

CERN / PPE 94-185  
November 21, 1994

# Averaging Correlated Data

Michael Schmelling / CERN

## Abstract

The problem of averaging strongly correlated data is addressed for the case that the exact correlation pattern is unknown. A procedure is proposed to estimate the effective size of the correlations from the data themselves and to take them properly into account when forming the average. The properties of the procedure are illustrated by using it for averaging measurements of the strong coupling constant and QCD colour-factor ratios.

Submitted to Physica Scripta

# 1 Introduction

When trying to average experimental results from different experiments, one is sometimes faced with the problem that the individual results are known to be correlated, with correlations which are exceedingly difficult to quantify. Typical examples are the measurements of the strong coupling constant or the colour-factor ratios of QCD, where the dominant uncertainties are theoretical errors. This article proposes a scheme for averaging such data and illustrates its properties by applying it to these examples.

The suggested procedure is based on the standard weighted average ignoring all correlations between the measurements. In the limiting case of independent measurements with gaussian errors this average has the smallest possible error. If correlations are present this is no longer the case, but in return for a non-optimal error one has a robust average which is independent of the details of the correlation pattern. The size of the correlations is important for a correct determination of the error of the average. In the scheme proposed in this paper the impact of the correlations is estimated in an effective way from the data, based only on the measurements and their total errors. Correlation coefficients or a breakdown of the individual errors into globally correlated and uncorrelated terms are not needed.

## 2 The 1-dimensional Case

The input information for the averaging procedure are published values  $x$  and errors  $\delta x$ . Following [1] the errors will be treated as gaussian errors with the range  $x \pm \delta x$  specifying a 68.3% confidence interval. The measurements themselves are considered to be unbiased estimates for the true physical value.

If all errors are truly gaussian and if also the correlations between the individual measurements are described correctly by a covariance matrix  $C$ , the optimal procedure for finding the average  $a$  of a set of results  $\vec{x}$  is obtained by minimizing the  $\chi^2$ -function

$$\chi^2 = \sum_{i,j} (x_i - a)(C^{-1})_{ij}(x_j - a) \quad (1)$$

with respect to  $a$ . The minimum is obtained for

$$a = \left( \sum_{i,j} (C^{-1})_{ij} \right)^{-1} \left( \sum_{i,j} (C^{-1})_{ij} x_j \right) \quad \text{with} \quad \sigma^2(a) = \left( \sum_{i,j} (C^{-1})_{ij} \right)^{-1}. \quad (2)$$

This solution  $a$  is optimal in the sense, that it is the unbiased estimate for the true underlying value of all measurements  $\vec{x}$  which has the smallest possible error.

Although optimal in a certain sense, the above procedure requires a precise knowledge of the correlation terms in the covariance matrix  $C$ . If the correlations are large, the average tends to converge to the single most precise measurement entering the procedure. This becomes a severe problem when the errors of the individual measurements, as e.g. for the case of the strong coupling constant, to some extent have to be based on subjective judgement. It then is certainly unacceptable if a global average is pulled towards the single most optimistic result.

A way out is to abandon the requirement of the smallest possible error for the average. Since any linear combination of the individual measurements

$$a = \sum_i w_i x_i \quad \text{with} \quad \sum_i w_i = 1 \quad (3)$$

is an unbiased estimate of the true value, the question remains how to choose the weights  $w_i$ . The “optimal” way is according to eq.(2) – with the inherent dangers discussed above. Following [1] a more robust estimate is obtained by ignoring all correlation terms in eq.(2) and using the simple weighted average instead:

$$a = \left( \sum_i^n x_i / \sigma_i^2 \right) / \left( \sum_i^n 1 / \sigma_i^2 \right) \quad \text{and} \quad \sigma^2(a) = 1 / \left( \sum_i^n 1 / \sigma_i^2 \right). \quad (4)$$

If the data are uncorrelated this estimate is optimal, i.e. giving the smallest possible error. In case they are correlated, optimality is traded for stability.

The quality of the average  $a$  can be judged by means of a  $\chi^2$ -variable. For uncorrelated data one expects  $\chi^2$  to be equal to the number of degrees of freedom  $n_{df}$ ,

$$\langle \chi^2 \rangle = \left\langle \sum_i^n \frac{(x_i - a)^2}{\sigma_i^2} \right\rangle = n_{df} = n - 1. \quad (5)$$

If the actual  $\chi^2$  value differs significantly from this expectation either the error estimates  $\sigma_i$  are wrong or there are strong correlations between the measurements. While the average  $a$  in any case is a valid estimate of the true mean value, a  $\chi^2$  which deviates significantly from its expectation value implies that the error estimate  $\sigma^2(a)$  is not reliable. For  $\chi^2 > n - 1$  it is recommended [1] to conservatively assume that all errors are underestimated by a common factor and scale  $\sigma^2(a)$  by  $\chi^2/n_{df}$ . This is conservative since a large  $\chi^2$  could also be due to negative correlations between the measurements. Then already the original error estimate would be too large.

If the  $\chi^2$  is significantly smaller than its expectation value the measurement errors are either overestimated or positively correlated. In the first case also the error of the average is overestimated, in the latter one it is underestimated. The key point is, that if positive correlations are known to be present, then the  $\chi^2$  can be used to estimate the size of the correlations and reevaluate the error estimate for the average.

In order to achieve this, the full covariance matrix of the measurements needs to be specified. Not knowing any details about its detailed structure, it is assumed that different measurements  $i$  and  $j$  are correlated with a fixed fraction  $f$  of the maximum possible correlation  $C_{ij}^{max}$

$$C_{ij} = f \cdot C_{ij}^{max} \quad i \neq j, \quad \text{with} \quad C_{ij}^{max} = \sqrt{C_{ii}} \sqrt{C_{jj}}. \quad (6)$$

For  $f = 0$  the measurements are treated as uncorrelated, for  $f = 1$  as 100% correlated entities. The value  $f$  then is adjusted such that

$$\chi^2(f) = \sum_{i,j} (x_i - a)(x_j - a)(C^{-1})_{ij} = n - 1. \quad (7)$$

Having thus constructed an effective global covariance matrix  $C_{ij}$  the error estimate for the average  $a$  according to eq.(4) becomes

$$\sigma^2(a) = \sum_{i,j} w_i w_j C_{ij}, \quad (8)$$

i.e. it is increased due to the presence of positive correlations.

Some remarks are in order here. First, it should be noted that the inclusion of correlations as proposed here is not the only possible way. Alternatively one might consider making  $f$  a function of the difference  $x_i - x_j$ , reasoning that the correlation should be smaller when two measurements are further apart. However, the objective is to estimate in a consistent way the amount to which correlations affect the size of the error of a weighted average and this turns out to be rather insensitive to the details of the assumed correlation pattern. Therefore only the most simple ansatz is pursued here. It also has to be emphasized that the scaling procedure defined above should only be applied if correlations are known to exist. Due to statistical fluctuations  $\chi^2$  values smaller than their expectation values are not unlikely to exist, especially if the number of degrees of freedom is small. An unconditional application of the scaling procedure thus will bias the error estimate towards large values.

The properties of the method are most easily illustrated by the simple case of two measurements  $x_1$  and  $x_2$  with the same error  $\sigma$ . The average is  $a = (x_1 + x_2)/2$ . For a given correlation coefficient  $\rho$  the  $\chi^2$  for this average is  $\chi^2 = \chi_0^2/(1 - \rho)$  with  $\chi_0^2 = (x_1 - x_2)^2/(2\sigma^2)$ . If the two measurements are correlated one expects  $\chi_0^2 < 1$ , and the condition  $\chi^2 = 1$  yields  $\rho = 1 - \chi_0^2$  as an estimate for the size of the correlation coefficient. With this the variance of the average  $a$  becomes  $\sigma^2(a) = \sigma^2(1 + \rho)/2 = \sigma^2(2 - \chi_0^2)/2$ , i.e. it grows linearly with the correlation coefficient. In the extreme case that the two measurements are much closer than one would expect them to be for uncorrelated errors, averaging does not lead to any error reduction.

## Numerical Example

The numerical behaviour of the proposed procedure will be illustrated by using it to average measurements of the strong coupling constant  $\alpha_s$ , done at LEP/SLC with global event shape variables. As the basic concept is a scaling procedure for the error estimate based on the value of a  $\chi^2$  test variable, errors are implicitly assumed to be gaussian. The experimental errors were treated as follows: If more than one error is quoted, then all uncertainties are added in quadrature and the combination used in the average. Asymmetric errors are treated such, that the quadratic sum of all positive and all negative errors is formed independently. Of the final errors the bigger one is symmetrized. This is more conservative than the practice adopted in [1], but the existence of asymmetric errors means that the likelihood function definitely is non-gaussian – which suggests that some caution should be exercised.

The input data are collected in table 1. Section (a) contains a compilation of measurements based on single event shape variables, section (b) averages from several event shape variables as published by the different experiments. Details about the origin of the data can be found in [2]. All numbers are highly correlated, where the partial averages in (b) already take the correlation between their respective inputs into account. Both sections thus essentially contain the same amount of information, although the exact correlation pattern is different. However, the

averaging procedure for obtaining the correlated average of all data should be insensitive to this and give the same result when applied to either section of table 1.

Doing a simple weighted average of the numbers given in table 1(a) yields  $\alpha_s = 0.1219 \pm 0.0011$  with  $\chi^2/n_{df} = 37.4/71$ . The small  $\chi^2$  is a clear reflection of the correlations between the results. Using the scaling procedure to bring the  $\chi^2$  up to its expectation value increases the error for  $\alpha_s$  by more than a factor of five and one obtains  $\alpha_s = 0.1219 \pm 0.0059$ . The simple weighted mean of all single-experiment averages, table 1(b), yields  $\alpha_s = 0.1215 \pm 0.0022$  with  $\chi^2/n_{df} = 3.0/10$ . Accounting for correlations this becomes  $\alpha_s = 0.1215 \pm 0.0059$ , which is very close to the value obtained before from all single results and illustrates nicely the robustness of the procedure.

a) Individual measurements for $\alpha_s(M_Z)$					
0.119 $\pm$ 0.029,	0.186 $\pm$ 0.058,	0.112 $\pm$ 0.030,	0.136 $\pm$ 0.017,	0.142 $\pm$ 0.015,	0.121 $\pm$ 0.014
0.123 $\pm$ 0.015,	0.115 $\pm$ 0.012,	0.124 $\pm$ 0.012,	0.118 $\pm$ 0.011,	0.126 $\pm$ 0.005,	0.126 $\pm$ 0.007
0.124 $\pm$ 0.008,	0.126 $\pm$ 0.012,	0.122 $\pm$ 0.008,	0.123 $\pm$ 0.010,	0.124 $\pm$ 0.009,	0.121 $\pm$ 0.010
0.110 $\pm$ 0.010,	0.134 $\pm$ 0.009,	0.123 $\pm$ 0.012,	0.121 $\pm$ 0.025,	0.122 $\pm$ 0.011,	0.122 $\pm$ 0.009
0.119 $\pm$ 0.008,	0.116 $\pm$ 0.009,	0.108 $\pm$ 0.008,	0.114 $\pm$ 0.009,	0.123 $\pm$ 0.008,	0.123 $\pm$ 0.005
0.126 $\pm$ 0.008,	0.117 $\pm$ 0.006,	0.133 $\pm$ 0.006,	0.132 $\pm$ 0.005,	0.121 $\pm$ 0.012,	0.115 $\pm$ 0.009
0.115 $\pm$ 0.013,	0.121 $\pm$ 0.009,	0.119 $\pm$ 0.008,	0.120 $\pm$ 0.014,	0.110 $\pm$ 0.006,	0.132 $\pm$ 0.008
0.122 $\pm$ 0.007,	0.123 $\pm$ 0.009,	0.128 $\pm$ 0.016,	0.122 $\pm$ 0.047,	0.129 $\pm$ 0.017,	0.129 $\pm$ 0.010
0.120 $\pm$ 0.013,	0.128 $\pm$ 0.010,	0.119 $\pm$ 0.012,	0.120 $\pm$ 0.009,	0.125 $\pm$ 0.008,	0.122 $\pm$ 0.009
0.128 $\pm$ 0.016,	0.118 $\pm$ 0.009,	0.112 $\pm$ 0.010,	0.119 $\pm$ 0.008,	0.116 $\pm$ 0.007,	0.118 $\pm$ 0.008
0.124 $\pm$ 0.012,	0.123 $\pm$ 0.010,	0.125 $\pm$ 0.011,	0.128 $\pm$ 0.021,	0.118 $\pm$ 0.013,	0.116 $\pm$ 0.009
0.114 $\pm$ 0.008,	0.108 $\pm$ 0.009,	0.126 $\pm$ 0.007,	0.136 $\pm$ 0.015,	0.122 $\pm$ 0.007,	0.122 $\pm$ 0.012
b) Single experiment averages for $\alpha_s(M_Z)$					
0.117 $\pm$ 0.010,	0.113 $\pm$ 0.007,	0.118 $\pm$ 0.010,	0.122 $\pm$ 0.006,	0.123 $\pm$ 0.010,	0.118 $\pm$ 0.011
0.125 $\pm$ 0.005,	0.123 $\pm$ 0.006,	0.124 $\pm$ 0.009,	0.120 $\pm$ 0.006,	0.126 $\pm$ 0.007	

Table 1: Individual  $\alpha_s$  measurements based on single global event shape variables (a) and averages combining information from different event shape variables (b). All errors are the combined experimental and theoretical uncertainties.

### 3 The n-dimensional Case

The basic formalism introduced above for 1-dimensional data shall now be generalized for higher dimensional cases where a single measurement consists of a vector  $\vec{x}$  and an error matrix  $C$ . Again the error matrix will be treated as if it were a covariance matrix of a multi-variate gaussian probability density function around a central value  $\vec{x}$ .

As in the 1-dimensional case, the optimal procedure to average several measurements  $\vec{x}_i$  with gaussian covariance matrices  $C_{ii}$  is given by minimizing a  $\chi^2$ -function eq.(11). The formal treatment is completely analogous. The only technical difference is, that now the variances are

described by matrices and the measurements and averages by vectors. With this replacement all expressions derived for 1-dimensional data also hold in the general  $n$ -dimensional case. In particular the weighted average, assuming that the different measurements are uncorrelated, becomes

$$\vec{a} = \sum_j M_j \vec{x}_j \quad \text{with} \quad M_j = \left( \sum_i (C_{ii})^{-1} \right)^{-1} (C_{jj})^{-1}. \quad (9)$$

The sum of all weight matrices  $M_j$  is the unit matrix, i.e. if the individual measurements are unbiased estimates for the true value, so is  $\vec{a}$ .

As before, the  $n$ -dimensional weighted average eq.(9) defines the combined result of various measurements, independent of whether or not correlations between the input data are present. Correlations between the measurements again only affect the error estimate for  $\vec{a}$ . One obtains

$$C(\vec{a}) = \sum_{i,j} M_i C_{ij} (M_j)^T \quad (10)$$

with the matrices  $M$  as introduced in eq.(9). Note that each element  $C_{ij}$  now stands for an  $n \times n$ -covariance matrix. The  $\chi^2$  for the combined average  $\vec{a}$  is

$$\chi^2(\vec{a}) = \sum_{i,j} (\vec{x}_i - \vec{a})^T W_{ij} (\vec{x}_j - \vec{a}) \quad (11)$$

where  $W$  is the inverse of the global covariance matrix  $C$ , and  $W_{ij}$  the  $n \times n$  sub-matrix of  $W$  at the same position as  $C_{ij}$  in  $C$ . As before the  $\chi^2$  eq.(11) serves as an indicator. If the  $\chi^2$  is significantly larger than its expectation value the errors may be scaled up by a corresponding factor. For too small  $\chi^2$ -values and if correlations are known to be present, the effective size of the correlations can be estimated from the requirement that  $\chi^2$  be equal to its expectation value. Formally this can be done exactly as before, by taking the correlation matrix between two different measurements  $i$  and  $j$  to be a certain fraction  $f$  of the maximal correlation  $C_{ij}^{max}$ , i.e.  $C_{ij} = f C_{ij}^{max}$ .

It remains to generalize the notion ‘‘maximally correlated’’ to the  $n$ -dimensional case. If all single covariance matrices are diagonal, the problem reduces to a set of independent 1-dimensional measurements and the maximal correlation matrix between two measurements  $\vec{x}_i$  and  $\vec{x}_j$  is given by  $C_{ij}^{max} = C_{ii}^{1/2} C_{jj}^{1/2}$ . The square-roots are understood to be taken element by element. The same prescription works if the off-diagonal elements are the same. A difficulty only arises when the correlations terms in  $C_{ii}$  and  $C_{jj}$  are different. However, arguing that the result of averaging any two measurements should be independent under any linear transformation of the variables, one can always chose a basis such, that the off-diagonal elements of the two covariance matrices become identical. In this basis the  $C_{ij}^{max}$  can be defined as before and transforming back to the original variables yields the wanted matrix. The sought for transformation always exists and is given by the orthogonal matrix  $Q$  which diagonalizes the difference  $C_{ii} - C_{jj}$ . One thus obtains the following general expression

$$C_{ij}^{max} = Q^T \left[ \sqrt{Q C_{ii} Q^T} \sqrt{Q C_{jj} Q^T} \right] Q \quad \text{where} \quad Q \text{ diagonalizes } C_{ii} - C_{jj}. \quad (12)$$

## Numerical Example

A practical application for averaging 2-dimensional correlated data is the combination of measurements of the QCD colour-factor ratios as done by the LEP collaborations. The results are based on an analysis of kinematical correlations in 4-jet events, and again the highly correlated theoretical errors are the dominant uncertainties. The data taken from [2] are listed in table 2 together with the uncorrelated and the correlated averages. Also here the presence of correlations between the results manifest itself through the small  $\chi^2$  of the uncorrelated average. Taking the correlations into account the uncertainties of the correlated average are close to the errors of the most precise single measurement. The decrease in the size of the error-ellipse is only 2.3%.

$C_A/C_F$	$T_F/C_F$	correlation
$2.24 \pm 0.40$	$0.58 \pm 0.29$	0.043
$1.95 \pm 0.37$	$0.23 \pm 0.14$	0.
$2.32 \pm 0.25$	$0.27 \pm 0.15$	-0.242
$2.11 \pm 0.32$	$0.40 \pm 0.17$	-0.450
uncorrelated average ( $\chi^2/n_{df} = 2.20/6$ )		
$2.195 \pm 0.156$	$0.316 \pm 0.082$	-0.208
correlated average ( $\chi^2/n_{df} = 6.00/6$ )		
$2.195 \pm 0.264$	$0.316 \pm 0.138$	-0.220

Table 2: Individual results for colour-factor ratios from the analysis of 4-jet events, together with the uncorrelated and the correlated averages. The errors are the combined experimental and theoretical uncertainties.

## 4 Summary and Conclusions

A procedure for averaging correlated measurements is proposed for the case, that the correlations are not known quantitatively. The central value is taken to be the simple weighted average of the individual results, which gives the minimal error when the input data are uncorrelated, and is a stable estimator also in the presence of large correlations. In the latter case optimality is traded for robustness. In order to calculate an error estimate for the average, the effective size of the correlations has to be known. Here it is inferred from the data, using the requirement that the (correlated)  $\chi^2$  for the average is equal to its expectation value. The robustness of the method has been demonstrated in averaging actual measurements of QCD parameters.

## References

- [1] Particle Data Group: M. Aguilar-Benitez *et al.*, Phys. Rev. D 45, Part 2, June (1992).
- [2] M. Schmelling, “QCD Results from the Study of Hadronic  $Z$  Decays”, to be published in Physica Scripta.