Fisher Matrix

Lifan Wang

The Fisher Matrix is useful in estimating the errors of a parameter set of an experiment *without* actually knowing or fitting the the parameter values. This is widely used in defining the observational strategies of an experiment.

1 The Basics

Consider first a simple case: suppose we observe a series of quantities y_b, b $\in \{1, \ . \ . \ . \ \mathrm{B}\}$, each of which has Gaussian uncertainties σ_b. Suppose in addition that each observable should be described by a function f_b of some parameters p. The common χ^2 value is (we will assume y_b for different b are uncorrelated)

$$\chi^2 \ = \Sigma_b^B \frac{(f_b(p) - y_b)^2}{\sigma_b^2}.$$

If the parameters p describe the true universe, then the likelihood of a given set of observations is

$$P(y) \propto \exp{(-\frac{\chi^2}{2})}.$$

The goal is to estimate parameters p given a realization of the data y. Using the Bayes' theorem with uniform prior on p, we have $P(p|y) \propto P(y|p)P(p) = P(y|p)$, so that the likelihood of a parameter estimate can be described as a Gaussian with the same χ^2, now viewed as a function of parameters. If we expand about the true values of the parameters, $p^i = p_0^i + \delta p^i$, and average over realizations of the data, and average over the realizations of the data,

$$< \chi^2(p) > = < \chi^2 > \ + \ \left\langle \frac{\partial \chi^2}{\partial p^j} \right\rangle \delta p^j + \frac{1}{2} \left\langle \frac{\partial^2 \chi^2}{\partial p^j \partial p^k} \right\rangle \delta p^j \delta p^k \ + \ ...$$

where the expectation values are taken at the true values p_0. The mean value of observable y_b is indeed $f_b(p_0)$, so the second term vanishes. The distribution of errors in the measured parameters is thus in the limit of high statistics proportional to

$$\exp\left(-\frac{1}{2}\chi^2\right) \propto \exp\left(-\frac{1}{4}\left\langle\frac{\partial^2\chi^2}{\partial p^j\partial p^k}\right\rangle\delta p^j\delta p^k\right) \;=\; \exp\left(-\frac{1}{2}F_{jk}\delta p^j\delta p^k\right),$$

where the Fisher matrix is

$$F_{ij} \;=\; \Sigma_b\frac{1}{\sigma_b^2}\frac{\partial f_b}{\partial p^j}\frac{\partial f_b}{\partial p^k}.$$

From this expression it follows that

$$< \delta p^j\delta p^k > = \;(F^{-1})^{jk},$$

or, the covariance matrix is simply the inverse of the Fisher Matrix (and vice versa).

A more general form of Fisher Matrix with covariant errors is given by

$$F_{ij} \;=\; \Sigma_{ab}\frac{\partial f_a}{\partial p^j}V_{ab}^{-1}\frac{\partial f_b}{\partial p^k}$$

More generally, if one can create a probability $P(p^i|y_b)$ of the model parameters given a set of observed data, e.g., by Bayesian methods, then one can define the Fisher matrix components via

$$F_{ij} \;=\; -\left\langle\frac{\partial^2\ln P}{\partial p^i\partial p^j}\right\rangle$$

and the Cramer-Rao theorem states that any unbiased estimator for the parameters will deliver a covariance matrix on the parameters that is no better than F^{-1}.

2 Prior

A Gaussian prior with width ? can be placed on the i^{th} parameter by adding to the appropriate diagonal element of the Fisher matrix:

$$F_{kl} \;--> \;F_{kl} + \frac{\delta_{ki}\delta_{li}}{\sigma^2}$$

or

$$F \;--> \;F + F^p$$

, with F^p being an extremely simple matrix with a single non-zero diagonal element $1/\sigma^2$ in the i_{th} row and column.

3 Tranformation to new variables

A common operation is to transform the Fisher matrix in some variable set p^i into a constraint on a new variable set q^i. The formula for this is to generate the derivative matrix M with $M_{ij} = \partial p^i / \partial q^j$. In the simple case of Gaussian errors on observables, the new Fisher matrix is

$$F' = M^T F M.$$

If one wants to add a prior on some quantity that is not a single parameter of the Fisher matrix, one can work in variables where it is a single parameter and then make the transformation using the above equation.

4 Marginalization

On many occasions we need to produce a Fisher matrix in a smaller parameter space by marginalizing over the uninteresting ?nuisance? parameters. This amounts to integrating over the nuisance parameters without assuming any additional priors on their values.

Suppose the full parameter vector set is $\vec{p}$, which is a union of two parameter sets: $\vec{p} = \vec{q} \cup \vec{r}$, and we are really only interested in the Fisher matrix for the parameter set $\vec{q}$. The Fisher matrix F' for parameters $\vec{q}$ after marginalization over $\vec{r}$ can be expressed as

$$F' = F_{qq} - F_{qr} F_{rr}^{-1} F_{rq}.$$

Here F_{rr}, F_{qr}, and F_{rq} are submatrices of F.

A common, but numerically unstable procedure, is to

1 invert F,

2 remove the rows and columns, corresponding to $\vec{r}$, that are being marginalized over,

3 then invert the result to obtain the reduced Fisher matrix

This is easy in operation but can fail for ill-conditioned F.

5 Combining observations

When we are combining the constraints from experiments A and B, we will be summing their Fisher matrices. In general any marginalization over nuisance parameters must be done *after* summation of the two Fisher matrices.

If, however, the nuisance parameters of A are disjoint from those of B, then the two data sets have independent probability distributions over the set of nuisance parameters, and it is permissible to marginalize before summation.

6 An Example: Measuring Baryon Acoustic Oscillations with Millions of Supernovae

From Zhan, H., Wang, L., Pinto, P., and Tyson, J. A. (2008), ApJL, 675, 1

Since Type Ia supernovae (SNe) explode in galaxies, they can, in principle, be used as the same tracer of the large-scale structure as their hosts to measure baryon acoustic oscillations (BAOs). To realize this, one must obtain a dense integrated sampling of SNe over a large fraction of the sky, which may only be achievable photometrically with future projects such as the Large Synoptic Survey Telescope. The advantage of SN BAOs is that SNe have more uniform luminosities and more accurate photometric redshifts than galaxies, but the disadvantage is that they are transitory and hard to obtain in large number at high redshift. We find that a half-sky photometric SN survey to redshift z = 0.8 is able to measure the baryon signature in the SN spatial power spectrum. Although dark energy constraints from SN BAOs are weak, they can significantly improve the results from SN luminosity distances of the same data, and the combination of the two is no longer sensitive to cosmic microwave background priors.

Since SNe explode in galaxies, their distribution bears the BAO imprint as well. To measure the SN spatial power spectrum, one needs the angular position and redshift of each SN, not its luminosity. Hence, the SN BAO technique does not suffer from uncertainties in the SN standard candle. Finally, SNe have rich and time-varying spectral features for accurate estimation of photometric redshifts (photo-zs) (Pinto et al. 2004; Wang 2007; Wang et al. 2007), which is helpful for measuring BAOs from a photometric survey.

For the BAO technique to be useful, one must survey a large volume at a sufficient sampling density as uniformly as pos- sible. Although SN events are rare, the spatial density of SNe accumulated over several years will be comparable to the den- sities targeted for future spectroscopic galaxy BAO surveys.

6.1 The sample

We assume two photo-z SN survey models: a shallow one (S20k) that covers 20,000 $\deg^2$ to $z = 0.8$ for 10 years, and a deep one (D2k) that covers 2000 $\deg^2$ to $z = 1.2$ for 5 years.