Weighting for Bins of a Hadron Identification or Energy Variable

Jim Linnemann
Michigan State University

April 17, 2006

Abstract

In this memo I (re-)derive Andy's weighting formula for binned data and find that with weighting, a $Q \approx 1.5$ improvement is possible compared to estimating the flux with the simple summed excess, given a Crab-like spectrum. This improvement in significance is more important than the estimated $Q = 1.1$ from using a likelihood analysis on the sky map (or using psf weighting) reported in Alexandreas et. al NIM A328 (1993). The significance gain is larger still for a harder spectrum. The derivation rests on two key elements: first, you are using each individual bin to independently estimate the flux, and second, the background fluctuations dominate the uncertainty of the photon excess. However, this improvement in significance is not without cost: if the excess vs. bin is not accurately reproduced when the weights are calculated, systematic errors in the flux determination result. A harder-than-assumed spectrum overestimates the flux. This systematic error can be checked for either by comparing the weighted estimate of the excess with the directly calculated excess, or a chi-squared fit of the bin excesses. If you are using bin weighting only for the significance calculation, not for the flux determination, the problem is much less, as sub-optimal weights merely reduce the significance achievable.

1 Weighted Flux Estimate from Binned Data

In Andy's memo of June 2005, "A Weighting Analysis of Crab Data", he used weights of the form $\langle s\rangle/\langle b\rangle$ for combining different bins of data, where $\langle s\rangle$ is the expected signal and $\langle b\rangle$ is the expected background in the bin (of X2, A4, Energy, or whatnot). At first I found this very surprising, as the background had already been subtracted, so why should I be allowed to apply expected signal to background all over again as a weight?

Alas, Andy didn't derive the formula. Here I do. The answer turns out to rest on two key elements: first, you are using each individual bin to estimate a flux, and second, the background fluctuations dominate the uncertainty of the photon excess. It is not, however, some likelihood-ratio of the photon excess being more likely to be a photon than a background event by the given factor, because the background has already been subtracted. Thanks to Andy for catching an algebra error I made in an earlier draft which prevented me from reproducing his formulae. I'm guessing that someone has shown the theoretical justification for these formulae before (unless Andy just has ridiculously good intuition), but if references exist, I'm not aware of them. Let me know if you do have a published reference.

Call $s_i = N_i - \hat{b_i}$ the photon excess measured in a bin. The expected value of the bin content is proportional to an incident overall flux by

$$\langle s_i \rangle = e_i \langle \Phi \rangle. \tag{1}$$

The whole weighting scheme depends on our wishing to estimate an overall flux linearly related to each bin content. Summing across bins, the flux is proportional to the total excess:

$$\langle S \rangle = \Sigma_i \langle s_i \rangle = \langle \Phi \rangle \Sigma_i e_i = \langle \Phi \rangle E. \tag{2}$$

We could choose $E = 1$ and define the photon excess as the "flux", or define E as the total effective area (and e_i the spectral-weighted effective area for the bin) to arrive at the more usual flux. As we will see later, either choice produces the same weights.

The most straightforward measure of the flux is just the total photon excess, $S = \Sigma_i s_i$, that is, weighting each bin equally. But this does not give the lowest variance estimate, as we will see.

Suppose instead I want to measure flux by a different linear combination of measurements from each bin. The intuition is that some bins have better fractional uncertainty, and should be weighted more heavily to give the best estimate. My flux estimate from bin i is just

$$\phi_i = s_i / e_i, \tag{3}$$

where e_i is the efficiency factor to convert counts into flux. For these estimates to make sense, they all have to be estimating the same thing, so from (1) linearity implies:

$$\langle \Phi \rangle = \langle \phi \rangle \text{ and thus } e_i = \langle s_i \rangle / \langle \Phi \rangle \tag{4}$$

Now combine the independent bin estimates of the flux:

$$\Phi = \Sigma_i(w_i\ \phi_i)/\ \Sigma_i(w_i) \tag{5}$$

A Maximum Likelihood estimator (with minimum variance), weights each estimate by the inverse of its (expected) variance. If the background fluctuations dominate, then

$$\sigma(\phi_i) = \frac{1}{e_i}\sqrt{\langle b_i\rangle},\ \text{leading to} \tag{6}$$

$$w_i = e_i^2/\langle b_i\rangle \tag{7}$$

When I substitute that back into the expression for the flux estimate Φ I find one of the e_i cancel so

$$\Phi = \Sigma_i\ \frac{e_i}{\langle b_i\rangle}\ s_i\ /\ \Sigma_i\ w_i \tag{8}$$

Substituting into this equation the expression (4) for e_i, the numerator is proportional to $\frac{\langle s_i\rangle}{\langle b_i\rangle}$. Defining

$$v_i = \langle s_i\rangle/\langle b_i\rangle, \tag{9}$$

I arrive at the weight Andy used in his memo, a constant (independent of i) times $\langle s\rangle/\langle b\rangle$:

$$\Phi = \langle\Phi\rangle\Sigma_i\ v_i\ s_i/\ \Sigma_i\ v_i\langle s_i\rangle = k\ \Sigma_i\ v_i\ s_i \tag{10}$$

$$k = \frac{\langle\Phi\rangle}{\Sigma_i\ v_i\ \langle s_i\rangle}. \tag{11}$$

Obviously from (10), a multiplicative constant[1] applied to the weights v_i doesn't affect the flux estimate. Andy defined the weights in the first bin to 1.0 by choosing new weights

$$u_i = v_i/v_1. \tag{12}$$

Note, however, that the v_i are not an i-independent multiple of the w_i. This causes the denominator of (10) and (11) to differ from a simple sum of the v_i.

[1]If you combine data samples which should have particular relative weights, you don't have that freedom for each sample separately.

2 Effects of Mis-weighting on Flux Estimate

This weighting clearly presumes you have a good handle on the spectrum so that the $\langle s_i \rangle$ are appropriate. If not, you will mis-weight the results. The optimized weighted estimate was designed so that each bin estimate had the same mean, the estimated flux: $\langle \phi_i \rangle = \langle \Phi \rangle$. For clarity, define [ ] as averages with respect to the true spectrum, in contrast to $\langle\ \rangle$ averages with respect to the spectrum assumed when defining the weights. Suppose the true spectrum gives for the photon excess $[s_i] = \lambda_i \langle s_i \rangle$ The individual estimates will then not average the same flux:

$$[\phi_i] = \lambda_i \langle \Phi \rangle \text{ and } [\Phi] = \langle \Phi \rangle \lambda_w \text{ where, by (5)} \tag{13}$$

$$\lambda_w = \Sigma_i w_i \lambda_i / \Sigma_i w_i \tag{14}$$

We can compare the flux estimate for the actual spectrum with its target value for the actual spectrum. In analogy with (2) we desire $[S] = [\Phi]E$. But

$$[S] = \Sigma_i [s_i] = \Sigma_i \lambda_i \langle s_i \rangle \text{ while} \tag{15}$$

$$[\Phi]E = \lambda_w \langle \Phi \rangle E = \lambda_w \Sigma_i \langle s_i \rangle \tag{16}$$

Taking the ratio, we find it is not guaranteed to be one, since

$$\frac{[S]}{[\Phi]E} = \lambda_s / \lambda_w \tag{17}$$

$$\text{where} \quad \lambda_s = \Sigma_i \ \lambda_i \langle s_i \rangle / \Sigma_i \langle s_i \rangle \tag{18}$$

In other words, the flux estimate is wrong unless λ weighted by the weights w_i and by $\langle s_i \rangle$ is the same. In the case that the shape of the spectrum used to estimate the weights was correct, but the normalization was wrong, $\lambda_i = \lambda$, and there is no error in estimating the final flux because $\lambda_s = \lambda_w = \lambda$. But if the shape was wrong, there will be a systematic error in estimating the final flux unless $w_i \propto \langle s_i \rangle$ where the constant is independent of i, and this is inconsistent with (3), (4) and (7).

You would also bias the significance by mis-weighting, because the variance would be larger than the minimum you could achieve with the correct weighting.

How large might the λ_i be? If the shape is a power law, the power is badly wrong by, say by .5 units, and the bins are in energy ranging from 1 to 25 TeV, the λ_i might be vary by a factor of 5 (assuming here the bins are roughly in log E).

3 Irrelevance of a Constant in a Weight and the Effective Number of Events

For the interested student, consider calculating any quantity as a weighted average (for example $\langle r^2 \rangle$, or a variance):

$$\langle f_w \rangle = \Sigma_i \; w_i f_i \; / \Sigma_i \; w_i \tag{19}$$

Now consider new weights $v_i = \lambda \; w_i$. Rewrite $\langle f \rangle$ in terms of the new weights:

$$\langle f_w \rangle = \Sigma_i \; \lambda \; v_i f_i \; / \Sigma_i \lambda \; v_i \tag{20}$$

On cancelling the common factor λ from top and bottom, we recognize that we get exactly the same average for either weight:

$$\langle f_w \rangle = \Sigma_i \; v_i f_i \; / \Sigma_i \; v_i = \langle f_v \rangle \tag{21}$$

As another example, consider the effective number of events:

$$N_{\text{eff}} = (\Sigma_i \; w_i)^2 / \Sigma_i \; w_i^2 \tag{22}$$

Notice that a common factor λ also cancels out in this expression.

One other comment about this formula: it warns you that a large dynamic range of weights decreases your effective statistical power. If some weights are much larger than others, only those events matter much in the sums; that will show up as a small number of effective events, and as a large fractional variance in the quantities you wish to calculate. However, we will see that we have chosen this emphasis, as many bins contain events with large variance, and they should indeed be given low weight: in the case of Φ, we are better off with a smaller effective number of events.

4 Calculation of Variance

The variance of a weighted sum of independent variables $X = \Sigma_i \; a_i x_i$ can be written as

$$\langle Var(X) \rangle = \Sigma_i \; a_i^2 \; \langle Var(x_i) \rangle \tag{23}$$

For the simplest estimate of $S = \Sigma_i s_i$, the expected variance is $Var(S) = \Sigma_i \; \langle b_i \rangle = B$ and the expected significance is $Z_S = \langle S \rangle / \sqrt{B}$.

For the weighted sum (with $E = 1$), from (10) the expected value is $k \Sigma v_i \langle s_i \rangle$ and the expected variance is $k^2 \Sigma v_i^2 \langle b_i \rangle$, so the expected significance is $Z_\Phi = \Sigma \; v_i \langle s_i \rangle / \sqrt{\Sigma_i v_i^2 \langle b_i \rangle}$.

The individual bin significance $\langle z_i \rangle = \langle s_i \rangle / \sqrt{\langle b_i \rangle}$ is of course independent of whether the bin is thought of as a member of a sum in an excess, or weighted by $1/e_i$ to form a ϕ estimate.

5 Results for Crab Spectrum Data

I built a spreadsheet based on the formulas above to investigate the implications. I used Andy's Crab "data": real background estimates, and MC signal estimates, to understand the expected behavior. This means I calculated expected performance, not actual performance with a specific data set. I scaled the background estimate to match the MC signal until I got a reasonable significance estimate: 3.00 for the last X2 bin alone, and 4.32 for the whole sample (Andy's bins 1-7). Applying Andy's weights to the sample then gives a Q factor of 1.49, or an expected significance for the whole sample of 6.43 sigmas when using the weighted flux estimate instead of just the excess itself. This enhancement of significance is certainly worthwhile, whatever the significance for the simple sum, because the fractional statistical error on the flux is just $\sigma(\hat{\Phi})/\hat{\Phi} = 1/Z$ where Z is the significance of the measurement.

With Andy's data, the straight excess of the highest bins (hardest cuts) was not actually strictly increasing in significance as the worse (lower X2) bins were added: the best significance for the excess came from the top 3 bins, which gave $Z = 4.47$, a bit better than the $Z = 4.32$ from including all bins (Q = 1.03, nothing to write home about). On the contrary, adding each lower bin improved the significance with the weighting scheme, though a rough estimate based on the Crab paper shows that adding the bin below the first bin would be expected to have lowered the significance for the simple excess and possibly the weighting scheme as well. To check the program, I also put in $\lambda = 2$, just estimating the wrong flux with the correct spectrum. As expected, the results were identical to the having the correct spectrum and normalization, and in particular the flux is estimated correctly.

6 Results for Harder or Softer Spectra

I then investigated the effects of having mis-specified the spectrum while choosing the weights. I defined lambda factors by hand to roughly simulate having a spectrum power in error by $\pm.1, \pm.5$ units, by imagining that the X2 bins were really energy bins between 1 and 25 TeV, so that λ (or $1/\lambda$) ranged from 1 to 5 for the $\pm.5$ error, and 1 to 1.5 for the $\pm.1$ error in spectral index. Rather than doing an exact calculation (the bins aren't really energy), I just cooked up some plausibly spaced λ values in the range above. The results are shown in Table 1.

In all cases but a spectrum *much* softer than assumed, the significance for the weighted estimate is better than the significance for the un-weighted summed excess. A real spectrum softer than the assumed spectrum is of

$\delta(\alpha)$	Q	$[S]/[\hat{\Phi}]$
0 (nominal)	1.49	1.00
.1 softer	1.30	1.15
.1 harder	1.72	0.87
.5 softer	.93	1.60
.5 harder	2.67	0.56

Table 1: Expected significance improvement and flux bias for spectra different than that assumed when choosing weights.

course more difficult to detect with a bin variable correlated with energy; further, the actual excess (flux) is larger than the excess estimated by the weighted method for a soft spectrum, and the significance gain is less than that available were weights to have been calculated with the soft spectrum. If the real spectrum is harder than the spectrum used in the weight assignment, the significance improvement is larger than would have been attained with correct weights for the nominal spectrum, though somewhat less than the $Q = 2.99$ if the correct harder spectrum had been used to calculate weights. The hard spectrum is easier to detect, but results in an overestimate of the excess (flux) when the weights come from a softer spectrum. The flux systematic errors are about 15% for a spectral index off by .1, and around 50% for spectral index error of .5 units.

To summarize, the achievable Q factor (gain from emphasizing bins with harder cuts) depends on the spectral index of the actual spectrum, and is best when the correct spectrum is used for determining the weights. The significance is usually better for a weighted estimate than for the simple excess unless the true spectrum is much softer than that used in the weight calculation.

To minimize the systematic dependence of the estimated flux on the assumed spectrum, there seem to be two steps. The first is diagnosis, most simply by comparing the computed excess with its weighted estimate and seeing whether they are statistically compatible; or more elaborately by fitting the binned distribution of excesses to the predicted spectrum and checking the chi-squared for their agreement. Second, if the agreement appears to be unsatisfactory, a spectral determination should be pursued, either by adjusting the assumed spectrum until better agreement is found, or by measuring the spectrum by other means and using that as the input to the weighting analysis.