

A Method to Simulate the Trials Factor

Chuan Chen

When we use Milagro smoothed map to search around EGRET sources and get the maximum significance in the searching region, we need to correct for the trials factor. If we use the total number of searching points as the trials factor, then we overestimate it. For example if the searching region is a circle with 1 degree in radius, since we search on 0.10 grid, the total number of searching points in it is roughly 314. But the real trial factor is obviously less than 314. The method I used to simulate trials factors for different significances is below.

I simulated 1 billion (10^8) sky maps with RA from -2.5 to 2.5 degrees and DEC from -2.5 to 2.5 degrees with 0.1 by 0.1 degree grid. Then I populated signals and backgrounds on each sky map. After that I smoothed the maps with 1.5 degrees. I searched the region around the center of the map within 1.0 degree in radius and saved the maximum significance.

The equation I used to calculate the trials factor for n sigma sources is

$$\frac{N_m}{N_t} = 1 - (1 - P_n)^{N_{Trials}} \quad (1)$$

Here N_m is the number of maps with the maximum significance greater than n sigmas in the searching region.

N_t is the total maps (1 billion).

P_n is the probability greater than n sigmas. (e.g. $P_3 = 0.0013499$)

N_{Trials} is the number of trials.

If $P_n \ll 1$, then $\frac{N_m}{N_t} \approx N_{Trials} \times P_n$

From equation (1) we can get
$$N_{Trials} = \frac{\ln(1 - \frac{N_m}{N_t})}{\ln(1 - P_n)} \quad (2)$$

For example from the results of the simulation, in 1 billion maps there are 6,133,710 maps with the maximum significance greater than 3 sigmas. Plugging numbers into equation (2) we can get the trials factor for 3 sigma sources is 46.86. Similarly we can get the trials factors for other significances.

I simulated 4 situations. The bin sizes for smoothing are 1.5 and 2.1 degrees and the radii of searching regions are 0.5 and 1.0 degree respectively. Results are showed on next page. One is for normal scale and the other for log scale.

Conclusion:

From the simulation results we can see that the trials factors increases with the significance.

Above 5.0 sigmas there aren't enough statistics, so I can fit the plots between 2 to 5 sigmas and extrapolate them to high significance region.



