Covariance Matrix for Unnormalized Moments

S. U. Chung
Physics Department, Brookhaven National Laboratory, Upton, NY 11973

A definition of the covariance matrix for unnormalized moments is given. In order to show consistency of this definition, a derivation is given of the error on a mean from the covariance matrix.

1 Covariance Matrix

Consider a sample of N events for which two functions f and g have been defined. The resulting unnormalized moments, F and G, are defined by

$$F = \sum_i^N f_i \quad \text{and} \quad G = \sum_i^N g_i \tag{1}$$

One can use the notations ϵ_f and ϵ_g to indicate the statistical errors, i.e.

$$\epsilon_f^2 = \overline{(\delta F)^2} = \sum_i^N f_i^2 \quad \text{and} \quad \epsilon_g^2 = \overline{(\delta G)^2} = \sum_i^N g_i^2 \tag{2}$$

The bar over a quantity is meant to signify that one has taken an 'average' or a 'mean' over a large statistical ensemble, each of whose elements consists of a sample of events. The sample of N events under consideration here is an element in the ensemble. The correlation error between the moments F and G is given by

$$E_{fg} = \overline{\delta F \, \delta G} = \sum_i^N f_i \, g_i \tag{3}$$

This defines the general covariance matrix involving any number of unnormalized moments, two of which have been denoted F and G here. It is seen that $\epsilon_f^2 = E_{ff}$ and $\epsilon_g^2 = E_{gg}$.

The errors given in (2) may be considered a generalization of the Poisson distribution, in the sense that, if $g_i = 1$ for all i for example, then one finds

$$\epsilon_1^2 = \overline{(\delta N)^2} = N \tag{4}$$

The correlation error reads

$$\overline{\delta F \, \delta N} = F \tag{5}$$

if $g_i = 1$ for all i in the sample.

2 Derivation of the Error on a Mean

One is now ready to derive an important formula concerning the error on any mean. For the purpose, one may introduce the following notations for the mean of f and of f^2:

$$\bar{f} = \frac{1}{N} \sum_{i}^{N} f_i = \frac{F}{N} \tag{6}$$

$$\overline{f^2} = \frac{1}{N} \sum_{i}^{N} f_i^2 \tag{7}$$

The error on $\bar{f}$ can be calculated by differentiating first

$$\delta\bar{f} = \frac{\delta F}{N} - \frac{F\,\delta N}{N^2} = \frac{1}{N}(\delta F - \bar{f}\delta N) \tag{8}$$

and then squaring the result

$$\overline{(\delta\bar{f})^2} = \frac{1}{N^2}[\overline{(\delta F)^2} + \bar{f}^2\ \overline{(\delta N)^2} - 2\bar{f}\ \overline{\delta F \delta N}] \tag{9}$$

One sees that the error on the mean is given by, substituting (2), (4), (5) and (6),

$$\overline{(\delta\bar{f})^2} = \frac{1}{N}(\overline{f^2} - \bar{f}^2) = \frac{1}{N}\overline{(f - \bar{f})^2} \tag{10}$$

This shows that the error decreases as a function of $1/\sqrt{N}$—a familiar result.

Acknowledgments

This research was supported by the U.S. Department of Energy under contract number DE-AC02-76CH00016.