The Gamma Ray Event Agglomeration Technique (GREAT)

Jeffrey D. Scargle

Space Science Division, NASA Ames Research Center

Jay Norris

Laboratory for High Energy Astrophysics
Code 661, NASA Goddard Space Flight Center

Draft of September 14, 2005.

Abstract

Description of a approximate way to compute the posterior prob-
ability that a given photon belongs to a source, given the energies of
the photons already known to belong to the source. This quantity
can be used to compare the probabilities that a given photon belongs
to two or more sources, aiding in the decision as to which source to
assign the photon to. The method does not require binning.

1 Introduction

At least two factors influence the probability that a photon belongs to
a source:

• the distance on the sky between the photon and the source
• the energies of the photon, compared to those in the source

2 Location, Location, Location

The probability that a photon belongs to a known source, based on
location alone, is simple to compute from the point spread function
(PSF), interpreted as the probability per unit area that the photon actually belongs to the source in question. One simply evaluates the value of the PSF, centered at the source centroid, at the photon position. The centroid is computed as the average position of the source’s photons, weighted inversely with the width of the PSF’s of these photons.

3 How Consistent is the Photon’s Energy with the Source Spectrum?

Here we are concerned with only the spectral information. Since the GREAT method deals with one photon at a time, it would seem that there is not much useful spectral information. However some improvement should result from the use of such information, however small it might be.

The direct approach would be:

1. estimate the spectrum of the source
2. interpret the spectrum as a probability distribution
3. evaluate this distribution at the energy of the given photon

Step 1 is based of course on the energies of the photons already assigned to the source. The number of such photons will be small at first, but will grow as the method progresses. Step 2 is merely a normalization.

We propose a simple, efficient way to carry out the above program. Figure 1 was constructed from a source of 66 photons, with energies ranging from about 500 MeV to 22.3 GeV. Only the 16 photons in the range from 500 to 580 MeV are plotted for ease of visualization. The ordinary histogram (yellow bars) is shown only for comparison.
purposes, and is not used in the algorithm. In blue is shown the simple histogram based the intervals between midpoints of successive data point pairs. The heights of the bars shown are equal to 1 divided by the width of the interval. This can be interpreted as the density estimate connected with Voronoi tessellation of data in any dimension (local density is proportional to the reciprocal of the volume of the Voronoi cell), or a non-equally spaced histogram in which each bin has the same number of data points, here 1. Note that this estimator does not disagree with the binned histogram, but can be considered superior in that it does not depend on an arbitrary choice of bin sizes and locations.

The figure also shows how one would evaluate this probability distribution at the energy of a new photon. The dotted line indicates such a photon (at 532 MeV), and the height of the bar containing the energy (see the asterisk) gives the corresponding probability. Hence we have this rule:

The probability that photon $P$ belongs to the spectrum corresponding to a set $S$ of $N$ photons is equal to the reciprocal of the width of the interval, in the ordered set of energies, which contains the energy of $P$, divided by $N$.

The division by $N$ arises from elementary probability theory: each event (photon) represents one of $N$ events, and therefore has a *probability mass* of $\frac{1}{N}$. This rule is implemented in the MatLab code shown below. Note that there is a need to consider the end cases in which the energy of the new photon is outside the range of source photon energies.
3 HOW CONSISTENT IS THE PHOTON’S ENERGY WITH THE SOURCE SPECTRUM?

function prob_energy = spectral_probability( ee_phot, ee_source )
%--------------------------------------------------------------------
% Input: ee_phot -- energy of a given photon
% ee_source -- set of energies of photons in a source
%--------------------------------------------------------------------
% Output: probability that the photon belongs to the source, based
% on an estimate of the source spectrum
%
% Use (the reciprocal of) the nearest-energy-distance
% to measure this probability.
%--------------------------------------------------------------------

num_phot = length( ee_source ); if num_phot == 0
    prob_energy = 1;
    return
end

ee_source = sort( ee_source ); % be sure energies are ordered
ii = find( ee_source > ee_phot ); % locate interval containing ee_phot

if isempty( ii )
    % energy is greater than any in the source
    nearest_energy_dist = 2 * ( ee_phot - ee_source( num_phot ) );
else
    ii = ii(1);
    if ii == 1
        % energy is less than any in the source
        nearest_energy_dist = 2 * ( ee_source(1) - ee_phot );
    else
        % average of the two bracketing energies
        nearest_energy_dist = ee_source(ii) - ee_source(ii-1);
    end
end

if nearest_energy_dist <= eps
    prob_energy = 1 / num_phot;
else
    % each photon contributes probability mass 1/N
    prob_energy = 1 / ( nearest_energy_dist * num_phot );
end

Note that this quantity is not to be taken as an absolute probability, but must be used in assessing relative probabilities that a photon belongs to one of 2 or more sources.
Figure 1: Spectrum estimates for a small sample of photons, with energies of individual photons represented by vertical red lines along the energy-axis. The single photon per bin estimate is superior to a grossly binned histogram, and is extremely simple to compute.