Unbinned Maximum Likelihood for LAT Data

J. Chiang (GSFC-UMBC)

Original version September 27, 2002;
Last Revised October 28, 2002

1 Basic Ideas

Although this memo is intended to address various practical issues for implementing a particular likelihood calculation, for the sake of clarity, it is useful to keep certain details somewhat abstract, at least initially. These details will be fleshed out in §2.

Similar notation and many of the ideas appearing in Pat's glastlike-paper (Nolan 2000) are used. The instrument response functions are denoted by

$$D(E'; E, \hat{p}, \vec{L}(t)) \equiv \text{Energy Dispersion} \quad (1)$$

$$P(\hat{p}'; E, \hat{p}, \vec{L}(t)) \equiv \text{Point Spread Function} \quad (2)$$

$$A(E, \hat{p}, \vec{L}(t)) \equiv \text{Effective Area} \quad (3)$$

where $\hat{p}'$ and E' are the apparent photon direction and energy, $\hat{p}$ and E are the true photon direction and energy, and $\vec{L}(t)$ is a generalized vector describing, as a function of time, the telescope attitude and location in space as well as internal LAT degrees-of-freedom such as instrument mode.

1.1 Likelihood Definition

The likelihood model, M, describes the expected distribution of photons,

$$M(E', \hat{p}', t) = \int dE\, d\hat{p}\, D(E'; E, \hat{p}, \vec{L}(t)) P(\hat{p}'; E, \hat{p}, \vec{L}(t)) A(E, \hat{p}, \vec{L}(t)) S(E, \hat{p}) \quad (4)$$

$$\equiv \int dE\, d\hat{p}\, R(E', \hat{p}', t; E, \hat{p}) S(E, \hat{p}). \quad (5)$$

The latter relation defines the function R, referred to hereafter as the "total response". The source model consists of point sources and diffuse emission,

$$S(E, \hat{p}) = \sum_i s_i(E) \delta(\hat{p} - \hat{p}_i) + S_G(E, \hat{p}) + S_{\rm eg}(E, \hat{p}). \quad (6)$$

The index i labels the individual point sources; $s_i(E)$ is the true energy spectrum of source i; and $\hat{p}_i$ is its location on the sky. S_G is the Galactic diffuse component, and $S_{\rm eg}$ is the extragalactic diffuse component. Note that the $s_i(E)$ have dimensions of $dN/dEdtdA$ while S_G and $S_{\rm eg}$ have dimensions of $dN/dEdtdAd\Omega$.

Labeling individual photon events with the index j, the logarithm of the Poisson likelihood is

$$\log \mathcal{L} = \sum_j \log M({E'}_j, \hat{p}'_j, t_j) - N_{\rm pred}, \quad (7)$$

where the predicted number of photons is

$$N_{\rm pred} = \int dE'\, d\hat{p}'\, dt\, M(E', \hat{p}', t). \tag{8}$$

1.2 Storing Parts of $\log \mathcal{L}$

As Pat observed, because of size constraints, we will likely want to calculate the total response, R, on the fly. However, the preceding formulation of the log-likelihood does lend itself to be broken up into constituent parts that we may wish to store separately in memory. In particular, if we define

$$a_{ij} \equiv \int dE\, s_i(E)\, R(E'_j, \hat{p}'_j, t_j; E, \hat{p}_i) \tag{9}$$

$$b_j \equiv \int dE\, d\hat{p}\, [S_G(E, \hat{p}) + S_{\rm eg}(E, \hat{p})]\, R(E'_j, \hat{p}'_j, t_j; E, \hat{p}) \tag{10}$$

$$c_i \equiv \int dE\, s_i(E) \int dE'\, d\hat{p}'\, dt\, R(E', \hat{p}', t; E, \hat{p}_i) \tag{11}$$

$$d \equiv \int dE\, d\hat{p}\, [S_G(E, \hat{p}) + S_{\rm eg}(E, \hat{p})] \int dE'\, d\hat{p}'\, dt\, R(E', \hat{p}', t; E, \hat{p}), \tag{12}$$

the log-likelihood is

$$\log \mathcal{L} = \sum_j \log \left(\sum_i a_{ij} + b_j \right) - \sum_i c_i - d. \tag{13}$$

It will be desirable to break-up b_j and d further into separate Galactic and extragalactic components so that their magnitudes, and perhaps also their overall spectral indices (see §1.4), can be included as fit parameters, and the integrals need only be done once at the start of the analysis.

The sizes of the a_{ij}, b_j, c_i and d arrays should not be prohibitive. For $N_{\rm phot} \sim 10^6$ photon events and $N_{\rm src} \sim 10^2$ point sources in the model, these data comprise $\sim 10^8$ values. However, for a typical analysis, only a relatively small number of the point sources, say 10 at most, will be subject to having their parameters varied for any stretch of time, so the number of stored values will likely be closer to $\sim 10^7$.

The advantage of storing these parts of the log-likelihood in this manner is that the model parameters that are less apt to be correlated in the optimization space are separated. This permits a relatively small fraction of the log-likelihood to be recomputed for each trial set of fit parameters and thus greatly increases the overall efficiency of the optimization calculation. For example, changing the flux or spectral index of a single source results in $N_{\rm phot} + 1 \sim 10^6$ terms, about 1% of the total, that need to be recomputed.

1.3 Region-of-Interest vs Source Region

In practice, the angular integrals over $\hat{p}$ and $\hat{p}'$ are not performed over the whole sky but rather over smaller regions that are determined by the point spread function and by the complexity of the local source model owing to structure in the diffuse emission or a high concentration of point sources. It is expected that queries to the photon database will specify an acceptance cone around the central location of the region to be analyzed. For this discussion, this nominally circular region will be referred to as the region-of-interest (ROI), and we take the ROI as the space over which the $\hat{p}'$ integrals are evaluated. Therefore, for an optimum set of model parameters, the number of selected photon events should be approximately $N_{\rm pred}$ (see eq. [8]), agreeing within Poisson uncertainties. Because of the breadth of the point spread function, sources fairly far outside of the ROI can contribute significantly to the observed number of photons found within the ROI. Therefore, the range of integration for $\hat{p}$ must be significantly larger than the ROI itself. We refer to this larger region as the "source region".

1.4 The Integrals over t

Because they have to be evaluated over the constantly changing orbit and attitude of the telescope, experience has shown that the integrations over t contained in the expressions for c_i and d (eqs. 11 & 12) can be extremely time-consuming, particularly when coupled with the energy and angular integrals for those quantities. Since the response functions will likely be indexed by a small number, say ten, of true energy values E_k, we can break-up the c_i integrals further,

$$\tilde{c}_i(E_k) \equiv \int dE'\, d\hat{p}'\, dt\, R(E', \hat{p}', t; E_k, \hat{p}_i), \tag{14}$$

so that the c_i values can be re-evaluated as necessary using an appropriate quadrature formula with weights w_k[1]:

$$c_i = \sum_k w_k \tilde{c}_i(E_k) s_i(E_k). \tag{15}$$

This allows for the fluxes or spectral indices of the point sources to be varied without incurring exorbitant computational cost. A similar scheme can be implemented for d.

This formulation assumes that the positions of the sources are not allowed to vary. If they are to be varied, then we can define

$$\hat{c}(E_k, \hat{p}) = \int dE'\, d\hat{p}'\, dt\, R(E', \hat{p}', t; E_k, \hat{p}). \tag{16}$$

For a source region defined on a 100×100 grid of pixels, the number of $\hat{c}$ values is $10^5 (= 10 \times 100 \times 100)$.

1.5 Source Time Dependence

For simplicity, I've assumed that all the sources are constant. One can take into account any time dependence by making the substitution

$$S(E, \hat{p}) \to S(E, \hat{p}, t) = \sum_i s_i(E, t)\delta(\hat{p} - \hat{p}_i) + S_G(E, \hat{p}, t) + S_{\rm eg}(E, \hat{p}, t) \tag{17}$$

for equation 6. However, this generalization prevents the decompositions of c_i and d discussed in the previous section. Because of the paucity of photons, it is expected that non-periodic time-dependence in form of flares or longer time scale secular changes in flux or spectral shape will be better characterized by performing constant-source likelihood analyses using sub-intervals of the data rather than trying to parameterize and fit a time-dependent model.

[1] Actually, all of the integrals in equations (9–12) will be performed using a similar sort of quadrature method.

2 Peeling Away the Layers of Abstraction

Even though the preceding discussion has stressed practical considerations such as storage requirements and computational efficiency, certain notational abstractions were made. In this section, those abstractions are made more concrete.

2.1 Telescope Degrees-of-Freedom and Livetime

The main simplification was including all of the telescope information in the generalized vector $\vec{L}(t)$. Here is a more complete description of the information contained therein:

- $\hat{x}_{\rm sc}(t)$, the unit vector along the spacecraft x-axis.
- $\hat{z}_{\rm sc}(t)$, the unit vector along the spacecraft z-axis.
- $m(t)$, instrument mode.

Note that all these are functions of time. Presumably, there will be different response functions for each mode, so substitutions along the lines of $R \to R^m$, $a_{ij} \to a_{ij}^m$, $b_j \to b_j^m$, $c_i \to c_i^m$, and $d \to d^m$ should be made in the equations of §1. The log-likelihood then has an additional sum over the different instrument modes:

$$\log \mathcal{L} = \sum_m \left[\sum_j \log\left(\sum_i a_{ij}^m + b_j^m \right) - \sum_i c_i^m - d^m \right]. \tag{18}$$

A livetime factor was implicit in the time integrals of the previous sections. With the specification of the instrument modes, the associated livetime factors must be made explicit. Define

$$\begin{aligned} \Theta^m(t) &= 1, \qquad \text{when live in mode } m && (19)\\ &= 0, \qquad \text{otherwise.} && (20) \end{aligned}$$

The likelihood model becomes

$$M(E', \hat{p}', t) = \sum_m \Theta^m(t) \int dE\, d\hat{p}\, R^m(E', \hat{p}', \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t); E, \hat{p}) S(E, \hat{p}); \tag{21}$$

and the expressions for c_i^m and d^m are then

$$c_i^m \equiv \int dE\, s_i(E) \int dE'\, d\hat{p}'\, dt\, \Theta^m(t)\, R^m(E', \hat{p}', \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t); E, \hat{p}_i) \tag{22}$$

$$\begin{aligned} d^m \equiv & \int dE\, d\hat{p}\, [S_G(E, \hat{p}) + S_{\rm eg}(E, \hat{p})] \\ & \times \int dE'\, d\hat{p}\, dt\, \Theta^m(t)\, R^m(E', \hat{p}', \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t); E, \hat{p}), \end{aligned} \tag{23}$$

2.2 Coordinate Systems

Both for notational and conceptual reasons, it is very useful to write photon, source, and telescope directions as unit vectors, such as $\hat{p}'_j$, $\hat{p}_i$, and $\hat{x}_{\rm sc}(t)$, rather than in terms of a specific coordinate system such as right ascension and declination (α, δ). This notation also fits in well with the object-oriented approach of the SkyDir class used by Gleam.

However, from a practical standpoint, one must still express these directions in terms of a coordinate system at some point.

There are three coordinate systems that will be used for any given analysis. The first two are fairly obvious. The photon events themselves will be stored in terms of celestial coordinates[2], while the instrument response functions are more conveniently and properly stored in terms of satellite coordinates, i.e., coordinates that are referenced to the directions of the spacecraft axes, $\hat{x}_{\rm sc}$ and $\hat{z}_{\rm sc}$. The third coordinate system we will use is defined specifically for a given analysis and is referenced to a fiducial direction, the nominal center of the ROI. Crudely speaking, this coordinate system corresponds to the one that is obtained by rotating the celestial coordinate system by a set of Euler angles that map the fiducial direction to the (0,0) direction (e.g., to the vernal equinox in equatorial coordinates). Although this is not absolutely essential, it provides a useful representation of the ROI and source region and allows for convenient storage of intermediate maps for the spatial parts of the b_j^m integrals and the spatial decompositions of c_i^m and d^m. After transforming to this system, the poles (i.e., $\delta = \pm 90^\circ$ in equatorial coordinates) are avoided, and maps that have constant pixel sizes in longitude and latitude are nearly tangent projections at the center of the ROI and thus have minimal distortion.

Given $\hat{x}_{\rm sc}$ and $\hat{z}_{\rm sc}$, it is straight-forward to express the direction of the ith source, $\hat{p}_i$, in satellite coordinates (θ_i, ϕ_i):

$$\begin{aligned}
\theta_i &= \cos^{-1}(\hat{p}_i \cdot \hat{z}_{\rm sc}) && (24)\\
\phi_i &= \cos^{-1}\left(\frac{\hat{p}_i \cdot \hat{x}_{\rm sc}}{\sqrt{1-(\hat{p}_i \cdot \hat{z}_{\rm sc})^2}}\right), \qquad \hat{p}_i \cdot \hat{y}_{\rm sc} \geq 0 && (25)\\
&= 2\pi - \cos^{-1}\left(\frac{\hat{p}_i \cdot \hat{x}_{\rm sc}}{\sqrt{1-(\hat{p}_i \cdot \hat{z}_{\rm sc})^2}}\right), \qquad \hat{p}_i \cdot \hat{y}_{\rm sc} < 0, && (26)
\end{aligned}$$

where $\hat{y}_{\rm sc} \equiv \hat{z}_{\rm sc} \times \hat{x}_{\rm sc}$. Writing the dependence of the response functions on source inclination θ and azimuth ϕ explicitly, we have

$$\begin{aligned}
R^m(E', \hat{p}', \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t); E, \hat{p}) &= D^m(E'; E, \theta(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)), \phi(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)))\\
&\quad \times P^m(\hat{p}'; E, \theta(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)), \phi(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)))\\
&\quad \times A^m(E, \theta(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)), \phi(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t))). \qquad (27)
\end{aligned}$$

The point spread function P has a dependence on the orientation of $\hat{p}'$ relative to $\hat{p}$ that can also be made explicit. In the minimal case, P depends only on $\hat{p}' \cdot \hat{p}$:

$$P^m = P^m(\hat{p}' \cdot \hat{p}; E, \theta(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t)), \phi(\hat{p}, \hat{x}_{\rm sc}(t), \hat{z}_{\rm sc}(t))). \qquad (28)$$

The unwieldiness of these expressions illustrates the utility of the abstract notation used in §1.

2.3 Other Considerations

- Cuts on zenith angle, inclination, etc.. These are straight-forward to handle in the code and are essentially incorporated into the livetime factors, $\Theta^m(t)$.
- Other LAT "degrees-of-freedom" such as those regarding the quality of the reconstruction: tracker conversion layer, noise hits, crossing of tower boundaries. Does this information all get swept into the response functions?

[2] For the present discussion, "celestial" refers to either equatorial (of some epoch) or Galactic coordinates.

3 Numerical Considerations

3.1 Response Function Representation: Interpolation versus Parameterization

3.2 Quadrature Accuracy