

Review

# Generalisations of Fisher Matrices

Alan Heavens

Imperial Centre for Inference and Cosmology (ICIC), Imperial College London, Blackett Laboratory,  
Prince Consort Road, London SW7 2AZ, UK; a.heavens@imperial.ac.uk

Academic Editor: Takuya Yamano

Received: 3 May 2016; Accepted: 18 June 2016; Published: 22 June 2016

**Abstract:** Fisher matrices play an important role in experimental design and in data analysis. Their primary role is to make predictions for the inference of model parameters—both their errors and covariances. In this short review, I outline a number of extensions to the simple Fisher matrix formalism, covering a number of recent developments in the field. These are: (a) situations where the data (in the form of  $(x, y)$  pairs) have errors in both  $x$  and  $y$ ; (b) modifications to parameter inference in the presence of systematic errors, or through fixing the values of some model parameters; (c) Derivative Approximation for Likelihoods (DALI) - higher-order expansions of the likelihood surface, going beyond the Gaussian shape approximation; (d) extensions of the Fisher-like formalism, to treat model selection problems with Bayesian evidence.

**Keywords:** fisher matrices; statistics; experimental design

---

## 1. Introduction

Fisher information matrices are widely used for making predictions for the errors and covariances of parameter estimates. They characterise the expected shape of the likelihood surface in parameter space, subject to an assumption that the likelihood surface is a multivariate Gaussian when viewed as a function of the model parameters. Diagonal terms are the inverse variances of the parameters, conditional on all others being known, and non-zero off-diagonal terms indicate correlations between inferred parameters. Diagonal terms of the inverse Fisher matrix yield the variances of parameters when all others are marginalised over. The Cramér–Rao inequality shows that the variances deduced from the Fisher matrix are lower bounds.

Fisher matrices have been extensively used in cosmology, where future experiments have been designed in order to deduce as precisely as possible the parameters of the standard cosmological model, so-called  $\Lambda$ CDM (Cold Dark Matter, with a cosmological constant  $\Lambda$ ), and are routinely used to give “figures-of-merit” [1] for the power of each experiment. Normally, these studies are standard applications of Fisher matrix theory, often simplified by an approximation (which is very good for observations of the Early Universe) that the data are Gaussian-distributed.

In this article, I review a number of generalisations of the Fisher matrix approach. In Section 2 the derivation of the Fisher matrix for Gaussian data is sketched out; in Section 3 we consider Fisher matrices for data pairs that have errors in both  $x$  and  $y$ ; in Section 4 we show how Fisher matrices may be used to estimate biases when some parameters are fixed at incorrect values; in Section 5 we explore better approximations for the likelihood surface (“DALI”), from expansions to higher order in derivatives, and in Section 6 we generalise the use of Gaussian likelihood surfaces to model selection and Bayesian evidence.

## 2. Gaussian Fields

In cosmology, one is very often dealing with Gaussian random fields, which are characterised statistically entirely by their mean and covariance. A pedagogical derivation for the Fisher matrix when the data  $\vec{y}$  are Gaussian appears in [2]. The negative log-likelihood  $\mathcal{L} \equiv -\ln L$  is

$$2\mathcal{L}(\vec{\theta}) = \ln \det 2\pi C + (\vec{y} - \vec{\mu})^T C^{-1} (\vec{y} - \vec{\mu}), \tag{1}$$

where in general both the mean vector  $\vec{\mu}$  and the covariance matrix  $C = \langle (\vec{y} - \vec{\mu})(\vec{y} - \vec{\mu})^T \rangle$  depend on the model parameters  $\vec{\theta}$ . If  $\vec{y}$  represents 1-point statistics, such as Fourier coefficients, then typically  $\vec{\mu} = \vec{0}$ , and all the parameter dependence is in  $C$ . If  $\vec{y}$  represents 2-point statistics, then for Gaussian fields they have only approximately a Gaussian distribution, and the analysis is only approximately correct. In this case, the covariance matrix has some parameter dependence through the 4-point function, which for Gaussian fields can be written as products of the 2-point function.

The Gaussian assumption is widely applicable in cosmology, since the quantum fluctuations that are thought to give rise to the density and radiation fields should ensure this, and limits on departures from gaussianity are very tight [3]. Defining the data matrix  $D \equiv (\vec{y} - \vec{\mu})(\vec{y} - \vec{\mu})^T$  and using the matrix identity for positive definite square matrices  $\ln \det C = \text{Tr} \ln C$ , where  $\text{Tr}$  indicates trace, we can re-write Equation (1) as

$$2\mathcal{L} = \text{Tr} \left[ \ln C + C^{-1} D \right]. \tag{2}$$

Using standard comma notation for partial derivatives,  $Z_{,\alpha} = \partial Z / \partial \theta_\alpha$ , and using the matrix identities  $(C^{-1})_{,\alpha} = -C^{-1} C_{,\alpha} C^{-1}$  and  $(\ln C)_{,\alpha} = C^{-1} C_{,\alpha}$ , we find after taking two derivatives and then the expectation value,

$$F_{\alpha\beta} = \langle \mathcal{L}_{,\alpha\beta} \rangle = \frac{1}{2} \text{Tr} [C^{-1} C_{,\alpha} C^{-1} C_{,\beta} + C^{-1} (\vec{\mu}_{,\alpha} \vec{\mu}_{,\beta}^T + \vec{\mu}_{,\beta} \vec{\mu}_{,\alpha}^T)], \tag{3}$$

The great advantage of the Fisher matrix approach is seen in this example: no data (real or simulated) are required to compute the expected log-likelihood surface, only the statistical properties of the data. This can be a big advantage if simulation is computationally expensive.

## 3. Fisher Matrix with Errors in $x$ as Well as $y$

The previous section gives the standard analysis where only the covariance of the  $y$  values is considered. Let us now consider the fairly general case where the data consists of data pairs  $(X, Y)$ , where we have errors in *both*  $X$  and  $Y$ . We can compute the Fisher matrix via the application of a Bayesian hierarchical model, provided that the errors in  $X$  are small (this will be defined later). The full analysis is given in [4].

We assume  $\vec{X}$  and  $\vec{Y}$  are length  $m$  and  $n$  vectors (for data pairs,  $m = n$ , but in fact the analysis is more general), and have Gaussian errors, around true values  $\vec{x}, \vec{y}$ , with a covariance matrix  $C$ , which also allows correlations between  $\vec{X}$  and  $\vec{Y}$ .  $\vec{x}$  and  $\vec{y}$  are not observed, being latent variables, and are essentially nuisance parameters. In fact the  $\vec{y}$  are not independent nuisance parameters as they are assumed to be related precisely to  $\vec{x}$  through a deterministic theoretical model  $\vec{y} = \vec{\mu}(\vec{x})$  (however, a stochastic element could easily be included). Given the observed data,  $\vec{X}, \vec{Y}$ , we seek the posterior  $p(\vec{\theta} | \vec{X}, \vec{Y})$ . With a uniform prior for  $\vec{\theta}$ , this is proportional to the likelihood  $L = p(\vec{X}, \vec{Y} | \vec{\theta})$ . We write this as the marginalised distribution over  $\vec{x}$  and  $\vec{y}$  as

$$\begin{aligned} L &= \int p(\vec{X}, \vec{Y}, \vec{x}, \vec{y} | \vec{\theta}) d\vec{x} d\vec{y} = \int p(\vec{X}, \vec{Y} | \vec{x}, \vec{y}, \vec{\theta}) p(\vec{x}, \vec{y} | \vec{\theta}) d\vec{x} d\vec{y} \\ &= \int p(\vec{X}, \vec{Y} | \vec{x}, \vec{y}, \vec{\theta}) p(\vec{y} | \vec{x}, \vec{\theta}) p(\vec{x} | \vec{\theta}) d\vec{x} d\vec{y}. \end{aligned} \tag{4}$$

A deterministic  $\vec{y}(\vec{x})$  relation gives a delta function,  $p(\vec{y}|\vec{x}, \vec{\theta}) = \delta(\vec{y} - \vec{\mu}(\vec{x}))$ , and assuming a uniform prior for  $\vec{x}$  (a more general prior is considered in [4]), integration over  $\vec{y}$  gives

$$L \propto \int p(\vec{X}, \vec{Y}|\vec{x}, \vec{\mu}(\vec{x}), \vec{\theta}) d\vec{x}. \tag{5}$$

We now assume that the errors in  $\vec{X}$  are small, for which we require that we can truncate at the linear term of the Taylor expansion of  $\vec{\mu}$ :

$$\vec{\mu}(\vec{x}) = \vec{\mu}(\vec{X}) + \mathbf{T}(\vec{X}) (\vec{x} - \vec{X}), \tag{6}$$

where

$$\mathbf{T}_{ij} \equiv \left. \frac{\partial \mu_i}{\partial x_j} \right|_{\vec{x}=\vec{X}}. \tag{7}$$

$\mathbf{T}$  is diagonal for data consisting of  $\vec{X}, \vec{Y}$  pairs.

We assume a multivariate Gaussian for  $p(\vec{X}, \vec{Y}|\vec{x}, \vec{y})$  (independent of  $\vec{\theta}$ ), and write the covariance matrix of the data in block form as

$$\mathbf{C} = \begin{matrix} & \vec{X} & \vec{Y} \\ \begin{matrix} \vec{X} \\ \vec{Y} \end{matrix} & \begin{pmatrix} \mathbf{C}_{XX} & \mathbf{C}_{XY} \\ \mathbf{C}_{XY}^T & \mathbf{C}_{YY} \end{pmatrix} \end{matrix}. \tag{8}$$

Note that  $\mathbf{C}_{XY} \equiv \langle (\vec{X} - \bar{\vec{X}})(\vec{Y} - \bar{\vec{Y}})^T \rangle$  is not symmetrical, nor invertible or even square in general; although  $\mathbf{C}_{XX}$  and  $\mathbf{C}_{YY}$  are. The covariance matrix may include a number of elements, such as intrinsic scatter and measurement noise, with individual covariance matrices adding to give the final  $\mathbf{C}$ . We also assume that the function  $\vec{\mu}(\vec{x})$  is linear across the width of the Gaussian error distribution of  $\vec{x}$ , in which case the likelihood may be integrated analytically, as follows. We write

$$L \propto \int \frac{1}{\sqrt{\det \mathbf{C}}} \exp\left(-\frac{Q}{2}\right) d\vec{x} \tag{9}$$

where  $Q \equiv (\vec{Z} - \vec{z})^T \mathbf{C}^{-1} (\vec{Z} - \vec{z})$ , and  $\vec{z}$  and  $\vec{Z}$  are  $m + n$ -dimensional vectors:  $z_i = x_i$  and  $Z_i = X_i$  for  $i \leq m$ ,  $Z_{m+j} = Y_j$  and  $z_{m+j} = \mu_j(\vec{X}) + [\mathbf{T}(\vec{X})(\vec{x} - \vec{X})]_j$ . The inverse of  $\mathbf{C}$  in block form is

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{G} & -\mathbf{H} \\ -\mathbf{H}^T & \mathbf{E} \end{pmatrix} \tag{10}$$

where

$$\begin{aligned} \mathbf{G} &= \mathbf{C}_{XX}^{-1} + \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY} \mathbf{E} \mathbf{C}_{XY}^T \mathbf{C}_{XX}^{-1} \\ \mathbf{H} &= \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY} \mathbf{E} \\ \mathbf{E} &= (\mathbf{C}_{YY} - \mathbf{C}_{XY}^T \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY})^{-1}. \end{aligned} \tag{11}$$

Defining  $\tilde{\vec{x}} \equiv \vec{X} - \vec{x}$ , and  $\tilde{\vec{Y}} \equiv \vec{Y} - \vec{\mu}(\vec{X})$ , we find that  $Q$  has the quadratic form

$$Q = \tilde{\vec{x}}^T \mathbf{A} \tilde{\vec{x}} - \mathbf{B}^T \tilde{\vec{x}} - \tilde{\vec{x}}^T \mathbf{B} + Q', \tag{12}$$

where

$$\begin{aligned} \mathbf{A} &= \mathbf{G} + \mathbf{T}^T \mathbf{E} \mathbf{T} - \mathbf{H} \mathbf{T} - \mathbf{T}^T \mathbf{H}^T \\ \mathbf{B} &= (\mathbf{H} - \mathbf{T}^T \mathbf{E}) \tilde{\vec{Y}} \equiv \mathbf{P} \tilde{\vec{Y}} \\ Q' &= \tilde{\vec{Y}}^T \mathbf{E} \tilde{\vec{Y}}. \end{aligned} \tag{13}$$

With the definition of  $Q$  in Equation (12), the Gaussian integral of Equation (9) can be performed, using

$$\int \exp\left(-\frac{1}{2}\tilde{x}^T A \tilde{x} + B^T \tilde{x}\right) d\tilde{x} = \frac{(2\pi)^{n/2}}{\sqrt{\det A}} \exp\left(\frac{1}{2}B^T A^{-1}B\right), \quad (14)$$

and noting that  $Q'$  is independent of  $\tilde{x}$ . The likelihood then simplifies to

$$L \propto \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2}\tilde{Y}^T R^{-1}\tilde{Y}\right), \quad (15)$$

where the inverse of the marginal covariance matrix of  $\tilde{Y}$  is  $R^{-1} = E - P^T A^{-1}P$ . This is obtained using the Woodbury formula [5]  $(K + UWV)^{-1} = K^{-1} - K^{-1}U(W^{-1} + VK^{-1}U)^{-1}VK^{-1}$ , giving

$$R = C_{YY} - C_{XY}^T T^T - TC_{XY} + TC_{XX}T^T, \quad (16)$$

This is a key result. We see that this looks just like a standard Gaussian (in terms of data) likelihood, but with the covariance matrix  $C$  ( $C_{YY}$  in our current notation) replaced by  $R$ . Hence to compute the Fisher matrix, we can use the standard formula found in Equation (3) and Equation (15) of [2], and simply replace  $C$  by  $R$ :

$$F_{\alpha\beta} = \frac{1}{2}\text{Tr}\left[R^{-1}R_{,\alpha}R^{-1}R_{,\beta} + R^{-1}(\vec{\mu}_{,\alpha}\vec{\mu}_{,\beta}^T + \vec{\mu}_{,\beta}\vec{\mu}_{,\alpha}^T)\right]. \quad (17)$$

Note that  $R$  depends not only on the standard covariance, but also on the covariance in the independent variable,  $C_{XX}$ , the meta-covariance,  $C_{XY}$ , and the first partial derivatives of the model function  $\vec{\mu}$ . In the case of uncorrelated data pairs, the result reduces to that found in [6]. For the simple case of no correlations between  $\vec{X}$  and  $\vec{Y}$  values  $R = C_{YY} + T^T C_{XX}T$ , and with diagonal covariance matrices  $C_{YY}$  and  $C_{XX}$  we recover the propagation of error result that the variance of  $f \equiv Y - \mu(X)$  for each data point is effectively

$$\sigma_f^2 = \sigma_Y^2 + \mu'(X)^2 \sigma_X^2, \quad (18)$$

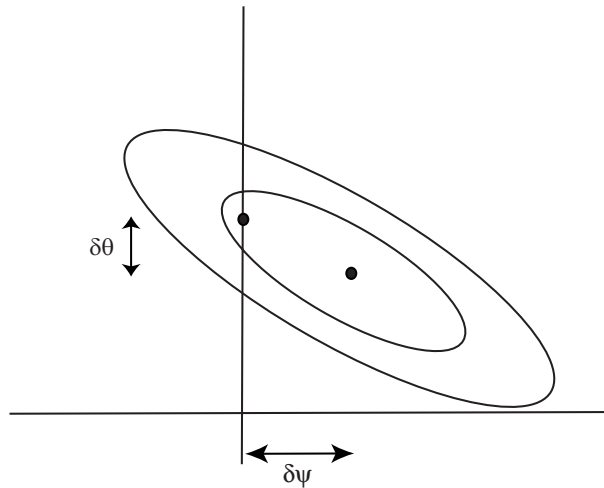
where  $\mu' = \partial\mu/\partial x$  and  $C$  can be replaced in the standard Fisher expression, Equation (3), by a diagonal  $n \times n$  matrix with these enhanced entries.

#### Generalising Still Further

The analysis above is applicable not just to the simple case of data with errors in  $x$  as well as  $y$ , but to any system where the “data”  $\vec{y}$  depend (in a locally linear way) on *any* parameters  $\vec{x}$  that have some error associated with them.

#### 4. Systematic Errors, or Errors from Simplified Nested Models

The Fisher matrix can also be useful to determine the errors in parameter inference that arise if one parameter is fixed at an erroneous value. This could arise in a number of contexts, such as a nuisance parameter (e.g., a calibration setting) being fixed at an incorrect value, or when considering nested models. An example of the latter would be cosmological models where the Universe is assumed to be flat. This is an example of a nested model, being a subset of a more general model, but with the curvature parameter (usually given the symbol  $\Omega_k$ ) set to zero. In these cases, the maximum likelihood values of all the other parameters are, in general, shifted from their maximum likelihood values in the more general model. See Figure 1 for an illustration of this in two dimensions. With the usual Fisher assumption that the likelihood surface is a Gaussian function of the parameters, these shifts can be computed using the Fisher matrix.



**Figure 1.** Illustration of the shift in the maximum likelihood value of a parameter, if another parameter is set at a specific value (e.g., a calibration parameter, or a nested model which does not allow a parameter to vary from some fixed value). The Fisher matrix may be used to determine the shift. Reproduced from Figure 1 of “On model selection forecasting, Dark Energy and modified gravity” published in Mon. Not. Roy. Astron. Soc. [7].

We consider two models,  $M$ , which has more  $(n' + p)$  parameters than a simpler nested model  $M'$ , which has  $n'$ . The extra parameters are designated  $\psi_\zeta$ , and these are fixed in  $M'$  at values that are  $\delta\psi_\zeta$  from their maximum likelihood values in  $M$ . In this case, the maximum likelihood values of all other parameters of  $M'$ ,  $\vec{\theta}$ , are systematically shifted by [7,8]

$$\delta\theta_\alpha = -(\mathbf{F}'^{-1})_{\alpha\beta} \mathbf{G}_{\beta\zeta} \delta\psi_\zeta \quad \alpha, \beta = 1 \dots n', \zeta = 1 \dots p \tag{19}$$

where

$$\mathbf{G}_{\beta\zeta} = \frac{1}{2} \text{Tr} \left[ \mathbf{C}^{-1} \mathbf{C}_{,\beta} \mathbf{C}^{-1} \mathbf{C}_{,\zeta} + \mathbf{C}^{-1} (\vec{\mu}_{,\zeta} \vec{\mu}_{,\beta}^T + \vec{\mu}_{,\beta} \vec{\mu}_{,\zeta}^T) \right], \tag{20}$$

which we recognise as a subset of the Fisher matrix.

### 5. Beyond the Gaussian Approximation—DALI

The Fisher matrix approach assumes that the likelihood surface is a multivariate Gaussian, which will be asymptotically true near the peak, but may not be a good approximation over the range of parameter values of interest. A generalisation of the Fisher matrix is DALI, Derivative Approximation for Likelihoods [9], which expands the likelihood surface to include higher-order derivatives than the second. This is a rather elegant expansion, in derivatives rather than parameters, that ensures that the approximate distribution is a genuine probability distribution—*i.e.*, it is non-negative and normalisable, non-divergent and asymptotically approaches the true likelihood.

The starting point is a Taylor expansion of the likelihood:

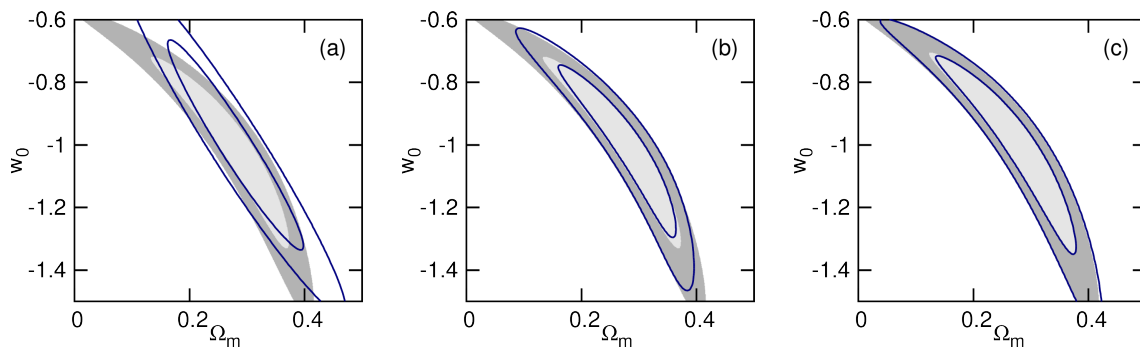
$$\begin{aligned} -\mathcal{L} \equiv \ln L \approx \ln L_0 + \frac{1}{2!} \mathbf{F}_{\alpha\beta} \Delta\theta_\alpha \Delta\theta_\beta + \frac{1}{3!} \mathbf{S}_{\alpha\beta\gamma} \Delta\theta_\alpha \Delta\theta_\beta \Delta\theta_\gamma \\ + \frac{1}{4!} \mathbf{Q}_{\alpha\beta\gamma\delta} \Delta\theta_\alpha \Delta\theta_\beta \Delta\theta_\gamma \Delta\theta_\delta + \dots, \end{aligned} \tag{21}$$

where  $L_0$  is a normalization constant and  $\mathbf{F}_{\alpha\beta} = \mathcal{L}_{,\alpha\beta}$ ,  $\mathbf{S}_{\alpha\beta\gamma} = \mathcal{L}_{,\alpha\beta\gamma}$  and  $\mathbf{Q}_{\alpha\beta\gamma\delta} = \mathcal{L}_{,\alpha\beta\gamma\delta}$ .

If the expansion is arranged in order of derivatives, the expressions are normalisable and positive-definite. For example, to second order in the  $\vec{\mu}$  derivatives, and assuming C is independent of  $\vec{\theta}$ , we have

$$L = L_0 \exp \left[ -\frac{1}{2} \vec{\mu}_{,\alpha}^T C^{-1} \vec{\mu}_{,\beta} \Delta\theta_\alpha \Delta\theta_\beta - \left( \frac{1}{2} \vec{\mu}_{,\alpha\beta}^T C^{-1} \vec{\mu}_{,\gamma} \Delta\theta_\alpha \Delta\theta_\beta \Delta\theta_\gamma + \frac{1}{8} \vec{\mu}_{,\delta\gamma}^T C^{-1} \vec{\mu}_{,\beta\alpha} \Delta\theta_\alpha \Delta\theta_\beta \Delta\theta_\gamma \Delta\theta_\delta \right) + \mathcal{O}(3) \right]. \tag{22}$$

This is apparently true at every order (see [9] for the third-order expansion, and [10] for the case where the parameter dependence is in C). Figure 2 shows the improvement in the expected likelihood surfaces for a supernova cosmology experiment.



**Figure 2.** The extended Fisher likelihood of DALI as applied to a sample of supernovae, as a function of the matter density parameter of the Universe,  $\Omega_m$ , and the equation of state parameter  $w_0 = p/\rho$  of dark energy. From left-to-right this shows the Fisher approximation, the “doublet-DALI” expansion, and the “triplet-DALI” expansion, all compared with the likelihood evaluated on a grid. Reproduced from Figure 1 of “*Breaking the Spell of Gaussianity: Forecasting with Higher Order Fisher Matrices*” published in Mon. Not. Roy. Astron. Soc. [9].

### 6. The Expected Bayesian Evidence—Generalising Fisher Matrices to Model Selection

At the root of the Fisher matrix formalism is the Laplace approximation, *i.e.*, the assumption that the likelihood surface is a multivariate Gaussian when viewed as a function of the model parameters. We can generalise this to the higher-level question of *model selection*, where we compute the posterior probabilities of different models, given the data collected, but regardless of the model parameters  $\vec{\theta}$ . The ratio of these probabilities is the ratio of the prior model probabilities, multiplied by the “Bayes factor”, which is the ratio of the marginal likelihoods (or Bayesian evidence) of the models, where the evidence for a model  $M$  is

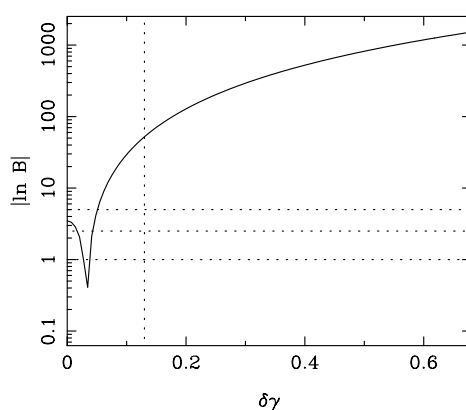
$$p(\vec{y}|M) = \int p(\vec{y}|\vec{\theta}, M) p(\vec{\theta}|M) d\vec{\theta}. \tag{23}$$

With the Laplace approximation for the first, likelihood term, and a uniform prior (which can be generalised to a Gaussian prior), we can compute the expected evidence (conditional on some fiducial set of parameters) by performing Gaussian integrals. For nested models (with  $n'$  and  $n = n' + p$  parameters respectively), the considerations of Section 4 on the locations of the peak likelihood is relevant, and the result depends on the shifts of the fiducial parameters away from the values that are fixed in the lower-dimensional model,  $\delta\psi_\zeta$ . If we further approximate that the expected Bayes factor is the ratio of the expected evidences, then the expected Bayes factor is (see [7] for details)

$$\langle B \rangle = (2\pi)^{-p/2} \frac{\sqrt{\det \bar{F}}}{\sqrt{\det F'}} \exp \left( -\frac{1}{2} \delta\theta_\alpha F_{\alpha\beta} \delta\theta_\beta \right) \prod_{q=1}^p \Delta\theta_{n'+q} \tag{24}$$

where  $\Delta\theta_\alpha$  are the prior ranges of the additional  $p$  parameters in the extended model, and the offsets  $\delta\theta_\alpha$  are given by Equation (19). Note that  $F$  is an  $n \times n$  matrix,  $F'$  is  $n' \times n'$ , and  $G$  is an  $n' \times p$  block of the full  $n \times n$  Fisher matrix  $F$ , given by Equation (20). The expression we find is a specific example of the Savage-Dickey density ratio [11]; here we explicitly use the Laplace approximation to compute the offsets in the parameter estimates which accompany the wrong choice of model.

Figure 3 shows the ratio of expected evidences, assuming the Laplace approximation (as the Fisher matrix does), for nested cosmological models. Details are in the caption, but essentially one parameter is fixed in the simpler model, but allowed to vary in the more complex model. If the more complex model applies, then the data will favour the simpler model if the parameter is close to the fixed value. This is shown in the figure by the cusp in the graph.  $\ln B$  is positive to the left of the cusp, and negative to the right.



**Figure 3.** The ratio of expected evidences  $B$  for two cosmological models. One is based on Einstein Gravity; the other is a more general model where the growth rate of perturbations is allowed to be a free parameter, rather than fixed. The graph shows the ratio as a function of the true shift of the growth rate away from the General Relativity value, for weak lensing data expected from ESA's Euclid satellite. If the growth rate is close to Einstein's (left of the figure;  $\ln B > 0$ ), Bayesian evidence is expected to favour Einstein gravity, whereas if the deviation is large enough (right of the cusp;  $\ln B < 0$ ), it favours the more complex model. Adapted from Figure 2 of "On model selection forecasting, Dark Energy and modified gravity" published in Mon. Not. Roy. Astron. Soc. [7].

## 7. Discussion

This article reviews some recent developments in Fisher matrix theory, which have been motivated by cosmology. The Fisher matrix for data consisting of pairs that have errors in both  $x$  and  $y$  is derived, as a specific example of a general result where the data can depend on arbitrary variables  $x$  that may be measured with some error. The Fisher matrix is shown to be able to determine biases in some parameters when others are set to fixed values (such as in nested models where the simpler model does not allow some parameters to vary). DALI, which goes beyond the Laplace approximation by using higher-order derivatives, is found to allow much more accurate predictions for the expected shape of the likelihood surface. Finally, the concept of expected probabilities in the Laplace approximation is generalised to model selection, by computing the expected Bayesian evidence.

**Acknowledgments:** The author thanks Elena Sellentin for useful discussions about DALI.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Albrecht, A.; Bernstein, G.; Cahn, R.; Freedman, W.L.; Hewitt, J.; Hu, W.; Huth, J.; Kamionkowski, M.; Kolb, E.W.; Knox, L.; *et al.* Report of the Dark Energy Task Force. **2006**, arXiv:astro-ph/0609591.
2. Tegmark, M.; Taylor, A.N.; Heavens, A.F. Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets? *Astrophys. J.* **1997**, doi:10.1086/303939.
3. Ade, P.A.R.; Aghanim, N.; Arnaud, M.; Arroja, F.; Ashdown, M.; Aumont, J.; Baccigalupi, C.; Ballardini, M.; Banday, A.J.; Barreiro, R.B.; *et al.* Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. **2015**, arXiv:1502.01592.
4. Heavens, A.F.; Seikel, M.; Nord, B.D.; Aich, M.; Bouffanais, Y.; Bassett, B.A.; Hobson, M.P. Generalised Fisher Matrices. *Mon. Not. Roy. Astron. Soc.* **2014**, doi:10.1093/mnras/stu1866.
5. Woodbury, M.A. Inverting Modified Matrices (Statistical Research Group, Memorandum Report 42). Available online: <http://www.ams.org/mathscinet-getitem?mr=38136> (accessed on 20 June 2016).
6. March, M.C.; Trotta, R.; Berkes, P.; Starkman, G.D.; Vaudrevange, P.M. Improved constraints on cosmological parameters from SNIa data. *Mon. Not. Roy. Astron. Soc.* **2011**, doi:10.1111/j.1365-2966.2011.19584.x.
7. Heavens, A.F.; Kitching, T.D.; Verde, L. On model selection forecasting, Dark Energy and modified gravity. *Mon. Not. Roy. Astron. Soc.* **2007**, doi:10.1111/j.1365-2966.2007.12134.x.
8. Taylor, A.N.; Kitching, T.D.; Bacon, D.J.; Heavens, A.F. Probing dark energy with the shear-ratio geometric test. *Mon. Not. Roy. Astron. Soc.* **2007**, doi:10.1111/j.1365-2966.2006.11257.x.
9. Sellentin, E.; Quartin, M.; Amendola, L. Breaking the spell of Gaussianity: Forecasting with higher order Fisher matrices. *Mon. Not. Roy. Astron. Soc.* **2014**, *441*, 1831–1840.
10. Sellentin, E. A fast, always positive definite and normalizable approximation of non-Gaussian likelihoods. *Mon. Not. Roy. Astron. Soc.* **2014**, doi:10.1093/mnras/stv1671.
11. Dickey, J.M. The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Stat.* **1971**, *42*, 204–223.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).