Methodology and code used in HFAG semileptonic averages

Bob Kowalewski
U. of Victoria
June 25, 2005

Abstract

The method used to average semileptonic results is described after giving a brief motivation. The modifications to the existing COMBOS program are documented, and some sample results given.

1 Averaging results in the Gaussian limit

The underlying assumption made when averaging measurements is that they are drawn from distributions that have a common mean value. To the extent that the uncertainties are Gaussian distributed, one can form an estimator for the common mean by minimizing an appropriate χ^2, as shown below. For simplicity, consider the case for N uncorrelated measurements y_i with variances $\bar{\sigma}_i^2$:

$$\chi^2(t) = \sum_i^N \frac{(y_i - t)^2}{\bar{\sigma}_i^2} \ . \tag{1}$$

Minimizing w.r.t. t gives the well-known solution:

$$\begin{aligned} \hat{t} &= \frac{\sum_i^N y_i/\bar{\sigma}_i^2}{\sum_i^N 1/\bar{\sigma}_i^2} \\ \sigma_{\hat{t}}^2 &= \left[\sum_i^N 1/\bar{\sigma}_i^2\right]^{-1} \ . \end{aligned} \tag{2}$$

This estimator is unbiased if the individual measurements are drawn from normal distributions with common mean $\bar{t}$ and r.m.s. $\bar{\sigma}_i$, where the bar is used to indicate the true, unknown values of these quantities, i.e. the "theory" values that correspond to the PDFs from which the measurements were drawn. Clearly $\hat{t}$ is an estimator for $\bar{t}$, and $\sigma_{\hat{t}}$ is an estimator for the uncertainty in $\bar{t}$.

The χ^2 formula given above depends on $\bar{\sigma}_i$. Unfortunately, we don't have access to these, but only to estimators for them based on the errors σ_i

assigned to the measurements. In cases where the assigned errors σ_i depend on the measured values y_i, we need to account for this dependence in order to obtain unbiased estimators for the $\bar{\sigma}_i$ (and, as we'll see below, for $\bar{t}$); i.e. we must use $\hat{\sigma}_i = \sigma_i(\hat{t})$, not σ_i ($\equiv \sigma_i(y_i)$), as our estimator for $\bar{\sigma}_i$. A few simple examples that illustrate this point are given in Appendix A.

Now the bad news: in practice we often don't know the dependence of σ_i on y_i. For example consider the errors on a BF, i.e. $(N-B)/\epsilon$: knowing how σ_i depends on y_i in this case requires knowledge of individual error contributions, which are not always readily available. Of course, treating correctly the errors we *do* know about reduces the bias in the average. In addition, it makes sense to vary the assumptions used in the averaging and to quote the spread of results obtained under different (a priori reasonable) assumptions as a **systematic error on the average itself.**

2 Theoretical errors

(Thanks for Frank Porter for raising my sensitivity to this issue.)

We often do not have any good knowledge of the pdf representing theoretical uncertainties on a result. As a result, it is useful to keep track of the impact of theoretical uncertainties on averaged results separately from experimental uncertainties. This can be done as follows. Assume we have a vector of N measurements $\mathbf{y}$ for which we're trying to determine the average value $\hat{t}$. The uncertainties associated with the measurements are described in a covariance matrix $\mathbf{V}$. Let's assume the covariance is the sum of 3 separate pieces: $\mathbf{V_{stat}}$ (statistics), $\mathbf{V_{sys}}$ (experimental systematics) and $\mathbf{V_{th}}$ (theory). The χ^2 can be written

$$\chi^2(t) = \mathbf{\Delta}^{\mathrm{T}}\,\mathbf{V}^{-1}\mathbf{\Delta} = \mathbf{\Delta}^{\mathrm{T}}\,\left(\mathbf{V_{stat}} + \mathbf{V_{sys}} + \mathbf{V_{th}}\right)^{-1}\mathbf{\Delta} \tag{3}$$

where $\mathbf{\Delta} = \mathbf{y} - \mathbf{t}$ ($\mathbf{t}$ is a vector with each element equal to t) and $\mathbf{V}^{-1}$ is just the weight matrix $\mathbf{W}$. The minimization of the χ^2 is as described previously. The solution for $\hat{t}$ and $\sigma^2_{\hat{t}}$ can be written as

$$\hat{t} = \frac{\sum_i (\mathbf{W}\,\mathbf{y})_i}{\sum_{i,j} W_{ij}} \tag{4}$$

$$\sigma^2_{\hat{t}} = \frac{\sum_{i,j} (\mathbf{W}\,\mathbf{V}\,\mathbf{W})_{ij}}{\left(\sum_{i,j} W_{ij}\right)^2} \tag{5}$$

$$\sigma^2_{\hat{t}}(\mathrm{th}) = \frac{\sum_{i,j} (\mathbf{W}\,\mathbf{V_{th}}\,\mathbf{W})_{ij}}{\left(\sum_{i,j} W_{ij}\right)^2}. \tag{6}$$

Written in this form, one can readily determine the separate contribution of each source of uncertainty, and it's clear that $\sigma^2_{\hat{t}} = \sigma^2_{\hat{t}}(\text{stat}) + \sigma^2_{\hat{t}}(\text{sys}) + \sigma^2_{\hat{t}}(\text{th})$. Note that the weight with which measurement i enters the average (namely $\sum_j W_{ij} \Big/ \sum_{k,l} W_{kl}$, the coefficient of the measured value y_i in equation 4) can be negative in the presence of correlations.

3 Treating correlations

If the exact value of a correlation coefficient is known (as it often is in the case of zero correlation) it should clearly be used as input in the construction of the covariance matrix used in averaging measurements. What is often less clear, however, is the extent of correlations amongst the systematic uncertainties. Consider as an example a set of measurements that use the same theory to calculate an acceptance correction (e.g. in converting a partial rate to a total rate). If the actual acceptance of each measurement is different, then they are not 100% correlated, since there is room for theoretical uncertainty on the difference between the acceptances. Again, one could try to evaluate the actual correlation by varying parameters of the theory, etc. In practice, this is at best labor intensive and at worst impossible (if enough details of the acceptance calculation used were not recorded). In these cases we can proceed with a simple ansatz: choose the correlation coefficient for each pair of measurements that maximizes the uncertainty one would obtain considering just those two measurements and the source of uncertainty under study. To be concrete, consider measurements x and y with uncertainties σ_x and σ_y and unknown correlation ρ. The uncertainty on the average of x and y is

$$\sigma^2_{\hat{t}} = \frac{\sigma_x^2\, \sigma_y^2\, (1-\rho^2)}{\sigma_x^2 + \sigma_y^2 - 2\,\rho\,\sigma_x\,\sigma_y} \tag{7}$$

Maximizing this w.r.t. ρ gives

$$\begin{aligned} \rho &= \frac{\sigma_x^2 + \sigma_y^2}{2\,\sigma_x\,\sigma_y} \left(1 - \sqrt{1 - \left(\frac{2\,\sigma_x\,\sigma_y}{\sigma_x^2 + \sigma_y^2} \right)^2} \right) \\ &= \min\left(\frac{\sigma_x}{\sigma_y}, \frac{\sigma_y}{\sigma_x} \right) \end{aligned} \tag{8}$$

This guarantees that the error on the combined result due to this particular source of uncertainty will never be smaller than the error on the best measurement entering the average.

4 Implementation in COMBOS

The Heavy Flavor Averaging Group (HFAG) uses a program called COMBOS, which was developed by Olivier Schneider and Henry Seywerd in the LEP heavy flavor steering group. COMBOS has many useful features, in particular a flexible mechanism for the specification of measurements, common systematic uncertainties and adjustments of prior measurements in light of modified values of parameters on which they rely. Until now, however, it did not include a combination routine that allows the user to specify the dependence of a quoted uncertainty on the measured value, i.e. that deals with the problem outlined in Section 1. A new combination routine, `CHI2_SYM_RVK`, has been added for this purpose. It solves the χ^2 problem where σ_i depends on the measured value y_i by iteration. The dependence of σ_i on y_i can be specified as $\hat{\sigma}_i = \sigma_i\,(\hat{t}/y_i)^\eta$, where $\eta = 0$, 1/2 or 1, separately for each error source. The mechanism for doing this is appending nothing (0), a "\$" (1/2) or a "&" (1) to the name of each error, e.g.

```
stat$  bkg    eff&   th+&   th-&
 0.35  -0.25  -0.45  0.02  0.01
```

These can also be used in conjunction with the %, which indicates that the error is specified as a percentage of the central value.[1] The input line below is equivalent to the one above for a measurement whose central value is 10.0:

```
stat%$  bkg%   eff%&  th+%&  th-%&
 3.5    -2.5   -4.5   0.2    0.1
```

This limited range of possibilities was chosen to cover the most pressing use cases and for ease of implementation within the existing COMBOS program. An improved averaging tool would give the user more control over this. In principle there are error sources for which the quoted error depends on some inverse power of the measured value, but in the obvious case (the error due to a background subtraction) the dependence is usually weak enough to ignore.

In the case of asymmetric errors, the error used in building the covariance is

$$\sigma = \sigma_- + \left(1 + \tanh\left(2\frac{\hat{t} - y}{\sigma_+ + \sigma_-}\right)\right)\frac{\sigma_+ - \sigma_-}{2} \tag{9}$$

[1] A bug which in the past forced either all or none of the errors to be specified as a percentage has been fixed.

where y is the measured value and $\hat{t}$ is the value of the average based on previous iteration.[2] The output of CHI2_SYM_RVK is a value and covariance matrix, i.e. output errors are symmetric. Nevertheless, this procedure allows the right "sign" of each input error to be used in calculating the average and its uncertainty.

The correlation coefficients used for correlated systematic errors are determined as in Section 3.

In addition, the CHI2_SYM_RVK output includes a breakdown of the errors into individual sources as discussed in Section 2.

By default the & and $ in the input file determine how each individual error contribution is treated.[3] When testing the sensitivity of an average to the assumptions on how σ_i depends on y_i, it is convenient to override these with a global switch. For this purpose four new SWITCH values have also been added:

SYSREL - treat ALL systematic errors for a measurement as proportional (equivalent to appending & to the name of each systematic error)

SYSABS - treat ALL systematic errors for a measurement as absolute (equivalent to removing & from the name of each systematic error)

STATSQRT - treat the statistical error as proportional to $\sqrt{\text{value}}$ (equivalent to appending $ to the STAT error for each measurement)

STATABS - treat the statistical error as absolute (equivalent to removing $ from the STAT error for each measurement)

5 Recommendations

Averaging the results of a set of measurements requires some care. In particular, one should (beyond what has been standard practice in past HFAG semileptonic averages)

- account for the dependence of the quoted error on the measured value to avoid bias and provide reliable confidence levels;
- provide a breakdown of the errors on the average into components, as would be done on an individual measurement: at a minimum statistical, experimental systematic, and theory.

[2] The initial iteration is done using symmetric errors.

[3] COMBOS "lumps" the uncorrelated systematic uncertainties for each measurement; the dependence of the largest component of the error on $\hat{t}$ is assigned to the entire lump.

- include a systematic uncertainty on the average to account for unknowns in the averaging procedure (e.g. uncertainties in the degree to which the quoted errors on a set of measurements are correlated, in the dependence of the quoted errors on the measured values, ...) in cases where it is not negligeable;

While these recommendations don't address all issues that arise in averaging results (e.g. whether to scale the errors in cases where the confindence level is very low, or how to treat non-Gaussian errors), their use would be an improvement over much of what is current practice. Since most quoted errors depend on the measured value with a power between 0 and 1, a reasonable prescription would be to run COMBOS twice, once with the STATSQRT and SYSREL switches on, and once with the STATABS and SYSABS switches on, and quote the mean $(t_1 + t_2)/2$ of these trials as the average and $|t_1 - t_2|/\sqrt{12}$ as the uncertainty due to the assumptions in the averaging. Of course, such a prescription should be adjusted to what is or is not known in each particular case.

6 Example: $\mathcal{B}(B \to \pi \ell \nu)$ average for BaBar

As an example, consider the average of the following measurements:[4] This is meant purely as an illustration; these are not official HFAG numbers.

```
Corrected measurements:
BABAR NU_RECO PREL  1.3763+/-  0.1000+/-  0.1763 Tot Err:  0.2027
BABAR HAD_TAG PREL  0.8900+/-  0.3400+/-  0.1198 Tot Err:  0.3605
BABAR HAD_TAG0 PRE  1.6800+/-  0.5200+/-  0.2423 Tot Err:  0.5737
BABAR SL_TAG PRELI  1.0193+/-  0.2500+/-  0.1262 Tot Err:  0.2800
BABAR SL_TAG0 PREL  3.3100+/-  0.6800+/-  0.4207 Tot Err:  0.7996
```

The correlated systematics in this case are small compared to the statistical and uncorrelated systematic uncertainties. The average using STATSQRT and SYSREL is

```
CHI2_SYM_RVK    : BR=    1.462  +-     0.1729          CL=0.0124
```

For what it's worth, the breakdown of the errors in this case is

```
STAT    USYS    BBREID  BBRMUID BXLNU   BUSF    BBRBCNT FB0     BBRTRK
0.1247  0.1061  0.0163  0.0210  0.0423  0.0182  0.0134  0.0077  0.00496
```

[4] This is copied from the output of COMBOS; for those who care, a bug in the DUMP_MASTER routine was fixed, so the "Tot Err" listed there is now correct.

where USYS means "uncorrelated systematic errors" and the remaining contributions are correlated amongst at least 2 measurements. As indicated above, the correlated systematics are small in this case. With STATABS and SYSABS the average comes out as

```
CHI2_SYM_RVK     : BR=     1.285  +-     0.1452            CL=0.0553
```

From these I would quote $\mathcal{B}(B \to \pi\ell\nu) = 1.37 \pm 0.16 \pm 0.05$, where the last error comes from the averaging itself.

7 Appendix A

Consider a set of measurements where the dominant uncertainty is on a multiplicative factor (e.g. an efficiency), so that the uncertainty is directly proportional to the value measured. In this case it's easy to verify that Eq. 2 will produce an average that is biased towards zero w.r.t. the true value. In this case it's also obvious that we should be weighting the measurements by their fractional errors, i.e. $1/(\sigma_i/y_i)^2$, and not simply by $1/\sigma_i^2$. If we apply the prescription given above, i.e. take $\hat{\sigma}_i = \sigma_i\,\hat{t}/y_i$ as our estimator for $\bar{\sigma}_i$ (where $\hat{t}$ is the value we're trying to determine), we solve

$$\chi^2(t) = \sum_i^N \frac{(y_i - t)^2}{(\sigma_i\,\hat{t}/y_i)^2} \tag{10}$$

$$= (1/\hat{t}^2)\sum_i^N \frac{(y_i - t)^2}{(\sigma_i/y_i)^2}\ , \tag{11}$$

and find the value of t that minimizes the χ^2:

$$\hat{t} = \frac{\sum_i^N y_i^3/\sigma_i^2}{\sum_i^N y_i^2/\sigma_i^2} \tag{12}$$

$$\sigma_{\hat{t}}^2 = \frac{\hat{t}^2}{\sum_i^N y_i^2/\sigma_i^2}\ . \tag{13}$$

It's easy to verify that this gives unbiased results for the case where $\sigma_i \propto y_i$.

Consider the following simple example. Suppose you are measuring a yield per luminosity of some process, and that you split your data up into two equal-luminosity samples, where you measure N_1 and N_2 events. The average yield per luminosity in this case is trivially obtained by summing the yields and dividing by 2: $\langle N \rangle = (N_1 + N_2)/2$. The appropriate weight factors in this case are given by the equality of the luminosity (i.e. are equal) and not by the observed numbers of events. Neither of the weighted averages given above produces this (correct) answer. To be concrete, let $N_1 = 36$ and $N_2 = 64$. The correct average is 50 ± 5, with $P(\chi^2) = 0.0035$, while Eq. 2 gives 46.1 ± 4.8 with $P(\chi^2) = 0.078$ and Eq. 13 gives 53.9 ± 5.4 with $P(\chi^2) = 0.0026$. Using the prescription advocated above in this case leads to

$$\chi^2(t) = \sum_i^N \frac{(y_i - t)^2}{\sigma_i^2\, \hat{t}/y_i} \tag{14}$$

$$= (1/\hat{t}) \sum_i^N \frac{(y_i - t)^2}{\sigma_i^2/y_i} \ , \tag{15}$$

and

$$\hat{t} = \frac{\sum_i^N y_i^2/\sigma_i^2}{\sum_i^N y_i/\sigma_i^2} \tag{16}$$

$$\sigma_{\hat{t}}^2 = \frac{\hat{t}}{\sum_i^N y_i/\sigma_i^2} \ , \tag{17}$$

which gives the correct results. Note that not only are the other estimators biased, they can give **very** different χ^2 probabilities in cases where the measurements do not agree well.

I've done numerical tests for cases where the dependence of σ_i on y_i is more complicated. In general this leads to a $d\chi^2/dt = 0$ equation that is non-linear in $\hat{t}$; however it is readily solved by iteration. The prescription advocated here always gave unbiased results, at least to the accuracy (0.001σ) tested (see table 1).

Table 1: Results of toy MC for 3 estimators of mean and sigma. The values with overbars (e.g. $\bar{\sigma}$) refer to true values. The notation σ_x is used for the uncertainty on the mean value x. The uncertainties on the measurements were taken as $\sigma_i = \sqrt{0.5 * [2^2 + (0.2 * y_i)^2]}$.

	Eq. 2	Eq. 13	Using $\sigma_i(\hat{t})$
$\bar{t}$	10	10	10
$\hat{t}$	9.812	10.219	10.012
$(\hat{t} - \bar{t})/\sigma_{\hat{t}}$	-13.3	15.5	0.85
$\bar{\sigma}_{\bar{t}}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$
$\hat{\sigma}$	1.411	1.457	1.414
$(\hat{\sigma} - \bar{\sigma})/\sigma_{\hat{\sigma}}$	-0.25	4.25	0.00