Notes on Information Matrix

Eric V. Linder

A brief introduction to the information matrix.

I. INTRODUCTION

The information matrix provides a lower limit on parameter estimation uncertainty, taking into account covariances between the dependence of observables on the different parameters. It is an expansion around the maximum likelihood, and so a quadratic (Gaussian) approximation to the likelihood.

In the cases we will be concerned with, the matrix is defined by

$$F_{ij} = \sum_{\rm obs} \frac{\partial O(z)}{\partial p_i}\,{\rm COV}^{-1}(z,z')\,\frac{\partial O(z')}{\partial p_j}\ , \tag{1}$$

where we have a set of parameters $\{p_k\}$, observable quantities (e.g. cosmic distances) $O(z)$, and uncertainties on the measurements of the observables given by a "noise" matrix ${\rm COV}^{-1}$. Note the information matrix is symmetric, and of rank equal to the number of parameters. The derivatives are evaluated at the fiducial model, which must be defined, i.e. the fiducial values of the parameters chosen. While taking the partial derivative, the other parameters are fixed at their fiducial values.

The name information matrix makes sense: we have more information if the sensitivity of an observation with respect to a parameter, $\partial O(z)/\partial p_i$ is large, more if there are many observations entering the sum, and more if the measurement noise is small.

We will deal with cases where the measurements of observations are independent of each other, so the noise matrix is diagonal, ${\rm COV}(z,z') = \sigma^2(z)\delta(z,z')$. This gives the simplified expression

$$F_{ij} = \sum_{\rm obs} \frac{\partial O(z)}{\partial p_i}\,\frac{1}{\sigma^2(z)}\,\frac{\partial O(z)}{\partial p_j}\ . \tag{2}$$

Given the sensitivities (partial derivatives) and the noises $\sigma(z)$, we can compute the information matrix.

If there is more than one type of observable (say lensing distances and supernova distances, or distances and cosmic microwave background data), one simply adds the information matrices coming from each (as long as the measurements are independent):

$$F_{ij}^{\rm total} = \sum_{\rm types} F_{ij}^{{\rm type}\,k}\ . \tag{3}$$

Sometimes one has information directly on a parameter, or combination of parameters. This is known as a prior (as in prior information). A common type is a Gaussian prior, where some parameter p is known to be some value within some uncertainty $\sigma(p)$. For example, $\Omega_m = 0.3 \pm 0.02$. In this case the fiducial is $\Omega_m = 0.3$, and the uncertainty is $\sigma(\Omega_m) = 0.02$. This prior information is added to the information matrix (information adds) by adding $1/\sigma^2(\Omega_m)$ to the entry $F_{\Omega_m\Omega_m}$. The tighter the prior, the more information.

Once we have the information matrix we are ready to do the parameter estimation. At this point you may decide you don't want to include a parameter, that is, you take it to be its fiducial value and don't vary it. This is called fixing the parameter, and is accomplished by removing it from the information matrix – simply deleting its row and column, reducing the rank of the matrix by one. For example, if you have an information matrix for parameters $\{\Omega_m, w_0, w_a\}$ you may decide not to allow time variation of the dark energy equation of state. You would set $w_a = 0$ (or whatever its fiducial value is) and remove its row and column from the information matrix. If you also fix w_0 (e.g. setting it to $w_0 = -1$, the cosmological constant value), then your only parameter is Ω_m. Thus you can reduce your full information matrix to the case of wCDM (constant w) cosmology, and even to ΛCDM (making the assumption that dark energy is the cosmological constant).

However, generally we will use all the parameters we created the information matrix for. At this point, we invert the information matrix to create the covariance matrix $C = F^{-1}$, which is also symmetric. The diagonal entries of C give the uncertainties on each parameter, and the offdiagonal entries give the covariances between parameter estimation:

$$\sigma(p_i) = \sqrt{C_{ii}}\,, \qquad {\rm cov}(p_i, p_j) = C_{ij}\ . \tag{4}$$

The more information in F, the smaller the entries in C, and the smaller the parameter uncertainties.

Because the information matrix is a quadratic approximation, the parameter joint constraints will define an ellipsoid in the N dimensional parameter space. For visualization, we often want to project this down onto two dimensions (parameters), e.g. w_0 and w_a. This projection of the probability density surface is called marginalization, and is equivalent to integrating over the probabilities of all parameters except those you want to remain. To marginalize, we omit the rows and columns of the unvisualized parameters in the covariance matrix C. Note the important difference: to fix parameters we remove them from F, but to marginalize over them we remove them from its inverse C. Generally we want to marginalize.

It is convenient to marginalize over all parameters except two, so as to leave a two dimensional confidence contour (ellipse) for plotting. One has to choose what confidence level to plot (how much probability to include): common choices are 68.3% and 95.4%, which correspond to 1σ and 2σ limits for Gaussian probabilities. Given the marginalized (2x2) covariance matrix and the fiducial parameter values it is straightforward to plot the confidence ellipse. See for example https://arxiv.org/abs/0906.4123.