Jim Wiss

August 1995

Errors on Weighted Averages, and Weighted Histograms

In this memo I discuss the method used to compute errors on weighted histograms, and averages. Special cases of this method such as the errors on unweighted averages, and asymmetries will also be discussed. In the SOAR data reduction program, the errors in are obtained by accumulating various moments of the weights as one increments the weighted histogram or plot. I will conclude by a discussing a test of the errors for a weighted average using a Monte Carlo simulation.

Weighted Histogram

This is the simplest example which illustrates the general principle. A given bin of weighted histogram has a sum (or weighted yield) given by $S = \sum W$. We are looking for a way of computing $\sigma^2(S)$ due to counting statistic fluctuations of the data. Usually the weight for a given entry will depend on some well measured property of the given event such as a kinematic property. Often the weight is a continuous variable, but in order to make the problem tractable, we cheat and consider a discrete set of possible weights, $W \in \{W_\alpha\}$, where α is an integer which we will call the "weight class". Let n_α be the entries in the weight class α. The weighted yield is then:

$$S = \sum_\alpha n_\alpha \; W_\alpha \tag{1}$$

The advantage of using is a discrete weight set $\{W_\alpha\}$ is that we can attribute fluctuations in the sum S to Poisson fluctuations of the number of events in each weight class n_α. We assume that these Poisson fluctuations obey:

$$< \; \delta n_\alpha \; \delta n_\beta \; >= \delta_{\alpha\beta} \; n_\alpha \tag{2}$$

which means that different bins are uncorrelated and the usual counting statistics applies: $\sigma = \sqrt{n}$. To calculate the variance on S we autocorrelate the fluctuations on δS due to fluctuations on the number of weight class entries δn_α

$$\delta S = \sum_\alpha W_\alpha \ \delta n_\alpha \quad , \quad \sigma^2(S) = \sum_\alpha \sum_\beta W_\alpha \ < \ \delta n_\alpha \ \delta n_\beta \ > \ W_\beta$$

$$\sigma_S^2 = \sum_\alpha W_\alpha^2 n_\alpha = \sum W^2 \tag{3}$$

The last equality says that sum of squared of each weight when multiplied by the number of events in each weight class is equivalent to just accumulating squared weights for each entry one at a time. We thus have the well known result:

$$\sigma_S = \sqrt{\sum W^2} \tag{4}$$

If all the weights are one (such as in an unweighted histogram), Eqn. (4) says that error is just the $\sqrt{N}$. In weighted histogrammers such as SOAR, ODDPACK, and others, one simply increments $S \rightarrow S + W$ and $\sigma_S^2 \rightarrow \sigma_S^2 + W^2$ as each entry is added to each bin.

How much does weight variation inflate errors?

Let us define the average weight, $< W >$, and weight variance σ_W^2 through the equations:

$$< W >= \frac{\sum W}{N} = \frac{\sum_\alpha n_\alpha W_\alpha}{N} = \frac{S}{N}$$

$$\sigma_W^2 = \frac{\sum (W - < W >)^2}{N} = \frac{\sum_\alpha n_\alpha \ (W_\alpha - < W >)^2}{N} \tag{5}$$

where $N = \sum_\alpha n_\alpha$. It is straighforward to show and well known that $\sigma_W^2 =< W^2 > - < W >^2$ where

$$< W^2 >= \frac{\sum_\alpha n_\alpha W^2}{N} = \frac{\sum W^2}{N} = \frac{\sigma_S^2}{N} \tag{6}$$

We can manipulate Equations (5) and (6) to obtain:

$$\frac{\sigma_S}{S} = \sqrt{\frac{N\sigma_W^2}{N^2 < W >^2} + \frac{N \ < W >^2}{N^2 < W >^2}} = \frac{1}{\sqrt{N}} \sqrt{1 + \left(\frac{\sigma_W}{< W >}\right)^2} \tag{7}$$

Equation (7) says that there will be an inflation in the relative error of a weighted histogram beyond the $1/\sqrt{N}$ errors due to counting statistics on number of events in a given bin which is related to the rms spread of the contributing weights. Our experience is that often this inflation factor is fairly small **unless** one has negative weights. When a histogram has nearly canceling positive and negative weights due to *eg* severe background subtraction, or severely smeared deconvolution, one can have $\sigma_W \gg < W >$. The fractional error on such weighted averages can greatly exceed the fractional error on an unweighted histogram. We have seen this effect in both Glen and Karen's measurement of the 87/88 cross section and the recent resonance analysis of the KK$\pi\pi$ final state performed by the Milan group.

Errors on a Weighted Average

Consider the weighted average $\overline{X}$ of a quantity X subject to a weight W:

$$\overline{X} = \frac{\sum W \ X}{\sum W} = \frac{\sum_\alpha \sum_i \ n_{\alpha i} \ W_\alpha \ X_i}{\sum_\alpha \sum_i \ n_{\alpha i} \ W_\alpha} = \frac{A}{B} \tag{8}$$

A few notational comments are in order. We have introduced both a weight class α and X class i and are again thinking of discrete outcomes $W \in \{W_\alpha\}$ and $X \in \{X_i\}$. The variables $A \equiv \sum W \ X$ and $B \equiv \sum W$ are just shorthand for the numerator and denominator of Eqn. (8). In this picture the source of fluctuations on $\overline{X}$ are the Poisson fluctuations on $n_{\alpha i}$ or the number of events in both the weight class α and the X class i.

Unlike the case of the weighted histogram, we must Taylor expand the fluctuation $\delta\overline{X}$ to first order in fluctuation $\delta n_{\alpha i}$:

$$\delta\overline{X} = \left(\frac{1}{B}\frac{\partial A}{\partial n_{i\alpha}} - \frac{A}{B^2}\frac{\partial B}{\partial n_{i\alpha}}\right)\ \delta\ n_{i\alpha} = \left(\frac{W_\alpha\ X_i}{B} - \frac{A\ W_\alpha}{B^2}\right)\delta\ n_{i\alpha} \tag{9}$$

Autocorrelating and averaging we have:

$$<\ \delta\overline{X}\delta\overline{X}\ \ >= \sum_{\alpha i\beta j}\left(\frac{W_\alpha\ X_i}{B} - \frac{A\ W_\alpha}{B^2}\right)\ <\ \delta n_{i\alpha}\delta n_{j\beta}\ \ >\ \left(\frac{W_\beta\ X_j}{B} - A\ \frac{W_\beta}{B^2}\right) \tag{10}$$

Invoking independent Poisson fluctuations

$$<\ \delta n_{i\alpha}\delta n_{j\beta}\ \ > = n_{i\alpha}\ \delta_{\alpha\beta}\ \delta_{ij} \tag{11}$$

we have:

$$\sigma^2_{\overline{X}} = \frac{1}{B^2}\ \sum\ (WX)^2 + \frac{A^2}{B^4}\sum\ W^2 - \frac{2A}{B^3}\ \sum\ W^2X$$

$$\sigma^2_{\overline{X}} = \frac{\sum\ W^2X^2}{\left(\sum\ W\right)^2} + \frac{\left(\sum\ WX\right)^2\ \sum\ W^2}{\left(\sum\ W\right)^4} - 2\ \frac{\left(\sum\ WX\right)\ \sum\ W^2X}{\left(\sum\ W\right)^3} \tag{12}$$

Special Cases

As a first test we consider the special case of an unweighted average where $W = 1$ throughout. Using $\sum\ W = N = \sum\ W^2$ where N the number of events in the average, we have:

$$\sigma^2_{\overline{X}} = \frac{\sum X^2}{N^2} + \frac{\left(\sum\ X\right)^2}{N^3} - 2\ \frac{\left(\sum\ X\right)^2}{N^3} = \frac{1}{N}\left(<\ X^2 > - < X >^2\right) \tag{13}$$

where $< X >= \sum\ X/N$ and $< X^2 >= \sum\ X^2/N$.

We next consider the more limiting case of an unweighted asymmetry where $W = 1$ and :

$$X = -1 \;\text{ for } N\ \frac{1-\alpha}{2} \text{ events and } X = +1 \;\text{ for } N\ \frac{1+\alpha}{2} \text{ events} \tag{14}$$

This means $\sum\ X = N\ \alpha$ and $\sum\ X^2 = N$ and thus:

$$\overline{X} = \frac{\sum X}{N} = \alpha \quad , \quad \sigma_x = \sqrt{\frac{\sum X^2}{N^2} - \frac{(\sum\ X)^2}{N^3}} = \sqrt{\frac{1-\alpha^2}{N}} \tag{15}$$

How much does weight variation inflate errors on asymmetry?

We re-cast Eqn. (12) for the case of an asymmetry measurement where the X values are described by Eqn. (14). In this case:

$$\frac{\sum\ WX}{\sum W} = \alpha \quad , \quad \sum\ W^2 X^2 = \sum\ W^2 \tag{16}$$

We thus have:

$$\sigma_\alpha^2 = \frac{\sum\ W^2}{(\sum W)^2} + \alpha^2\ \frac{\sum\ W^2}{(\sum W)^2} - 2\alpha\ \frac{\sum\ W^2\ X}{(\sum\ W)^2} \tag{17}$$

The last term is a bit tricky because of the evaluation $\sum\ W^2\ X$ if we allow for the possibilty that the events with with $X = \pm 1$ have different average weights ($ie < W_+^2 > \neq < W_-^2 >$). Let us denote:

$$< W_+ >= V + \Delta \quad , \quad < W_- >= V - \Delta$$

$$< W^2 >= \frac{1+\alpha}{2}(V+\Delta) + \frac{1-\alpha}{2}(V-\Delta) = V + \alpha\ \Delta \tag{18}$$

$$\sum\ W^2\ X = N\ \frac{1+\alpha}{2}(V+\Delta) - N\ \frac{1-\alpha}{2}(V-\Delta) = N\Delta + N\alpha V$$

$$\sum\ W^2\ X = N\ \alpha\ \ < W^2 > + N(1-\alpha^2)\ \Delta$$

$$\sum W^2 X = \left(\sum W^2\right) \left(\alpha + (1-\alpha^2) \frac{< W_+^2 > - < W_-^2 >}{2 < W^2 >}\right) \tag{19}$$

Inserting Eqn. (19) into Eqn. (17) we get:

$$\sigma_\alpha^2 = (1-\alpha^2) \frac{\sum W^2}{\left(\sum W\right)^2} \left(1 - \alpha \left(\frac{< W_+^2 > - < W_-^2 >}{< W^2 >}\right)\right) \tag{20}$$

Using $\sigma_W^2 =< W^2 > - < W >^2$ we have:

$$\frac{\sum W^2}{\left(\sum W\right)^2} = \frac{< W^2 >}{N \ < W >^2} = \frac{1}{N}\left(1 + \left(\frac{\sigma_W}{< W >}\right)^2\right) \tag{21}$$

Hence:

$$\sigma_\alpha = \sqrt{\frac{1-\alpha^2}{N}} \sqrt{1 + \left(\frac{\sigma_W}{< W >}\right)^2} \sqrt{1 - \alpha \left(\frac{< W_+^2 > - < W_-^2 >}{< W^2 >}\right)} \tag{22}$$

For the case where $< W_+^2 > \approx < W_-^2 >$ we note that that the error bars for an asymmetry are inflated by the weight spread in a way analogous to Eqn. (7). A large inflation occurs only if $< W > \ll \sigma_W$.

Monte Carlo Tests

I performed a series of Monte Carlo tests to examine the error formula given in Eqn. (12) for the case of the errors on a weighted average. The test was inspired by Rob's charm asymmetry measurements. I tied a toy model asymmetry to a variable x according to

$$\alpha = 0.5 \ (.5 - \sqrt{x}) + 0.2 \ (\tilde{R} - .5) - .2 \tag{23}$$

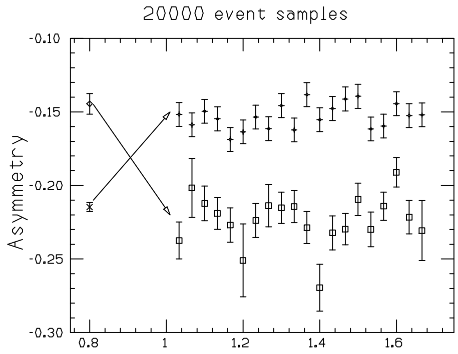
where $\tilde{R}$ is a uniformly drawn random number from 0 to 1 designed to prevent α from being completely correlated x. The variable x (meant to be a parton distribution) was drawn according to

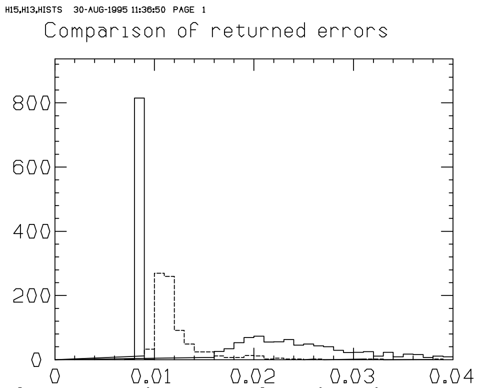
$$dN/dx \propto (1-x)^\eta \quad ,0 < x < 1 \tag{24}$$

We wanted to investigate the inflation of errors when one weights a Monte Carlo generated with $\eta = q$ in order to mimic a Monte Carlo which would have been generated with $\eta = p$. Specifically I would try to recover the simulated asymmetry for data drawn with a power $\eta = p$ by computing the asymmetry using a sample events simulated according to a power $\eta = q$ by making averaging the asymmetry variable (1 or -1) with a weight:

$$W = (1-x)^p/(1-x)^q \tag{25}$$

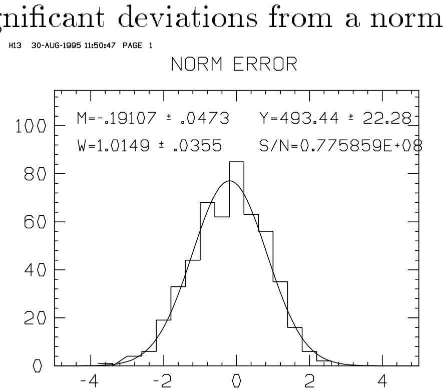
The figure on the left illustrates this weighting investigation. For the case of $\eta = 1$ this model gives an asymmetry of $\alpha = -0.21$; while for $\eta = 3$ the asymmetry is $\alpha = -0.14$. The upward arrow shows the use of a weight $W(p,q) = W(3,1)$ used to reweight data generated according to $(1-x)^1$ to compute the asymmetry for data which would have been generated according to $(1-x)^3$. The downward arrow shows the use of a weight $W(p,q) = W(1,3)$ used to reweight data generated according to $(1-x)^3$ to compute the asymmetry for data which would have been generated according to $(1-x)^1$. As you can see for this set of 20 experiments consisting of 20,000 simulated events, the weighted asymmetry does cluster about the appropriate asymmetry computed using unweighted events (the leftmost points). We also note that the returned error bars using Eqn. (12)for $W(1,3)$ are much larger than for $W(3,1)$. One might expect this since for $W(1,3)$ one has generated a narrow x distribution to infer the properties of a broad distribution and must very heavily weight events at large x. Specifically the $W(1,3)$ weights are given by $(1-x)^{-2}$ and range from $1 < W < \infty$. By way of contrast $W(3,1)$ weights are much less wild since they given by $(1-x)^2$ which ranges from $0 < W < 1$.

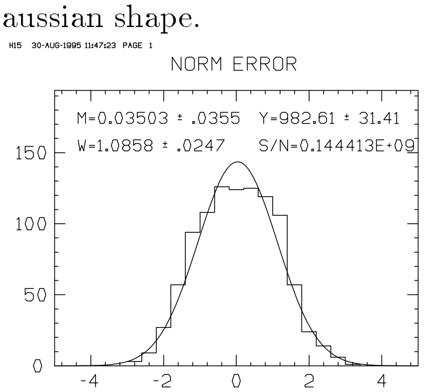




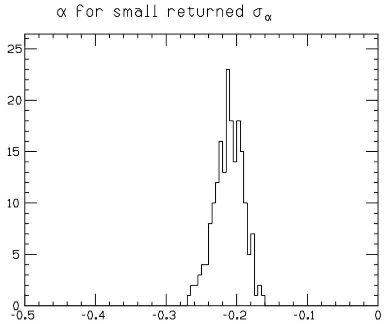
The figure on the right shows a distribution of returned errors for the three cases (left to right) $W(3,1)$, $W(1,3)$, and $W(1,5)$. The spectrum of returned errors is surprisingly broad and appears to have a tail extending out to ∞ for the case of $W(1,5)$ where the weights $(1-x)^{-4}$ have the most wild variation.

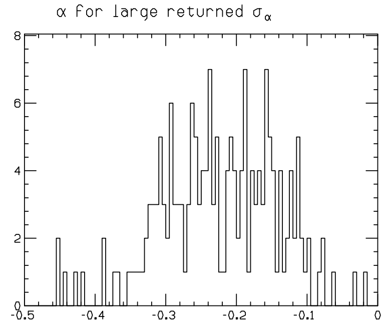
The below two figures compares the normalized errors on the asymmetry for the case $W(1,3)$ and $W(1,5)$. The normalized error is the deviation from the true asymmetry divided by the error returned from Eqn. (12). In both cases (and in the case W(3,1) not shown) the normalized error has the appropriate unit width although the normalized error distribution for the pathological $W(1,5)$ shows significant deviations from a normalized Gaussian shape.





The below two figures shows the distribution of returned α estimates for the $W(1,5)$ case when $\sigma_\alpha < 0.02$ (left) and $\sigma_\alpha > 0.04$ (right). Indeed events with a large computed σ_α appear to be very badly measured.





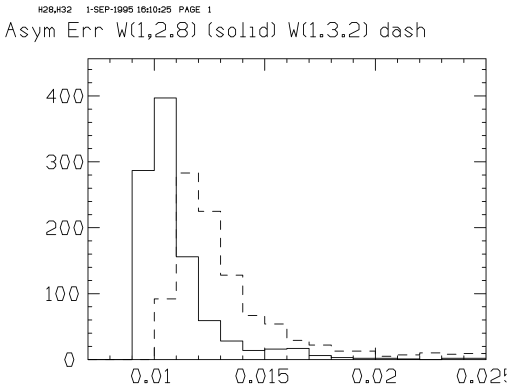
A Speculation

I believe that the distribution of weights for the cases $W(1,3)$ and $W(1,5)$ falls off too slowly as $W \to \infty$ to have a finite second moment and hence σ_W technically infinite. Eqn. (22) would say one would expect to observe infinite errors on α for these cases since the weights are so wild. Of course in any given trial with a finite number of generated events, one will never actually get an infinite predicted error since Eqn. (12) will produce a finite result. A reasonable conjecture might be that the returned error distribution will pick up significant high side tail once σ_W becomes divergent.

For the case at hand, one can show that the distribution of weights is of the form:

$$\frac{dN}{dW} \propto W^{\frac{1+2q-p}{p-q}} \tag{26}$$

The condition that σ_W is finite is that dN/dW falls off faster than W^{-3}. I examined whether or not a tail developed dramatically when one chose p and q about values giving $dN/dW \approx W^{-3}$. The below plot shows the distribution of returned errors for the cases $W(p,q) = W(1,2.8)$ and $W(2,3.2)$ which have powers of $dN/dW \propto W^{-3.111}$ and $W^{-2.909}$ respectively.



Although there is a significant broadening of the returned estimator distribution once one falls off slower than $dN/dW \propto W^{-3}$, it does not appear that there is a dramatic "phase transition" once σ_W diverges. This might again reflect the fact that a given experiment with a finite number of trial never senses that the σ_W has become divergent since it never actually produces events with divergent weights.