10.5 Fisher forecasting

In many situations we want to know the error on our parameters that an experiment can achieve before having taken any data. Theory papers need to estimate whether their effect is observable, and experiments need to be designed to meet specified sensitivity goals. These forecasts are commonly made using the Fisher forecasting formalism (a different option is running MCMC on synthetic data). We first discuss Fisher forecasting for Gaussian likelihoods, but the formalism also generalizes to other likelihoods.

If a given observed variable $O_{\boldsymbol{a}}$ is characterized by Gaussian distributed errors, then its likelihood is

$$\mathcal{L} \propto e^{\chi^2/2}, \tag{10.25}$$

where the χ^2 statistic is defined as:

$$\chi^2 = \sum_{\boldsymbol{a}} \frac{\left[O_{\boldsymbol{a}}(\lambda) - \hat{O}_{\boldsymbol{a}}(\lambda)\right]^2}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]}, \tag{10.26}$$

where $\hat{O}_{\boldsymbol{a}}$ are the measured values of our observable, for example the power spectrum bins $\hat{P}(k_\alpha)$. To find the best fit parameters $\hat{\lambda}$ we minimize χ^2 (which is equivalent to maximizing the likelihood). We assume here that the variance is not parameter dependent and thus we don't need the determinant term in the likelihood.

If we first work in the 1-dimensional case with only one variable λ we can expand the χ^2 around its minimum

$$\chi^2(\lambda) = \chi^2(\bar{\lambda}) + \frac{1}{2} \left.\frac{\partial^2 \chi^2}{\partial \lambda^2}\right|_{\lambda=\bar{\lambda}} (\lambda - \bar{\lambda})^2 \,. \tag{10.27}$$

The linear term vanishes at the minimum. The quadratic term describes the **local curvature of the likelihood**. It tells us how narrow or wide the minimum is, and thus what it's error bar is. If we define

$$\mathcal{F} \equiv \frac{1}{2} \left.\frac{\partial^2 \chi^2}{\partial \lambda^2}\right|_{\lambda=\bar{\lambda}} \,, \tag{10.28}$$

then we can estimate the minimum possible error on λ as $1/\sqrt{F}$. Note that the Fisher matrix depends on where we have assumed the minimum to be, i.e. it depends on the **fiducial parameters** $\bar{\lambda}$ of our forecast.

If we compute $\mathcal{F}$ explicitly we get

$$\mathcal{F}_{\lambda\lambda} = \sum_{\alpha} \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda}\right)^2 + \left(O_{\boldsymbol{a}} - \hat{O}_{\boldsymbol{a}}\right) \frac{\partial^2 O_{\boldsymbol{a}}}{\partial \lambda^2} \right]. \tag{10.29}$$

To forecast F we will not have observed data. Rather we should be taking the expectation value, which simplifies our expression because $\left\langle O_{\boldsymbol{a}} - \hat{O}_{\boldsymbol{a}} \right\rangle = 0$ at the minimum (because the measurements will fluctuate around the truth). Thus

$$F_{\lambda\lambda} = \langle \mathcal{F}_{\lambda\lambda} \rangle \tag{10.30}$$

$$= \sum_{\alpha} \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda}\right)^2 \right] \tag{10.31}$$

This quantity is called the **Fisher Information** F. For several variables, this generalizes to the **Fisher information matrix**:

$$F_{\lambda\lambda'} = \langle \mathcal{F}_{\lambda\lambda} \rangle \tag{10.32}$$

$$= \sum_a \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda} \right) \left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda'} \right) \right] \tag{10.33}$$

If the variables are correlated, the Fisher matrix is

$$F_{\lambda\lambda'} = \sum_{a,b} \left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda} \right) \mathrm{Cov}^{-1}\left(O_{\boldsymbol{a}}, O_{\boldsymbol{b}}\right) \left(\frac{\partial O_{\boldsymbol{b}}}{\partial \lambda'} \right) \tag{10.34}$$

From the Fisher matrix one can obtain two different errors. If we have several parameters and we assume all parameters except λ are known then

$$\sigma_\lambda = \frac{1}{\sqrt{F_{\lambda\lambda}}} \qquad \text{unmarginalized} \tag{10.35}$$

More commonly, we want to know the error on λ if all other parameters are marginalized over. This is obtained by inverting the Fisher matrix as follows

$$\sigma_\lambda = \sqrt{(F^{-1})_{\lambda\lambda}} \qquad \text{marginalized} \tag{10.36}$$

Often the marginalized errors are significantly larger than the unmarginalized ones. An illustration of this in the 2-parameter case is shown in Fig. 13.

10.5.1 Non-Gaussian likelihoods and the Rao-Cramer bound

The Fisher matrix for **any likelihood** (even non-Gaussian ones) is defined as

$$F_{\lambda\lambda'} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda \, \partial \lambda'} \right\rangle \Big|_{\lambda=\hat{\lambda}} \tag{10.37}$$

In general, the Fisher matrix sets a lower bound on the possible error bar, called the **Rao-Cramer bound**. The bound is

$$\sigma_\lambda \geq \frac{1}{\sqrt{F_{\lambda\lambda}}} \qquad \text{unmarginalized} \tag{10.38}$$

$$\sigma_\lambda \geq \sqrt{(F^{-1})_{\lambda\lambda}} \qquad \text{marginalized} \tag{10.39}$$

For maximum likelihood estimators and large enough data sets the Rao Cramer bound is saturated, which is why we wrote an equal sign in the previous section. Some details about the Rao-Cramer bound can be found in Appendix A of 1001.4707. In cosmology we usually assume that the Rao-Cramer bound is saturated in our forecasts.

10.5.2 Priors, subsets, and combining Fisher matrices

If we want to combine the Fisher forecast of two experiments, we can add their Fisher matrices. This has to be done before marginalization over nuisance parameters (unless these are independent for the two experiments). If we want to add a Gaussian prior to a parameter in the Fisher matrix

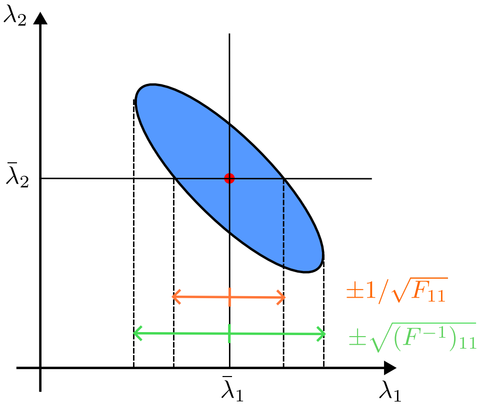

Figure 13. Marginalized and unmarginalized error on parameter λ_1 in a 2 parameter Fisher matrix.

(for example from a different measurements), we add a term to the corresponding diagonal Fisher matrix element

$$F_{\lambda\lambda} \to F_{\lambda\lambda} + \frac{1}{\sigma_\lambda^2}. \tag{10.40}$$

Sometimes we want to marginalize over a subset of the parameters only. This can be done as follows:

- invert F
- remove the rows and columns of parameters we want to marginalize over, to arrive at a smaller matrix which we call G^{-1}
- invert this smaller matrix to get the new Fisher matrix G

A code that helps automatize these operation is pyfisher (`https://pyfisher.readthedocs.io/en/latest/`). Note that numerical inversion of a Fisher matrix can fail if it is not well conditioned, for example due to numerical inaccuracies.

A common practice is to marginalize all but 2 parameters and then plot their **Fisher ellipses**. An illustration is shown in Fig. 13. A review of Fisher forecasting that explains drawing ellipses is given in 0906.4123.

10.5.3 Fisher matrix for a general Gaussian distribution

Above we have given expressions for the Fisher matrix of a Gaussian distribution with a parameter independent covariance matrix. This is not always a correct treatment (in particular not for the likelihood of a Gaussian field as in Sec. 10.3.2). For a general Gaussian

$$L = \frac{1}{(2\pi)^{n/2}\det \mathbf{C}^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\lambda}))^T\mathbf{C}^{-1}(\boldsymbol{\lambda})(\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\lambda}))\right] \tag{10.41}$$

one can show that the Fisher matrix is

$$F_{ij} = \boldsymbol{\mu}_{,i}^T\mathbf{C}^{-1}\boldsymbol{\mu}_{,j} + \frac{1}{2}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}] \tag{10.42}$$