Physics 600 : Fisher Matrices

Nikhil Padmanabhan

2023-09-14

These notes introduce the Fisher matrix formalism that is very common in cosmology for forecasting the statistical reach of an experiment. While I am going to use the problem of constraining cosmological parameters from distance measurements as a motivating example, the formalism is quite general.

Introduction

Review of Homogeneous Expansion

Let us start by writing out the cosmological model we want to consider.

The key starting equation is the first Friedmann equation :

$$H^2 = H_0^2 \left[\Omega_{m,0}(1+z)^3 + \Omega_{DE,0}(1+z)^{3(1+w)} + \Omega_{k,0}(1+z)^2 \right] \equiv H_0^2 E^2(z)$$

where, as an excellent approximation, to low-ish redshifts, we ignore the radiation contribution (Remember that matter-radiation equality is at $z \sim 3000$, so even by $z \sim 1000$ when the CMB last scatters, the matter component is three times the radiation component). We also are assuming a constant equation of state for the dark energy, but this can be generalized using the continuity equation. Also recall that

$$\Omega_{m,0} + \Omega_{DE,0} + \Omega_{k,0} = 1$$

Our simplified cosmological model therefore has four free parameters : H_0, $\Omega_{m,DE}$ and w. At certain times, we might simplify further by limiting only to a cosmological constant ($w = -1$) or a flat Universe ($\Omega_k = 0$).

For simplicity, we will imagine that our observables are the comoving distances as a function of redshift. The comoving distance is

$$\chi = \frac{c}{H_0} \int dz \, \frac{1}{E(z)}$$

These notes make one simplification, and that is to ignore the sin and sinh terms that arise when $\Omega_k \neq 0$. I did this just to simplify algebra/code, but it is easy to add this back in.

The Problem

Suppose an experiment measures the comoving distances χ_i at a set of redshifts z_i with errors σ_i. You might worry that the errors are correlated, and you would be right to worry - you can replace the errors as given above with a covariance matrix C. But I will just assume uncorrelated errors for simplicity here.

The question before us is to estimate how well this experiment would measure the cosmological parameters. If you can answer this, you can optimize the design of your experiment.

Conceptually, this is straightforward. You just need to search the parameter space to find the best set of parameters that fit the data, and then see how the fit degrades as you move off this best set. The methodology outlined here effectively just does this process.

The Likelihood, and the Gaussian Approximation

Above, we discussed finding the set of parameters that best fit the data. How do we define the "best fit"? This is the job of the likelihood function $\mathcal{L}(\mathbf{d}|\mathbf{p})$ where $\mathbf{d}$ is the data (imagined to be organized as a vector) and $\mathbf{p}$ are the model parameters (again organized into a vector). You should read this as the "likelihood of the data given a set of parameters".

Note that this organization of data and parameters as a vector is a convenience, since it allows us to use matrix notation to simplify many steps.

Finding the best parameters is then just the problem of maximizing $\mathcal{L}(\mathbf{d}|\mathbf{p})$ over the space of parameters $\mathbf{p}$.

One needs to choose an appropriate likelihood function. In many cases, the simple choice of a multivariate Gaussian is a good approximation, driven in part by the central limit theorem. We therefore have

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

where   is the model for the data and $\mathbf{C}$ is the covariance matrix of the data. Note that both the model and covariance matrix can be functions of the parameters $\mathbf{p}$.

The Fisher Matrix

The Fisher matrix is defined by the Hessian of the log-likelihood at the peak of the likelihood function

$$F_{ij} = \left\langle -\frac{\partial^2 \ln \mathcal{L}}{\partial p_i \partial p_j} \right\rangle$$

and the inverse of the Fisher matrix gives an estimate of the covariance of parameter errors. This makes intuitive sense; if the curvature of the likelihood surface about the maximum is large, then the errors should be small and vice-versa.

The usefulness of the Fisher matrix comes from the Cramer-Rao bound, which states the the error on parameter p_i is bounded by

$$\sigma_{p_i}^2 \geq (F^{-1})_{ii}$$

when marginalizes over all other parameters and

$$\sigma_{p_i}^2 \geq (F_{ii})^{-1}$$

if holding all parameters fixed. The former case is the one that we consider the most often. We also assume that we saturate the Cramer-Rao bound.

Notice that the Fisher matrix assumes choosing a set of parameters to compute the Hessian about.

For the Gaussian case, we can show that

$$F_{ij} = \mu_{,i}^T \mathbf{C}^{-1} \mu_{,j} + Tr\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\right]$$

where the comma notation represents partial derivatives. A nice derivation of this result is in Tegmark et al 1997.

Constraining Cosmology through Distances : Examples

Let us now consider some examples. For simplicity, we will assume that the errors specified are relative errors, and so, we will consider $\ln\chi$ to be the observable quantity.

Furthermore, we assume that our errors are independent of the cosmological parameters. So in this case, we have

$$F_{ij} = \mu_{,i}^T \mathbf{C}^{-1} \mu_{,j}$$

where

$$\mu(H_0, \Omega_m, \Omega_{DE}, w) = \ln\chi(H_0, \Omega_m, \Omega_{DE}, w)$$

Measuring the Hubble constant

Suppose that the only unknown was H_0. In this case, we see that

$$\mu_{,H_0} = -\frac{1}{H_0}$$

and is independent of redshift. Now suppose the fractional distance error at a single redshift is σ. Then, we see that the error on H_0 is simply

$$\sigma^2_{H_0} = H_0^2 \sigma^2$$

which makes intuitive sense since all distances scale as H_0^{-1}.

Note that the error is independent of what redshift the measurements are made at, as expected.

Unknown calibration

Let us add an additional twist that the observed distances equal the comoving distances but for an unknown constant factor α. This can (and does occur) when the standard candle/ruler has a fixed, but unknown, size.

We have
$$\mu(\ln H_0, \ln \alpha) = \ln \alpha - \ln H_0 + ...$$

where we have dropped terms that are independent of the parameters of interest.

Let us now imagine we have measured this at two redshifts, each with error σ (the redshifts again do not matter).

The Fisher matrix in this case is

$$F = \frac{2}{\sigma^2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

Notice that this matrix has zero determinant, which means that its inverse (the covariance matrix) is not defined. Physically, this just means that the errors on the individual parameters blow up, which is what we would have expected - there is no way to break the degeneracy between α and H_0. Looking at the eigenvectors, we see something similar. We can constrain the combination

$$\ln \alpha - \ln H_0$$

but not

$$\ln \alpha + \ln H_0$$

4

Additional Reading

- Huterer, 10.6
- Tegmark, Taylor, Heavens, 1997, ApJ 480, 1, astro-ph/9603021
- Albrecht et al, arXiv:0901.0721