Notes on COMBOS

Let $M(x_1)$ and $M(x_2)$ be measurements of 2 correlated quantities $x_1 \pm \sigma_1$ and $x_2 \pm \sigma_2$, respectively, with correlation coefficient $\rho \equiv Cov(x_1, x_2)/(\sigma_1\sigma_2)$. The Error Matrix is:

$$E = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{1}$$

$$\Rightarrow E^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \tag{2}$$

$$\text{With } w_i = \frac{\sum_j (E^{-1})_{ij}}{\sum_j \sum_k (E^{-1})_{jk}}, \sum_i w_i = 1 \Rightarrow \langle x \rangle = \sum_i w_i x_i, \tag{3}$$

$$\sigma = \sqrt{\sum_i \sum_j w_i E_{ij} w_j} = \sqrt{\frac{1}{\sum_i \sum_j (E^{-1})_{ij}}} = \sqrt{\frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \tag{4}$$

$$w_1 = \frac{(E^{-1})_{11} + (E^{-1})_{12}}{(E^{-1})_{11} + (E^{-1})_{12} + (E^{-1})_{21} + (E^{-1})_{22}} = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{5}$$

$$w_2 = \frac{(E^{-1})_{21} + (E^{-1})_{22}}{(E^{-1})_{11} + (E^{-1})_{12} + (E^{-1})_{21} + (E^{-1})_{22}} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{6}$$

A numerical example with un-correlated statistical errors and correlated systematic errors:

$$M(x_1) = x_1 \pm \sigma_{1u}(un-correlated) \pm \sigma_{1c}(correlated) = 1.2 \pm 0.1(stat) \pm 0.2(syst)$$
$$M(x_2) = x_2 \pm \sigma_{2u}(un-correlated) \pm \sigma_{2c}(correlated) = 0.8 \pm 0.1(stat) \pm 0.5(syst)$$

$$E = \begin{pmatrix} \sigma_{1u}^2 & 0 \\ 0 & \sigma_{2u}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{1c}^2 & \sigma_{1c}\sigma_{2c} \\ \sigma_{1c}\sigma_{2c} & \sigma_{2c}^2 \end{pmatrix} \tag{7}$$

$$\text{Following Eqn(1), we have } \sigma_1^2 = \sigma_{1u}^2 + \sigma_{1c}^2, \sigma_2^2 = \sigma_{2u}^2 + \sigma_{2c}^2, \rho\sigma_1\sigma_2 = \sigma_{1c}\sigma_{2c} \tag{8}$$

Here $\sigma_1 = \sqrt{(0.1)^2 + (0.2)^2}, \sigma_2 = \sqrt{(0.1)^2 + (0.5)^2}, \rho = \dfrac{(0.2)(0.5)}{\sqrt{(0.1)^2 + (0.2)^2}\sqrt{(0.1)^2 + (0.5)^2}} = 0.877058.$

$$\text{Thus, } w_1 = 1.45455, w_2 = 1 - w_1 = -0.45455 \Rightarrow \langle x \rangle = 1.38182, \sigma = 0.165145.$$

This is the derivation of average of Eqn (21) and (22) on page 33 of `http://www.slac.stanford.edu/xorg/hfag/docs/combos.pdf`: "one gets 1.38 ± 0.17".

If, instead we use $\rho = -0.877058$, we get $w_1 = 0.705882, w_2 = 0.294118 \Rightarrow \langle x \rangle = 1.08235 \pm 0.0766965$.

Example data-cards are in `swagato_example/example0` directory. To run, execute "../../combos average.input | tee average.output". The correlation between "experiment1" and "experiment2" is introduced via a change of "observable1" due to change in "parameter1", as explained below.

Eqn (17) on page 32 of `http://www.slac.stanford.edu/xorg/hfag/docs/combos.pdf` describes this problem as minimization of

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i + \sum_{\alpha=1}^{m} \Delta_{\alpha i} Y_\alpha - X}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{m} Y_\alpha^2, \text{ where } Y_\alpha = \frac{P_\alpha - \bar{P}_\alpha}{\delta_\alpha}.$$

Here, X_i are n measurements of X, σ_i are the un-correlated systematic and statistical errors on X_i. For m sources of correlation, $\Delta_{\alpha i}$ are the correlated systematics from parameter P_α and measurement i, determined as the change in the measured quantity when the parameter value is changed from $\bar{P}_\alpha$ to $\bar{P}_\alpha + \delta_\alpha$.

Note the sign of the Δ-term in Eqn (17). Due to this convention, "experiment1.input" and "experiment2.input" require the lines: "DATA parameter1 -0.2" and "DATA parameter1 -0.5". If we change it to "DATA parameter1 +0.2" only in "experiment1.input", we simulate the case of negative correlation, eg. $\rho = -0.877058 \Rightarrow \langle x \rangle = 1.08235 \pm 0.07670$ with COMBOS.

Note that in the above example, the "parameter1" has not been explicitly defined. The Δ_α specifying dependence of "observable1" and "observable2" on "parameter1" is the change when "parameter1" increases by a common value δ_α, which has also not been explicitly defined.

Next, let us explore the case when the parameters have been assumed to have different values in the two experiments. Example data-cards are in `swagato_example/example1` directory.

First, consider the case when the parameter is known without any uncertainty. In the files: "experiment1.param", "experiment2.param" and "common.param", this case is simulated using the line "parameter1 1.0 0.0 0.0". This gives $\langle x \rangle = 1.38182 \pm 0.16514$.

Same result is also obtained by using the line "parameter1 1.0 +1.0 -1.0" in "experiment1.param", "experiment2.param" and "common.param", which is probably the default usage.

```
$> grep param *.input *.param
average.input:INCLUDE     common.param
experiment1.input:INCLUDE     experiment1.param
experiment1.input:DATA        parameter1 -0.2
experiment2.input:INCLUDE     experiment2.param
experiment2.input:DATA        parameter1 -0.5
common.param:      parameter1   1.0          +1.0        -1.0
experiment1.param:      parameter1   1.0          +1.0        -1.0
experiment2.param:      parameter1   1.0          +1.0        -1.0
```

Same result is also obtained by using this combination:

```
$> grep param *.input *.param
average.input:INCLUDE     common.param
experiment1.input:INCLUDE     experiment1.param
experiment1.input:DATA        parameter1 -0.4
experiment2.input:INCLUDE     experiment2.param
experiment2.input:DATA        parameter1 -1.0
common.param:      parameter1   1.0          +1.0        -1.0
experiment1.param:      parameter1   1.0          +2.0        -2.0
experiment2.param:      parameter1   1.0          +2.0        -2.0
```

In the above example, both experiments use a value of the common systematic source "parameter1" with a factor of two worse error than what has been used in the averaging procedure. Note the corresponding changes of Δ_α to compensate for the change of "parameter1" in "experiment1.param" and "experiment2.param" in this illustrative example.

Now, if the common "parameter1" is known with factor of two better precision than what has been used by the two experiments through for example an "a-posterio" measurement of "parameter1", this can be used to reduce the error on the averaged value of "observable1". As an illustration, consider the following data-cards:

```
$> grep param *.input *.param
average.input:INCLUDE     common.param
experiment1.input:INCLUDE     experiment1.param
experiment1.input:DATA        parameter1 -0.2
experiment2.input:INCLUDE     experiment2.param
experiment2.input:DATA        parameter1 -0.5
common.param:      parameter1   1.0          +0.5        -0.5
experiment1.param:      parameter1   1.0          +1.0        -1.0
experiment2.param:      parameter1   1.0          +1.0        -1.0
```

This gives $\langle x \rangle = 1.24706 \pm 0.13933$, corresponding to $\rho = \frac{(0.1)(0.25)}{\sqrt{(0.1)^2+(0.1)^2}\sqrt{(0.1)^2+(0.25)^2}} = 0.656532$.

The same conclusion of reduction of averaged error in the case of negative correlation, eg. $\rho = -0.656532$ corresponding to $\langle x \rangle = 1.07368 \pm 0.07609$, is illustrated below:

```
$> grep param *.input *.param
average.input:INCLUDE     common.param
experiment1.input:INCLUDE     experiment1.param
experiment1.input:DATA        parameter1 +0.2
experiment2.input:INCLUDE     experiment2.param
experiment2.input:DATA        parameter1 -0.5
common.param:      parameter1   1.0          +0.5        -0.5
experiment1.param:      parameter1   1.0          +1.0        -1.0
experiment2.param:      parameter1   1.0          +1.0        -1.0
```

Let us re-visit the case with $\rho = \frac{(0.1)(0.25)}{\sqrt{(0.1)^2+(0.1)^2}\sqrt{(0.1)^2+(0.25)^2}} = 0.656532$ from above example in `swagato_example/example1a` directory, but this time changing the Δ_α instead of scaling the error of parameter1, as illustrated below:

```
$> grep param *.input
average.input:            parameter1  1.0          +1.0        -1.0
experiment1.input:            parameter1  1.0          +1.0        -1.0
experiment1.input:DATA        parameter1 0.1
experiment2.input:            parameter1  1.0          +1.0        -1.0
experiment2.input:DATA        parameter1 0.25
```

The log file (produced with `CALL DUMP_MASTER_INC`) shows that $\sigma_1 = \sqrt{(0.1)^2 + (0.1)^2}$ and $\sigma_2 = \sqrt{(0.1)^2 + (0.25)^2}$ have also changed along with the new value of $\rho = 0.656532$, which gives $\langle x \rangle = 1.24706 \pm 0.13933$. [This was also true in the previous case, when parameter1 was re-scaled via the "common.param" file.]

The desired behaviour of *only* changing the ρ without changing σ_1 and σ_2 is illustrated in `swagato_example/example1b` directory. This is acheived by introducing 2 un-correlated parameters in "experiment1.input" and "experiment2.input".

In this example, the systematic error is broken into correlated and un-correlated components. The correlated components each have half the original value as in `swagato_example/example1a`, but the un-correlated component is correctly evaluated such that $\rho = \frac{(0.1)(0.25)}{\sqrt{(0.1)^2+(0.2)^2}\sqrt{(0.1)^2+(0.5)^2}} = 0.219264504826757 \Rightarrow \langle x \rangle = 1.16154 \pm 0.21817$. This result from COMBOS can be cross-checked using a simple script:

```
$> ../Common/average.perl  1.2 .05 .8 .26 .025
x1 = 1.2 +- 0.223606797749979 x2 = .8 +- 0.509901951359279 rho = 0.219264504826757
w1 = 0.903846153846154 w2 = 0.0961538461538461 w1+w2 = 1
<x> = 1.16153846153846 +- 0.218165427706027
```

In `swagato_example/example1c` directory, we investigate the effect of scaling of error on "parameter1" in "average.input" w.r.t `swagato_example/example1b`.

```
$> grep param *.input
average.input:            parameter1  1.0          +0.5        -0.5
experiment1.input:            parameter1  1.0          +1.0        -1.0
experiment1.input:DATA        parameter1 0.1
experiment2.input:            parameter1  1.0          +1.0        -1.0
experiment2.input:DATA        parameter1 0.25
```

This is probably the default usage, where scaling the error on "parameter1" to a common value reduces the error on "observable1", eg. $\sigma_1 = \sqrt{.1^2 + .2^2 - .1^2 + 0.05^2}$ and $\sigma_2 = \sqrt{.1^2 + .5^2 - .25^2 + .125^2}$ have changed $\Rightarrow \rho = 0.0656559912465273, \langle x \rangle = 1.14033414026455 \pm 0.192635502522062$.

`swagato_example/example1c` is reproduced in `swagato_example/example1d`, by changing the Δ_α but without scaling the error on "parameter1" in "average.input". Note: "Tot Err" in output file is different from input file.

Appendix: Recipe for calculating Correlation between BR

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y) \text{ with } a = b = 1, c = 0 \Rightarrow$$

$$\rho(BR_1, BR_2) = \frac{Cov(BR_1, BR_2)}{\sqrt{Var(BR_1)Var(BR_2)}} = \frac{Var(BR_1 + BR_2) - Var(BR_1) - Var(BR_2)}{2\sqrt{Var(BR_1)Var(BR_2)}} \quad (9)$$

Simplest recipe is to adopt a Toy Monte Carlo approach: firstly, parameterize all statistical and systematic errors on the branching ratio as variations of each component used to evaluate the branching ratio (for example, using Eqn (13, 14) as given below); for each of N trials, vary each of these components within their error and evaluate BR_1, BR_2, $BR_1 + BR_2$ and their squares. Then for this ensemble of N trials, calculate $Var(BR_1)$, $Var(BR_2)$ and $Var(BR_1 + BR_2)$ using $Var(X) = \langle X^2 \rangle - \langle X \rangle^2$. Finally, estimate $\rho(BR_1, BR_2)$ using Eqn (11).

To validate the Toy Monte Carlo, cross-check the result with a simple derivation of the result for statistical correlation between the channels, assuming Poisson errors on the measurements.

Inverting the efficiency matrix that gives

$$\begin{pmatrix} Data_1 - Bkg_1 \\ Data_2 - Bkg_2 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} \begin{pmatrix} Sig_1 \\ Sig_2 \end{pmatrix} \quad (10)$$

$$BR_1 = \frac{1}{2\mathcal{L}\sigma} Sig_1 = \frac{1}{2\mathcal{L}\sigma}(\epsilon_{11}^{-1}(Data_1 - Bkg_1) + \epsilon_{12}^{-1}(Data_2 - Bkg_2)) \quad (11)$$

$$BR_2 = \frac{1}{2\mathcal{L}\sigma} Sig_2 = \frac{1}{2\mathcal{L}\sigma}(\epsilon_{21}^{-1}(Data_1 - Bkg_1) + \epsilon_{22}^{-1}(Data_2 - Bkg_2)) \quad (12)$$

$$BR_1 + BR_2 = \frac{1}{2\mathcal{L}\sigma}((\epsilon_{11}^{-1} + \epsilon_{21}^{-1})(Data_1 - Bkg_1) + (\epsilon_{12}^{-1} + \epsilon_{22}^{-1})(Data_2 - Bkg_2)) \quad (13)$$

Noting $Var(Data_i) = \sigma^2(Data_i) = Data_i$, and $Cov(Data_1, Data_2) = 0$, we get:

$$Var(BR_1) = (\frac{\epsilon_{11}^{-1}}{2\mathcal{L}\sigma})^2 Data_1 + (\frac{\epsilon_{12}^{-1}}{2\mathcal{L}\sigma})^2 Data_2 \quad (14)$$

$$Var(BR_2) = (\frac{\epsilon_{21}^{-1}}{2\mathcal{L}\sigma})^2 Data_1 + (\frac{\epsilon_{22}^{-1}}{2\mathcal{L}\sigma})^2 Data_2 \quad (15)$$

$$Var(BR_1 + BR_2) = (\frac{\epsilon_{11}^{-1} + \epsilon_{21}^{-1}}{2\mathcal{L}\sigma})^2 Data_1 + (\frac{\epsilon_{12}^{-1} + \epsilon_{22}^{-1}}{2\mathcal{L}\sigma})^2 Data_2 \quad (16)$$

$$\rho(BR_1, BR_2) = \frac{\epsilon_{11}^{-1}\epsilon_{21}^{-1} Data_1 + \epsilon_{12}^{-1}\epsilon_{22}^{-1} Data_2}{\sqrt{(\epsilon_{11}^{-1})^2 Data_1 + (\epsilon_{12}^{-1})^2 Data_2}\sqrt{(\epsilon_{21}^{-1})^2 Data_1 + (\epsilon_{22}^{-1})^2 Data_2}} \quad (17)$$

This is the statistical correlation between two measurements of branching ratios. Clearly, if more there are more than two measurements or if we are to evaluate the correlation including systematic errors on the branching ratios, the formula becomes complicated.