Gaussian Statistics

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

\[ \Theta(\hat{n}) = \sum_{\ell m} \Theta_{\ell m} Y_{\ell m}(\hat{n}) \]

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

- 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

\[ \langle N^*_{\ell m} N_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C^{NN}_{\ell} \]

- Construct an unbiased estimator of the power spectrum

\[ \langle \hat{C}^{\Theta \Theta}_\ell \rangle = C^{\Theta \Theta}_\ell \]

\[ \hat{C}^{\Theta \Theta}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \hat{\Theta}^*_{\ell m} \hat{\Theta}_{\ell m} - C^{NN}_{\ell} \]

- Variance in estimator

\[ \langle \hat{C}^{\Theta \Theta}_\ell \hat{C}^{\Theta \Theta}_\ell \rangle - \langle \hat{C}^{\Theta \Theta}_\ell \rangle^2 = \frac{2}{2\ell + 1} \left( C^{\Theta \Theta}_\ell + C^{NN}_{\ell} \right)^2 \]
Cosmic and Noise Variance

- RMS in estimator is simply the total power spectrum reduced by \( \sqrt{2/N_{\text{modes}}} \) where \( N_{\text{modes}} \) 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

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

where \( d_T \) can be thought of as a noise level per steradian of the temperature measurement, \( \sigma \) 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 $\pi_\alpha$.
- Define $F = C^{-1}$ [formalized as the Fisher matrix later]. Making an infinitesimal transformation to a new set of parameters $p_\mu$

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

- In our case $\pi_\alpha$ are the $C_\ell$ the covariance is diagonal and $p_\mu$ 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

\[ \langle n_t n_{t'} \rangle = C_{d,tt'} \]
The pointing matrix $P$ 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.

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

More generally incorporates differencing, beam, rotation (for polarization)
Maximum Likelihood Mapmaking

- What is the best estimator of the underlying map $\Theta_i$?

- Likelihood function: the probability of getting the data given the theory $\mathcal{L} \equiv P[\text{data}|\text{theory}]$. In this case, the theory is the set of parameters $\Theta_i$.

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

- Bayes theorem says that $P[\Theta_i|d_t]$, the probability that the temperatures are equal to $\Theta_i$ 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 $\Theta_i$ is simple since the log-likelihood is quadratic.

- Differentiating the argument of the exponential with respect to $\Theta_i$ and setting to zero leads immediately to the estimator

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

where $C_N \equiv (P^{tr}C_d^{-1}P)^{-1}$ 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).
The next step in the chain of inference is the power spectrum extraction. Here the correlation between pixels is modelled through the power spectrum

\[ C_{S,ij} \equiv \langle \Theta_i \Theta_j \rangle = \sum_{\ell} \Delta_{T,\ell}^2 W_{\ell,ij} \]

Here \( W_{\ell} \), the window function, is derived by writing down the expansion of \( \Theta(\hat{n}) \) in harmonic space, including smoothing by the beam and pixelization.

For example in the simple case of a gaussian beam of width \( \sigma \) it is proportional to the Legendre polynomial \( P_{\ell}(\hat{n}_i \cdot \hat{n}_j) \) 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 $\ell$
- 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

$$
\mathcal{L}_B(\Theta_i) = \frac{1}{(2\pi)^{N_p/2} \sqrt{\det C_\Theta}} \exp \left( -\frac{1}{2} \Theta_i C_{\Theta,i,j}^{-1} \Theta_j \right)
$$

where $C_\Theta = C_S + C_N$ and $N_p$ is the number of pixels in the map.
As before, $L_B$ is Gaussian in the anisotropies $\Theta_i$, but in this case $\Theta_i$ 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

$$\hat{B}_a = \hat{B}_a^{(0)} + F_{B,ab} \frac{\partial \ln L_B}{\partial B_b}$$

$$= \hat{B}_a^{(0)} + \frac{1}{2} F^{-1}_{B,ab} \left( \Theta_i C^{-1}_{\Theta,ij} \frac{\partial C_{\Theta,kl}}{\partial B_b} C^{-1}_{\Theta,kl} \Theta_l - C^{-1}_{\Theta,ij} \frac{\partial C_{\Theta,ji}}{\partial B_b} \right),$$
Fisher Matrix

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

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

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

\[ F_{ab} = \frac{1}{2} \text{Tr}[C^{-1} C_{,a} C^{-1} C_{,b}] \]

- Kramer-Rao identity says that the best possible covariance matrix on a set of parameters is \( C = F^{-1} \)

- Thus, the iteration returns an estimate of the covariance matrix of the estimators \( C_B \)
Cosmological Parameters

- The probability distribution of the bandpowers given the cosmological parameters $c_i$ is not Gaussian but it is often an adequate approximation

$$L_c(B_a) \approx \frac{1}{(2\pi)^{N_c/2} \sqrt{\det C_B}} \exp \left[ -\frac{1}{2} (\hat{B}_a - B_a) C_{B}^{-1} (\hat{B}_b - B_b) \right]$$

- 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

\[ F_{c,ij} = \frac{\partial B_a}{\partial c_i} C_{B,ab}^{-1} \frac{\partial B_b}{\partial c_j} \]

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:

\[ F_{\mu\nu} = \sum_{\ell} \frac{(2\ell + 1)f_{\text{sky}}}{2(C^{\Theta\Theta}_\ell + C^{NN}_\ell)^2} \frac{\partial C^{\Theta\Theta}_\ell}{\partial p_\mu} \frac{\partial C^{\Theta\Theta}_\ell}{\partial p_\nu} \]

where the sky fraction \( f_{\text{sky}} \) quantifies the loss of independent modes due to the sky cut.