GLAST Likelihood Calculations

Draft 1 NOV 2000

P. L. Nolan
Stanford University

ABSTRACT

Analyzing the data from the GLAST LAT will present some challenges which are not found in EGRET work or other branches of astronomy. The small number of photons means that approximate statistical methods such as χ^2 tests are not applicable. It will be necessary to use the full Poisson likelihood. The broad field of view and scanning mode of viewing mean that the data will not come packaged neatly in discrete observations. Any analysis, no matter how limited, will probably require stitching together selected parts of the whole data stream. The width of the point spread function varies with energy, angle of incidence, and internal detector variables. In effect, each photon will have its own PSF.

Calculation of likelihood values will involve multidimensional integrals. Some of these integrals are displayed, and some simplifying techniques are given.

1. This is not X-ray astronomy

Before proceeding with more formal topics, I feel the need to point out some ways in which the GLAST LAT (Large Area Telescope) [1] produces data which are very different from that produced by X-ray astronomy instruments or even EGRET [2]. I have heard a number of ideas circulating in the high-energy astronomy community which indicate that these differences are not sufficiently appreciated.

- GLAST won't stare at a fixed position on the sky. Most of the time it will point at the earth zenith, sweeping around the sky once per ~ 90 minute orbit. Any particular point source will be visible for ~ 40 minutes at at time. How to combine these intervals is a choice the data analyst should be able to make at will.
- The LAT field of view is huge, > 2 steradians. Lots of interesting sources will be in view at any time.

- The point-spread function (PSF) is broad, a few degrees for 20-30 MeV photons, but about 0.1° at 10 GeV (normal incidence). Sources will tend to overlap. There is a strong, structured background from the Galactic plane. These factors mean that one can't cut out a single point source and analyze it in isolation. Any analysis must involve a fairly large portion of the sky and it must simultaneously fit multiple point sources and the background.

- There are not many photons. The Vela Pulsar, the brightest steady point source, will produce a few dozen detected photons in a single orbital pass. We will be dealing with many sources three orders of magnitude fainter. As with EGRET, we will be analyzing sources with 50-100 photons collected over months of observations.

- Individual photon events are not simple. There will be thousands of bits of information from the tracker and calorimeter. These need to be analyzed to distinguish photons from cosmic rays and to estimate energy and direction. The software which does this is part of the system, just as much as the detector hardware.

In summary, the data stream does not naturally divide into small portions which can be studied in isolation. Each analysis will require selecting a fairly large fraction of the database, generally discontinuous in time.

2. Likelihood

During its expected lifetime, the GLAST LAT will observe hundreds of millions of photons, but for most analyses we will be interested in a subset of only a few hundred or a few thousand photons. In general the data will be too sparse to allow the use of χ^2 as a statistical test; it will be necessary to calculate the Poisson likelihood. Each photon is recorded individually, with characteristics such as energy E, direction $\mathbf{p}$ (α, δ or ℓ, b), and time t. There will also be other characteristics peculiar to the GLAST instrument, such as conversion layer, earth zenith angle, and instrument polar angle. We can label all these characteristics by $\boldsymbol{\lambda}$, a vector of coordinates in an abstract parameter space. When we establish quality cuts and regions of interest in the data, this will specify a subset $\boldsymbol{\Lambda}$ of the parameter space.

For a likelihood analysis [3], there must be a model $M(\boldsymbol{\lambda})$ which predicts the density of detected photons in $\boldsymbol{\Lambda}$. M includes both a sky model S, which describes the point and diffuse sources of gamma-ray emission, and the instrument response functions. The sky model S will include parameters such as source locations and fluxes which will be adjusted to maximize the likelihood.

Suppose $\mathbf{\Lambda}$ contains N detected photons indexed by the label i. The logarithm of the likelihood of observing this set of photons is

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

where

$$N_{\text{pred}} = \int_{\mathbf{\Lambda}} M(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \tag{2}$$

is the number of photons predicted by the model M. To calculate N_{pred} requires no access to the photon data.

We can produce an approximate likelihood by dividing $\mathbf{\Lambda}$ into small bins indexed by the label ν. Let the νth bin have n_ν detected photons, an average model density $\bar{M}_\nu$, a central parameter value $\boldsymbol{\lambda}_\nu$, and a volume $\Delta\boldsymbol{\lambda}_\nu$. The log likelihood of the model is

$$\ln \mathcal{L} = \sum_{\nu} \left[n_\nu \ln \bar{M}_\nu - \bar{M}_\nu \Delta\boldsymbol{\lambda}_\nu \right]. \tag{3}$$

It is easy to see that equation (1) is recovered in the limit $\Delta\boldsymbol{\lambda}_\nu \to 0$.

3. Instrument Response Functions

The instrument response of a gamma-ray telescope is usually characterized by three functions or matrices. This begs the question whether this decomposition is an acceptable approximation. It must be done this way, or we won't be able to do any computations.

- The effective area A is equal to the detector's actual projected area multiplied by the probability that a photon will react in the detector and produce a recognizable shower of particles. A is a function of the photon energy E and the angle of incidence. I assume that there is no azimuthal dependence.

  In the GLAST analysis, each photon-induced particle shower will be classified by some internal variables which describe the quality of the software reconstruction. These will probably include such things as the tracker layer in which the conversion occurred, the number of resolved tracks, the number of "noise hits" in the tracker, and whether the tracks cross tower boundaries. For now, we can lump all these extra factors into a single index k. There will be a different effective area for each value of k: $A_k(E, \cos\theta)$, where θ is the detector angle of incidence. If $\mathbf{z}$ is the detector axis unit vector, then the effective area for a sky direction $\mathbf{p}$ can be written $A_k(E, \mathbf{p}\cdot\mathbf{z})$.

- The energy dispersion D reflects the imperfection of the energy measurement. Some energy leaks out the sides of the tracker or through the calorimeter and leakage corrections are only approximate. There are fluctuations in the calorimeter light output. Some energy is absorbed by passive material. If the true energy of a photon is E, then the probability density of the estimated energy E' is $D_k(E'; E, \cos\theta)$. As with the effective area there are different D functions for different types of events.

- The point spread function (PSF) P reflects the imperfection of the direction measurement. The direction must be estimated by a weighted sum of the directions of the secondary charged particle tracks, which are perturbed by scattering in the tracker planes. If the true photon direction is $\mathbf{p}$ and the energy is E, the probability density of the estimated direction $\mathbf{p}'$ is $P_k(\mathbf{p}'; \mathbf{p}, E, \cos\theta)$. If the PSF is azimuthally symmetric, it can be written as $P_k(|\mathbf{p}' - \mathbf{p}|; E, \cos\theta)$. It is worth emphasizing that the width of the PSF varies strongly with E and $\cos\theta$, and that it will also be different for each k. In effect, there is a different PSF for each detected photon.

4. A Toy Problem

Let's work on a toy problem to illustrate how real calculations could be done. The sky model has a single point source at a fixed position $\mathbf{p}$ with no diffuse background. The source has an energy spectrum $s(E)$ with no time variation. The $\boldsymbol{\lambda}$ variables are estimated photon energy (E'), estimated direction ($\mathbf{p}'$), time (t), and the internal index (k). There is also a history of the detector pointing direction $\mathbf{z}(t)$.

$$M(E', \mathbf{p}', t, k) = \int_0^\infty D_k(E'; E, \mathbf{p}\cdot\mathbf{z}(t)) P_k(|\mathbf{p}' - \mathbf{p}|; E, \mathbf{p}\cdot\mathbf{z}(t)) A_k(E, \mathbf{p}\cdot\mathbf{z}(t))\, s(E)\, dE \quad (4)$$

$$\ln\mathcal{L} = \sum_i \ln M(E_i', \mathbf{p}'_i, t_i, k_i) - N_{\rm pred} \quad (5)$$

$$N_{\rm pred} = \sum_k \iiint M(E', \mathbf{p}', t, k)\, dE' d\mathbf{p}' dt \quad (6)$$

The goal is to adjust the parameters of $s(E)$ to maximize $\mathcal{L}$.

In practice, we don't use continuous functions. The response functions and the source energy spectrum are stored as matrices or vectors to be interpolated as necessary. Integrals are replaced by sums. The D and P functions are 4-dimensional objects, while A is only 3-dimensional. The combination

$$R_k(E', |\mathbf{p}' - \mathbf{p}|; E, \cos\theta) \equiv D_k(E'; E, \cos\theta) P_k(|\mathbf{p}' - \mathbf{p}|; E, \cos\theta) A_k(E, \cos\theta) \quad (7)$$

which appears in equation (4) is a 5-dimensional object. If each index has 100 values, then a 4-dimensional object can be stored in 400 megabytes. This is not an unreasonable burden. However, a 5-dimensional object will require 40 GB, which is something of a challenge to store and use. Thus the required values of R will probably be calculated on the fly.

Suppose that the E integration requires evaluating our functions at N_E different values of the photon energy. With N detected photons, $N \cdot N_E$ evaluations of the functions D, P, and A are required to produce the first term of (5).

The $N_{\rm pred}$ term is subject to some simplification for this particular problem. If the region $\mathbf{\Lambda}$ is large enough, $\int D_k(E'; E, \cos\theta) dE' = 1$ and $\int P_k(|\mathbf{p}' - \mathbf{p}|; E, \cos\theta) d\mathbf{p}' = 1$, so

$$N_{\rm pred} = \sum_k \int\int A_k(E, \mathbf{p} \cdot \mathbf{z}(t)) s(E) dE dt. \quad (8)$$

The number of function evaluations is proportional to the length of the observation. Since the 4-dimensional objects D and P have been integrated out, the cost of each evaluation should be relatively low.

Note that the likelihood calculation can be written as

$$\ln \mathcal{L} = \sum_i [\ln \int G_i(E) s(E) dE] - \int H(E) s(E) dE. \quad (9)$$

Most of the effort goes into evaluating $G_i(E)$ and $H(E)$. Once this is done, the values can be stored. It will then be relatively easy to optimize the parameters of s(E).

5. Binned Likelihood

Equation (3) shows a way to produce an approximate likelihood value. This method has an obvious drawback: Information is lost when photons are treated as if their individual $\boldsymbol{\lambda}$ values are equal to $\boldsymbol{\lambda}_\nu$. Furthermore, the approximation $\bar{M}_\nu \approx M(\boldsymbol{\lambda}_\nu)$ is accurate only for small bins. Otherwise a more elaborate calculation is needed to estimate $\bar{M}_\nu = (\Delta\boldsymbol{\lambda}_\nu)^{-1} \int_{\Delta\boldsymbol{\lambda}_\nu} M(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$.

However, practical experience with EGRET data has shown that the binned calculation (3) can have an advantage over the "exact" formula (1) in terms of accuracy. The specific difficulty comes about because of the normalization of the PSF. The energy-dependent PSF is available out to 20° from the true photon direction. Any tail beyond 20° is considered negligible. This turns out to be a false assumption at low energy. When the two terms of (1) are calculated separately, each is much larger than their difference. A tiny error in normalization (which affects the $N_{\rm pred}$ term more seriously) can have a large effect on the result. The binned calculation (3) is relatively immune to such problems because the same value of $\bar{M}_\nu$ is used in both terms. Taking a term-by-term difference also tends to reduce roundoff error.

The binned calculation will always require less computation than the unbinned version. The unbinned $N_{\rm pred}$ integration (2) is almost the same thing as the sum (3) since most of the work goes into finding $\bar{M}_\nu$. The additional sum over detected photons in (1) may or may not be a major addition, depending on the number of photons.

To do a binned calculation it is necessary to decide how to divide up $\mathbf{\Lambda}$. In general, detailed studies have not yet been done to determine the necessary accuracy or how to achieve it. Some work has been done, though, on energy binning.

This work was inspired by a shortcoming in the EGRET data analysis. The most common science task was to find point sources, determining their fluxes and locations. This was done by considering a single broad energy bin from 100 MeV to $\sim$ 30 GeV. Since the width of the PSF varies greatly over this range, there was a considerable loss of precision in the position determination. Bill Tompkins did simulations with different numbers of energy bins [4]. He concluded that 4 or 5 energy bins would be adequate for GLAST data, giving almost the same results as an unbinned analysis. Of course, more bins will be needed for spectroscopic work.

There is no explicit time dependence in this model. Nevertheless the scanning of the detector axis makes the response functions for any $\mathbf{p}$ vary with time. Thus to do an integral such as (8), time will need to be binned on a scale determined by the orbital period and the width of the field of view. In scanning mode the pointing direction will change about 4° per minute. This suggests that 1-minute time bins should be adequate for everything except studies of rapid pulsations.

The size of the angular bins should be energy-dependent, like the PSF. At the lowest end of the GLAST energy range, bins of 1° or so would be adequate. For high-energy photons, something at least an order of magnitude finer (on a linear scale) will be needed.

6. Multiple point sources

Consider a more realistic problem. Suppose our sky model contains multiple point sources and diffuse emission, both galactic and extragalactic. Assume that the sources are steady emitters and that the "shapes" of the diffuse components are known. Write $S(E,t,\mathbf{p}) = \sum_j F_j \phi_j(E,\mathbf{p})$. The index j numbers the point sources and the background components. Normalize ϕ_j so that $\int\int \phi_j(E, \mathbf{p}) dE d\mathbf{p} = 1$. Thus F_j is the flux due to each source or component. Now $\mathbf{p}$ is a variable location on the sky, not a fixed position as it was before.

Now

$$M(E', \mathbf{p}', t, k) = \sum_j F_j \int_0^\infty \int D_k(E'; E, \mathbf{p} \cdot \mathbf{z}(t)) P_k(|\mathbf{p}' - \mathbf{p}|; E', \mathbf{p} \cdot \mathbf{z}(t)) \times A_k(E, \mathbf{p} \cdot \mathbf{z}(t)) \phi_j(E, \mathbf{p}) d\mathbf{p}\, dE. \quad (10)$$

Comparing this to the simpler equation (4) shows that new features have been added, a sum over the sources and background components (j) and a two-dimensional integral over the chosen portion of the sky ($\mathbf{p}$). Whatever the complexity of our toy calculation, this seems to be worse by a large factor.

Things are not as bad as they might seem. For all the point sources, $\phi_j(E, \mathbf{p}) = h_j(E)\delta(\mathbf{p} - \mathbf{p}_j)$, where $\mathbf{p}_j$ is the position of source j and $h_j(E)$ is its normalized energy spectrum. This removes most of the $\mathbf{p}$ integrals. Only the (typically two?) background components must be integrated. In the EGRET analysis the background components were treated as known, fixed quantities, except for their overall normalizations, so pre-integrated versions were made available. Thus the complexity can be made to be proportional to the number of point sources + background components.

Take note of a special case. If the point source positions are all considered to be fixed, then the ϕ_j contain no adjustable parameters. Only the F_j values are adjustable. In this case it can be proven that there is one global maximum maximum of $\mathcal{L}$ with no local maxima, and the Hessian matrix

$$H_{jk} = \frac{\partial^2 \ln \mathcal{L}}{\partial F_j \partial F_k} = -\sum_i \frac{\phi_j(\boldsymbol{\lambda}_i)\phi_k(\boldsymbol{\lambda}_i)}{M(\boldsymbol{\lambda}_i)^2} \quad (11)$$

is always negative-definite. These conditions can simplify the fitting greatly.

7. Miscellaneous parting notes

No distinction has been made between the tasks of finding sources, estimating fluxes in chosen energy bands, and fitting model spectra. All involve adjusting model parameters to maximize $\mathcal{L}$. In EGRET analysis, the first two tasks were done by one program, and the third by another which accepted data from the first. This led to difficulties because the data passed between programs was not really suitable for the job. Approximations were made and information was lost. I strongly recommend that the GLAST software tools be designed with the goals in mind from first to last.

One drawback of likelihood analysis is that it doesn't produce a goodness-of-fit measure. Those of us who grew up using χ^2 for everything are bound to mourn this loss. It is lovely to be sure that your model is adequate to explain the data. Pure likelihood methods can answer questions like "Will the fit be improved significantly if I put another point source at x, y?" Unfortunately there are often a multitude of such questions available. Bayesian methods offer a way out, perhaps at the cost of many more likelihood calculations.

REFERENCES

[1] `http://www-glast.stanford.edu/`

[2] `http://lheawww.gsfc.nasa.gov/docs/gamcosray/EGRET/egret.html`

[3] Mattox, J.R. et al. 1996, ApJ 461, 396. "The Likelihood Analysis of EGRET Data"

[4] Tompkins, William 1999, PhD dissertation, Stanford University. "Applications of Likelihood Analysis in Gamma-Ray Astrophysics". Available electronically at `http://giants.stanford.edu/~billt/thesis.ps`

This preprint was prepared with the AAS LaTeX macros v4.0.