Nuclear Instruments and Methods in Physics Research A270 (1988) 110–117
North-Holland, Amsterdam

HOW TO COMBINE CORRELATED ESTIMATES OF A SINGLE PHYSICAL QUANTITY

Louis LYONS and Duncan GIBAUT
Nuclear Physics Laboratory, Keble Road, Oxford, UK

Peter CLIFFORD
Mathematical Institute, St. Giles, Oxford, UK

Received 4 January 1988

Experiments to measure a single physical quantity often produce several estimates based on the same data, and which are hence correlated. We describe how to combine these correlated estimates in order to provide the best single answer, and also how to check whether the correlated estimates are mutually consistent.

We discuss the properties of our technique, and illustrate its application by using it for a specific experiment which measured the lifetime of charmed particles.

1. Introduction

It is common for a single experiment which aims to measure some physical quantity y to produce more than one estimate of this quantity. This could arise, for example, from using different techniques to analyse the same data, from having several slightly different data sets determined by various possible selection criteria aimed at improving the quality of the data sample, etc.

It is clearly desirable to quote a *single* number at the end of the paper, as *the* result of the experiment, rather than to leave it as an exercise for the reader to decide which if any of the several estimates he prefers. One possibility is to select the result with the smallest fractional error, but this is almost equivalent to ignoring the other approaches. Another technique is to produce some sort of average of the different answers in order to obtain a better one. It is this approach that we discuss in this paper.

The standard method for combining different experiments consists in weighting each result $y_i \pm \sigma_i$ by a factor inversely proportional to that measurement's variance, i.e.

$$\hat{y} = \sum \left(y_i/\sigma_i^2 \right) \Big/ \sum \left(1/\sigma_i^2 \right). \tag{1}$$

The corresponding error σ is given by

$$1/\sigma^2 = \sum \left(1/\sigma_i^2 \right). \tag{2}$$

This, however, applies only when the individual measurement errors are uncorrelated. In the situation we have described this is unlikely to be so, because the different estimates are based on much or all of the same data. Thus if, for example, the common events result in values of the parameter which are larger than average, then all the methods will tend to produce results which are too high; this is the source of the correlation. We here present an alternative to eqs. (1) and (2) for the case where the errors on the different results are correlated; this makes use of the complete error matrix **E** of the estimates.

It is common for an experiment to claim that alternative methods of determining a parameter agree within their errors, and hence are satisfactory. When the same data is used for more than one determination, the errors are not independent and so it is not clear how well we should really expect the separate estimates to agree. We provide a prescription for dealing with this problem as well.

The plan of this paper is as follows. The rather simple mathematics of the method is given in section 2, while section 3 contains a discussion of some special cases, in order to provide more insight into what this approach achieves. In section 4 we outline a particular high energy physics experiment which determined the mean lifetime of charmed particles by 4 different methods, and for which the combination technique described in this paper is applicable. Details of the way we use our technique for the charmed particle lifetime are given in section 5, together with an account of how we use a Monte Carlo method to estimate the error matrix **E**. Our conclusions appear in section 6.

0168-9002/88/$03.50

2. The method

We assume that our experiment has provided us with several unbiassed estimates y_i $(i = 1 \ldots n)$ of our parameter, together with their error matrix $\mathbf{E}$. The diagonal elements of $\mathbf{E}$ give the variances of the individual estimates, while the off-diagonal elements describe the correlations between pairs of estimates. The standard procedure for combining different results (eqs. (1) and (2)) applies when the off-diagonal elements are zero.

We employ a BLUE technique [1] (i.e., Best Linear Unbiassed Estimate) which consists in looking for an estimate $\hat{y}$ which

(1) is a linear combination of the individual estimates;
(2) provides an unbiassed estimate of y; and
(3) has the minimum possible variance σ^2.

The first condition is that

$$\hat{y} = \sum \alpha_i y_i, \tag{3}$$

where the α_i are constants, the weighting factors for the various estimates y_i, and which we want to determine.

For eq. (3) to be unbiassed, we require

$$\sum \alpha_i = 1; \tag{4}$$

this follows from our assumption that the y_i are themselves unbiassed.

From eq. (3), we deduce the variance of $\hat{y}$ as

$$\sigma^2 = \tilde{\boldsymbol{\alpha}} \mathbf{E} \boldsymbol{\alpha} \tag{5}$$

where $\boldsymbol{\alpha}$ is the vector of the weighting factors α_i, and $\tilde{\boldsymbol{\alpha}}$ is its transpose. In terms of components

$$\sigma^2 = \sum_i \sum_j E_{ij} \alpha_i \alpha_j. \tag{5'}$$

The BLUE technique consists simply of finding the n values of α_i which minimise σ^2, subject to the constraint (4). This can be achieved by using eq. (4) to eliminate one particular α_i (e.g. α_1) and then minimising σ^2 with respect to the remaining $(n-1)$ α's, regarded as independent. Alternatively, we can use the method of Lagrangian multipliers to give

$$\boldsymbol{\alpha} = \mathbf{E}^{-1} U / (\tilde{U} \mathbf{E}^{-1} U), \tag{6}$$

where U is a vector whose n components are all unity, and $\mathbf{E}^{-1}$ is the inverse error matrix.

These values of α_i can then be substituted back into eq. (3) to provide our best value $\hat{y}$, and into eq. (5) to obtain its variance.

This method is equivalent to constructing a weighted sum of squares

$$S = \sum_i \sum_j (y' - y_i)(y' - y_j)(\mathbf{E}^{-1})_{ij}, \tag{7}$$

which measures the extent to which the individual y_i are consistent with a single value y'. We then minimise S with respect to our one parameter y' in order to obtain the best estimate $\hat{y}$ of our physical quantity y.

Thus the final step in the procedure is to use eq. (7) to determine whether our individual estimates are self consistent. We expect $S_{\min}$ to be distributed as χ^2 with $n-1$ degrees of freedom. This then is the procedure to be adopted in judging how well the correlated results from a single experiment agree with each other.

An essential ingredient of the BLUE technique is to have knowledge of the error matrix $\mathbf{E}$ of the individual estimates. In the discussion of applying this method for a specific experiment in section 5, we describe in detail how we use Monte Carlo simulation to obtain an estimate of $\mathbf{E}$. We are thus applying BLUE theory in an approximate sense.

3. Discussion of the method

The above procedure is guaranteed to provide us with an estimate $\hat{y}$ whose variance is smaller than or equal to the smallest variance of the individual estimates y_i. Since these separate estimates are all based on more or less the same set of data, it is hard to avoid the suspicion that we may be in danger of double counting, and hence that the reduction in the variance of $\hat{y}$ is invalid. We show that this is not so by considering two specific examples for the simplest nontrivial case of two measurements y_1 and y_2.

Our first example is the case where the two individual methods happen to be identical. They thus provide estimates which are guaranteed to be the same, and hence the variances are equal and the errors are completely correlated. Thus

$$\mathbf{E} \begin{pmatrix} \sigma_1^2 & \text{cov} \\ \text{cov} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 \end{pmatrix}.$$

Then from eq. (3), our best estimate

$$\begin{aligned} \hat{y} &= \alpha y_1 + (1-\alpha) y_2 \\ &= \alpha y_1 + (1-\alpha) y_1 \\ &= y_1, \end{aligned}$$

with variance

$$\begin{aligned} \sigma^2 &= \alpha^2 \sigma_1^2 + 2\alpha(1-\alpha)\,\text{cov} + (1-\alpha)^2 \sigma_2^2 \\ &= \alpha^2 \sigma_1^2 + 2\alpha(1-\alpha)\sigma_1^2 + (1-\alpha)^2 \sigma_1^2 \\ &= \sigma_1^2. \end{aligned}$$

Thus, not surprisingly and completely satisfactorily, the best value is equal to the individual estimate(s), and the procedure does not produce a spurious improvement in the error estimate.

We now consider the situation where again the identical data sample is used for the two methods, but in the second case there is some additional source of noise

which makes the second estimate worse than the first. Thus the variance for the first estimate is σ_1^2, while that of the second is the somewhat larger σ_2^2. Since the covariance arises because of the identity of the data samples for the two methods and is unaltered by the noise affecting the second estimate, the covariance is still σ_1^2. (This argument is confirmed numerically by the Monte Carlo calculations described below.)

Thus for this case, our error matrix

$$\mathbf{E} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_2^2 \end{pmatrix}, \tag{8}$$

with $\sigma_2 > \sigma_1$. Then

$$\sigma^2 = \alpha^2\sigma_1^2 + 2\alpha(1-\alpha)\sigma_1^2 + (1-\alpha)^2\sigma_2^2 \tag{9}$$

and this minimises for $\alpha = 1$, with $\sigma^2 = \sigma_1^2$. Hence our best estimate is $y_1 \pm \sigma_1$, and for this specific example the lower accuracy measurement is completely ignored. Again we see that including the extra information does not result in an unphysical improvement in the final answer.

After these two rather special examples, we now investigate the properties of our technique for the general case of two measurements. We write the error matrix as

$$\mathbf{E} = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{10}$$

where $\sigma_2 \geq \sigma_1$, and r is the correlation coefficient which must satisfy

$$|r| \leq 1. \tag{11}$$

We scale our measurements so that $\sigma_1 = 1$. As usual our best estimate is

$$\hat{y} = \alpha y_1 + \beta y_2, \tag{12}$$

with

$$\beta = 1 - \alpha. \tag{13}$$

We choose α to minimise the variance on $\hat{y}$.

In fig. 1a we plot β as a function of r for various values of σ_2. Several features are apparent.

(1) For $r = 0$, the weight β is simply $\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, as it should be according to eq. (1) for the uncorrelated situation.

(2) The weight β becomes zero for a correlation coefficient $r = \sigma_1/\sigma_2$. The second case considered above was a specific example of this situation. We can obtain some insight into why the lower accuracy measurement is ignored in this case by considering the situation in which the distribution of results is a 2-dimensional Gaussian in the measurements y_1 and y_2. The effect of the correlation is that, for a given value of y_1, the distribution of y_2 is centred on

$$y_{\text{true}} + r\frac{\sigma_2}{\sigma_1}(y_1 - y_{\text{true}}),$$

rather than on the correct value y_{true} (see fig. 2 and also, for example, p. 60 of ref. [4]). Thus for our specific choice of $r = \sigma_1/\sigma_2$, the lower accuracy measurement is distributed about y_1, and hence adds no new information.

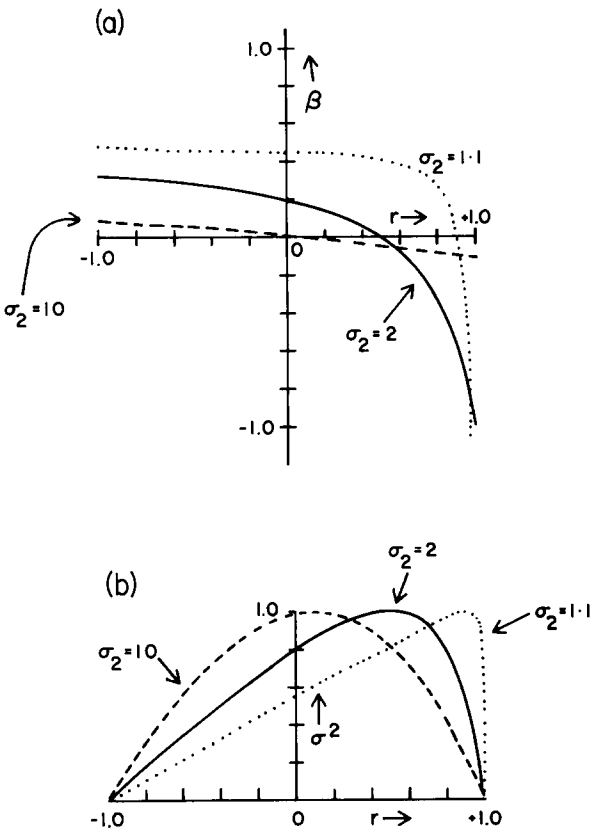

Fig. 1. Graphs showing the properties of the BLUE solution (eq. (12) in the text) for an experiment producing two estimates. (a) The fraction β of the second estimate, as a function of the correlation coefficient r, for three different values of σ_2 ($\sigma_2 = 1.1$ for the dotted curve, 2 for the solid curve, and 10 for the dashed one; the error σ_1 on the first measurement is 1). (b) The variance σ^2 of the BLUE solution as a function of r, for the same values of σ_2 as used in (a).

(3) For a correlation coefficient r larger than σ_1/σ_2, β is negative and α is greater than 1. This apparently surprising situation can be explained as follows. When the correlation between the two estimates of y becomes large and positive, the individual measurements are likely to be on the same side of the true value, with y_1 being closer to it than y_2 is (since $\sigma_1 < \sigma_2$). Thus our best estimate $\hat{y}$ involves extrapolating from y_2 past y_1 (see fig. 2); this implies that α is larger than unity and β is negative. In contrast, a negative correlation always implies interpolation, with both α and β positive and less than 1.

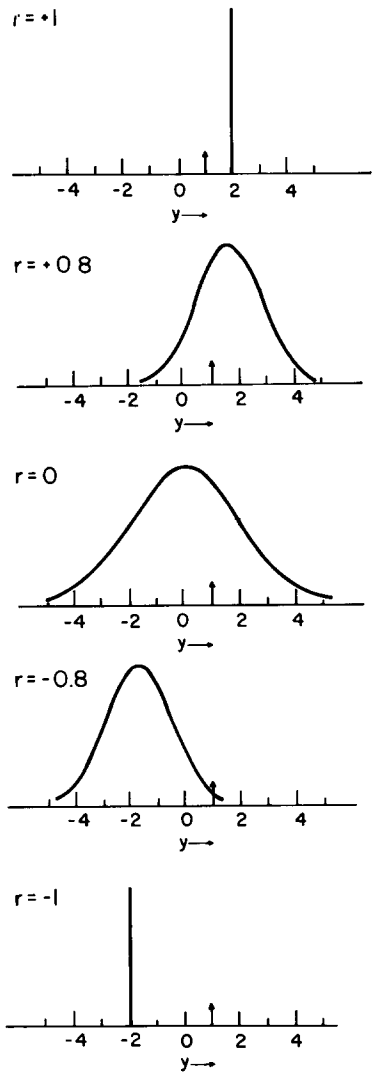

Fig. 2. The distribution of y_2, assuming that the (y_1, y_2) distribution is a 2-dimensional Gaussian centred on the origin with $\sigma_1 = 1$ and $\sigma_2 = 2$, and that y_1 has been measured as 1 (see arrow). The different curves are for various values of the correlation coefficient r, as labelled. For $r = +0.5$ (not shown), the y_2 distribution peaks at 1, and the information it provides does not improve the estimate $\hat{y}$ obtained from y_1 alone. For $r > \sigma_1/\sigma_2$ (e.g. $r = 0.8$ or 1.0), the y_2 distribution is centred further away from the correct value than the measurement y_1. The BLUE approach then involves extrapolating from y_2 beyond y_1, and the fraction α is larger than unity.

Fig. 1b shows the variance σ^2 on our best estimate as a function of r, for the same set of σ_2 as were used in fig. 1a. The main points to note are:

(1) For $r = 0$, our value of σ^2 agrees with that of eq. (2).

(2) For $r = \sigma_1/\sigma_2$, $\sigma^2 = 1$ since we here have $\beta = 0$ (see above), and the second measurement is ignored.

(3) For r close to -1 or very close to $+1$, σ^2 tends to zero. This is because, if the two measurements are completely anticorrelated or completely correlated with known variances, we can reconstruct the true value exactly from the two measurements y_1 and y_2. The only exception to this is the case where $\sigma_2 = \sigma_1$ and $r = +1$ (i.e. the first example considered above) in which we achieve no improvement as compared with each single measurement.

We also note that, while $\beta = 0$ means that the lower accuracy measurement is ignored, a negative weight implies that this particular result does contribute to lowering the variance on the final answer. Thus, for example, for $\sigma_2 = 5$ and $r = 0.8$, the weight $\beta = -1/6$ results in a variance $\sigma^2 = 0.5$ for the best estimate of y, as compared with $\sigma^2 = 1$ if the second measurement is completely ignored.

One point to beware of in practice is the situation in which the individual weights become numerically very large. As seen in fig. 1a, this happens when σ_1/σ_2 is close to $+1$, and r_2 even more so. If we have slightly mis-estimated the elements of the error matrix, or if our measurements are slightly biassed, the effect of the large weights with different signs can be to drive our solution $\hat{y}$ far away from the correct value. For the experiment we describe later, the weights were all reasonable in magnitude, and so this danger does not exist.

Having discussed the performance of the technique for the case of two measurements, we now describe in detail how we apply it to the case of a real experiment involving four correlated measurements of the same quantity.

4. Charm particle lifetime experiment

Data on the production and decay of charmed particles were obtained by exposing the European Hybrid Spectrometer [2] (EHS) to beams of high energy π^- and of protons at the CERN Super Proton Synchrotron (SPS). The essential features of the detector are a bubble chamber which provides a picture of the charm particle production and decay, and a downstream spectrometer from which the momenta of most of the charged and some of the neutral particles can be determined. Some information on the nature of the charged particles (i.e. whether they are electrons, pions, kaons or protons) is also available.

This experiment [3] studied both neutral and charged charmed particles, D° and $D^\pm$ respectively. The essentials of the analysis are very similar for these two types of charmed particles, and for the purposes of this paper we shall restrict our attention to the $D^\pm$. Fig. 3a is a symbolic picture of the information available for a typical $D^\pm$ event.

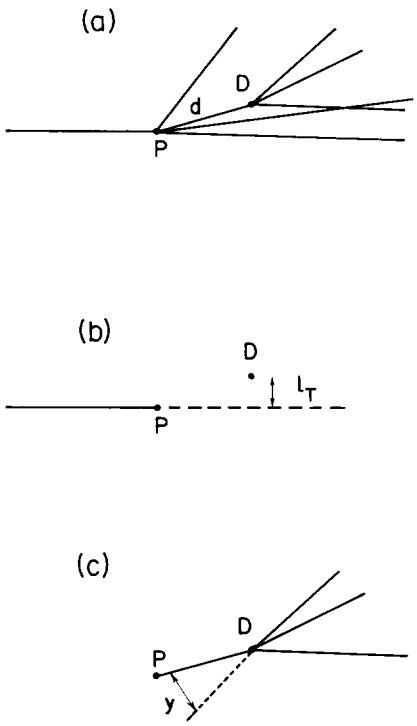

Fig. 3. (a) Schematic diagram showing an event in which a charm particle is produced at P, and decays into three charged particles at D, after travelling a distance d. (b) Part of (a), showing the definition of the transverse length l_T. (c) Part of (a), showing the definition of the impact parameter y for one of the three decay tracks.

We now describe the four different approaches that were used to extract the $D^{\pm}$ lifetime from such events. Further details can be found in ref. [3].

4.1. Fully identified decays

In order to determine the lifetime τ, we would ideally like to know the individual lifetimes t_i of a large set of events; in the absence of visibility and efficiency cuts, the required lifetime is just the average of the individual t_i. From the bubble chamber pictures, we can measure the distance d_i that the ith $D^{\pm}$ travels before it decays. In order to convert this to a time t_i, we need to know the $D^{\pm}$ momentum, or equivalently the vector sum of the momenta of its decay products. For a subsample of our events, all the decay products are measured in our apparatus, and we can check that there are no other missing neutral particles. For these events we can then use the direct method outlined above. In practice, because of the various cuts we impose on our data, we determine τ by a likelihood method, rather than simply as the average of the t_i. We denote this directly determined lifetime as τ_1.

Since the majority of the events are such that we do not know the $D^{\pm}$ momentum, we resort to indirect methods for extracting τ for a larger sample of $D^{\pm}$ particles. We now outline these other three approaches.

4.2. Transverse length method

The transverse length l_T is defined as the distance from the $D^{\pm}$ decay position to the continuation of the incident beam direction (see fig. 3b). For a $D^{\pm}$ produced at a given angle, the longer it lives for the larger will be l_T. The average value of l_T thus provides information on τ. The relationship between them requires knowledge of $\langle p_T \rangle$, the mean transverse momentum of the $D^{\pm}$ mesons; p_T is defined as the component of the $D^{\pm}$ momentum perpendicular to the beam direction. It is an experimental observation that, for a given type of particle, $\langle p_T \rangle$ varies little from one experiment to another. We use a value of $\langle p_T \rangle$ obtained from our experiment, and which is consistent with other similar experiments.

In practice, we determined τ by a likelihood fit to the l_T distribution. We denote this estimate by τ_2.

4.3. Impact parameter method

The impact parameter y of a decay track is defined as the distance of the point at which the $D^{\pm}$ was produced from the backwards extrapolation of that decay track's line of flight (see fig. 3c). The longer the $D^{\pm}$ lives before it decays, the larger will be the impact parameters of its decay tracks. The expected relationship between τ and the mean value of y depends on the branching ratios for the various decay modes of the $D^{\pm}$, but to a good approximation it is independent of the momentum distribution of the $D^{\pm}$; it is determined by a Monte Carlo programme. The experimental value of $\langle y \rangle$ then provides us with our estimate τ_3 of the lifetime.

4.4. Momentum estimator method

If all the decay particles of the $D^{\pm}$ were measured and identified, the $D^{\pm}$ momentum would be known, and we could also calculate its mass, which within experimental errors should agree with the known mass m_D. With missing particles, however, the observed mass and momentum m_i and q_i will both be below their correct values. We then use the observed m_i in order to scale up q_i in order to provide an estimate of the true momentum p_i. Again a Monte Carlo technique is used to help deduce the nature of this procedure. Finally the p_i provide estimates t_i for the individual events as in the method described in section 4.1, and the lifetime τ_4 is deduced by a maximum likelihood method.

Our experiment thus provides us with four estimates of τ. The subsamples of events that we use for the

various determinations overlap (and hence the estimates certainly are not independent). They are not identical, however, because of the more stringent requirements on the available information for some methods as compared with others. In particular, the numbers of events available for the indirect methods are larger than for the direct procedure of section 4.1. It is for this reason that we wish to combine the various estimates, since if the same events were involved in each method, the direct determination τ_1 would be preferable to any of the other estimates (compare the second example in section 3).

We note in passing that correlations can arise even for completely independent data samples, because of common effects in the analysis techniques. Thus for the methods we have described, any uncertainty in the $D^{\pm}$ decay properties would induce correlated errors in the impact parameter and the momentum estimator techniques, because both make use of the $D^{\pm}$'s decay tracks, whereas the other two methods would be almost unaffected. In simpler situations, the error matrix could be deduced directly; for more complicated cases, Monte Carlo techniques could again be used. In the discussion below, we ignore these effects and concentrate on the correlations which arise from the common events in the various analysis samples.

5. Application to charm particle lifetime experiment

The results of the individual methods described in the previous section are as follows [7]

$$\tau_1 = \left(9.5^{+1.7}_{-1.2}\right) \times 10^{-13} \text{ s},$$
$$\tau_2 = \left(11.9^{+1.5}_{-1.3}\right) \times 10^{-13} \text{ s},$$
$$\tau_3 = \left(11.1^{+1.8}_{-1.2}\right) \times 10^{-13} \text{ s},$$
$$\tau_4 = \left(8.9^{+1.6}_{-1.2}\right) \times 10^{-13} \text{ s}. \qquad (14)$$

We are now going to combine these by the technique described in section 2, in order to produce our best estimate $\hat{\tau}$.

The results (14) contain the error estimates on the individual measurements but not their correlations, which arise from the fact that the same events contribute to the separate determinations. Thus if a particular $D^{\pm}$ lives for much longer than the average, it tends to raise each of the estimates to which it contributes. The correlation is not complete, however, for the following reasons:

(1) The data samples for the different determinations are not identical, because of the different selection criteria for the various methods.

(2) Different features of the $D^{\pm}$ are important for the separate methods. Thus τ_2 depends almost entirely on the production characteristics of the D^{+}, and τ_3 on its decay properties.

We have estimated the full error matrix by a Monte Carlo method. We have created a series of 100 "experiments", in which the $D^{\pm}$ are generated in accordance with the known production and decay properties, and are subjected to the same selection criteria as were used for the real data. We also ensured that the numbers of events available for each of the 4 analyses and the numbers which were common to each possible combination of methods were identical to those in the actual data. We then analysed each of these "experiments" in exactly the same way as described in section 4 for the real data. In this way we obtained 100 sets of four lifetime estimates τ_{ik}, where the subscript k denotes the Monte Carlo experiment number and i refers to the method used to determine τ (i.e. i goes from 1 to 4, and k from 1 to 100).

We then calculate the elements of the error matrix $\mathbf{E}$ as

$$E_{ij} = \frac{1}{100} \sum_{k=1}^{100} (\tau_{ik} - \bar{\tau}_i)(\tau_{jk} - \bar{\tau}_j), \qquad (15)$$

where $\bar{\tau}_i$ is the Monte Carlo average for the ith method, i.e.

$$\bar{\tau}_i = \frac{1}{100} \sum_{k=1}^{100} \tau_{ik}. \qquad (16)$$

The above procedure provides us with the following error matrix

$$\mathbf{E} = \begin{pmatrix} 2.66 & 1.15 & 0.86 & 1.31 \\ 1.15 & 1.45 & 0.82 & 1.32 \\ 0.86 & 0.82 & 1.06 & 1.05 \\ 1.31 & 1.32 & 1.05 & 2.56 \end{pmatrix}, \qquad (17)$$

where each element is expressed in units of 10^{-26} s^2. Before we use this as described in section 2, we need to take into account the fact that the diagonal elements of $\mathbf{E}$ as given in eq. (17) are not identical with the squares of the errors of the estimates based on the real data (eqs. (14)). The reasons for this, and the corrective action we take, are as follows:-

(1) Lifetime determinations $\tau \pm \epsilon$ are such that the fractional error ϵ/τ is inversely proportional to the square root of the number of events used. Thus, for a given number of events, if our estimate of τ by one technique is low, then so will be the corresponding error ϵ. We do not want to give this particular method a larger weight because of this statistical fluctuation. The relative weights are there to do justice to a method with a genuinely higher accuracy and not because the estimates of τ and ϵ happen to be too low.

Since we are assuming that all the techniques are measuring the identical parameter, we adjust the errors on the actual data to the values they would have had if the estimates τ_i had all been identical. This value, which does not sensitively affect our answer, we take to be 11×10^{-13} s.

Since the upper and lower errors are not equal, we simply take their geometric mean at this stage. The errors are then 1.7, 1.3, 1.5 and 1.7×10^{-13} s respectively. Other possibilities are described in (3) below.

(2) The errors on the actual data include contributions for possible systematic effects, which differ from method to method. Thus for example, the transverse length estimate τ_2 is sensitive to the mean transverse momentum $\langle p_T \rangle$ as deduced from other data. On the other hand the impact parameter estimate τ_3 depends somewhat on the branching ratios for the various decay modes of the $D^{\pm}$. To a good approximation these types of contributions to the errors are independent of each other and hence the error matrix for these possible systematic effects is essentially diagonal. The overall error matrix is the sum of those for the Monte Carlo estimated random errors and the diagonal systematic ones.

Our corrective procedure is thus simply to leave unchanged the off-diagonal elements of $\mathbf{E}$, but to increase the diagonal elements to be equal to the square of the errors as given in (1) above. We thus obtain

$$\mathbf{E} = \begin{pmatrix} 2.74 & 1.15 & 0.86 & 1.31 \\ 1.15 & 1.67 & 0.82 & 1.32 \\ 0.86 & 0.82 & 2.12 & 1.05 \\ 1.31 & 1.32 & 1.05 & 2.93 \end{pmatrix}. \qquad (17')$$

This error matrix determines the weights of the various estimates as

$$\alpha_1 = 0.14,$$

$$\alpha_2 = 0.47,$$

$$\alpha_3 = 0.35,$$

$$\alpha_4 = 0.04. \qquad (18)$$

Together with the individual results (14), these give a best estimate

$$\tau = (11.2 \pm 1.1) \times {}^{-13} \text{ s}. \qquad (19)$$

The weighted sum of squares (7) is 6.0 for 3 degrees of freedom; the χ^2 probability for exceeding such a value is about 12%. This implies that the consistency of the measurements is satisfactory.

(3) Another feature of lifetime determinations is that the errors, as estimated for example by finding when the likelihood is reduced by a factor of 0.6 from its maximum, tend to be asymmetric – the upper error is larger than the lower one (compare our actual data in eqs. (14); or see ref. [5]). Our method of determining the error matrix necessarily neglects any such asymmetry.

We allow for this by calculating the sum of squares (7) for a series of values of the lifetime τ. As τ changes, we alter the elements of the error matrix such that the errors on individual determinations vary smoothly from σ_1 at $\tau - \sigma_1$ to σ_2 at $\tau + \sigma_2$; the off-diagonal elements are such as to maintain the correlation coefficients as in the matrix (17′). Where the sum of squares (7) increases from its minimum by 1 determines the asymmetric errors as $+1.3$ and -1.0×10^{-13} s.

We regard this as a rather unsatisfactory "patch up", which is not a general procedure. As an alternative, when the goal is to obtain a single overall maximum likelihood estimate, it may be possible to use the method of Dingle and Gratton [6], which involves more computation.

Finally we would like to emphasise that the complications caused by the asymmetric errors are not specific to our problem of combining correlated estimates, but also need to be addressed in the standard situation of combining uncorrelated results by eqs. (1) and (2).

6. Conclusions

We have described a simple method for combining different correlated measurements. The properties of our technique were discussed for the simple-to-understand case of two correlated measurements.

We have applied our method to a real experiment which produced 4 estimates of the lifetime of the charmed meson $D^{\pm}$. This involves using the full 4×4 error matrix for our estimates. We have described how we obtained this error matrix by a Monte Carlo calculation.

With the individual results (in units of 10^{-13} s) of

$9.5^{+1.7}_{-1.2}$, $11.9^{+1.5}_{-1.3}$, $11.1^{+1.8}_{-1.2}$ and $8.9^{+1.6}_{-1.2}$

our technique provided us with a best estimate of

$11.2^{+1.3}_{-1.0}$

and with a satisfactory χ^2 probability that the individual estimates are consistent.

We believe that such a technique can have widespread application in providing the best single number for experiments which produce several different but not independent estimates of a single physical quantity, and for checking their consistency.

Acknowledgements

L.L. and D.G. are grateful to our colleagues of the NA27 collaboration, whose data is used here. The discussions we had with them about the technique described here made us think more about its significance and validity.

References

[1] See, for example, K.V. Mardia, J.T. Kent and J.M. Bibby, Multivariate Analysis (Academic Press, 1979).

[2] M. Aguilar-Benitez et al., Nucl. Instr. and Meth. 205 (1983) 79.
[3] M. Aguiler-Benitez et al., Z. Phys. C34 (1987) 143; and Phys. Lett. B 193 (1987) 140.
[4] L. Lyons, Statistics for Nuclear and Particle Physicists (Cambridge University Press, 1986).
[5] L. Lyons, The errors on maximum likelihood lifetime estimates, Oxford preprint 53/86 unpublished.
[6] P.J. Diggle and R.J. Gratton, J.R. Statis. Soc. B46, no. 2, 193.
[7] D. Gibaut, D. Phil Thesis, Oxford (1986) unpublished.