

Statistical Model for the CtU2015 Hack: ‘The Music of the Sphere’

Michael Schneider, Roland de Putter, Ryan Keisler, Phil Marshall

July 29, 2015

Our goal is to draw posterior samples of the large-scale 3D cosmological mass density (or gravitational potential) in our universe, given a map of the CMB.

1 Preliminary Notes

The observed CMB temperature \mathbf{d} is modeled as a sum of a signal \mathbf{s} and \mathbf{n} , neglecting foregrounds,

$$\mathbf{d} = \mathbf{s} + \mathbf{n}, \quad (1)$$

where both \mathbf{s} and \mathbf{n} are Gaussian distributed with zero means and covariances \mathbf{S} and \mathbf{N} .

We want to sample realizations of the 3D gravitational potential Ψ given the observed 2D CMB map \mathbf{d} . We can model the relationship between the potential and the CMB temperature as a linear transfer function matrix \mathbf{R} ,

$$\mathbf{s} \equiv \mathbf{R}\Psi. \quad (2)$$

The likelihood is a Gaussian distribution given our assumption of a Gaussian noise model,

$$\Pr(\mathbf{d}|\mathbf{s}, \mathbf{N}) = \mathcal{N}_{\mathbf{d}}(\mathbf{s}, \mathbf{N}) = \mathcal{N}_{\mathbf{d}}(\mathbf{R}\Psi, \mathbf{N}) \quad (3)$$

The prior on the gravitational potential is defined as

$$\Pr(\Psi|\Psi) = \mathcal{N}_{\Psi}(0, \Psi) \quad (4)$$

1.1 Conditional posteriors for Gibbs sampling

[Wandelt et al. \(2004\)](#) shows that the conditional distribution of the signal or gravitational potential is a Gaussian with a mean and covariances (also known as the ‘Wiener filter’),

$$\Pr(\Psi|\mathbf{d}, \Psi, \mathbf{N}) = \mathcal{N}_{\Psi}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (5)$$

where,

$$\boldsymbol{\mu} \equiv \Psi(\Psi + \mathbf{C}_N)^{-1} \mathbf{m} \quad (6)$$

$$\boldsymbol{\Sigma} \equiv \Psi(\Psi + \mathbf{C}_N)^{-1} \mathbf{C}_N, \quad (7)$$

$$\mathbf{m} \equiv (\mathbf{R}^T \mathbf{N}^{-1} \mathbf{R})^{-1} \mathbf{R}^T \mathbf{N}^{-1} \mathbf{d} \quad (8)$$

is the maximum likelihood estimator of the signal and,

$$\mathbf{C}_N \equiv (\mathbf{R}^T \mathbf{N}^{-1} \mathbf{R})^{-1}. \quad (9)$$

1.2 Summary for code implementation

Our goal is to sample conditional Gaussian draws from Equation 5. An important point is that the vector μ can be extended from just those voxels that encompass the last scattering surface to also those voxels filling the rest of the universe. Then, posterior inference of the mass density away from the last scattering surface is described by a conditional multivariate Gaussian distribution with covariance matrix given by the Schur complement of Σ .

The chief computational challenge is inverting the matrices in Equation 6 and Equation 7. Wandelt et al. (2004) and Wandelt (2004) describe a conjugate gradient based method to evaluate the Wiener filter terms for Gibbs sampling.

A secondary computational issue is specification of the discretized matrix R which represents the transfer function relating the 3D gravitational potential to the 2D CMB anisotropies.

2 Detailed Notes

We will jump, not always in a well-organized way, between the general notation from Zaroubi et al. (1995) and a more practical notation for the problem at hand (the simplest version of it).

Let's start with the following toy model. We have a 3D potential field on a grid

$$\Phi(\mathbf{x}_i). \quad (10)$$

Here i labels the grid points so that we can treat this field as a “theory vector”,

$$\mathbf{s} \equiv \Phi, \quad (11)$$

with components

$$s_i = \Phi_i = \Phi(\mathbf{x}_i) \quad i = 1, \dots, N. \quad (12)$$

Moreover, we have a M -dimensional data vector, given by some linear operator on the theory vector, plus noise,

$$\mathbf{d} = R\mathbf{s} + \varepsilon. \quad (13)$$

Here R is the transformation matrix defining the data, and the noise has covariance

$$N_{kl}^\varepsilon \equiv \langle \varepsilon_k \varepsilon_l \rangle. \quad (14)$$

I will try to use the indices i, j on the theory (i.e. Φ) side (so that these indices run from 1 to N) and use k, l on the data side (so that they run from 1 to M). More concretely, let's say the data are simply measurements of $\Phi(\mathbf{x}_i)$ at a subset of specific grid locations $\{i_1, \dots, i_M\}$,

$$d_k \equiv \hat{\Phi}(\mathbf{x}_{i_k}) = \Phi(\mathbf{x}_{i_k}) + \varepsilon_k = s_{i_k} + \varepsilon_k. \quad (15)$$

Then, \mathbf{R} is a $M \times N$ matrix with entries

$$R_{ki} \equiv \delta_{i_k i}^K. \quad (16)$$

For simplicity, we can start with a diagonal noise matrix,

$$N_{kl}^\varepsilon = \sigma_\varepsilon^2 \delta_{kl}^K \quad (17)$$

Now, given a measurement of \mathbf{d} , i.e. given a noisy measurement of the potential at a subset of grid locations, for instance on a sphere, we want to generate realizations of the full field. There is one more ingredient we need: the prior covariance matrix of the theory vector. This is a $N \times N$ matrix \mathbf{S} , with entries

$$S_{ij} \equiv \langle s_i s_j \rangle = \langle \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_j) \rangle = \xi_\Phi(|\mathbf{x}_i - \mathbf{x}_j|). \quad (18)$$

Now, following e.g. Rybicki & Press (1992); Zaroubi et al. (1995), we can first construct the maximum likelihood estimator for the field $\Phi(\mathbf{x}_i)$.

$$\mathbf{s}^{\text{MV}} = \mathbf{F} \mathbf{d}, \quad (19)$$

with

$$\mathbf{F} = \langle \mathbf{s} \mathbf{d}^\dagger \rangle \langle \mathbf{d} \mathbf{d}^\dagger \rangle^{-1} = \mathbf{S} \mathbf{R}^\dagger (\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N}^\varepsilon)^{-1} \quad (20)$$

Since we assume Gaussianity, this estimator gives the posterior expectation value for the field and the maximum likelihood solution.

The above Eqs. (19)-(20) give the general result. Things become clearer when we apply it to our specific example. For instance,

$$\langle d_k d_l \rangle = (\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N}^\varepsilon)_{kl} = \xi_\Phi(|\mathbf{x}_{i_k} - \mathbf{x}_{i_l}|) + \sigma_\varepsilon^2 \delta_{kl}^K \equiv (\xi_\Phi + \sigma_\varepsilon^2 \mathbf{I})_{kl} \quad (21)$$

(\mathbf{I} is the identity matrix and ξ_Φ is the correlation function matrix restricted to the grid points where we have measurements, e.g. the CMB last-scattering surface) and

$$\langle s_i d_k \rangle = (\mathbf{S} \mathbf{R}^\dagger)_{ik} = \xi_\Phi(|\mathbf{x}_i - \mathbf{x}_{i_k}|). \quad (22)$$

We can't proceed further analytically, because we need to invert the matrix $\xi_\Phi + \sigma_\varepsilon^2 \mathbf{I}$, which is simply the total covariance matrix of the observable $d_k \equiv \hat{\Phi}(\mathbf{x}_{i_k})$. We can now write for the solution,

$$s_i^{\text{MV}} = \Phi^{\text{MV}}(\mathbf{x}_i) = \sum_{kl} \xi_\Phi(|\mathbf{x}_i - \mathbf{x}_{i_k}|) (\xi_\Phi + \sigma_\varepsilon^2 \mathbf{I})_{kl}^{-1} \hat{\Phi}(\mathbf{x}_{i_l}) \quad (23)$$

2.1 Generating realizations of $\Phi(\mathbf{x}_i)$

Now that we have the *mean* of the conditional posterior of $\mathbf{s} = \Phi$, given in Eq. (23), there is a straightforward method for generating realizations that follow the posterior distribution:

- Generate an unconstrained Gaussian random realization of the theory field, $\tilde{\mathbf{s}} = \tilde{\Phi}$ (i.e. the 3D potential), that follows the prior correlation function of the field, i.e. the same way of generating random realizations we use for the true field, that is our input to the whole calculation.
- Generate a (Gaussian) realization of the measurement noise ε , that follows the covariance given by \mathbf{N}^ε .
- Construct a realization of the data:

$$\tilde{\mathbf{d}} = \mathbf{R} \tilde{\mathbf{s}} + \varepsilon. \quad (24)$$

Specifically, for us this means

$$\tilde{\Phi}(\mathbf{x}_{i_k}) = \tilde{\Phi}(\mathbf{x}_{i_k}) + \tilde{\varepsilon}_k. \quad (25)$$

- Write the realization of the field, *conditional on the observations*, as

$$\mathbf{s}_{\text{post}} = \mathbf{s}^{\text{MV}} + \tilde{\mathbf{s}} - \mathbf{F}\tilde{\mathbf{d}}. \quad (26)$$

For us this means

$$\Phi_{\text{post}}(\mathbf{x}_i) = \Phi^{\text{MV}}(\mathbf{x}_i) + \tilde{\Phi}(\mathbf{x}_i) - \tilde{\Phi}^{\text{MV}}(\mathbf{x}_i), \quad (27)$$

where I have written

$$\tilde{\Phi}^{\text{MV}}(\mathbf{x}_i) \equiv (\mathbf{F}\tilde{\mathbf{d}})_i = \sum_{kl} \xi_{\Phi}(|\mathbf{x}_i - \mathbf{x}_{i_k}|) (\xi_{\Phi} + \sigma_{\varepsilon}^2 \mathbf{I})_{kl}^{-1} \tilde{\Phi}(\mathbf{x}_l) \quad (28)$$

This quantity has the appropriate statistics given the measurement \mathbf{d} and the prior statistics of \mathbf{s} and of the noise.

Possibly more advanced method that would be interesting to explore: [Jasche & Lavaux \(2015\)](#).

2.2 Informing simulations with existing data

Consider the common scenario that we want to constrain cosmological parameters θ given a data set \mathbf{d} , which is related to the cosmological mass density as in [Equation 13](#). If we have no knowledge about the realized mass density Φ in our observed portion of the universe then we must marginalize over all possible mass density realizations in computing the marginal posterior of θ .

However, if we have performed previous surveys of some volumes of the universe, we should be able to use the information from those surveys to constrain the mass density marginalization in the analysis of our current survey. Specifically, the joint marginal likelihood of our current data \mathbf{d} and an ‘external’ or past data set \mathbf{d}_{ext} is,

$$P(\mathbf{d}, \mathbf{d}_{\text{ext}} | \theta) = P(\mathbf{d} | \mathbf{d}_{\text{ext}}, \theta) P(\mathbf{d}_{\text{ext}} | \theta). \quad (29)$$

That is, we want a *conditional* inference of the data \mathbf{d} given the external data \mathbf{d}_{ext} , which factors the joint data likelihood.

The posterior on the cosmological parameters θ factors as,

$$\begin{aligned} P(\theta | \mathbf{d}, \mathbf{d}_{\text{ext}}) &= \frac{P(\mathbf{d}, \mathbf{d}_{\text{ext}} | \theta) P(\theta)}{P(\mathbf{d}, \mathbf{d}_{\text{ext}})} \\ &= \frac{P(\mathbf{d} | \mathbf{d}_{\text{ext}}, \theta) P(\mathbf{d}_{\text{ext}} | \theta) P(\theta)}{P(\mathbf{d} | \mathbf{d}_{\text{ext}}) P(\mathbf{d}_{\text{ext}})} \\ &= \frac{P(\mathbf{d} | \mathbf{d}_{\text{ext}}, \theta) P(\theta | \mathbf{d}_{\text{ext}})}{P(\mathbf{d} | \mathbf{d}_{\text{ext}})}. \end{aligned} \quad (30)$$

So, we can infer posterior constraints on θ given both past \mathbf{d}_{ext} and current \mathbf{d} data sets by evaluating only the likelihood of the current data conditioned on the external data $P(\mathbf{d} | \mathbf{d}_{\text{ext}}, \theta)$ as long as we update our prior on θ from the posterior inferences given only the external data sets.

If we assume the observational noise is uncorrelated between the current \mathbf{d} and past \mathbf{d}_{ext} data sets, the likelihood conditioned on the mass density factors,

$$P(\mathbf{d}, \mathbf{d}_{\text{ext}} | \Phi, \theta) = P(\mathbf{d} | \Phi, \theta) P(\mathbf{d}_{\text{ext}} | \Phi, \theta). \quad (31)$$

But, after marginalizing over the mass density the joint likelihood no longer factors,

$$P(\mathbf{d}, \mathbf{d}_{\text{ext}} | \theta) = \int d\Phi P(\mathbf{d} | \Phi, \theta) P(\mathbf{d}_{\text{ext}} | \Phi, \theta) P(\Phi | \theta). \quad (32)$$

We can evaluate Equation 32 and thereby understand the impact of adding information from a past data set \mathbf{d}_{ext} by assuming Gaussian distributions for each conditional distribution in the integrand,

$$P(\Phi|\theta) \equiv \mathcal{N}_{\Phi}(0, \mathbf{S}(\theta)) \quad (33)$$

$$P(\mathbf{d}|\Phi, \theta) \equiv \mathcal{N}_{\mathbf{d}}(\mathbf{R}(\theta)\Phi, \mathbf{N}^{\epsilon}) \quad (34)$$

$$P(\mathbf{d}_{\text{ext}}|\Phi, \theta) \equiv \mathcal{N}_{\mathbf{d}_{\text{ext}}}(\mathbf{R}_{\text{ext}}(\theta)\Phi, \mathbf{N}^{\epsilon_{\text{ext}}}). \quad (35)$$

The Gaussian distribution in Equation 33 comes from our assumption that the initial conditions for the cosmological mass density perturbations are Gaussian with mean zero and a covariance $\mathbf{S}(\theta)$ given by the two-point correlation function as in Equation 18. The Gaussian distributions in Equation 34 and Equation 35 come from the assumed noise models for each data set.

By first rewriting the likelihood as a function of Φ ,

$$\mathcal{N}_{\mathbf{d}}(\mathbf{R}\Phi, \mathbf{N}^{\epsilon}) = |\mathbf{R}\mathbf{R}^T|^{1/2} \mathcal{N}_{\Phi}\left(\mathbf{R}^{-1}\mathbf{d}, \left(\mathbf{R}^T(\mathbf{N}^{\epsilon})^{-1}\mathbf{R}\right)^{-1}\right) \quad (36)$$

and then multiplying by the prior on Φ ,

$$P(\mathbf{d}|\Phi, \theta)P(\Phi|\theta) = |\mathbf{R}\mathbf{R}^T|^{1/2} \mathcal{N}_{\mathbf{R}\mathbf{d}}(0, \mathbf{S} + \tilde{\mathbf{N}}^{\epsilon}) \mathcal{N}_{\Phi}\left(\mathbf{C}^{\epsilon}(\tilde{\mathbf{N}}^{\epsilon})^{-1}\mathbf{R}\mathbf{d}, \mathbf{C}^{\epsilon}\right) \quad (37)$$

we get another Gaussian distribution in Φ with maximum given by Equation 20. And, in Equation 37 we defined,

$$\tilde{\mathbf{N}}^{\epsilon} \equiv \left(\mathbf{R}^T(\mathbf{N}^{\epsilon})^{-1}\mathbf{R}\right)^{-1}, \quad (38)$$

$$\mathbf{C}^{\epsilon} \equiv \left(\mathbf{S}^{-1} + (\tilde{\mathbf{N}}^{\epsilon})^{-1}\right)^{-1}. \quad (39)$$

Performing similar manipulations with the Gaussian likelihood of \mathbf{d}_{ext} , we can perform the integral in Equation 32 as the integral of Gaussian distribution to get,

$$P(\mathbf{d}, \mathbf{d}_{\text{ext}}|\theta) = \left[|\mathbf{R}_{\text{ext}}\mathbf{R}_{\text{ext}}^T|^{1/2} \mathcal{N}_{\mathbf{R}_{\text{ext}}\mathbf{d}_{\text{ext}}}(0, \mathbf{S} + \tilde{\mathbf{N}}^{\epsilon_{\text{ext}}})\right] \times \left[|\mathbf{R}\mathbf{R}^T|^{1/2} \mathcal{N}_{\mathbf{R}\mathbf{d}}\left(\mathbf{C}^{\epsilon_{\text{ext}}}(\tilde{\mathbf{N}}^{\epsilon_{\text{ext}}})^{-1}\mathbf{R}_{\text{ext}}\mathbf{d}_{\text{ext}}, \tilde{\mathbf{N}}^{\epsilon} + \mathbf{C}^{\epsilon_{\text{ext}}}\right)\right], \quad (40)$$

where the term in the first square brackets is recognized as $P(\mathbf{d}_{\text{ext}}|\theta)$ while the second set of square brackets gives $P(\mathbf{d}|\mathbf{d}_{\text{ext}}, \theta)$.

The result for $P(\mathbf{d}_{\text{ext}}|\theta)$ in Equation 40 is familiar. The marginal likelihood for a mass density tracer without any extra prior information has mean zero and a covariance that is the sum of the signal and noise covariance matrices. Although, the prefactor depending on the transfer function \mathbf{R}_{ext} is not usually kept even though it does, in general, depend on the cosmological parameters θ .

The result for $P(\mathbf{d}|\mathbf{d}_{\text{ext}}, \theta)$ in Equation 40 is a Gaussian distribution in the data with a mean determined by the external data \mathbf{d}_{ext} , suitably filtered by the signal and noise covariances. The covariance of \mathbf{d} is a sum of the noise covariance for the data \mathbf{N}^{ϵ} and the matrix $\mathbf{C}^{\epsilon_{\text{ext}}}$ as defined in Equation 39 (inserting the past data noise covariance). Because the *inverse* covariances are summed in Equation 39,

$$\mathbf{C}^{\epsilon_{\text{ext}}} \approx \tilde{\mathbf{N}}^{\epsilon_{\text{ext}}} \quad (41)$$

when the past data \mathbf{d}_{ext} is measured with high signal-to-noise ratio (i.e., as the observational noise on \mathbf{d}_{ext} becomes arbitrarily small). So, even though measurements of \mathbf{d}_{ext} may be sample variance dominated, the conditional inferences from \mathbf{d} given \mathbf{d}_{ext} need not be limited by sample variance. This is our main result.

2.2.1 Algorithm for simulations

This suggests a procedure for running simulations of cosmological large-scale structure for parameter inference,

1. Infer $P(\theta|\mathbf{d}_{\text{ext}})$ given past or external data sets \mathbf{d}_{ext} as is typically done.
2. Use $P(\Phi, \theta|\mathbf{d}_{\text{ext}})$ as in [subsection 2.1](#) to generate initial conditions for cosmological simulations to aid in the inferences from a new data set \mathbf{d} that at least partially overlaps in volume with \mathbf{d}_{ext} .
3. Infer cosmological parameter constraints given \mathbf{d} given the likelihood function obtained from the previous step and replacing the usual prior on θ with $P(\theta|\mathbf{d}_{\text{ext}})$.

From [Equation 30](#), this procedure is guaranteed to yield correct inferences from $P(\theta|\mathbf{d}, \mathbf{d}_{\text{ext}})$ (i.e., as if we had analyzed both data sets jointly). But, we obtain the computational savings in running *conditional* simulations given the first data set \mathbf{d}_{ext} .

References

- Jasche, J. & Lavaux, G. 2015, MNRAS, 447, 1204
- Rybicki, G. B. & Press, W. H. 1992, ApJ, 398, 169
- Wandelt, B. D. 2004, arXiv.org
- Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, arXiv, D70, 083511
- Zaroubi, S., Hoffman, Y., Fisher, K. B., & Lahav, O. 1995, ApJ, 449, 446