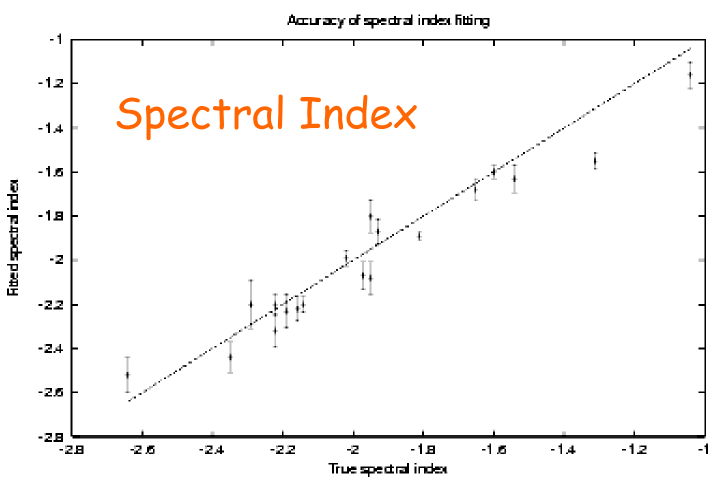
The Optimization Algorithms Agree

The Minuit and LBFGS optimization algorithms converge to nearly equal results in about the same amount of CPU time. If the tolerances were tighter, the agreement between the two probably would have been better. Δ ln(L) was still about 2.

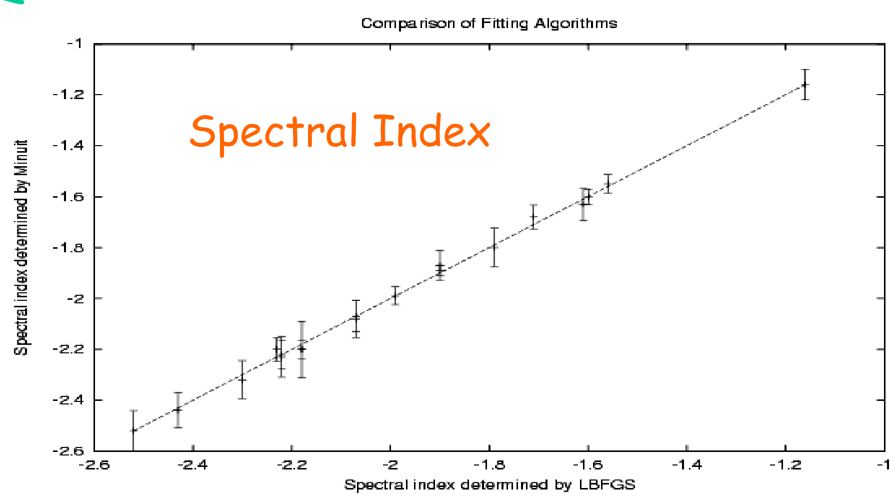


Minuit

Fitted



Input



LBFGS