Astro 448

Set 2: Theoretical Data Analysis

Wayne Hu

Time Ordered Data

- Beyond idealizations like $|\Theta_{\ell m}|^2$ type C_ℓ estimators and $f_{\rm sky}$ mode counting, basic aspects of data analysis are useful even for theorists
- Starting point is a string of "time ordered" data coming out of the instrument (post removal of systematic errors, data cuts)
- Begin with a model of the time ordered data as (implicit summation or matrix operation)

$$d_t = P_{ti}\Theta_i + n_t$$

  where i denotes pixelized positions indexed by i, d_t is the data in a time ordered stream indexed by t e.g. number of time ordered data numbers up to 10^{10} whereas number of pixels $10^6 - 10^7$.
- Noise n_t is drawn from distribution with known power spectrum

$$\langle n_t n_{t'} \rangle = C_{d,tt'}$$

Design Matrix

- The design, pointing or projection matrix $\mathbf{P}$ is the mapping between pixel space and the time ordered data
- Simplest incarnation: row with all zeros except one column which just says what point in the sky the telescope is pointing at that time

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 1 & \ldots & 0 \end{pmatrix}$$

- If each pixel were only measured once in this way then the estimator of the map would just be the inverse of $\mathbf{P}$
- More generally encorporates differencing, beam, rotation (for polarization) and unequal coverage of pixels

Maximum Likelihood Mapmaking

- What is the best estimator of the underlying map Θ_i?
- Likelihood function: the probability of getting the data given the theory $\mathcal{L} \equiv P[\mathrm{data}|\mathrm{theory}]$. In this case, the *theory* is the set of parameters Θ_i.

$$\mathcal{L}_\Theta(d_t) = \frac{1}{(2\pi)^{N_t/2}\sqrt{\det \mathbf{C}_d}} \exp\left[-\frac{1}{2}\left(d_t - P_{ti}\Theta_i\right) C^{-1}_{d,tt'}\left(d_{t'} - P_{t'j}\Theta_j\right)\right] .$$

- Bayes theorem says that $P[\Theta_i|d_t]$, the probability that the temperatures are equal to Θ_i given the data, is proportional to the likelihood function times a *prior* $P(\Theta_i)$, taken to be uniform

$$P[\Theta_i|d_t] \propto P[d_t|\Theta_i] \equiv \mathcal{L}_\Theta(d_t)$$

Maximum Likelihood Mapmaking

- Maximizing the likelihood of Θ_i is simple since the log-likelihood is quadratic.
- Differentiating the argument of the exponential with respect to Θ_i and setting to zero leads immediately to the estimator

$$\hat{\Theta}_i = C_{N,ij} P_{jt} C^{-1}_{d,tt'} d_{t'} \ ,$$

where $\mathbf{C}_N \equiv (\mathbf{P}^{\mathrm{tr}} \mathbf{C}_d^{-1} \mathbf{P})^{-1}$ is the covariance of the estimator

- Given the large dimension of the time ordered data, direct matrix manipulation is unfeasible. A key simplifying assumption is the stationarity of the noise, that $C_{d,tt'}$ depends only on $t - t'$ (temporal statistical homogeneity)

Foregrounds

- Maximum likelihood mapmaking can be applied to the time streams of multiple observations frequencies N_ν and hence obtain multiple maps
- A cleaned CMB map can be obtained by modeling the maps as

$$\hat{\Theta}_i^\nu = A_i^\nu \Theta_i + n_i^\nu + f_i^\nu$$

  where $A_i^\nu = 1$ if all the maps are at the same resolution (otherwise, embed the beam as in the pointing matrix; f_i^ν is the noise contributed by the foregrounds
- Again, a map making problem. Given a covariance matrix for foregrounds noise (a prior from other data), same solution. Alternately, can derive weights from stats of the recovered maps
- 5 foregrounds: synchrotron, free-free, radio pt sources, at low frequencies and dust and IR pt sources at high frequencies.

Power Spectrum

- The next step in the chain of inference is the power spectrum extraction. Here the correlation between pixels is modelled through the power spectrum

$$C_{S,ij} \equiv \langle \Theta_i \Theta_j \rangle = \sum_\ell \frac{\ell(\ell+1)C_\ell}{2\pi} W_{\ell,ij}$$

- W_ℓ, the window function, is derived by writing down the expansion of $\Theta(\hat{\mathbf{n}})$ in harmonic space, including smoothing by the beam and pixelization

- For example in the simple case of a gaussian beam of width σ it is proportional to the Legendre polynomial $P_\ell(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j)$ for the pixel separation multiplied by $b_\ell^2 \propto e^{-\ell(\ell+1)\sigma^2}$

Bandpowers

- In principle the underlying theory to extract from pixel data is the power spectrum at every ℓ
- However with a finite patch of sky, multipoles separated by $\Delta\ell < 2\pi/L$ where L is the dimension of the survey will fully-covary and not supply independent information
- So consider instead a theory parameterization of $\ell(\ell+1)C_\ell/2\pi$ constant in bands of $\Delta\ell$ chosen to match the survey forming a set of bandpowers B_a
- The likelihood of the bandpowers given the pixelized data is

$$\mathcal{L}_B(\Theta_i) = \frac{1}{(2\pi)^{N_p/2}\sqrt{\det \mathbf{C}_\Theta}} \exp\left(-\frac{1}{2}\Theta_i C^{-1}_{\Theta,ij}\Theta_j\right)$$

where $\mathbf{C}_\Theta = \mathbf{C}_S + \mathbf{C}_N$ and N_p is the number of pixels in the map.

Bandpower Estmation

- As before, $\mathcal{L}_B$ is Gaussian in the anisotropies Θ_i, but in this case Θ_i are *not* the parameters to be determined; the theoretical parameters are the B_a, upon which the covariance matrix depends.
- The likelihood function is not Gaussian in the parameters, and there is no simple, analytic way to find the maximum likelihood bandpowers or their covariance
- In principle one can still use Bayes' Theorem to find the posterior joint probability of the bandpowers or the cosmological parameters that parameterize them
- In practice the exact likelihood is expensive to compute
- Need fast approximation to the likelihood function and a fast way of exploring it

Bandpower Estmation

- One example is to find the maximum likelihood bandpowers by iteration
- Take a trial point $B_a^{(0)}$ and improve estimate based a Newton-Rhapson approach to finding zeros

$$
\begin{aligned}
\hat{B}_a &= \hat{B}_a^{(0)} + \hat{F}_{B,ab}^{-1} \frac{\partial \ln \mathcal{L}_B}{\partial B_b} \\
&= \hat{B}_a^{(0)} + \frac{1}{2} \hat{F}_{B,ab}^{-1} \left(\Theta_i C_{\Theta,ij}^{-1} \frac{\partial C_{\Theta,jk}}{\partial B_b} C_{\Theta,kl}^{-1} \Theta_l - C_{\Theta,ij}^{-1} \frac{\partial C_{\Theta,ji}}{\partial B_b} \right) ,
\end{aligned}
$$

- Still need the covariance matrix of the bandpowers

Fisher Matrix

- The expectation value of the local curvature is the Fisher matrix

$$F_{B,ab} \equiv \left\langle -\frac{\partial^2 \ln \mathcal{L}_B}{\partial B_a \partial B_b} \right\rangle = \frac{1}{2} C^{-1}_{\Theta,ij} \frac{\partial C_{\Theta,jk}}{\partial B_a} C^{-1}_{\Theta,kl} \frac{\partial C_{\Theta,li}}{\partial B_b} \,.$$

- This is a general statement: for a gaussian distribution the Fisher matrix

$$F_{ab} = \frac{1}{2} \mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{,a}\mathbf{C}^{-1}\mathbf{C}_{,b}]$$

- Kramer-Rao identity says that the best possible covariance matrix on a set of parameters is $\mathbf{C} = \mathbf{F}^{-1}$
- Thus, the iteration returns an estimate of the covariance matrix of the estimators $\mathbf{C}_B$

Cosmological Parameters

- The probability distribution of the bandpowers given the cosmological parameters c_i is not Gaussian but central limit theorem says at high ℓ it is often an adequate approximation

$$\mathcal{L}_c(\hat{B}_a) \approx \frac{1}{(2\pi)^{N_c/2}\sqrt{\det \mathbf{C}_B}} \exp\left[-\frac{1}{2}(\hat{B}_a - B_a) C^{-1}_{B,ab}(\hat{B}_b - B_b)\right]$$

- Again Bayes' theorem gives the joint posterior of the cosmological parameters from the bandpower likelihood
- With this or other more sophisticated approximations to the bandpower likelihood, still need a fast approach to exploring the bandpower likelihood function

MCMC

- Monte Carlo Markov Chain (MCMC)
- Start with a set of cosmological parameters $\mathbf{c}^m$, compute likelihood
- Take a random step in parameter space to $\mathbf{c}^{m+1}$ of size drawn from a multivariate Gaussian (a guess at the parameter covariance matrix) $\mathbf{C}_c$ (e.g. from the crude Fisher approximation or the covariance of a previous short chain run). Compute likelihood.
- Draw a random number between 0,1 and if the likelihood ratio exceeds this value take the step (add to Markov chain); if not then do not take the step (add the original point to the Markov chain). Repeat.
- Given Bayes' theorem the chain is then a sampling of the joint posterior probability density of the parameters

Parameter Errors

- Can compute any statistic based on the probability distribution of parameters
- For example, compute the mean and variance of a given parameter

$$\bar{c}_i = \frac{1}{N_M} \sum_{m=1}^{N_M} c_i^m$$

$$\sigma^2(c_i) = \frac{1}{N_M - 1} \sum_{m=1}^{N_M} (c_i^m - \bar{c}_i)^2$$

- Trick is in assuring burn in (not sensitive to initial point), step size, and convergence
- Usually requires running multiple chains. Typically tens of thousands of elements per chain.

Radical Compression

- Started with time ordered data $\sim 10^{10}$ numbers for a satellite experiment
- Compressed to a map assuming a CMB spectrum (and time independent fluctuations) $\sim 10^7$ numbers
- Compressed to a power spectrum (Gaussian statistics) independent of m (statistical isotropy) $\sim 10^3$ numbers
- Compressed to cosmological parameters (a cosmological model) $\sim 10^3$
- A factor of 10^9 reduction in the representation. Nature is very efficient.

Parameter Forecasts

- Now connect this discussion with the crude approximations from previous set of notes.
- Gaussian approximation says Fisher matrix of the cosmological parameters becomes

$$F_{c,ij} = \frac{\partial B_a}{\partial c_i} C^{-1}_{B,ab} \frac{\partial B_b}{\partial c_j}$$

  which is the error propagation formula discussed above
- The bandpower covariance can be computed from the Fisher approximation of the pixel likelihood.
- In the crude approximation one takes the covariance to be given by the number of independent modes going into each bandpower estimate

Parameter Forecasts

- For bandpowers being C_ℓ itself, i.e. estimating every ℓ approximate covariance with an increased variance:

$$F_{ij} = \sum_\ell \frac{(2\ell+1) f_{\rm sky}}{2(C_\ell^{\Theta\Theta} + C_\ell^{NN})^2} \frac{\partial C_\ell^{\Theta\Theta}}{\partial c_i} \frac{\partial C_\ell^{\Theta\Theta}}{\partial c_j}$$

  where the sky fraction $f_{\rm sky}$ quantifies the loss of independent modes due to the sky cut

- This is the form we previously derived from just thinking about the simple estimator