The Fisher Information matrix and cosmological applications

Daniel Burkey

January 2002

1 Introduction

The Fisher information matrix can be used to quantify the information content of a data set with regard to model parameters. The Cramer Rao inequality can be used to predict a minimum variance bound on estimates of parameters from the data set. The Fisher matrix when inverted can, under some assumptions, be shown to give the covariance matrix of the set of model parameters and so can be used to highlight degeneracies between them. Using the Fisher matrix it is possible to optimize the design of large scale surveys in the sense of maximizing their information content. The following notes give a concise introduction to Fisher information analysis and give two examples of its applications involving the forthcoming 6DF galaxy redshift and peculiar velocity surveys.

2 The Fisher Information matrix

2.1 Definition

$$\mathbf{F}_{ij} = \langle -\frac{\partial^2 \log \mathrm{L}(\mathbf{x}|\theta)}{\partial\theta_i \partial\theta_j} \rangle \tag{1}$$

where $\mathbf{x}$ is a vector representing some data set, $\boldsymbol{\theta}$ is a vector representing a set of model parameters and L is the probability distribution function (pdf) of the data set. The angled brackets denote expectation values.

2.2 Cramer Rao inequality

The Cramer Rao inequality places a mathematical limit on the accuracy to which estimates can be made from a data set:

$$\Delta\theta_i \geq (\mathbf{F}_{ii})^{-1/2} \tag{2}$$

if all other parameters are known, and this increases to:

$$\Delta\theta_i \geq (\mathbf{F}^{-1}_{ii})^{1/2} \tag{3}$$

where more than one parameter is being estimated.

2.3 Correlations between model parameters

It can be shown that for a Gaussian Likelihood $\mathrm{L}(\boldsymbol{\theta}|\mathbf{x})$, the inverse of the Fisher matrix is equal to the parameter covariance matrix $\mathbf{M}_{ij} = \langle\theta_i\theta_j\rangle - \mu_i\mu_j$. Clearly if θ_i and θ_j are independent then $M_{ij} = 0$. If there are correlations between the model parameters then the non-diagonal components of the parameter covariance matrix will have non-zero values. By calculating the Fisher matrix and inverting, we can observe the correlations between parameters and find out how they change over different wavelength ranges.

3 An example data set - an overdensity field

3.1 The discrete case

The overdensity δ is defined as $\delta = \frac{\rho-\bar{\rho}}{\bar{\rho}}$. To start off with we can define a data set as a discrete set of Fourier modes of δ: $\mathbf{x} = (\delta(k_1), \delta(k_2), ...\delta(k_N))$. Assuming that the pdf is Gaussian:

$$\mathrm{L}(\mathbf{x}|\theta) = \frac{1}{(2\pi)^{\mathrm{N}/2}\sqrt{\det\mathbf{C}}}\exp -\frac{1}{2}\mathbf{x}\mathbf{C}\mathbf{x}^{\mathbf{t}} \tag{4}$$

where $\mathbf{C}$ is the covariance matrix of the data set $\mathbf{C}_{ij} = \langle\delta(k_i)\delta(k_j)\rangle$. Homogeneity and isotropy dictate that $\mathbf{C}$ is a diagonal matrix. Taking the log of the pdf and differentiating we get the Fisher matrix for this discrete data set:

$$\mathbf{F}_{ij} = \frac{1}{2}\mathrm{Tr}\left(\langle\delta\delta\rangle\frac{\partial^2}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\mathbf{C}^{-1} + \frac{\partial^2}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\log(\mathbf{C})\right). \tag{5}$$

3.2 Moving to the continuous limit

In reality the overdensity field $\delta(k)$ is a continous function. We will use the following definition for the density power spectrum:

$$\langle\delta(\mathbf{k}_1)(\mathbf{k}_2)\rangle = (2\pi)^3\mathrm{P}(\mathrm{k})\delta_\mathrm{D}(\mathbf{k}_1+\mathbf{k}_2). \tag{6}$$

In going to the continuous limit the trace becomes a double integral $\mathrm{Tr} \rightarrow \int \frac{\mathrm{d}^3\mathrm{k}}{(2\pi)^3}\int \mathrm{d}^3\mathrm{r}$ and $\mathbf{C} \rightarrow \mathrm{P}(\mathrm{k})$. So equation (4) becomes:

$$\mathbf{F}_{ij} = \frac{1}{2}\int \frac{d^3k}{(2\pi)^3}\int d^3r\left(\mathrm{P}(\mathrm{k})\frac{\partial^2}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\frac{1}{\mathrm{P}(\mathrm{k})} + \frac{\partial^2}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\log\mathrm{P}(\mathrm{k})\right) \tag{7}$$

which simplifies to

$$\mathbf{F}_{ij} = \frac{1}{2}\int \frac{d^3k}{(2\pi)^3}\int d^3r\left(\frac{\partial\log\mathrm{P}}{\partial\theta_i}\frac{\partial\log\mathrm{P}}{\partial\theta_j}\right). \tag{8}$$

3.3 Including noise

The power spectrum from a real life survey will consist of the sum of a theoretical model and a noise term. Including the noise $\mathrm{P}(\mathrm{k}) \rightarrow \mathrm{P}(\mathrm{k}) + \mathrm{N}(\mathrm{r})$ equation (8) becomes:

$$\mathbf{F}_{ij} = \frac{1}{2}\int \frac{d^3k}{(2\pi)^3}\int d^3r\left(\frac{\partial\log\mathrm{P}}{\partial\theta_i}\frac{\partial\log\mathrm{P}}{\partial\theta_j}\right)\frac{\mathrm{P}^2}{(\mathrm{P}+\mathrm{N})^2}. \tag{9}$$

The shot noise term $\mathrm{N}(\mathrm{r})$ incorporates the survey dependence into the analysis.

4 A more involved example - 2 fields

The analysis becomes more complicated with the inclusion of velocity field information. In this case the data set becomes enlarged:

$$\mathbf{x} = (\delta(k_1), ...\delta(k_N), u(k_1), ...u(k_N)). \tag{10}$$

Tegmark et al 1997 give the following expression for the Fisher matrix of a discrete data set:

$$\mathbf{F}_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\frac{\partial^2\mathbf{C}}{\partial\theta_\mathrm{i}}\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\theta_\mathrm{j}}\right]. \tag{11}$$

This time our data covariance matrix is much larger and involves non-zero off-diagonal elements:

$$\mathbf{C} = \left(\begin{array}{cccccc} \langle\delta(k_1)\delta(k_1)\rangle & \langle\delta(k_1)\delta(k_2)\rangle & ...... & \langle\delta(k_1)u(k_1)\rangle & \langle\delta(k_1)u(k_2)\rangle & ...... \\ \langle\delta(k_2)\delta(k_1)\rangle & \langle\delta(k_2)\delta(k_2)\rangle & ...... & \langle\delta(k_2)u(k_1)\rangle & \langle\delta(k_2)u(k_2)\rangle & ...... \\ ...... & ...... & ...... & ...... & ...... & ...... \\ \langle u(k_1)\delta(k_1)\rangle & \langle u(k_1)\delta(k_2)\rangle & ...... & \langle u(k_1)u(k_1)\rangle & \langle u(k_1)u(k_2)\rangle & ...... \\ \langle u(k_2)\delta(k_1)\rangle & \langle u(k_2)\delta(k_2)\rangle & ...... & \langle u(k_2)u(k_1)\rangle & \langle u(k_2)u(k_2)\rangle & ...... \\ ...... & ...... & ...... & ...... & ...... & ...... \end{array} \right) \tag{12}$$

Modes of different wavelength are uncorrelated and so have zero covariance, however terms such as $\langle\delta(k_1)u(k_1)\rangle$, $\langle\delta(k_2)u(k_2)\rangle$ etc. are non-zero so the covariance matrix looks something like:

$$\mathbf{C} = \left(\begin{array}{cccccc} \langle\delta(k_1)\delta(k_1)\rangle & 0 & ...... & \langle\delta(k_1)u(k_1)\rangle & 0 & ...... \\ 0 & \langle\delta(k_2)\delta(k_2)\rangle & ...... & 0 & \langle\delta(k_2)u(k_2)\rangle & ...... \\ ...... & ...... & ...... & ...... & ...... & ...... \\ \langle u(k_1)\delta(k_1)\rangle & 0 & ...... & \langle u(k_1)u(k_1)\rangle & 0 & ...... \\ 0 & \langle u(k_2)\delta(k_2)\rangle & ...... & 0 & \langle u(k_2)u(k_2)\rangle & ...... \\ ...... & ...... & ...... & ...... & ...... & ...... \end{array} \right) \tag{13}$$

This can be written in terms of four separate (diagonal) covariance matrices $\mathbf{C}_{11}$, $\mathbf{C}_{22}$ and $\mathbf{C}_{12} = \mathbf{C}_{21}$, where

$$\mathbf{C} = \left(\begin{array}{cc} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{array} \right) \tag{14}$$

and $\mathbf{C}_{11} = \mathrm{Cov}(\delta_\mathrm{i}\delta_\mathrm{j})$, $\mathbf{C}_{22} = \mathrm{Cov}(\mathrm{u_i u_j})$ and $\mathbf{C}_{12} = \mathrm{Cov}(\delta_\mathrm{i}\mathrm{u_j})$. Using these definitions equation (11) becomes:

$$\begin{aligned} \mathbf{F}_{ij} = \frac{1}{2}\mathrm{Tr}(&\mathbf{C}_{22}^2\partial_\mathrm{i}\mathbf{C}_{11}\partial_\mathrm{j}\mathbf{C}_{11} + [\mathbf{C}_{11}^2 + \mathbf{C}_{22}^2]\partial_\mathrm{i}\mathbf{C}_{12}\partial_\mathrm{j}\mathbf{C}_{12} + \mathbf{C}_{11}^2\partial_\mathrm{i}\mathbf{C}_{22}\partial_\mathrm{j}\mathbf{C}_{22} \\ &-\mathbf{C}_{11}\mathbf{C}_{12}[\partial_i\mathbf{C}_{12}\partial_j\mathbf{C}_{11} + \partial_i\mathbf{C}_{22}\partial_j\mathbf{C}_{12} + \partial_i\mathbf{C}_{11}\partial_j\mathbf{C}_{12} + \partial_i\mathbf{C}_{12}\partial_j\mathbf{C}_{22}] \\ &-\mathbf{C}_{22}\mathbf{C}_{12}[\partial_i\mathbf{C}_{11}\partial_j\mathbf{C}_{12} + \partial_i\mathbf{C}_{12}\partial_j\mathbf{C}_{11} + \partial_i\mathbf{C}_{22}\partial_j\mathbf{C}_{12} + \partial_i\mathbf{C}_{12}\partial_j\mathbf{C}_{22}] \\ &+\mathbf{C}_{12}^2[\partial_i\mathbf{C}_{11}\partial_j\mathbf{C}_{11} + 2\partial_i\mathbf{C}_{12}\partial_j\mathbf{C}_{12} + \partial_i\mathbf{C}_{22}\partial_j\mathbf{C}_{22}])/(2(\mathbf{C}_{11}\mathbf{C}_{22} - \mathbf{C}_{12}^2)^2) \end{aligned} \tag{15}$$

4.1 Going to the continuous limit

Again we can go to the continuous limit by substituting $\mathrm{Tr} \rightarrow \int \frac{\mathrm{d}^3\mathrm{k}}{(2\pi)^3} \int \mathrm{d}^3\mathrm{r}$ and

$$\mathbf{C}_{11} \rightarrow \mathrm{P}_{\mathrm{gg}}(\mathrm{k}) + \mathrm{N}_{\mathrm{gg}}(\mathrm{r})$$

$$\mathbf{C}_{22} \rightarrow \mathrm{P}_{\mathrm{uu}}(\mathrm{k}) + \mathrm{N}_{\mathrm{uu}}(\mathrm{r})$$

$$\mathbf{C}_{12} \rightarrow \mathrm{P}_{\mathrm{gu}}(\mathrm{k}) = \mathrm{r}_{\mathrm{g}}\sqrt{\mathrm{P}_{\mathrm{gg}}\mathrm{P}_{\mathrm{uu}}} \tag{16}$$

4.2 Linking P_{uu} and P_{gg}

From perturbation theory:

$$\mathrm{P}_{\mathrm{uu}} = \frac{\beta^2\mathrm{H}^2\mu^2}{\mathrm{k}^2}\mathrm{P}_{\mathrm{gg}}. \tag{17}$$

Putting all of this together allows you to calculate the Fisher matrix for the combined data set of density modes and radial velocity modes. This can be inverted to give minimum parameter uncertainties from the Cramer Rao inequality. These best possible error bars may be compared with those without velocity data.

5 Useful References

The Cramer Rao inequality is explained in : Kendall and Stuart, The Advanced theory of Statistics, Vol2 under the heading “minimum variance bound”. The form of the multivariate Gaussian distribution is explained at:

http://mathworld.wolfram.com/GaussianBivariateDistribution.html (this is a bivariate Gaussian the generalization of which is a multivariate) see also Kenney and Keeping, The Mathematics of Statistics. Equation (11) is derived in Tegmark et al (1997) ApJ 480, 22+.