PARTICLE DATA GROUP NOTES PDG–95–01

14 August 1995

Least Squares Fitting with Correlated Data Errors

Don Groom
Lawrence Berkeley National Laboratory, Berkeley, CA 94720

1. Introduction

Although multivariate Gaussian distributions and least squares fitting with data correlations are discussed in RPP94 and earlier editions, it is less than obvious for the reader to deciper exactly *how* to introduce his data correlations into the formalism. The purpose of this note is to outline, in long form, revisions to Sections 16.3.3 and 17.4 in RPP94. We fill in many steps in the algebra for reference purposes; the actual revision will be more succinct but hopefully not more cryptic.

2. Revisions to Table 16.1

An entry for the multivariate Gaussian distribution should be made just after the Gaussian entry:

Distribution	Probability density function f(variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},S)=\frac{1}{(2\pi)^{n/2}\sqrt{\lvert S\rvert}}$ $\times\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T S^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^T S\boldsymbol{u}\right]$	$\boldsymbol{\mu}$	S_{ij}

3. Revisions to Section 16.3.3 (long form)

The easiest generalization of the characteristic function for one random variable with a Gaussian distribution to many variables is

$$\phi\left(\boldsymbol{x};\boldsymbol{\mu},S\right)=\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u}-\tfrac{1}{2}\boldsymbol{u}^T S\boldsymbol{u}\right]\ . \tag{1}$$

To apply Eq. (12), we first calculate the first and second derivatives:

$$\frac{\partial\phi}{\partial u_j}=\left(i\mu_j-\sum S_{jl}u_l\right)\phi$$

$$\frac{\partial^2\phi}{\partial u_j\partial u_k}=-S_{jk}\phi+\left(i\mu_j-\sum S_{jl}u_l\right)\left(i\mu_k-\sum S_{km}u_m\right)\phi \tag{2}$$

Then

$$i^{-1}\frac{\partial\phi}{\partial u_j}\bigg|_{\boldsymbol{u}=0}=\mu_j=\text{mean for the } j\text{th variable}$$

$$i^{-2}\frac{\partial^2\phi}{\partial u_j\partial u_k}\bigg|_{\boldsymbol{u}=0}=S_{jk}+\mu_j\mu_k=\text{the } ij\text{th moment of the distribution.} \tag{3}$$

By Eq. (13), we see that higher moments vanish.

To make a little more sense out of Eq. (3),

(a) Consider the transformation $\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{\mu}$; in other words, measure the variables about the means. Then it is evident that S_{jk} is the covariance about the mean; that is, $E\big[(x_j - \mu_j)(x_k - \mu_k)\big]$. For example, the variance of the kth variable, σ_{kk}^2 (or σ_k^2) is S_{kk}.

(b) Consider the case when S is diagonal:

$$S = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \sigma_n^2 \end{pmatrix} \tag{4}$$

In this case, the variance for the kth variable is just σ_k^2, and the characteristic function is just the product of c.f.'s for independent Gaussians.

This approach leads to an easy inversion of the c.f.: Let U be an orthogonal matrix which diagonalizes S; *i.e.*, $(U^{-1}SU)_{jk} = S' = \sigma_k^2 \delta_{jk}$. Then if $\boldsymbol{y} \equiv U^{-1}\boldsymbol{x}' = U^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ we see that Eq. (1) becomes

$$\phi\left(\boldsymbol{v}; S'\right) = \exp\left[-\tfrac{1}{2}\boldsymbol{v}^T S' \boldsymbol{v}\right] = \prod \exp\left[-\tfrac{1}{2}\sigma_j^2 v_j^2\right] . \tag{5}$$

From Table 16.1 in RPP94, it is then evident that the corresponding probability density function is the product of Gaussians:

$$f\left(\boldsymbol{y}; S'\right) = \frac{1}{(2\pi)^{n/2} \prod \sigma_1 \sigma_2 \dots \sigma_n} \exp\left[-\tfrac{1}{2}\boldsymbol{y}^T S'^{-1} \boldsymbol{y}\right] \tag{6}$$

The product in the denominator is recognized as the square root of the determinant $|S'|$. It only remains to transform back to the original variables. Since a determinant is invariant under orthogonal transformation, this term becomes $\sqrt{|S|}$, and we have for the probability density function corresponding to Eq. (1)

$$f\left(\boldsymbol{x}; \boldsymbol{\mu}, S\right) = \frac{1}{(2\pi)^{n/2} \sqrt{|S|}} \exp\left[-\tfrac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^T S^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right] \tag{7}$$

It is also worth pointing out that the variance matrices (not their inverses) are additive. For example, suppose that we have three variables, and the first two have both independent statistical errors and a common error, which might be the result of a common baseline with its own statistical errors (variance s^2) which has been subtracted from each. Then

$$S = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} + \begin{pmatrix} s^2 & s^2 & 0 \\ s^2 & s^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{8}$$

This has all the right properties: The variance of the first variable is $\sigma_1^2 + s^2$, and the covariance between the first and second variables is s^2, as required.

If unequal amounts of the common baseline were subtracted from variables 1, 2, and 3—*e.g.*, fractions f_1, f_2, and f_3, then we would have

$$S = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} + \begin{pmatrix} f_1^2 s^2 & f_1 f_2 s^2 & f_1 f_3 s^2 \\ f_1 f_2 s^2 & f_2^2 s^2 & f_2 f_3 s^2 \\ f_1 f_3 s^2 & f_2 f_3 s^2 & f_3^2 s^2 \end{pmatrix} \tag{9}$$

Finally, we note that the covariance matrix S can be related to the correlation matrix (a sort of normalized covariance matrix) defined by Eq. (14). With the definition $\sigma_k^2 \equiv S_{kk}$, we have $\rho_{jk} = S_{jk}/\sigma_j \sigma_k$.

3. Revisions to Section 17.4 (long form)

Non-independent y_i's

Eq. (15) is based on the assumption that the likelihood function is the product of independent Gaussian distributions. More generally, the measured y_i's are not independent, and we must consider them as being from a multivariate distribution with nondiagonal covariance matrix S. As per the discussion in Sec. Normal (= 16.3.3), the generalization of Eq. (15) is

$$\chi^2 = \sum_{ij} [y_i - F(x_i;\boldsymbol{\alpha})]\, S^{-1}_{ij}\, [y_j - F(x_j;\boldsymbol{\alpha})] \ . \tag{10}$$

We again note that the basic quantity is S rather than its inverse, and that it may be constructed by adding pieces containing correlations to the diagonal matrix containing individual statistical errors, described in Sec. Normal (= 16.3.3).

In the case of a fitting function linear in the parameters, one may differentiate χ^2 to find the generalization of Eq. (16), and with the definitions

$$\begin{aligned} g_m &= \sum_{ij} y_i\, f_m(x_j)\, S^{-1}_{ij} \\ \left(V^{-1}_{\widehat{\alpha}}\right)_{mn} &= \sum_i f_n(x_i)\ f_m(x_j)\, S^{-1}_{ij} \end{aligned} \tag{11}$$

solve Eq. (17) for the estimators $\widehat{\boldsymbol{\alpha}}$.

Appendix: Equations from the main text referred to in this Note:

Eq. 16.13, p. 1272 in RPP94:

$$i^{-n} \left.\frac{d^n\phi}{du^n}\right|_{u=0} = \int_{-\infty}^{\infty} x^n f(x)\, dx = \alpha_n \ . \tag{12}$$

Eq. 16.18, p. 1272 in RPP94:

$$\begin{aligned} \kappa_1 &= \alpha_1 \ (= \mu, \text{ the mean}) \\ \kappa_2 &= m_2 \ = \alpha_2 - \alpha_1^2 \ \left(= \sigma^2, \ \text{the variance}\right) \\ \kappa_3 &= m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^2 \ . \end{aligned} \tag{13}$$

Eq. 16.9, p. 1271 in RPP94:

$$\rho_{xy} = E\left[(x-\mu_x)\left(y-\mu_y\right)\right] / \sigma_x\,\sigma_y = \mathrm{Cov}\,(x,y)\,/\sigma_x\,\sigma_y \ , \tag{14}$$

Eq. 17.10, p. 1276 in RPP94:

$$\chi^2 = -2\ln\mathcal{L} + \text{constant} = \sum_1^N \frac{[y_i - F(x_i;\boldsymbol{\alpha})]^2}{\sigma_i^2} \ . \tag{15}$$

Eq. 17.12, p. 1276 in RPP94:

$$-\frac{1}{2}\frac{\partial\chi^2}{\partial\alpha_m} = \sum_i f_m(x_i)\left(\frac{y_i - \sum_n \alpha_n\ f_n(x_i)}{\sigma_i^2}\right)$$

$$= \sum_i \frac{y_i\ f_m(x_i)}{\sigma_i^2} - \sum_n \alpha_n\ \sum_i \frac{f_n(x_i)\ f_m(x_i)}{\sigma_i^2}\ . \tag{16}$$

Eq. 17.14, p. 1277 in RPP94:

$$\widehat{\boldsymbol{\alpha}} = V_{\widehat{\alpha}}\ \boldsymbol{g}\ . \tag{17}$$