Likelihood Analysis of LAT Data

James Chiang

(GLAST SSC – GSFC/UMBC @ SLAC)

Maximum Likelihood for EGRET

- Maximum Likelihood analysis provides, in principle, the best possible estimate of source characteristics.
- EGRET used binned Poisson likelihood:

$$\mathcal{L} = \prod_k \frac{\theta_k^{n_k} e^{-\theta_k}}{n_k!}$$

  - θ_k = predicted counts, n_k = measured counts in map pixel k.
  - Events were binned in sky coordinates, e.g., RA, Dec, with 0.5° bins.
  - Count maps for E > 100 MeV (1000 MeV) were used for source detection, identification, and flux estimation.
  - Source count estimates from maps partitioned into standard energy bands (30, 50, 70, 100, 150, 300, 500, 1000, 2000, 4000, 10000 MeV) were used for spectral analyses.
  - A single effective psf (for a given measured energy range) was used for convolving the diffuse model and for estimating counts for each of the sources, regardless of intrinsic spectrum.

Unbinned Likelihood for LAT Data

- The nature of the LAT response functions motivates our use of unbinned likelihood:
  - Relatively broad PSFs ⇒ Emission from nearby point sources <u>always</u> overlaps. The amount of overlap is less severe at higher (> 1GeV) energies.
  - Strong PSF energy dependence ⇒ Intrinsic source spectrum affects the degree of source confusion.
  - Large FOV + strong variation of response as a function of incident angle (wrt instrument axes) + scanning mode ⇒ Each event effectively has its own response function.
- Unbinned likelihood is the limiting case of a binned analysis with infinitesimally small bins, each containing 0 or 1 count.
- The data space we consider includes an energy axis as well as photon direction. The spectral fitting is not decoupled from the source flux estimation.

Nuts and bolts of the Statistical Model

- Use a standard factoring of the total response, R:

$$R(E',\hat{p}';E,\hat{p},t) = A(E,\hat{p},\vec{L}(t))D(E';E,\hat{p},\vec{L}(t))P(\hat{p}';E,\hat{p},\vec{L}(t)),$$

A = effective area, D = energy dispersion, P = psf, E = photon energy, p = photon direction, L(t) represents the time variation of the instrument orientation and internal degrees of freedom, primes indicate measured quantities.

- The Source Model:

$$S(E,\hat{p},t) = \sum_i s_i(E,t)\delta(\hat{p}-\hat{p}_i) + S_G(E,\hat{p}) + S_{\mathrm{eg}}(E,\hat{p}) + \sum_l S_l(E,\hat{p},t),$$

This accounts for point sources, Galactic diffuse emission, extragalactic diffuse, and other diffuse and possibly time varying sources (e.g., LMC, Moon, SNRs, etc.).

Convolving with the Instrument Response

- The region-of-interest (ROI) is the extraction region for the data in measured energy, direction, and arrival time.
- Folding the source model through the instrument response yields the event distribution function, M, (i.e., the expected counts given the model) in the space of measured quantities:

$$M(E',\hat{p}',t) = \int_{\mathrm{SR}} dE\, d\hat{p}\, R(E',\hat{p}',t;E,\hat{p})S(E,\hat{p},t)$$

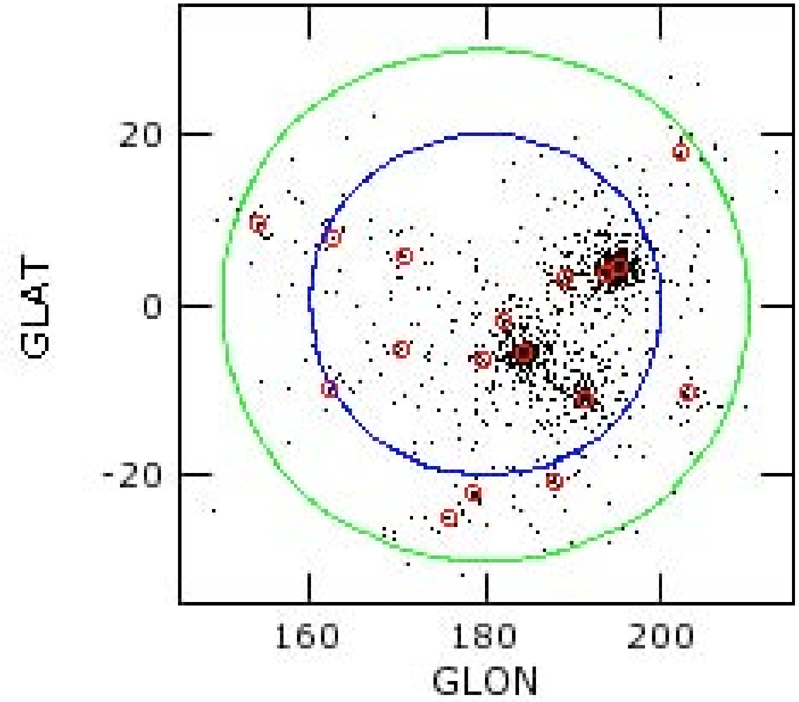
The "source region", SR, is the part of the sky defined to contain all sources that contribute significantly to the ROI.

- For standard analyses, we will treat "steady" sources, so that

$$S(E,\hat{p},t) \rightarrow S(E,\hat{p})$$

An Example Source Region and Region-of-Interest

- One day's worth of simulated events (black points) for the 3EG sources (red circles) in the Galactic anticenter region:



  - Here we have SR radius = 30° & ROI radius = 20°.
  - The DC1 IRFs were used to generate these data.

The Unbinned Likelihood

- Finally, the function we would like to maximize is

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

  - The sum is taken over all events, indexed by j, lying within the ROI.
  - Comparing this to the expression for the binned likelihood, the first term on the rhs can be identified with the factor $\Pi_k\theta_k$ and second term with $\Pi_k\exp(\_\theta_k)$.
- The predicted number of observed events is the integral of M over the ROI:

$$N_{\mathrm{pred}} = \int_{\mathrm{ROI}} dE'\, d\hat{p}'\, dt\, M(E', \hat{p}', t)$$

Likelihood's Exposure Maps

- The calculation of N_{pred} can be aided greatly by defining something similar to an exposure map:

$$\varepsilon(E,\hat{p}) \equiv \int_{\mathrm{ROI}} dE'\, d\hat{p}'\, dt\, R(E',\hat{p}',t;E,\hat{p})$$

This comprises quantities that are independent of any source model parameters and can thus be pre-computed.

- The predicted number of events in the ROI is then

$$N_{\mathrm{pred}} = \int_{\mathrm{SR}} dE d\hat{p}\, S(E,\hat{p})\, \varepsilon(E,\hat{p})$$

The extent of the exposure map should enclose the source region, SR

Modeling Sources

- For both point sources and diffuse sources, we assume that the spectral and spatial parts can be factored.
- Spectral Models:
  - Power-laws, broken power-laws, or power-laws with exponential cut-offs are typically, but any function that can be parameterized can be fit.
- Spatial Models:
  - FITS images can be used to describe the spatial distribution of emission from diffuse sources, e.g., the EGRET diffuse model. However, as with spectra, any parameterized function can be used.
  - Spectral variation across a diffuse source can be modeled by building it from smaller components.

Source Detection and Localization

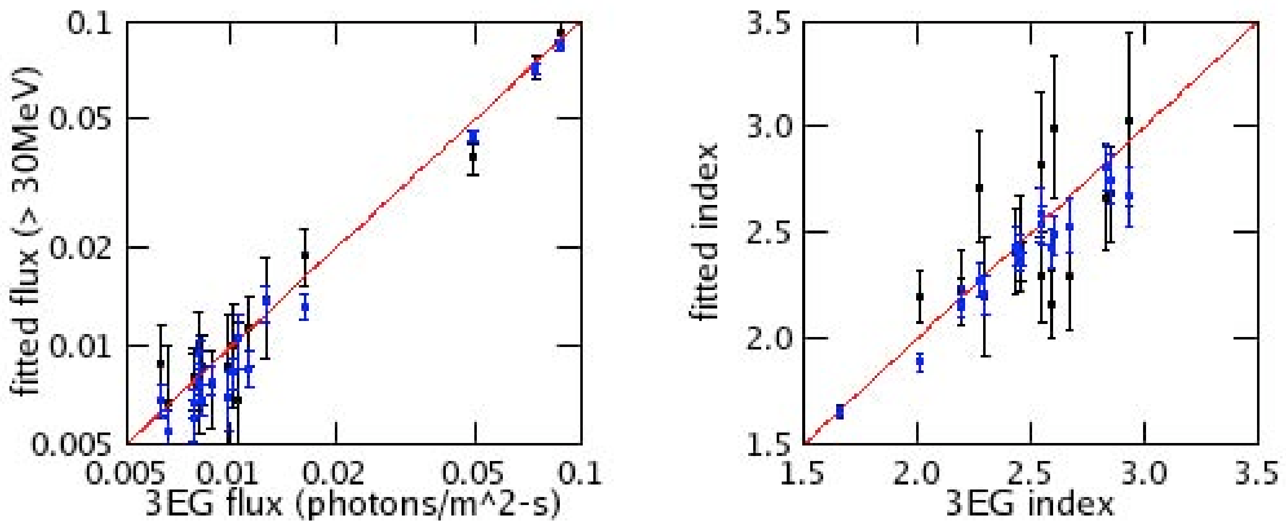
- Following EGRET analyses, we rely on "test-statistic" maps for detailed source detection and localization:

$$T_s = 2(\log \mathcal{L} - \log \mathcal{L}_0)$$

  - A point source is moved from each map location to the next and the maximum log-likelihood is evaluated.
  - The peak of the resulting T_s map is taken as the best fit location, and the 50, 68, 95, & 99% C.L. contours correspond to ΔT_s = 1.4, 2.3, 6.0, & 9.1 according to Wilks' Theorem (Mattox et al. 1996).
  - Accurate source positions rely on the other sources being accurately modeled. As with EGRET, an iterative "clean" algorithm will likely be required.
  - Rapid source detection methods based on wavelets or Bayesian Blocks may be used prior to Likelihood to find candidate sources.

Performance

- An example fit, the 17 strongest 3EG sources in the Galactic anticenter region (34 free parameters):



  - **black points**: 1 day simulation time, 1.7k events, 98 cpu secs on a 2.8 GHz Pentium 4 machine.
  - blue: 1 week, 11k events, 745 cpu secs.
  - Similar results are found when Galactic and extragalactic diffuse components are included (for a factor ~ 4 more events).
  - Execution time ~O(N_{events}).

Implementation

- Written in C++. All code is available from the SAS CVS repository: http://glast.stanford.edu/cgi-bin/cvsweb/
- The Likelihood package contains several applications:
  - likelihood: The fitting application itself.
  - expMap: Exposure map calculator.
  - diffuseResponses: Pre-computes the spatial integral part of N_{pred} for each diffuse component and each event.
  - TsMap: Creates test-statistic maps for source localization.
- Source model components and their parameters are specified in an xml file.
- The input data consist of event (FT1) and spacecraft data (FT2) files.
- The present implementation can be run interactively or in “batch” mode and can be easily driven by Python scripts for catalog analyses, light curve generation, etc..