Variance Propogation and Extraction Estimation in 2dfDR

m. birchall

June 19, 2015

1 Introduction

This document focuses on the propogation of variance as applied to the 2dfdr fibre extraction process, in particular optimal extraction.

2 Basics of Variance Propogation

Given two variates X and Y of some assumed distribution, it is easy to show, from the basic definition of variance, the following relationships:

$$VAR(\alpha X) = \alpha^2 VAR(X) \tag{1}$$

$$VAR(X + Y) = VAR(X) + VAR(Y) \tag{2}$$

$$VAR(X - Y) = VAR(X) + VAR(Y) \tag{3}$$

And in general:

$$VAR(\alpha X \pm \beta Y) = \alpha^2 VAR(X) + \beta^2 VAR(Y) \tag{4}$$

Extending this to any number of variates and by considering all covariant terms, it is straightforward to show that if a vector of m variates Y is related to a vector of n variates X by the linear relationship

$$Y = AX \tag{5}$$

Then the covariance matrix of Y is related to the covariance matrix of X by

$$\Sigma_{yy} = A\Sigma_{xx}A^T \tag{6}$$

With 2dfdr we are only concerned with the variance terms so the estimated input covariance matrix is the diagonal of the estimated variance of each variate and even though the propogated covariance matrix will not be diagonal in general, it is only the diagonal terms that will be extracted.

3 Linear Least Squares Fitting

The linear least squares problem is the determination of a "best" solution to the overdetermined linear system

$$Ax = y \tag{7}$$

where A is an mxn (m¿n) matrix.

The basic solution to finding a least squares fit is to use calculus to derive a minimum to the sum of the residuals squares. This results in the deterministic system referred to as the Normal Equations:

$$A^T A x = A^T y \tag{8}$$

for which the solution can be expressed as:

$$x = (A^T A)^{-1} A^T y \tag{9}$$

From the covariance propogation formula:

$$\Sigma_{xx} = (A^T A)^{-1} A^T \Sigma_{yy} ((A^T A)^{-1} A^T)^T \tag{10}$$

which simplifies to

$$\Sigma_{xx} = (A^T A)^{-1} A^T \Sigma_{yy} A (A^T A)^{-T} \tag{11}$$

4 Weighted Linear Least Squares Fitting

A weighted linear least squares fit is defined as minimising a weighted sum of the residuals squared, i.e.

$$\sum_i w_i r_i^2 \tag{12}$$

where

$$ri = \sum_k a_{ik} x_k - y_i \tag{13}$$

This is equaivalent to solving the linear overdetermined system

$$\sqrt(W) A x = \sqrt(W) y \tag{14}$$

where W is the diagonal matrix of the weight terms.

For this system the Normal Equations are:

$$A^T W A x = A^T W y \tag{15}$$

for which the solution can be expressed as:

$$x = (A^T W A)^{-1} A^T W y \tag{16}$$

and the covariance propogation formula is

$$\Sigma_{xx} = (A^T W A)^{-1} A^T W \Sigma_{yy} W A (A^T W A)^{-T} \tag{17}$$

5 Using variance recipricals as the weights

In 2dfdr and in particular the extraction process the weights are chosen to be the recipricals of the variance. In this circumstance

$$\Sigma_{yy} W = I \tag{18}$$

and after combining matrix inverse pairs, the covariance propogation formula reduces to

$$\Sigma xx = (A^T W A)^{-T} \tag{19}$$

Which due to symmetry means

$$\Sigma xx = (A^T W A)^{-1} \tag{20}$$

Thus for 2dfdr issues, it is required to estimate the diagonals of this inverse

6 Estimations

The problem of determining the diagonals of the inverse is a well researched problem, however, given that the solution covariance matrix is expected to be diagonal dominant, the simplest approximation for the variance of Y is the recipricals of the diagonal of the above matrix expression.

7 Non Linear Functions

If Z is a non linear of the two variates X and Y, i.e. Z = f(X,Y), then the variance of Z is generally derived from the taylor expansion of the function around the point (E(X),E(Y)), i.e.

$$VAR(Z) = \frac{\partial f}{\partial X}(E(X), E(Y))^2 * VAR(X) + \frac{\partial f}{\partial Y}(E(X), E(Y))^2 * VAR(Y) \tag{21}$$

8 Division of Variates

For the division of variates an approximate formula can be derived from the Talor serise expansion as:

$$VAR(Z) = \frac{VAR(X)}{E(Y)^2} + E(X)^2 \frac{VAR(Y)}{E(Y)^4} \tag{22}$$

9 Multiplication of Variates

For the multiplication of variates an approximate formula can be derived from the Talor serise expansion as:

$$VAR(Z) = E(X)^2 VAR(Y) + E(Y)^2 VAR(X) \tag{23}$$

However, there is an exact formula which is

$$VAR(Z) = E(X)^2 VAR(Y) + E(Y)^2 VAR(X) + VAR(X)VAR(Y) \tag{24}$$

For the most part, the only term of difference is a negligible fraction of the whole but can become more significant for smaller sample sizes.

10 Sample Distribution of Median

The statistics of any specified index of the sorting of a sampled distribution such as the minimum, the maximum, the quartiles and the median is covered in the field of Order Statistics. For the case of Normal distributions, a good introduction is given in "the Covariance Matrix of Normal Order Statistics" by CS Davis and MA Stephens. The main approxiamtion for the median is given by

$$VAR_{median} = (\pi/2N)VAR_{mean} \tag{25}$$