PRACTICALITIES OF COMBINING ANALYSES: W PHYSICS RESULTS AT LEP

C. Parkes
Department of Physics & Astronomy, Kelvin Building, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

Abstract
Two standard methods of combining results are discussed through case studies from W Physics at LEP: the BLUE method used in the W Mass combination and a likelihood method in the presence of correlated errors that is used for the estimate of gauge couplings. Difficulties encountered and points for consideration in performing these combinations are presented.

1 LEP W Mass Combination

1.1 Introduction

Improving the accuracy of the W Mass measurement was the principal aim of the CERN LEP2 collider measurements programme. Due to the excellent performance of the accelerator and the four LEP detectors the measurement is now dominated by systematic uncertainties. Results are provided by each experiment for each of the five years of data taking. W bosons are pair-produced at LEP and the mass is measured primarily from two WW decay modes: the first is where both W bosons decay hadronically and the other where one decays hadronically and the other leptonically.

1.2 BLUE Method

The BLUE method [1] for the combination of correlated measurements is used for the LEP W Mass determination [2] and several other high profile measurements at LEP2.

The aim is to produce the unbiased linear combination of n measurements, y_i, with the minimum variance. The errors on the measurements are assumed to be symmetric, Gaussian and independent of the value of the measured parameter.

$$\hat{y} = \sum_{i=0}^{n} w_i y_i$$

where w_i is the weight of the measurement i in the combination, and $\sum_{i=0}^{n} w_i = 1$.

It is simple to show that

$$w_i = \frac{\sum_j E_{ij}^{-1}}{\sum_i \sum_j E_{ij}^{-1}},$$

where E_{ij} is the measurements' covariance matrix.

The total error can be calculated, and broken down into its component contributions from each independent error type k

$$(\sigma^k)^2 = \sum_i \sum_j w_i w_j \sigma^{k_i} \sigma^{k_i} \rho^{k_{ij}}.$$

where $\rho^{k_{ij}}$ is the correlation coefficient for measurements i and j arising from error type k.

1.3 Remarks on Usage

The error categories may be defined in two ways, both of which are in common usage: by correlation, or by physical error source. Naturally, these two error division formats are identical for calculational purposes but not for sociological considerations.

In the former, all error sources with the same correlation source are quoted in a combined form. For example in the W Mass measurement, one category would be those error sources that are correlated between the four LEP experiments and between the results of different years but are uncorrelated between the two decay channels in which the measurement is made. Such a format can be problematic, and lead to unphysical situations which are not readily apparent: experiment A may decide that an error source is fully correlated with experiment B, B that it is fully correlated with C, but C that it is uncorrelated with A. The alternative method of specifying the values of each error source and its correlations is to be preferred for its transparency.

In practice, it is rarely necessary to specify the full correlation matrix for each error source: a short-hand form will suffice and aids clarity. In the W Mass combination each experiment specifies for each of fourteen error sources the correlations of measurements in their experiment from a given decay channel, and between the decay channels. The equivalent correlations between the experiments are discussed and, wherever possible, determined through tests.

Listing all the physical error sources has the additional benefit that it ensures that all experiments have considered the full range of error sources and flags any wildly varying estimates as points for further study. The values of the error components σ^k and measurement weights w_i are useful in directing people's time and available CPU power to address the areas that are critical to the final result.

1.4 Example Cases of Error Sources

In the example of a systematic uncertainty dominated measurement such as the W Mass, the estimate of the correlations as well as the error values are critical. However, there are often significant practical difficulties in the estimate of these correlations.

Consider the W Mass case of a measurement error source obtained from the comparison of several phenomenological models of particle hadronization. Sufficient Monte Carlo simulation events must be obtained with each model to ensure that the statistical error on this test is not a dominant error: this requires of order 10^6 minutes of CPU on a 1GHz processor. With these resource requirements, estimating the correlation between two models with several free parameters may simply not be practical. In practice, when other arguments cannot readily be applied, such cases are often resolved by making the most conservative assumptions for the correlations - for the W mass this turns out to be that the measurements are fully correlated, but when performing the cross-check of the difference in mass obtained from a measurement in two WW decay channels this error source should be considered uncorrelated.

When determining the optimal method of treating sources of uncertainty, conflicting pressures may occur depending on the type of estimate that you consider is being performed. The W Mass measurement can be considered both as a point estimate, in the sense that one would like to quote a central value that is the most probable W Mass value, and of an interval with a certain frequentist coverage. The case of the final state interaction errors illustrate why this can raise difficulties.

The decays of the two W bosons in the WW system may be cross-connected when both W's decay to hadronic final states (colour reconnection and Bose-Einstein effects). Many models of these effects may predict no shift in the reconstructed W Mass. A case could be made that the most probable shift in the central value due to an effect, say Bose-Einstein correlations, is zero. For a point estimate this may be the most reasonable procedure. However, it is not possible to make a credible case that no uncertainty should be ascribed to this source as several theoretically justified models predict significant shifts. Hence, the symmetric error could be minimised for the optimal interval estimate by applying half the shift of the most extreme model and quoting the uncertainty around this. Fortunately, in practice, independent experimental measurements of these final state effects are now becoming available which will help to clarify the situation.

1.5 Limitations

The BLUE method requires specified error components that are independent of the fitted value. For example, the theoretical error on the WW cross-section is 0.5% of its value, if significant (this example is not) this could be included through iteration: a similar situation is discussed in [3].

The formulation does not readily allow the treatment of the physically common case of asymmetric error sources. An example of a combination in which this method is unsuitable as the errors are highly non-Gaussian is discussed in the following section.

2 Gauge Coupling Combination

2.1 Introduction

The triple gauge boson couplings between two W bosons and a Z or a photon have been measured at LEP2 [4]. The Z and photon vertex couplings, which are zero in the Standard Model, have also been determined at LEP2. The measurements are still typically statistically dominated but the uncertainties from systematic sources are sufficiently significant that they must be included in the combination in an acceptable manner.

The measurements combine information from two sources. The total cross-section has a parabolic dependence on the gauge coupling parameters, with non Standard Model couplings leading to an increase in the observed cross-section. The differential cross-section also contains gauge coupling information, with the most powerful variable being the angle of the produced W bosons to the incoming electron beam: this information distinguishes between the signs of the couplings.

Simultaneous determinations of the three principal charged coupling parameters are performed. As the measurements are highly correlated, in the event of a discovery such multi-dimensional parameter determinations would be necessary to disentangle the nature of the non-Standard Model coupling structure.

2.2 Likelihood Curves

Due to the quadratic dependence of the total cross-section on the couplings, measuring a cross-section in excess of the Standard Model will typically produce a likelihood which has a double maxima structure as a function of the coupling parameter. An asymmetry will be created in the likelihood, so that one maximum is the preferred solution, from the differential cross-section information.

When estimating the systematic uncertainties, care must be taken that they are not over-estimated due to flips in the preferred maxima. For example a change in detector response calibration may affect the measured differential distribution but not the total cross-section and the position of a given maxima may change only slightly but the preferred maxima may flip to the other state. The systematic uncertainities should be determined simultaneously for all the coupling parameters considered. These vectors of systematic uncertainities $\sigma_{\mathbf{S}}^{\mathbf{j}}$ on measurements j and from sources S are used in the method described below.

In the absence of systematic uncertainties the triple gauge boson coupling parameters may be extracted by the maximum likelihood method from the sum of the individual results log likelihood curves. However, a method was required to include the effect of correlated systematics in the gauge coupling determination.

2.3 Method

The method adopted for this problem of including fully correlated systematic uncertainties in multi-dimensional parameter interval estimation with non-parabolic statistical errors was suggested in [5].

In the standard maximum likelihood method for combining results (j), the required parameters

$$\mathbf{x} \equiv (x_1, x_2, ..., x_{N_{coupl}}),$$

are obtained from the minimum, with respect to $\mathbf{x}$ of the likelihood function

$$-log\mathcal{L}(\mathbf{x}) = \sum_{\mathbf{j}} -\mathbf{log}(\mathcal{L}^{\mathbf{j}}(\mathbf{x}))$$

If we assume that the systematic uncertainties are Gaussian they can be included as parabolic terms in the likelihood curves through the introduction of nuisance parameters

$$\boldsymbol{\Delta} \equiv (\Delta_1, \Delta_2, ..., \Delta_{N_{syst.}}).$$

We define the uncertainties $\sigma^{\mathbf{j}}_{\mathbf{S}}$ as the systematic shifts observed in the parameters $\mathbf{x}$ for measurement j when the source S varies by one standard deviation. For the case, relevant to the gauge couplings, of including a set of independent error sources each of which is fully correlated between the measurements, we get

$$-log\mathcal{L}(\mathbf{x}, \boldsymbol{\Delta}) = \sum_j -log(\mathcal{L}^j(\mathbf{x} - \sum_S \sigma^{\mathbf{j}}_{\mathbf{S}} \Delta_S) + \frac{1}{2}\sum_S (\Delta_S)^2,$$

where the likelihood curves used are those obtained for the statistical and uncorrelated systematics.

The minimisation is then performed with respect to both the required coupling parameters $\mathbf{x}$ and the additional parameters $\boldsymbol{\Delta}$. The resulting uncertainties on $\mathbf{x}$ now take into account all sources of uncertainty, yielding a measurement of the couplings with the error representing statistical and systematic sources.

3 Closing Remarks

The fraction of the time of a ‘combination’ group spent on the mathematical form of the combination or indeed actually performing the combination is minimal. The W Mass case study illustrates this point, the combination is in principle mathematically very straightforward but significant decisions, with considerable impact on the final result, still need to be taken by the combination group.

Furthermore, the aim of an improved parameter estimate is achieved not just through the combination itself but through the increased dialogue between the measurement groups. These considerations suggest that a combination group should be set up early in the life of the measurement and concentrate as much on providing feedback on improvements in the measurement as on the actual combination.

References

[1] L. Lyons, D. Gibaut and P. Clifford, *Nucl. Instrum. Methods* **A 270** (1988) 110

[2] LEP Collaborations and the LEP W Working Group, LEPEWWG/MASS/20 01-02

[3] L. Lyons, A. J. Martin and D. H. Saxon, *Phys. Rev.* **D 41** (1990) 982

[4] LEP Collaborations and the LEP W Working Group, LEPEWWG/TGC/2002-01

[5] J. Alcaraz, *A proposal for the combination of TGC measurements*, L3 Note 2718 (2001)