Astro 448

Lecture Notes Set 7

Wayne Hu

Gaussian Statistics

- Statistical isotropy says two-point correlation depends only on the power spectrum

$$\Theta(\hat{\mathbf{n}}) = \sum_{\ell m} \Theta_{\ell m} Y_{\ell m}(\hat{\mathbf{n}})$$

$$\langle \Theta_{\ell m}^* \Theta_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell^{\Theta\Theta}$$

- Reality of field says $\Theta_{\ell m} = (-1)^m \Theta_{\ell(-m)}$
- For a Gaussian random field, power spectrum defines all higher order statistics, e.g.

$$\langle \Theta_{\ell_1 m_1} \Theta_{\ell_2 m_2} \Theta_{\ell_3 m_3} \Theta_{\ell_4 m_4} \rangle$$
$$= (-1)^{m_1+m_2} \delta_{\ell_1 \ell_3} \delta_{m_1(-m_3)} \delta_{\ell_2 \ell_4} \delta_{m_2(-m_4)} C_{\ell_1}^{\Theta\Theta} C_{\ell_2}^{\Theta\Theta} + \text{all pairs}$$

Idealized Statistical Errors

- Take a noisy estimator of the multipoles in the map

$$\hat{\Theta}_{\ell m} = \Theta_{\ell m} + N_{\ell m}$$

and take the noise to be statistically isotropic

$$\langle N^*_{\ell m} N_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell^{NN}$$

- Construct an unbiased estimator of the power spectrum $\langle \hat{C}_\ell^{\Theta\Theta} \rangle = C_\ell^{\Theta\Theta}$

$$\hat{C}_\ell^{\Theta\Theta} = \frac{1}{2\ell+1} \sum_{m=-l}^{l} \hat{\Theta}^*_{\ell m} \hat{\Theta}_{\ell m} - C_\ell^{NN}$$

- Variance in estimator

$$\langle \hat{C}_\ell^{\Theta\Theta} \hat{C}_\ell^{\Theta\Theta} \rangle - \langle \hat{C}_\ell^{\Theta\Theta} \rangle^2 = \frac{2}{2\ell+1} (C_\ell^{\Theta\Theta} + C_\ell^{NN})^2$$

Cosmic and Noise Variance

- RMS in estimator is simply the total power spectrum reduced by $\sqrt{2/N_{\rm modes}}$ where $N_{\rm modes}$ is the number of m-mode measurements
- Even a perfect experiment where $C_\ell^{NN} = 0$ has statistical variance due to the Gaussian random realizations of the field. This cosmic variance is the result of having only one realization to measure.
- Noise variance is often approximated as white detector noise. Removing the beam to place the measurement on the sky

$$N_\ell^{\Theta\Theta} = \left(\frac{T}{d_T}\right)^2 e^{\ell(\ell+1)\sigma^2} = \left(\frac{T}{d_T}\right)^2 e^{\ell(\ell+1){\rm FWHM}^2/8\ln 2}$$

where d_T can be thought of as a noise level per steradian of the temperature measurement, σ is the Gaussian beam width, FWHM is the full width at half maximum of the beam

Idealized Parameter Forecasts

- A crude propagation of errors is often useful for estimation purposes.
- Suppose $C_{\alpha\beta}$ describes the covariance matrix of the estimators for a given parameter set π_α.
- Define $\mathbf{F} = \mathbf{C}^{-1}$ [formalized as the Fisher matrix later]. Making an infinitesimal transformation to a new set of parameters p_μ

$$F_{\mu\nu} = \sum_{\alpha\beta} \frac{\partial \pi_\alpha}{\partial p_\mu} F_{\alpha\beta} \frac{\partial \pi_\beta}{\partial p_\nu}$$

- In our case π_α are the C_ℓ the covariance is diagonal and p_μ are cosmological parameters

$$F_{\mu\nu} = \sum_\ell \frac{2\ell + 1}{2(C_\ell^{\Theta\Theta} + C_\ell^{NN})^2} \frac{\partial C_\ell^{\Theta\Theta}}{\partial p_\mu} \frac{\partial C_\ell^{\Theta\Theta}}{\partial p_\nu}$$

Idealized Parameter Forecasts

- Polarization handled in same way (requires covariance)
- Fisher matrix represents a local approximation to the transformation of the covariance and hence is only accurate for well constrained directions in parameter space
- Derivatives evaluated by finite difference
- Fisher matrix identifies parameter degeneracies but only the local direction – i.e. all errors are ellipses not bananas

Beyond Idealizations: Time Ordered Data

- For the data analyst the starting point is a string of "time ordered" data coming out of the instrument (post removal of systematic errors!)
- Begin with a model of the time ordered data as

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

where i denotes pixelized positions indexed by i, d_t is the data in a time ordered stream indexed by t. Number of time ordered data will be of the order 10^{10} for a satellite! number of pixels $10^6 - 10^7$.

- The noise n_t is drawn from a distribution with a known power spectrum

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

Pointing Matrix

- The pointing 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

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

- More generally encorporates differencing, beam, rotation (for polarization)

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 $\mathbf{C}_N \equiv (\mathbf{P}^{\rm tr}\mathbf{C}_d^{-1}\mathbf{P})^{-1}$ 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)

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

$$C_{S,ij} \equiv \langle \Theta_i \Theta_j \rangle = \sum_\ell \Delta^2_{T,\ell} W_{\ell,ij}$$

- Here 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 $b_\ell^2 \propto e^{-\ell(\ell+1)\sigma^2}$

Band Powers

- In principle the underlying theory to extract from maximum likelihood is the power spectrum at every ℓ
- However with a finite patch of sky, it is not possible to extract multipoles separated by $\Delta\ell < 2\pi/L$ where L is the dimension of the survey
- So consider instead a theory parameterization of $\Delta^2_{T,\ell}$ constant in bands of $\Delta\ell$ chosen to match the survey forming a set of band powers 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.

Band Power Esitmation

- 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
- Iterative approach to maximizing the likelihood: take a trial point $B_a^{(0)}$ and improve estimate based a Newton-Rhapson approach to finding zeros

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

Fisher Matrix

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

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

- 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 it is often an adequate approximation

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

- Grid based approaches evaluate the likelihood in cosmological parameter space and maximize
- Faster approaches monte carlo the exploration of the likelihood space intelligently (“Monte Carlo Markov Chains”)
- Since the number of cosmological parameters in the working model is $N_c \sim 10$ this represents a final radical compression of information in the original timestream which recall has up to $N_t \sim 10^{10}$ data points.

Parameter Forecasts

- The 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 Fisher matrix can be more accurately defined for an experiment by taking the pixel covariance and using the general formula for the Fisher matrix of gaussian data
- Corrects for edge effects with the approximate effect of

$$F_{\mu\nu} = \sum_\ell \frac{(2\ell+1) f_{\rm sky}}{2(C_\ell^{\Theta\Theta} + C_\ell^{NN})^2} \frac{\partial C_\ell^{\Theta\Theta}}{\partial p_\mu} \frac{\partial C_\ell^{\Theta\Theta}}{\partial p_\nu}$$

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