Averaging Measurements with Hidden Correlations and Asymmetric Errors

Michael Schmelling / MPI for Nuclear Physics

Postfach 103980, D-69029 Heidelberg

Abstract

Properties of weighted averages are studied for the general case that the individual measurements are subject to hidden correlations and have asymmetric statistical as well as systematic errors. Explicit expressions are derived for an unbiased average with a well defined error estimate.

1 Introduction

Combining independent measurements of a physical quantity in order to summarize the information from different sources into one number is usually done by calculating a weighted average, where each measurement contributes with a weight which is inversely proportional to its variance.

In practical applications, however, one is often confronted with the problem that the measurement errors which are quoted define a 68% confidence level interval around a Maximum Likelihood estimate, whereas one would need an unbiased estimator with its standard deviation in order to form a weighted average. In addition, one usually has to deal with statistical and systematic uncertainties, and different results may be correlated, e.g. due to common uncertainties in some theoretical prediction, while a quantitative estimate for the size of the correlation does not exist.

This paper presents a systematic approach to form a meaningful weighted average under such conditions. After a short reminder of the properties of weighted averages, it first is shown how to deal with hidden correlations in the data. Then a detailed analysis of parameter estimates based on the Maximum Likelihood Method is performed, and the consequences when forming a weighted average are studied in detail. The results are illustrated by means of a numerical simulation. Finally, the issue of systematic errors and how to combine statistical and systematic errors is addressed.

2 Weighted Averages

Let $\{x_i\}$ be a set of n unbiased measurements of a parameter μ, with expectation values $\langle x_i \rangle$ and covariance matrix $C_{ij}(x)$ given by

$$\langle x_i \rangle = \mu \quad \text{and} \quad C_{ij}(x) = \langle (x_i - \mu)(x_j - \mu) \rangle = \langle x_i x_j \rangle - \mu^2. \tag{1}$$

From $\{x_i\}$ a new estimator $\bar{x}$ for μ can be constructed, which has a smaller variance than any of the x_i, by minimizing

$$\chi^2 = \sum_{i,j} (x_i - \bar{x})(x_j - \bar{x})C_{ij}^{-1}(x). \tag{2}$$

Note that the χ^2-function is constructed in such a way, that it is invariant with respect to arbitrary linear transformations of the variables with a regular transformation matrix. In case the individual x_i are independent, the covariance matrix becomes diagonal, $C_{ij}(x) = \sigma_i^2 \delta_{ij}$ and eq.(2) simplifies to

$$\chi^2 = \sum_i \frac{(x_i - \bar{x})^2}{\sigma_i^2}. \tag{3}$$

The value $\bar{x}$ which minimizes χ^2 and thus in the least squares sense is optimally consistent with all measurements x_i is found to be

$$\bar{x} = \sum_i w_i x_i \quad \text{with} \quad w_i = \frac{1/\sigma_i^2}{N} \quad \text{and} \quad N = \sum_k \frac{1}{\sigma_k^2}, \tag{4}$$

and the variance of $\bar{x}$ is obtained by standard error propagation as

$$C(\bar{x}) = \sigma^2(\bar{x}) = \sum_{ij} w_i w_j C_{ij} = \sum_i w_i^2 \sigma_i^2 = N^{-1}. \tag{5}$$

Another relevant quantity is the expectation value of χ^2 at the minimum

$$\left\langle \chi_{\min}^2 \right\rangle = \sum_i \left\langle \frac{(x_i - \bar{x})^2}{\sigma_i^2} \right\rangle = \sum_i \frac{\langle x_i^2 \rangle}{\sigma_i^2} - N \left\langle \bar{x}^2 \right\rangle = n - 1 \tag{6}$$

Note that although the value of $\langle \chi^2 \rangle$ has been derived for the case of a diagonal covariance matrix, the result is valid in general since one always can find a linear transformation of variables which diagonalizes C_{ij} while χ^2 by construction is invariant.

Minimizing $\sum_i w_i^2 \sigma_i^2$ with respect to w_i subject to the constraint $\sum_i w_i = 1$, one sees that the weights w_i used for the weighted average eq.(4) lead to the smallest possible variance for $\bar{x}$ if the individual measurements are uncorrelated.

In case that correlations are present, one can still calculate a weighted average according to eq.(4), which always is an unbiased estimator for μ, even if it no longer is the value which minimizes eq.(2). It also does not have the smallest possible variance, although the loss in precision usually is only marginal. In practical applications one thus might prefer the robust average eq.(4) to the formally optimal result from minimization of eq.(2). As this requires to calculate the inverse of C_{ij}, it can only be performed reliably if the off-diagonal elements of C_{ij} are known precisely. Otherwise numerical instabilities are likely to render the formally optimal result $\bar{x}$ meaningless. Note, however, that while it is perfectly meaningful to calculate a weighted average ignoring off-diagonal elements of the covariance matrix, the full matrix C_{ij} must be taken into account for the variance of the average.

For uncorrelated inputs the weighted average eq.(4) is the optimal way to combine measurements in order to minimize the variance of the result. It is independent of the actual shape of the probability density functions (PDFs) involved. All one needs to know are the values x_i and the variances σ_i^2 of the individual measurements. Note that the confidence levels associated to the variances may vary with the shape of the PDF. The weighted average, on the other

hand, has a PDF which is a convolution of the PDFs of the individual x_i, and according to the central limit theorem with an increasing number of contributions quickly converges towards a Gaussian, where the range $\bar{x} \pm \sigma(\bar{x})$ corresponds approximately to an 68% confidence level interval.

3   Hidden Correlations

In practical applications one sometimes is confronted with a situation where measurements have to be averaged which are suspected to be correlated, but where the correlations are very hard to quantify. While, as argued before, a weighted average can still be calculated, it is not obvious how to give a meaningful estimate for the error of the estimate. One possible approach to this problem has been discussed in [1]. The basic idea was to use the value of the χ^2 calculated according to eq.(2) as an indicator and adjust a global correlation coefficient between all measurements such that it becomes equal to its expectation value. Here a similar argument will be presented, which is more convenient as it leads to a much simpler scaling prescription.

A value χ^2 larger than its expectation value can result if anti-correlations are present, or if the measurement errors are underestimated. In case of the latter, N^{-1} from eq.(5) underestimates the variance of $\bar{x}$, in the former case it is overestimated. In order to be conservative, a common practice [2] is to assume that a large χ^2 is a consequence of too small measurement errors and apply a common scaling factor $\chi^2/(n-1)$ to all variances to obtain a χ^2 identical to its expectation value. Although one certainly can question this approach, it has to be pointed out, that unless there are severe discrepancies in the data, this procedure still gives an error estimate for $\bar{x}$ which is smaller than any of the individual errors σ_i of the data.

The other indication for a possible problem is a too small value for χ^2. This can either mean overestimated measurement errors or positive correlations between the data. While the former would imply that the error of $\bar{x}$ in reality is smaller than N^{-1}, the latter would require a larger error. Being conservative, again an estimate for the larger error shall be derived. As these considerations only affect error estimates, the calculations can be simplified without loss of generality by assuming $\langle x_i \rangle = \langle \bar{x} \rangle = 0$. This implies $\langle x_i^2 \rangle = \sigma_i^2$ and $\langle \bar{x}^2 \rangle = C(\bar{x})$, and the expectation value $\langle \chi^2_{\min} \rangle$ eq.(6) simplifies to

$$\left\langle \chi^2_{\min} \right\rangle = n - NC(\bar{x}). \tag{7}$$

If the individual measurements are uncorrelated, then one has $C(\bar{x}) = N^{-1}$, and as shown before $\langle \chi^2_{\min} \rangle = n - 1$. If there are reasons to suspect that the measurements x_i are not independent, then one can re-interpret eq.(7), assuming that the actual best fit χ^2-value is equal to its expectation value and derive a variance estimate

$$C(\bar{x}) = \frac{1}{N}(n - \chi^2). \tag{8}$$

This expression constitutes a simple scaling prescription for the case that the χ^2 of a weighted average is smaller than expected. Like the scaling procedure for the errors in case of too large χ^2 values, it defines an effective way to take imperfections or incomplete knowledge about the input data into account.

Like the weighted average, also the simple scaling eq.(8) of the uncorrelated variance estimate N^{-1} is readily generalized to m-dimensional vector-valued measurements $\vec{x}_i$. It will also be demonstrated explicitly that eq.(8) is equivalent to the assumption of a fixed common covariance between all measurements. In contrast to this, the more complicated procedure discussed in [1] was based on the assumption of a common correlation coefficient. In practice both scaling schemes give rather similar results. Numerically the method [1] yields slightly smaller errors, indicating that the assumption of a common correlation coefficient tends to assign a higher weight for more precise measurements than the assumption of a common covariance.

Ignoring correlations between measurements, the generalization of eq.(3) to vector valued measurements $\vec{x}_i$ with covariance matrices C_i is given by

$$\chi^2 = \sum_i \left(\vec{x}_i - \bar{x}\right)^T C_i^{-1} \left(\vec{x}_i - \bar{x}\right). \tag{9}$$

Note that in this expression the index i enumerates entire vectors rather than individual components. Those vectors $\vec{x}_i$ are assumed to be uncorrelated; correlations between the components of $\vec{x}_i$ are described by the covariance matrices C_i. The minimum of eq.(9) is obtained for

$$\bar{x} = N^{-1} \sum_i C_i^{-1} \vec{x}_i \quad \text{with} \quad N = \sum_i C_i^{-1}. \tag{10}$$

Note that now $\bar{x}$ is a vector of dimension m. Generalizing eq.(6) the value of χ^2 at the minimum then can be expressed as

$$\chi^2_{\min} = \mathrm{Tr}\left(\sum_i C_i^{-1} \vec{x}_i \vec{x}_i^T - N\bar{x}\bar{x}^T\right) \tag{11}$$

Without loss of generality one can again make the simplifying assumption $\langle \vec{x}_i \rangle = \langle \bar{x} \rangle = 0$. This leads to $\langle \vec{x}_i \vec{x}_i^T \rangle = C_i$ and $\langle \bar{x}\bar{x}^T \rangle = C(\bar{x})$, and the expectation value of $\chi^2_{\min}$ becomes

$$\left\langle \chi^2_{\min} \right\rangle = n \cdot m - \mathrm{Tr}\left(N \left\langle \bar{x}\bar{x}^T \right\rangle\right) = n \cdot m - \mathrm{Tr}\left(NC(\bar{x})\right). \tag{12}$$

At this point one could directly apply the same short argument as before and derive the generalization of eq.(8) for m-dimensional measurements. However, in order to illuminate a little bit the background of the scaling rule, here a different route shall be taken, making the explicit assumption that the correlation between the vectors $\vec{x}_i$ is such, that the covariance matrix C_i for each vector $\vec{x}_i$ is the sum of a common contribution C and a specific component S_i

$$C_i = C + S_i \quad \text{i.e.} \quad \left\langle \vec{x}_i \vec{x}_j^T \right\rangle = \begin{cases} C & \text{for } i \neq j \\ C_i & \text{for } i = j \end{cases} \tag{13}$$

One then obtains

$$C(\bar{x}) = \left\langle \bar{x}\bar{x}^T \right\rangle = N^{-1} \left[\mathbf{1} + NC - \left(\sum_i C_i^{-1} C C_i^{-1}\right) N^{-1}\right], \tag{14}$$

where $\mathbf{1}$ denotes the unit matrix, and

$$\left\langle \chi^2_{\min} \right\rangle = m \cdot (n-1) - \mathrm{Tr}\left[NC - \left(\sum_i C_i^{-1} C C_i^{-1}\right) N^{-1}\right]. \tag{15}$$

For $C = 0$ the covariance matrix of the average is given by N^{-1} and one has $\left\langle \chi^2_{\min} \right\rangle = m \cdot (n-1)$. A significant deviation of the $\chi^2_{\min}$ from this expectation value can be taken as an

indication for correlations C between the individual measurements. The simplest way to take these correlations into account for the error of the average, is by scaling the uncorrelated estimate N^{-1} by a factor $(1 + c)$. From eq.(14) then follows that C satisfies the relation

$$NC - \left(\sum_i C_i^{-1} C C_i^{-1} \right) N^{-1} = c \, \mathbf{1}. \tag{16}$$

For a given c this is a linear equation which can be solved numerically for C by iterating for example the fixed point condition

$$C_{n+1} = N^{-1} + N^{-1} \left(\sum_i C_i^{-1} C_n C_i^{-1} \right) N^{-1} \tag{17}$$

and setting $C = c \, C_\infty$. In practical applications this only has to be done if one wants to extract C in order to get a quantitative estimate for the size of the correlations between the measurements. Otherwise it is sufficient to fix the value of the parameter c by requiring the observed χ^2-value to be equal to the expectation value $\langle \chi^2_{\min} \rangle$. With eq.(16) and eq.(15) one obtains

$$\chi^2_{\min} = m(n-1) - m \, c \tag{18}$$

and a scaled variance estimate

$$C(\bar{x}) = N^{-1} \left(n - \frac{\chi^2}{m} \right). \tag{19}$$

For $m = 1$ the result eq.(8) is recovered.

To summarize, unless correlations between measurements are really well understood, it appears advisable to employ the weighted average ignoring all correlations as a robust procedure to combine several inputs into one number. Care, however, has to be exercised to give a realistic error estimate. In the absence of other information the χ^2 of the average is a useful indicator for potential problems. The conservative approach would be to interpret a χ^2 smaller than its expectation value as evidence for positive correlations between the measurements, and a χ^2 larger than its expectation value as evidence that the errors are underestimated. In both cases the simple estimate N^{-1} for the variance of the weighted average would have be to increased, which is conveniently done according to the following scheme

$$C(\bar{x}) = N^{-1} \begin{cases} \left(n - \dfrac{\chi^2}{m} \right) & \text{for} \quad \chi^2 \leq m(n-1) \\[2ex] \dfrac{\chi^2}{m(n-1)} & \text{for} \quad \chi^2 > m(n-1) \end{cases} \tag{20}$$

It has to be emphasized, that such a scaling should only be performed if one has reasons to believe that the variance estimate N^{-1} of the weighted average underestimates the true uncertainties, or if one wants to quote a conservative error. Otherwise, since the χ^2-distribution especially for a small number of degrees of freedom has a large relative width, one will systematically bias the error to large values.

4 Statistical Errors

In order to form a weighted average of independent data one needs unbiased measurements with known variances. On the other hand, in order to quantify the statistical precision of a

measurement, usually a 68% confidence level interval is quoted, which is constructed such that the true value of the parameter is inside this interval in 68% of all cases. For a Gaussian distribution the 68% confidence level interval is given by $\bar{x} \pm \sigma$, where $\bar{x}$ is the mean value and σ the rms-width of the distribution, i.e. the square of the error is the variance. In general one needs to know the shape of the PDF in order to extract the variance from the error interval.

If the shape of the PDF is not explictly specified, one has to resort to certain reasonable assumptions in order to proceed. For the following it will be assumed that in absence of other information the primary result of a measurement has a Gaussian PDF. Using the Maximum Likelihood Method this measurement then is interpreted in terms of a physical parameter, with a nominal result and an error range covering a 68% confidence level interval. Under these conditions a rigorous interpretation can be attached to asymmetric errors even if no further information about the PDF is given.

4.1   Interpretation of Asymmetric Errors

As explained above, a measurement x is assumed to scatter around a mean value $\langle x \rangle$ according to a Gaussian PDF with variance σ^2. The precision of the measurement σ is assumed to be known, e.g. from first principles or some calibration procedure. The mean value $\langle x \rangle$ is related to a physical parameter μ via a function $s(\mu)$, $\langle x \rangle = s(\mu)$. The likelihood function $p(x|\mu)$ to observe a value x given μ, then is

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - s(\mu))^2}{2\sigma^2}\right). \tag{21}$$

In the framework of the Maximum Likelihood Method the function eq.(21) is used to extract an estimate μ_{ml} for μ from a given measurement x as the parameter which maximizes the likelihood. The solution is evidently given by $\mu_{ml} = s^{-1}(x)$. Note that the likelihood function here is merely a tool to extract an estimate μ_{ml} for μ by looking for the value which maximizes $p(x|\mu)$, i.e. no attempt is made to interpret the likelihood function as a PDF for μ. Note that in a Bayesian approach using a flat prior distribution, one could actually consider eq.(21) as a probability density for μ and take e.g. the average $\langle \mu \rangle$ of $p(x|\mu)$ instead of μ_{ml} as an estimate for μ, although, as shown below, the Maximum Likelihood estimate μ_{ml} usually is the better choice.

In addition to the parameter estimate also an error interval has to be given. This is done by translating the error interval $[x_-, x_+]$ of the measurement to an error interval for the Maximum Likelihood estimate according to $[\mu_-, \mu_+] = [s^{-1}(x_-), s^{-1}(x_+)]$. By conservation of probability both intervals have the same probability content. If $p(x)$ denotes the PDF for the measurement x, then the PDF for the Maximum Likelihood estimate $q(\mu_{ml})$ is given by

$$q(\mu_{ml}) = \int_{-\infty}^{\infty} dx\, p(x)\, \delta(\mu_{ml} - s^{-1}(x)) \tag{22}$$

and one obtains for the probability content

$$\int_{\mu_-}^{\mu_+} d\mu_{ml}\, q(\mu_{ml}) = \int_{s^{-1}(x_-)}^{s^{-1}(x_+)} d\mu_{ml} \int_{-\infty}^{\infty} dx\, p(x)\, \delta(\mu_{ml} - s^{-1}(x)) = \int_{x_-}^{x_+} dx\, p(x). \tag{23}$$

Usually the error interval for the measurement is given by $[x_-, x_+] = [x - \sigma, x + \sigma]$ which in case of a Gaussian PDF includes the expectation value $\langle x \rangle$ in approximately 68% of all cases. The

same confidence level then also applies for $[\mu_-, \mu_+]$. From eq.(21) one sees that this definition of error interval corresponds to finding regions where the logarithm of the likelihood function stays within half a unit of the maximum, i.e. it is the shortest interval which can be constructed for a given probability content. This concludes the discussion of what will be assumed to be the connection between a measurement $x \pm \sigma$ and the corresponding Maximum Likelihood parameter estimate $\mu_{ml} \pm \sigma_\pm$.

It is now possible to study the expectation value and variance of μ_{ml}. With eq.(22) the expectation value becomes

$$\langle \mu_{ml} \rangle = \int d\mu_{ml} \, \mu_{ml} \, q(\mu_{ml}) = \int dx \, p(x) \, s^{-1}(x). \tag{24}$$

It is clearly not possible to determine this expectation value for arbitrary functions $s^{-1}(x)$, but one can address the problem in a systematic way by expanding $s^{-1}(x)$ around $\langle x \rangle = s(\mu)$. Since an asymmetric error will not come about for linear functions $s(\mu)$, the expansion should at least go to second order. Assuming that for practical applications the non-linearities will be small, the expansion will be truncated there.

$$s^{-1}(x) = \mu + \alpha y + \beta y^2 \quad \text{with} \quad y = x - \langle x \rangle. \tag{25}$$

Substituting this into eq.(24) and using $\langle y^2 \rangle = \sigma^2$ then yields

$$\langle \mu_{ml} \rangle = \mu + \beta \sigma^2. \tag{26}$$

One finds that the maximum likelihood estimate μ_{ml} is a biased estimator, with a constant bias proportional to the curvature of $s^{-1}(x)$. An unbiased estimator can be constructed by exploiting the information contained in the asymmetric errors. As discussed before, the error range $\mu_\pm = \mu_{ml} \pm \sigma_\pm$ is defined by the condition $\mu_\pm = s^{-1}(x \pm \sigma)$, which yields

$$\sigma_+ = \alpha\sigma + 2\beta\sigma y + \beta\sigma^2 \quad \text{and} \quad \sigma_- = \alpha\sigma + 2\beta\sigma y - \beta\sigma^2. \tag{27}$$

The difference of the two is proportional to the bias term in eq.(26) so that an unbiased estimator $\hat{\mu}$ for μ is given by

$$\hat{\mu} = \mu_{ml} - \frac{1}{2}(\sigma_+ - \sigma_-). \tag{28}$$

At first glance this result may be surprising, since it implies that in case the positive error is larger than the negative one, the unbiased estimate $\hat{\mu}$ for the parameter is in fact smaller than μ_{ml}. This underlines the fact that the asymmetric error does not describe a likelihood function for μ. Instead, a large positive error means that the value μ_{ml} on average will overestimate the true value, which is then compensated by subtracting half the difference of the errors.

In terms of the expansion eq.(25) the unbiased estimate $\hat{\mu}$ is given by

$$\hat{\mu} = \mu + \alpha y + \beta y^2 - \beta \sigma^2 \tag{29}$$

and with

$$\langle y^n \rangle = \frac{1}{\sqrt{2\pi}\sigma} \int dy \, y^n \exp\left(-\frac{y^2}{2\sigma^2}\right) = \begin{cases} 0 & n = 2k - 1 \\ 1 \cdot 3 \cdots (2k-1)\sigma^{2k} & n = 2k \end{cases} \tag{30}$$

one finds for the variance of $\hat{\mu}$

$$C(\hat{\mu}) = \alpha^2 \sigma^2 + 2\beta^2 \sigma^4. \tag{31}$$

From $\hat{\mu}$ and the asymmetric errors alone it is not possible to determine the value of the variance, since in addition to μ, α and β also $(x - \langle x \rangle)$ is an unknown, i.e. one is faced with three constraints for four variables. In the same spirit, however, as $\hat{\mu}$ is an unbiased estimator for μ, one can also construct an unbiased estimator $\hat{C}(\hat{\mu})$ for C. The solution is given by

$$\hat{C}(\hat{\mu}) = \frac{1}{4}(\sigma_+ + \sigma_-)^2 - \frac{1}{2}(\sigma_+ - \sigma_-)^2. \tag{32}$$

One easily verifies that the expectation value of eq.(32) reproduces eq.(31). The functional expression eq.(32) can be interpreted as the naive estimate for the variance of $\hat{\mu}$ which is reduced by a correction term that vanishes for symmetric errors. An important aspect is that $\hat{C}(\hat{\mu})$ is only an estimate for the variance of $\hat{\mu}$, even if the variance σ^2 of the measurement is known precisely. Only for $\beta = 0$, i.e. when the connection between measurement x and parameter μ is linear one recovers $\hat{C} = C$. As shown below, the fact that $\hat{C}(\hat{\mu})$ is subject to statistical fluctuation complicates matters a lot.

4.2 Weighted Averages with Asymmetric Errors

In the previous subsection a prescription was derived to quote an unbiased estimate for mean value and variance of a parameter μ. In terms of the expansion eq.(25) the results are

$$\hat{\mu} = \mu + \alpha y + \beta y^2 - \beta \sigma^2 \quad \text{and} \quad \hat{C} = (\alpha \sigma + 2\beta \sigma y)^2 - 2\beta^2 \sigma^4. \tag{33}$$

With eq.(31) this can be rewritten as

$$\hat{C} - C = \gamma(\hat{\mu} - \mu) \quad \text{with} \quad \gamma = 4\beta\sigma^2 = 2(\sigma_+ - \sigma_-). \tag{34}$$

It follows immediately that the variance of $\hat{C}$ is given by $\gamma^2 C$, i.e. it vanishes for symmetric errors. In case of asymmetric errors both $\hat{C}$ and $\hat{\mu}$ are random variables which scatter around their respective expectation values, and according to eq.(34) both are fully correlated. Although individually both variables are unbiased, their ratio $\hat{\mu}/\hat{C}$ is not. As a consequence also the weighted average $\bar{\mu}$

$$\bar{\mu} = \frac{\sum_i \hat{\mu}_i/\hat{C}_i}{\sum_i 1/\hat{C}_i} = \mu + \frac{\sum_i (\hat{\mu}_i - \mu)/\hat{C}_i}{\sum_i 1/\hat{C}_i}. \tag{35}$$

is no longer an unbiased estimator for the true parameter value μ. Introducing

$$d_i = \hat{\mu}_i - \mu \quad \text{and} \quad \hat{C}_i = C_i + \gamma_i d_i, \tag{36}$$

the expectation value of $\bar{\mu}$ can be written as

$$\langle \bar{\mu} \rangle = \mu + \left\langle \frac{\sum_i d_i/(C_i + \gamma_i d_i)}{\sum_i 1/(C_i + \gamma_i d_i)} \right\rangle = \mu + \left\langle \frac{Z_1}{Z_0} \right\rangle \quad \text{with} \quad Z_m = \sum_i \frac{d_i^m}{C_i + \gamma_i d_i}, \tag{37}$$

and the bias of the weighted average is given by the expectation value of the ratio Z_1/Z_0.

Using a Taylor expansion about the origin, the bias $\langle Z_1/Z_0 \rangle$ can be expressed through the moments of the deviates d_i. For a general function F of deviates d_i one has

$$\langle F \rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \left\langle \left(\sum_k d_k \frac{\partial}{\partial d_k} \right)^n F(0) \right\rangle. \tag{38}$$

8

The notation $F(0)$ implies that all derivatives of F are to be taken at $d_i = 0$. For the problem at hand the first moments are zero, $\langle d_i \rangle = 0$, and different deviates are independent, $\langle d_k^m d_l^n \rangle = \langle d_k^m \rangle \langle d_l^n \rangle$ for $k \neq l$. Under these conditions eq.(38) simplifies considerably and the leading order terms become

$$\langle F \rangle = \left(1 + \sum_k \frac{\langle d_k^2 \rangle}{2!} \frac{\partial^2}{\partial d_k^2} + \sum_k \frac{\langle d_k^3 \rangle}{3!} \frac{\partial^3}{\partial d_k^3} + \sum_k \frac{\langle d_k^4 \rangle}{4!} \frac{\partial^4}{\partial d_k^4} + \sum_{k<l} \frac{\langle d_k^2 \rangle}{2!} \frac{\langle d_l^2 \rangle}{2!} \frac{\partial^2}{\partial d_k^2} \frac{\partial^2}{\partial d_l^2} \cdots \right) F(0). \quad (39)$$

The evaluation of the moments $\langle d_i^n \rangle$ using eq.(30) is straightforward. The leading order results, expressed as function of the variance C_k and the asymmetry parameter γ_k, are

$$\left\langle d_k^2 \right\rangle = C_k, \quad \left\langle d_k^3 \right\rangle = \frac{3}{2} C_k \gamma_k - \frac{1}{16} \gamma_k^3 \quad \text{and} \quad \left\langle d_k^4 \right\rangle = 3 C_k^2 + 3 \gamma_k^2 C_k - \frac{3}{16} \gamma_k^4. \quad (40)$$

At this points all ingredients are available which are needed to construct an unbiased weighted average also in presence of asymmetric errors. To simplify the organization of terms it is convenient to define, in addition to Z_m from eq.(37), the following sums:

$$T_{lm} = \sum_k \frac{\gamma_k^l}{(C_k + \gamma_k d_k)^m} \quad \text{and} \quad S_{lm} = \sum_k \frac{\gamma_k^l}{C_k^m} = T_{lm}(0). \quad (41)$$

For the derivatives of Z_m and T_{lm}, $n > 0$, one finds

$$\frac{\partial^n}{\partial d_k^n} Z_m(0) = n! \frac{(-\gamma_k)^{n-m}}{C_K^{n-m+1}} \text{ if } n \geq m \quad \text{and} \quad \frac{\partial^n}{\partial d_k^n} T_{lm}(0) = \frac{(-\gamma_k)^{n+l}}{C_k^{n+m}} \frac{(n+m-1)!}{(m-1)!}. \quad (42)$$

Evidently every derivative picks up one power of γ_k and the expansion around $d_i = 0$ will be a power series in γ_i. The error of a truncated series will be quoted as $O(\gamma^n)$, meaning a sum where each term is a product of at least n factors γ_i. In general these factors will come from different measurements $\hat{\mu}_i$.

The actual calculations are straightforward but lengthy. Results of auxiliary calculations are collected in the appendix. For the bias on the weighted average one finds

$$\left\langle \frac{Z_1}{Z_0} \right\rangle = -\frac{S_{11}}{S_{01}} + \frac{S_{12}}{S_{01}^2} + O(\gamma^3). \quad (43)$$

Equation(43) shows that in case of asymmetric error the bias is of $O(\gamma)$, i.e. proportional to the asymmetry of the errors. After correction for the dominant effect the residual bias is only of $O(\gamma^3)$, because both the third derivative of Z_1/Z_0 and the third moments $\langle d_k^3 \rangle$ each provide one extra power of γ. The higher order terms also contribute at least two factors γ.

In practical applications a bias correction according to eq.(43) will be done by substituting the T_{lm} for the S_{lm}. This is permitted as long as the expectation value $\langle T_{lm} \rangle$ is equal to S_{lm} to the same order in γ as the terms that are anyhow neglected in the correction terms. As shown in the appendix, the additional bias from this substitution for eq.(43) is again of $O(\gamma^3)$. To leading order it thus is consistent to use T_{lm} instead of S_{lm} to obtain a bias corrected weighted average $\bar{\mu}_c$

$$\bar{\mu}_c = \bar{\mu} + \frac{T_{11}}{T_{01}} - \frac{T_{12}}{T_{01}^2} = \mu + \frac{Z_1}{Z_0} + \frac{T_{11}}{T_{01}} - \frac{T_{12}}{T_{01}^2} = \mu + \Delta \quad (44)$$

with $\bar{\mu}_c - \mu = O(\gamma^3)$.

The variance $C(\bar{\mu}_c)$ of the bias corrected average $\bar{\mu}_c$ is given by

$$
\begin{aligned}
C(\bar{\mu}_c) = \langle \Delta^2 \rangle - \langle \Delta \rangle^2 &=& \left\langle \frac{Z_1^2}{Z_0^2} \right\rangle - \left\langle \frac{Z_1}{Z_0} \right\rangle^2 &+& 2 \left(\left\langle \frac{Z_1}{Z_0} \frac{T_{11}}{T_{01}} \right\rangle - \left\langle \frac{Z_1}{Z_0} \right\rangle \left\langle \frac{T_{11}}{T_{01}} \right\rangle \right) \\
&+& \left\langle \frac{T_{11}^2}{T_{01}^2} \right\rangle - \left\langle \frac{T_{11}}{T_{01}} \right\rangle^2 &-& 2 \left(\left\langle \frac{Z_1}{Z_0} \frac{T_{12}}{T_{01}^2} \right\rangle - \left\langle \frac{Z_1}{Z_0} \right\rangle \left\langle \frac{T_{12}}{T_{01}^2} \right\rangle \right) \\
&+& \left\langle \frac{T_{12}^2}{T_{01}^4} \right\rangle - \left\langle \frac{T_{12}}{T_{01}^2} \right\rangle^2 &-& 2 \left(\left\langle \frac{T_{11}}{T_{01}} \frac{T_{12}}{T_{01}^2} \right\rangle - \left\langle \frac{T_{11}}{T_{01}} \right\rangle \left\langle \frac{T_{12}}{T_{01}^2} \right\rangle \right) .
\end{aligned} \tag{45}
$$

The individual contributions to this expression can be found in the appendix. Collecting all terms one obtains

$$
C(\bar{\mu}_c) = \frac{1}{S_{01}} + \frac{S_{22}}{S_{01}^2} - \frac{2 S_{23}}{S_{01}^3} + \frac{S_{12}^2}{S_{01}^4} + O(\gamma^4). \tag{46}
$$

In practical applications one again has to use T_{lm} instead of S_{lm}, which for the first term in eq.(46) introduces an additional bias of $O(\gamma^2)$ which must be compensated by some higher order terms. To leading order one has

$$
\frac{1}{S_{01}} = \frac{1}{T_{01}} + \frac{T_{22}}{T_{01}^2} - \frac{T_{23}}{T_{01}^3} + O(\gamma^4). \tag{47}
$$

For all other terms the substitution T_{lm} for S_{lm} is consistent, and the variance $\hat{C}(\bar{\mu}_c)$ expressed through known quantities becomes

$$
\hat{C}(\bar{\mu}_c) = \frac{1}{T_{01}} + \frac{2 T_{22}}{T_{01}^2} - \frac{3 T_{23}}{T_{01}^3} + \frac{T_{12}^2}{T_{01}^4} + O(\gamma^4). \tag{48}
$$

Note that in case of symmetric errors only the first term contributes, recovering the conventional estimate for the variance of a weighted average.

Finally, the expectation value of the χ^2-test-variable shall be examined for a weighted average based on measurements with asymmetric errors. Defined in the usual way, χ^2 is given by

$$
\chi^2 = \sum_k \frac{(\hat{\mu}_k - \bar{\mu})^2}{\hat{C}_k} = Z_2 - \frac{Z_1^2}{Z_0} \tag{49}
$$

where $\hat{\mu}_k$ and $\hat{C}_k$ are the unbiased estimates for mean value and variance of the individual measurements and $\bar{\mu}$ determined according to eq.(35). Up to second order the determination of the expectation value is easily performed, giving the result

$$
\left\langle \chi^2 \right\rangle = (n-1) + \frac{3}{2} S_{21} - \frac{S_{11}^2 + 4 S_{22}}{S_{01}} + \frac{4 S_{11} S_{12} + 7 S_{23}}{2 S_{01}^2} - \frac{2 S_{12}^2}{S_{01}^3} + O(\gamma^4), \tag{50}
$$

where n is the number of measurements contributing to the weighted average. For practical applications this result implies, that also in the presence of asymmetric errors the usual χ^2-variable can be used as a goodness-of-fit criterion. For a consistent set of measurements the value should be not too deviant from the number of degrees of freedom. One should also note the somewhat surprising result that a meaningful χ^2-variable is based on the non-bias-corrected average $\bar{\mu}$.

Before moving on, it is worth while to discuss the explicit expressions when averaging n data points which all are drawn from the same parent distribution, i.e. $C_k = C$ and $\gamma_k = \gamma$ for all

10

indices k, as then the inherent structure of the results becomes most evident. One finds the following expressions for the expectation values discussed above:

$$\langle \bar{\mu} \rangle = \mu - \gamma \left(1 - \frac{1}{n} \right) + O(\gamma^3) \tag{51}$$

$$\langle C(\bar{\mu}) \rangle = \frac{C}{n} + \frac{\gamma^2}{n} \left(1 - \frac{1}{n} \right) + O(\gamma^4) \tag{52}$$

$$\langle \chi^2 \rangle = (n-1) + \frac{\gamma^2}{2C} \left(n - 4 + \frac{3}{n} \right) + O(\gamma^4) \tag{53}$$

A simple cross check is obtained by setting $n = 1$ in which case the values $\langle \bar{\mu} \rangle = \mu$, $\langle C(\bar{\mu}) \rangle = C$ and $\langle \chi^2 \rangle = 0$ are recovered correctly. One also sees that for $\gamma = 0$ the usual expressions known for weighted averages are obtained. The corrections which are needed for non-linear relations between measurements and model parameters are an expansion in the non-linearity parameter γ and $1/n$.

4.3   A Numerical Example

Since some of the above results may seem counterintuitive they shall be illustrated by a numerical example. The model assumes that the relation between a true parameter μ_t and the expectation value $\langle x \rangle$ of a measurement x is given by $\langle x \rangle = \sqrt{\mu_t}$. Assuming that an experiment provides an unbiased measurement x with a Gaussian PDF around the mean, the likelihood function is given by

$$p(x|\mu) \sim \exp\left(-\frac{(x - \sqrt{\mu})^2}{2\sigma^2} \right). \tag{54}$$

To study the averaging procedure, measurements x were generated in a Monte Carlo simulation and treated in the usual way to obtain a Maximum Likelihood estimate μ_{ml} with asymmetric errors covering a 68% confidence level interval. The properties of single estimates as well as weighted averages will be discussed below.

For the simulation it was assumed that the true value is $\mu_t = 25$ and that the variance of the measurements is $\sigma^2 = 0.5$. The characteristics of the problem then are given by $\langle x \rangle = 5$, $\alpha = 10$, $\beta = 1$, $\gamma = 2$ and $C = 50.5$. A single value x yields a maximum likelihood estimate $\hat{\mu}_{ml} = x^2$ with asymmetric errors $\sigma_\pm = 2x\sigma \pm \sigma^2$.

The distribution of the measurements x is shown together with an example for the slightly asymmetric likelihood function, taking $x = 5$, in the top row of fig.(1). For this example the maximum likelihood parameter estimate becomes $\mu_{ml} = 25 \pm^{7.57}_{6.57}$. The two lower plots of fig.(1) display the distributions of the maximum likelihood estimates that result from the Gaussian distribution in x. The left hand plot shows the distribution of the plain estimate μ_{ml}, the right hand plot has the bias corrected distribution of $\hat{\mu}$. The expectation values are $\langle \mu_{ml} \rangle = 25.5$ and $\langle \hat{\mu} \rangle = 25$. The plain Maximum Likelihood estimate evidently is biased, consistent with the expected bias $(\sigma_+ - \sigma_-)/2 = \sigma^2 = 0.5$. The example nicely illustrates that an asymmetric error $\sigma_+ > \sigma_-$ really means that the maximum likelihood estimate tends to be larger than the true parameter value. To correct for the bias one has to reduce the value by the average asymmetry of the errors.

As an aside, it may be interesting to note that using the mean value $\langle \mu(x) \rangle$ of the likelihood function $\mu(x)$ rather than the maximum as an estimate for the unknown parameter actually
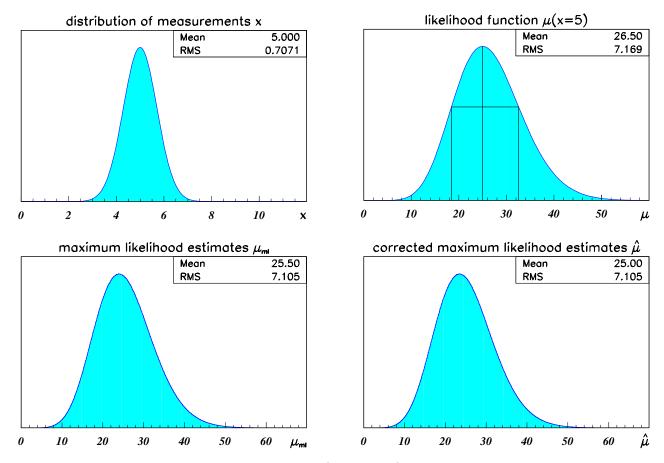
11

Figure 1: Distribution of measurements (upper left), and likelihood function for the parameter μ from a measurement $x = 5$ (upper right). The lower plots show the distribution of the uncorrected (left) and the bias corrected (right) Maximum Likelihood estimates, μ_{ml} and $\hat{\mu}$ (right), respectively.

makes things worse. If spill-over to negative values can be ignored, then one finds $\langle \mu(x) \rangle = x^2 + 3\sigma^2$, i.e. a bias three times larger than the bias of the Maximum Likelihood estimate μ_{ml}. Also this is illustrated in fig.(1) by the likelihood function $\mu(x = 5)$.

The Monte-Carlo simulation shows that the variance of the bias corrected maximum likelihood estimate $C(\hat{\mu})$ is consistent with the expectation $C = 50.5$. It is interesting to compare this value to a heuristic estimate for the variance, based on the average of the positive and the negative errors. The expectation value of this heuristic variance is slightly larger than C. One finds $\langle (\sigma_+ + \sigma_-)^2/4 \rangle = \langle 4\sigma^2 x^2 \rangle = 4\sigma^2(\langle x \rangle^2 + \sigma^2) = 51$. The term $(\sigma_+ - \sigma_-)^2/2$ in eq.(32) corrects this bias.

Having established unbiased estimates for μ and the variance of that estimate, the next step is to study the weighted average of results from independent measurements. As discussed before, using the inverse of the variance estimate as the weight, biases the average because of the correlation between parameter and variance estimate. The result of using eq.(35) for an average of 10 values is displayed in the upper left plot of fig.(2). The central value of the average is about 0.8 standard deviations lower than the true parameter value. The estimated bias is shown in the upper right. According to eq.(43), i.e. neglecting terms of $O(\gamma^4)$, one would expect a shift of 1.8 units, which is close to what is actually observed. The bias corrected
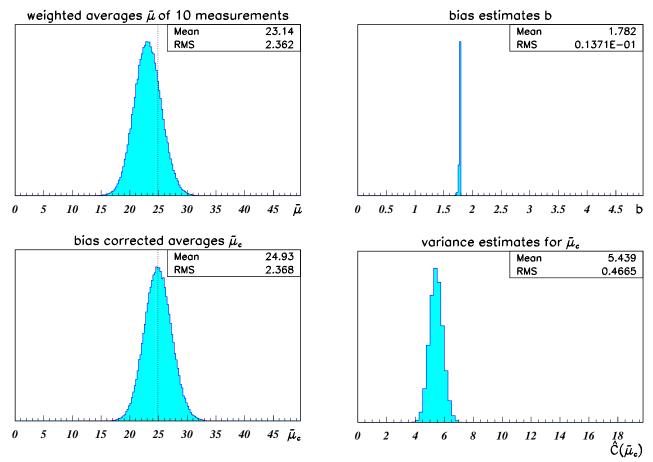
12

Figure 2: Weighted average of 10 bias-corrected maximum likelihood estimates $\hat{\mu}$ with unbiased variance estimates $\hat{C}(\hat{\mu})$. The upper left plot shows that the uncorrected average $\bar{\mu}$ is significantly biased. The bias estimate is displayed in the upper right. The distribution of the bias corrected averages $\bar{\mu}_c$ is shown in the lower left, and the distribution of the variance estimates for the corrected averages on the lower right.

average is shown in the lower left of fig.(2). The lower right finally shows the distribution of the variance estimate for the bias corrected weighted average. The distribution is rather narrow with a central value close to the actual variance found in the distribution of the averages. That the general picture is consistent is also shown in the χ^2-distribution of the weighted average, fig.(3). According to eq.(50), including all terms up to $O(\gamma^2)$ one would predict an expectation value $\langle \chi^2 \rangle \approx 9.25$, again very close to the observed value.

To illustrate further the properties of different schemes to form a weighted average from measurements with asymmetric errors, fig.(4) shows how the various methods converge when increasing the number n of measurements. Here n was varied from 1 to 50. The top plot shows the result of a simple heuristic method, where the maximum likelihood estimates μ_{ml} are averaged, using as weights the inverse of $(\sigma_+ + \sigma_-)^2/4$, the middle diagram was obtained by using unbiased, though correlated, estimates for mean value and variance, and the lower plot shows the result of the proper bias correction scheme. The first two schemes yield very similar results, the main difference being a shift corresponding to the bias correction for an individual measurement. The heuristic method has a large bias for $n = 1$, which is corrected in the second method. At large n both significantly underestimate the true parameter value. One also has
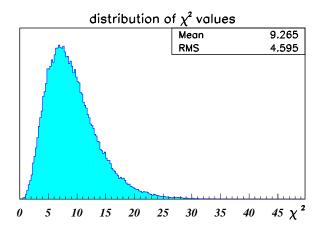
distribution of χ^2 values

Mean	9.265
RMS	4.595

Figure 3: Distribution of χ^2-values for the weighted average of 10 bias-corrected maximum likelihood estimates.

to mention that both provide error estimates which underestimate the actual uncertainty of an individual weighted average by roughly 10%. The proper bias correction scheme has both a stable expectation value for the average and a reliable error estimate.

evolution of weighted averages

heuristic average

average with unbiased inputs
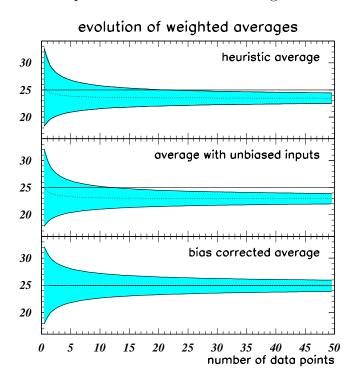
bias corrected average

number of data points

Figure 4: Evolution of mean value and error estimate as function of the number of measurements in the average. The top plot shows the behavior of a simple heuristic treatment of asymmetric errors, the middle one the result of using bias corrected estimates $\hat{\mu}$ and $\hat{C}(\hat{\mu})$ for mean value and variance, and the lower plot the results after proper bias correction for the average and its variance.

5   Systematic Uncertainties

The previous considerations apply to statistical uncertainties. Many measurements are also subject to systematic or theoretical uncertainties, which affect the model that is fitted to the experimental data. Formally the effect of these uncertainties can be described by an additional parameter a entering the model, where the knowledge about a is described by a PDF $h(a)$. The function $h(a)$ can be the result of a measurement or a purely Bayesian parameterization for the subjective degree of belief in a certain value a. The formal treatment will be unaffected

14

by this. With the parameter a eq.(21) becomes

$$p(x|\mu, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - s(\mu, a))^2}{2\sigma^2}\right). \tag{55}$$

The simplest and probably still very common approach to deal with the situation is to vary a according to $h(a)$ around a nominal value a_0 over a range covering a 68% confidence level interval, and quote the resulting variation in μ as a systematic error. The discussion presented here will deal with only this approach. As before, from the quoted errors the mean value and variance of the parameter estimate shall be inferred, since that is the information which is needed to form a weighted average.

For a fixed parameter a the maximum likelihood estimate μ_{ml} for μ according to eq.(55) is given by $\mu_{ml} = s^{-1}(x, a)$. The resulting PDF $\rho(\mu_{ml})$ for μ_{ml} depends on $s^{-1}(x, a)$ and the PDF $h(a)$ of a.

$$\rho(\mu_{ml}) = \int da\ h(a)\ \delta\left(\mu_{ml} - s^{-1}(x, a)\right). \tag{56}$$

From eq.(56) one finds for the moments of μ_{ml}

$$\left\langle \mu_{ml}^k \right\rangle = \int da\ h(a)\ \left(s^{-1}(x, a)\right)^k. \tag{57}$$

To continue, it will be assumed that $h(a)$ is a Gaussian with mean value a_0 and standard deviation τ, and that $s^{-1}(x, a)$ has an expansion around a_0

$$s^{-1}(x, a) = \mu_0(x) + \alpha(x)(a - a_0) + \beta(x)(a - a_0)^2 \tag{58}$$

where the higher order terms are negligible. To simplify the notation the x-dependence of the coefficient will be dropped. With

$$h(a) = \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(a - a_0)^2}{2\tau^2}\right) \tag{59}$$

one finds for the first and second moment of μ_{ml}

$$\left\langle \mu_{ml} \right\rangle = \mu_0 + \beta\tau^2 \quad \text{and} \quad \left\langle \mu_{ml}^2 \right\rangle = \mu_0^2 + (\alpha^2 + 2\mu_0\beta)\tau^2 + 3\beta^2\tau^4. \tag{60}$$

From eq.(60) one finds the variance of μ_{ml}

$$C(\mu_{ml}) = \left\langle \mu_{ml}^2 \right\rangle - \left\langle \mu_{ml} \right\rangle^2 = \alpha^2\tau^2 + 2\beta^2\tau^4. \tag{61}$$

For $a = a_0$ one has the nominal parameter estimate μ_0. The in general asymmetric errors are given by

$$\tau_+ = \alpha\tau + \beta\tau^2 \quad \text{and} \quad \tau_- = \alpha\tau - \beta\tau^2. \tag{62}$$

Combining the information from eq.(60) and eq.(62) the unbiased parameter estimate and its variance can be written as

$$\left\langle \mu_{ml} \right\rangle = \mu_0 + \frac{1}{2}(\tau_+ - \tau_-) \quad \text{and} \quad C(\mu_{ml}) = \frac{1}{4}(\tau_+ + \tau_-)^2 + \frac{1}{2}(\tau_+ - \tau_-)^2. \tag{63}$$

These expressions eq.(63) show how to deal with theoretical or systematic uncertainties when averaging independent measurements. In contrast to the statistical errors discussed before, now asymmetric errors translate into a shift of the parameter as one would naively expect. Also the variance is strictly positive. Although the derivation is formally very similar to the one of the statistical errors, here the PDF of the function of the parameter a is known a priori, which is not the case for the measurement x. Note that the expressions eq.(63) have been derived for the assumption that systematic uncertainties are Gaussian errors.

6   Combination of Statistical and Systematic errors

Except in the context of Bayesian statistics the combination of statistical and systematic errors is not a well defined concept. If one, however, adopts the attitude to model the influence of the two kinds of uncertainties by means of a Monte Carlo method, where the measurement x as well as the theory-parameter a is varied according to their respective probability density functions, then the total bias and variance for the parameter estimate $\hat{\mu}$ is given by the sum of the individual biases and the individual variances, i.e. the errors can be combined in quadrature:

$$C_k = C_k^{\text{stat}} + C_k^{\text{syst}}. \tag{64}$$

In the Bayesian sense this is the correct way of combining different sources of uncertainty. From a non-Bayesian point of view the ansatz eq.(64) at least has the qualitatively desired properties, that measurements with large uncertainty in either of the two components will get a low weight in any average.

When using the combined variances in a weighted average, one again needs to perform a bias correction based on the asymmetry parameters γ. As defined in eq.(34), only the difference of the statistical errors contributes to γ, since only asymmetries in the statistical uncertainties lead to a correlation between variance and mean value. No such correlation exists for systematic errors. Otherwise the definitions eq.(41) apply as before.

7   Summary and Conclusions

The weighted average with weights based only on the variance of the individual measurements is a robust estimator for a common mean. As a consequence of the central limit theorem, its PDF can be assumed to be a Gaussian even if the PDFs of the inputs are not known in detail. Although it would be preferable from a theoretical point of view to take correlations into account when forming the average, this can only be recommended if those are well understood. A simple effective way is proposed to include the impact of correlations between the measurements for the error estimate, using the χ^2 of the average as an indicator variable.

In a second step the problem of asymmetric 68% confidence level error intervals was addressed. The argument was based on the assumption of a measurement with a Gaussian PDF and the use of the Maximum Likelihood Method to infer the value of a physical parameter and its error. Explicit expressions were derived to convert the given information into an unbiased estimate for the value of the parameter and its variance. In addition, it has been shown that when dealing with asymmetric errors, correction terms for the weighted average and its variance have to be taken into account. In case of symmetric errors the usual expressions for the weighted average are recovered. The scaling prescription to account for correlations between measurements can be applied in the same way for measurements with symmetric and asymmetric errors.

Finally, systematic errors were studied, assuming that they enter as model parameters with a known PDF that has to be interpreted in the Bayesian sense. As such they behave differently from the statistical errors of a measurement. Also here explicit expressions are given how to infer mean value and variance and how to combine them with statistical uncertainties.

Acknowledgments

References

[1] M. Schmelling, Physica Scripta 51 (1995) 676.

[2] C. Caso, *et al.*, The Particle Data Group, Eur. Phys. J. C3 (1998) 1;

Appendix: Auxiliary Expectation Values

Here the complete set of expectation values is given which are needed to derive the results of section 4.2. In most cases a precision of at least $O(\gamma^2)$ for the expectation values is obtained already when truncating the expansion eq.(39) after the second term.

$$\langle F \rangle = F(0) + \sum_k \frac{\langle d_k^2 \rangle}{2!} \frac{\partial^2}{\partial d_k^2} F(0) + \dots \tag{65}$$

Elementary calculus with the derivatives and definitions given in eq.(37), eq.(41) and eq.(42) then yields the following results:

$$\left\langle \frac{1}{T_{01}} \right\rangle = \frac{1}{S_{01}} - \frac{S_{22}}{S_{01}^2} + \frac{S_{23}}{S_{01}^3} + O(\gamma^4) \tag{66}$$

$$\left\langle \frac{T_{11}}{T_{01}} \right\rangle = \frac{S_{11}}{S_{01}} + \frac{S_{32}}{S_{01}} - \frac{S_{11}S_{22} + S_{33}}{S_{01}^2} + \frac{S_{11}S_{23}}{S_{01}^3} + O(\gamma^5) \tag{67}$$

$$\left\langle \frac{T_{11}^2}{T_{01}^2} \right\rangle = \frac{S_{11}^2}{S_{01}^2} + \frac{2S_{11}S_{32} + S_{43}}{S_{01}^2} - \frac{2S_{11}^2 S_{22} + 4S_{11}S_{33}}{S_{01}^3} + \frac{3S_{11}^2 S_{23}}{S_{01}^4} + O(\gamma^6) \tag{68}$$

$$\left\langle \frac{T_{12}}{T_{01}^2} \right\rangle = \frac{S_{12}}{S_{01}^2} + \frac{S_{33}}{S_{01}^2} - \frac{2S_{12}S_{22} + 4S_{34}}{S_{01}^3} + \frac{3S_{12}S_{23}}{S_{01}^4} + O(\gamma^5) \tag{69}$$

$$\left\langle \frac{T_{12}^2}{T_{01}^4} \right\rangle = \frac{S_{12}^2}{S_{01}^4} + \frac{6S_{12}S_{33} + 4S_{45}}{S_{01}^4} - \frac{4S_{12}^2 S_{22} + 16S_{12}S_{34}}{S_{01}^5} + \frac{10S_{12}^2 S_{23}}{S_{01}^6} + O(\gamma^6) \tag{70}$$

$$\left\langle \frac{T_{11}T_{12}}{T_{01}^3} \right\rangle = \frac{S_{11}S_{12}}{S_{01}^3} + \frac{S_{12}S_{32} + 3S_{11}S_{33} + 2S_{44}}{S_{01}^3}$$
$$- \frac{3S_{11}S_{12}S_{22} + 3S_{12}S_{33} + 6S_{11}S_{34}}{S_{01}^4} + \frac{6S_{11}S_{12}S_{23}}{S_{01}^5} + O(\gamma^6) \tag{71}$$

$$\left\langle \frac{Z_1}{Z_0} \right\rangle = -\frac{S_{11}}{S_{01}} + \frac{S_{12}}{S_{01}^2} + O(\gamma^3) \tag{72}$$

$$\left\langle \frac{T_{11}Z_1}{T_{01}Z_0} \right\rangle = -\frac{S_{11}^2 + S_{22}}{S_{01}^2} + \frac{2S_{11}S_{12}}{S_{01}^3} + O(\gamma^4) \tag{73}$$

$$\left\langle \frac{T_{12}Z_1}{T_{01}^2 Z_0} \right\rangle = -\frac{S_{11}S_{12} + 2S_{23}}{S_{01}^3} + \frac{3S_{12}^2}{S_{01}^4} + O(\gamma^4) \tag{74}$$

The quantities where one has to go to fourth order in eq.(39) in order to collect all terms up to $O(\gamma^2)$ are $\langle Z_1^2/Z_0^2 \rangle$, $\langle Z_1^2/Z_0 \rangle$ and $\langle Z_2 \rangle$. For $F = Z_1^2/Z_0^2$ the partial derivatives up to fourth order are:

$$\frac{1}{2!}\frac{\partial^2}{\partial d_k^2}F(0) = \frac{1}{S_{01}^2}\frac{1}{C_k^2} \tag{75}$$

$$\frac{1}{3!}\frac{\partial^3}{\partial d_k^3}F(0) = -\frac{1}{S_{01}^2}\frac{2\gamma_k}{C_k^3} + \frac{1}{S_{01}^3}\frac{2\gamma_k}{C_k^4} \tag{76}$$

$$\frac{1}{4!}\frac{\partial^4}{\partial d_k^4}F(0) = \frac{1}{S_{01}^2}\frac{3\gamma_k^2}{C_k^4} - \frac{1}{S_{01}^3}\frac{6\gamma_k^2}{C_k^5} + \frac{1}{S_{01}^4}\frac{3\gamma_k^2}{C_k^6} \tag{77}$$

$$\frac{1}{2!2!}\frac{\partial^2}{\partial d_k^2}\frac{\partial^2}{\partial d_l^2}F(0) = \frac{1}{S_{01}^2}\frac{2\gamma_k\gamma_l}{C_k^2 C_l^2} - \frac{1}{S_{01}^3}\left(\frac{2\gamma_k^2}{C_k^3 C_l^2} + \frac{4\gamma_k\gamma_l}{C_k^2 C_l^3} + \frac{4\gamma_k\gamma_l}{C_k^3 C_l^2} + \frac{2\gamma_l^2}{C_k^2 C_l^3}\right) \tag{78}$$

$$+ \frac{1}{S_{01}^4}\left(\frac{3\gamma_k^2}{C_k^4 C_l^2} + \frac{12\gamma_k\gamma_l}{C_k^3 C_l^3} + \frac{3\gamma_l^2}{C_k^2 C_l^4}\right)$$

With the moments given in eq.(40) and using the identity

$$2\sum_{k<l} a_k b_l = \left(\sum_k a_k\right)\left(\sum_l b_l\right) - \left(\sum_k a_k b_k\right) \tag{79}$$

the expectation value $\langle Z_1^2/Z_0^2 \rangle$ finally becomes

$$\left\langle \frac{Z_1^2}{Z_0^2}\right\rangle = \frac{1}{S_{01}} + \frac{S_{11}^2 + 3S_{22}}{S_{01}^2} - \frac{4S_{11}S_{12} + 6S_{23}}{S_{01}^3} + \frac{6S_{12}^2}{S_{01}^4} + O(\gamma^4). \tag{80}$$

Similarly one finds for the other expectation values

$$\left\langle \frac{Z_1^2}{Z_0}\right\rangle = 1 + \frac{S_{11}^2 + 4S_{22}}{S_{01}} - \frac{4S_{11}S_{12} + 7S_{23}}{2S_{01}^2} + \frac{2S_{12}^2}{s_{01}^3} + O(\gamma^4) \tag{81}$$

$$\langle Z_2 \rangle = n + \frac{3}{2}S_{21} + O(\gamma^4) \tag{82}$$

where n is the number of measurements contributing to the average.