PHYS 644 Lecture # 25: Fisher, Mapmaking, Foregrounds

Last time we talked about the idea that in measuring parameters from data, we are trying to characterize the posterior distribution of parameters given the data $p(\vec{\theta}|\vec{d})$.

$$\Rightarrow \boxed{\text{Review Slides}}$$

Fisher Matrix Examples

Let's look at some concrete examples. With the Fisher matrix, we are modelling the pdf of parameters $\vec{\theta}$ as Gaussian. However, something that is not always appreciated is that as a function of data $\vec{d}$, the likelihood doesn't have to be Gaussian.

That said, suppose our data is Gaussian distributed, ie

$$\mathcal{L}(\vec{d};\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi C)}}\exp\left[-\frac{1}{2}(\vec{d}-\vec{\mu}(\vec{\theta}))^t C^{-1}(\vec{\theta})(\vec{d}-\vec{\mu}(\vec{\theta}))\right]$$

where $C \equiv \langle(\vec{d}-\vec{\mu})(\vec{d}-\vec{\mu})^t\rangle$ is the covariance of the input data, and $\vec{\mu} \equiv \langle \vec{d} \rangle$ is the ensemble average of the data.

(A little like "ideal" data with no noise or cosmic variance)

Notice that the covariance can depend on the parameters $\vec{\theta}$ too! In our example of measuring the CMB, we're measuring a random field, and the info is contained in the variance (after all, the power spectrum is the variance in spherical harmonic space). Eg for a CMB experiment we might have

$$C = S(\vec{\theta}) + N \leftarrow \text{Instrumental noise}$$

↖ "True" variance of cosmological signal

Hilary

Anyway, if the data is Gaussian distributed then there is a closed form for the Fisher matrix:

$$F_{ij} = \frac{\partial \vec{\mu}^t}{\partial \theta_i} C^{-1} \frac{\partial \vec{\mu}}{\partial \theta_j} + \frac{1}{2} \mathrm{tr}\left[C^{-1} \frac{\partial C}{\partial \theta_i} C^{-1} \frac{\partial C}{\partial \theta_j} \right].$$

Let's apply this to a CMB map. This is a map of anisotropies, so we've already subtracted off the mean and $\vec{\mu} = 0$. If one cranks through the algebra, the Fisher information looks like

$$F_{ij} = \sum_{\ell=2}^{\ell_{max}} \frac{1}{\sigma_\ell^2} \left(\frac{\partial C_\ell}{\partial \theta_i} \right) \left(\frac{\partial C_\ell}{\partial \theta_j} \right)$$

where σ_ℓ is the error bar with which we can measure a particular ℓ mode of the angular power spectrum C_ℓ. It is given by

"Knox formula" in some circles.

$$\sigma_\ell = \sqrt{\frac{2}{f_{sky}(2\ell+1)}} \left[C_\ell + \frac{4\pi \sigma_N^2}{N_{pix}} e^{\theta_b^2 \ell(\ell+1)} \right]$$

pixel noise variance σ_N^2

$\theta_b \equiv$ beam smearing "angular resolution"

Fraction of sky surveyed

Several features:

- The second term is the instrumental noise and notice that it blows up at small scales (high ℓ). And when it really starts blowing up around angular scales $\sim \theta_b$. This is a more subtle (and mathematically precise) way of talking about finite angular resolution.

- Even with no instrumental noise ($\sigma_N^2 = 0$), there's a non-zero error bar on our ability to measure C_ℓ! This is cosmic variance — the fact that with a finite sky we only have so many samples. Notice how because of the $(2\ell+1)^{-1/2}$ bit that this is worst on large scales (small ℓ) where we have fewer samples of the sky.
- The Fisher info has σ_ℓ^{-2}. Makes sense → lower measurement errors means more information content.
- The more the power spectrum changes with a parameter the more Fisher info on that parameter. If C_ℓ doesn't change when that parameter is varied, ie $\frac{\partial C_\ell}{\partial \theta_i} = 0$, there's no info content on that parameter.

This form of the Fisher matrix also teaches us about degeneracies! Let's define a vector $\vec{V}_i$ to be

$$\vec{V}_i = \left(\frac{1}{\sigma_{\ell=2}} \frac{\partial C_{\ell=2}}{\partial \theta_i}, \frac{1}{\sigma_{\ell=3}} \frac{\partial C_{\ell=3}}{\partial \theta_i}, \frac{1}{\sigma_{\ell=4}} \frac{\partial C_{\ell=4}}{\partial \theta_i}, \dots \right)$$

Then:

$$F_{ij} = \vec{V}_i \cdot \vec{V}_j$$

What did we gain from this? Each vector $\vec{V}_i$ is basically the (weighted) derivative of C_ℓ with respect to θ_i.

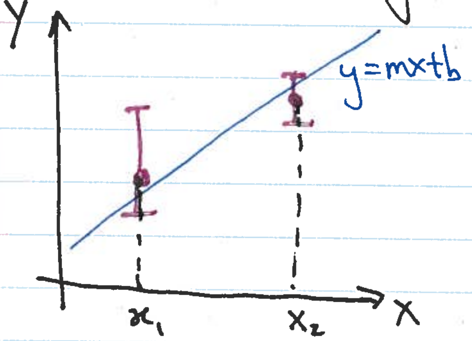
$\Rightarrow$ $\boxed{\text{Show CMB derivatives}}$

The more similar the curves look, the more degenerate the parameters are, because the more changes in one parameter can mimic the other. The off-diagonals are measuring this similarity via the dot product!

Take the extreme: if two parameters have identical vectors, then we have two identical rows in the Fisher matrix! This is the linear algebra recipe for a singular matrix, so the inverse blows up and the marginalized error for one of the parameter becomes infinite (recall $\Delta\theta_i = \sqrt{(F^{-1})_{ii}}$.)

Let's do another application: fitting a straight line.

We measure the "y" values at some predetermined "x" values:



Here, the only randomness is due to the measurement error, so our covariance is

$$C = N$$ ← noise covariance

Eg $N=\sigma^2 \mathbb{1}$ for uncorrelated equal errors

This covariance doesn't depend on our parameters of interest $\vec{\theta} = (~~m,b~~)$ (b,m). Therefore, $\frac{\partial C}{\partial \theta_i} = 0$ and the 2nd term in our Fisher expression vanishes.

For the first term, we need $\vec{\mu} = \langle \vec{d} \rangle$. Here,

$$\vec{d} = \begin{pmatrix} mx_1 + b \cancel{+n_1} \\ mx_2 + b \cancel{+n_2} \\ \vdots \end{pmatrix} + \vec{n}$$ ← noise realization

where $\langle \vec{n} \rangle = 0$ and $N = \langle \vec{n}\vec{n}^t \rangle$.

Then $$\vec{\mu} = \langle \vec{d} \rangle = \begin{pmatrix} mx_1+b \\ mx_2+b \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \end{pmatrix}}_{\equiv A} \underbrace{\begin{pmatrix} b \\ m \end{pmatrix}}_{\equiv \vec{\theta}}$$

(Same notation for later)

We can now evaluate the $\partial\vec{\mu}/\partial\theta_i$ that we need. Explicitly, $\frac{\partial\vec{\mu}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$ and $\frac{\partial\vec{\mu}}{\partial m} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix}$

If we assume $N = \sigma^2 \mathbb{1}$ to make our lives easier for this example, then

$$F_{bb} = (1, 1, \ldots) \, \sigma^{-2} \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix} = \sigma^{-2} N_{pts}.$$

$$F_{bm} = (1, 1, \ldots) \, \sigma^{-2} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \sigma^{-2} \sum_{i=1}^{N_{pts}} x_i$$

$$F_{mm} = (x_1, x_2, \ldots) \, \sigma^{-2} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \sigma^{-2} \sum_{i=1}^{N_{pts}} x_i^2$$

$$\Rightarrow \quad F = \begin{array}{c|cc} \theta = & b & m \\ \hline b & N_{pts} & \sum_{i=1}^{N_{pts}} x_i \\ m & \sum x_i & \sum x_i^2 \end{array} \; \sigma^{-2}$$

One can invert this, to get the theoretically attainable error bars on m and b.

The Fisher matrix can also be written more generally for this problem as

$$\frac{\partial\vec{\mu}}{\partial\vec{\theta}} = A \quad \Rightarrow \quad F = A^t N^{-1} A$$

Now, this tells us how good our error bars can be, without telling us how to achieve this optimal analysis!

Our likelihood is

$$\mathcal{L}(\vec{d};\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left[-\frac{1}{2}(\vec{d}-\overset{\substack{\text{This was } \vec{\mu} \text{ before}\\ \downarrow}}{A\vec{\theta}})^t N^{-1}(\vec{d}-A\vec{\theta})\right]$$

Maybe the max likelihood solution is the optimal one? Maximizing the likelihood is equivalent to minimizing

$$\chi^2 \equiv (\vec{d}-A\vec{\theta})^t N^{-1}(\vec{d}-A\vec{\theta})$$

This is precisely ~~$= \sum_{ij}(d_i - \sum_k A_{ik}\theta_k$~~ the χ^2 we normally try to minimize in a least-squares fit. The resulting best fit $\hat{\theta}$ for $\vec{\theta}$ is:

$$\hat{\theta} = [A^t N^{-1} A]^{-1} A^t N^{-1}\vec{d}.$$

What is the covariance of this solution?

$$\mathrm{Cov}(\hat{\theta}) \equiv \langle \hat{\theta}\hat{\theta}^t\rangle - \langle\hat{\theta}\rangle\langle\hat{\theta}\rangle^t = \langle(\hat{\theta}-\langle\hat{\theta}\rangle)(\hat{\theta}-\langle\hat{\theta}\rangle)^t\rangle$$

Now, $\langle\hat{\theta}\rangle = (A^tN^{-1}A)^{-1}A^tN^{-1}(A\vec{\theta} + \underset{\overset{\shortparallel}{0}}{\langle\vec{n}\rangle}) = \vec{\theta}$

$$\Rightarrow \hat{\theta} - \langle\hat{\theta}\rangle = (A^tN^{-1}A)^{-1}A^tN^{-1}\underbrace{(A\vec{\theta}+\vec{n})}_{\vec{d}} - \vec{\theta} = (A^tN^{-1}A)^{-1}A^tN^{-1}\vec{n}$$

Therefore, $\mathrm{Cov}(\hat{\theta}) = (A^tN^{-1}A)^{-1}A^tN^{-1}\underbrace{\langle\vec{n}\vec{n}^t\rangle}_{N}N^{-1}A(A^tN^{-1}A)^{-1}$

$$= (A^tN^{-1}A)^{-1}$$

This is the inverse of the Fisher matrix! This is only the case for optimal analyses, so we have proved that this solution gives unbeatable error bars!

Why did I waste so much time on fitting a straight line?

$\Rightarrow \boxed{\text{Slide Q}}$

We actually use this to make maps of the CMB! The basic operations of a telescope, like the blurring/convolution due to finite angular resolution is a linear operation.

Fundamentally because Maxwell's Equations are linear.

When we have time-ordered data and want to turn it into a map, the parameters we want to infer are the map pixel values:

$$\vec{\theta} = \left(T(\hat{r}_1), T(\hat{r}_2), \ldots \right)$$

and $\vec{d} = \left(d(t_1), d(t_2), \ldots \right)$

Time-ordered data as telescopes scans sky.

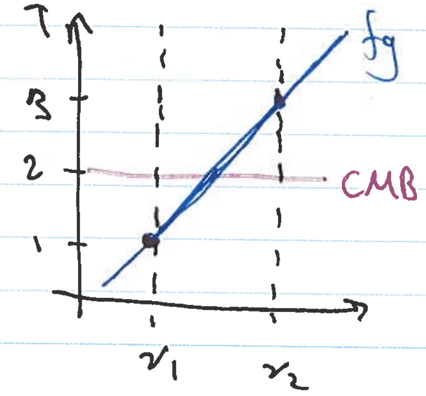
$$\vec{d} = A\vec{\theta} + \vec{n} \; ! \quad \text{Then } \hat{\theta} = (A^t N^{-1} A)^{-1} A^t N^{-1} \vec{d} .$$

Details of observation, angular res. etc.

Now suppose I make a raw map of CMB data....

$\Rightarrow \boxed{\text{Foregrounds and FIRAS slides and Q}}$

How do we get rid of foregrounds? We can take linear combinations of different frequencies:



Suppose I did $T = w_1 T(\nu_1) + w_2 T(\nu_2)$

Preserve the CMB if $w_1 + w_2 = 1$

and get rid of foregrounds if $w_1 + 3w_2 = 0$

$\Rightarrow w_1 = \frac{3}{2}$ and $w_2 = \frac{-1}{2}$

This required knowing the foregrounds ahead of time. If we don't want this, we can just minimize the variance in the maps. Any residual foregrounds add to the variance.

Can also do this in harmonic space, l by l:

$$a_{\ell m} \equiv \sum_{i=1}^{N_{freq}} w_\ell^i \, a_{\ell m}(\nu_i)$$

Minimize $\langle a_{\ell m}^2 \rangle$ subject to the constraint $\sum_i w_\ell^i = 1$

$\Rightarrow \boxed{\text{Show weights and cleaned maps}}$