

MCMC Solver on Planck 2015 High- ℓ C_ℓ^{TT}

Alexander Nie

Harvard University, 28 Dewolf St., Cambridge, MA 02138

(Dated: December 7, 2016)

In this project, an MCMC solver was implemented to estimate six cosmological parameters (H_0 , $\Omega_b h^2$, $\Omega_{CDM} h^2$, τ , n_s , and A_s) from the two-point angular power spectrum of temperature data from the Planck 2015 release. In contrast with standard MCMC solvers which use the Metropolis-Hastings algorithm, this project used an optimization known as **emcee** in order to optimize exploration of the possible values of the cosmological parameters. Using the high- ℓ data, the maximum likelihood estimates of the six-parameter set was $H_0 = 65.22 \pm 11.08$, $\Omega_b^2 = .02193 \pm .00437$, $\Omega_c h^2 = .1389 \pm .1389$, $\tau = .08178 \pm .03291$, $n_s = .9602 \pm .0698$, $A_s = 2.212 \times 10^9 \pm .204 \times 10^9$, in agreement with the existing literature.

I. INTRODUCTION

Modern cosmology faces a desirable, yet daunting problem: given the vast observational data made possible by increasingly large surveys, how can reliable statistical analyses be performed in a timely manner? One such solution to this problem are Markov-Chain Monte Carlo solvers, a class of algorithms which use random walks to sample the probability space of possible parameter values. In turn, this allows the solver to estimate aggregate properties of the sample without necessarily having to examine every single point in parameter space.

For this project, six cosmological parameters, H_0 , $\Omega_b h^2$, $\Omega_{CDM} h^2$, τ , n_s , A_s , are varied near their fiducial values in the standard Λ -CDM cosmological model. Using a Boltzmann solver to predict the angular power spectrum C_ℓ of the Cosmic Microwave Background at each given value of the cosmological parameters, the likelihood of each cosmology can be computed by comparing against existing data from the Planck 2015 data release. Aggregate quantities such as the mean, variance, and covariance of the parameters are then estimated on the samples.

II. THEORETICAL FOUNDATIONS

A. CMB Power Spectrum and Temperature

To nearly 1 in 10000 parts, the Cosmic Microwave Background is a homogeneous and isotropic source of radiation[1]. However, small deviations from this isotropy in frequency, temperature, and polarization provide insight into the underlying cosmology. At early, radiation-dominant times in the universe, temperature deviations from isotropy largely originated from fluctuations in photon density n_γ .

$$\left(\frac{\delta \dot{n}_\gamma}{n_\gamma}\right) = -\nabla \cdot (\mathbf{v}_\gamma) \quad (1)$$

$$\nabla p_\gamma = \frac{1}{3} \nabla \rho_\gamma \quad (2)$$

Then from the continuity relation (1), the equation of state on photons (2), and the fact that $n_\gamma \propto T^3$, it can be shown that the angular power spectrum of temperature Θ in Fourier space obeys the harmonic oscillator equation [1]

$$\ddot{\Theta} + \frac{1}{3} k^2 \Theta = 0 \quad (3)$$

Assuming minimal initial velocities, the temperature power spectrum thus obeys

$$\Theta(\eta_{recomb}) = \Theta(0) \cos(k s_{recomb}) \quad (4)$$

so peaks are expected to appear at $k s_{recomb} = 1$ or characteristic scale $\lambda = \frac{2s}{n}$.

Decomposing Θ over the projected observational sphere into the spherical harmonics then yields coefficients of contribution whose two-point correlation functions are the power-spectrum C_ℓ :

$$\Theta(\hat{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}) \quad (5)$$

$$\langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell \quad (6)$$

$$C_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \langle |a_{\ell m}|^2 \rangle \quad (7)$$

Thus, the C_ℓ encode the underlying dynamics of the cosmology expressed in temperature perturbations at a given scale in the CMB. For this project, $C_\ell^T T$ data was examined for $\ell > 30$, as non-primordial perturbations, namely the Sachs-Wolfe effect, contribute to anisotropy at low ℓ [5]

B. MCMC Solvers

In a general Monte Carlo simulation, random samples are repeatedly taken in order to estimate some underlying distribution also known as the target distribution.

In particular, the problem usually reduces to calculating some quantity

$$\langle A \rangle = \frac{\sum_c A(c) e^{-\beta E(c)}}{Z} \quad (8)$$

for a quantity A dependent on the total configurations c in the system, along with the weight of each configuration $0 < e^{-\beta E(c)} < 1$ and a normalization $Z = \sum_c e^{-\beta E(c)}$ [2]. In a uniform random sampling of all configurations, $\langle A \rangle$ can be estimated when Z is known; however this simply creates a circular problem.

In the Markov Chain solution to the problem, circularity is broken by sampling each configuration with dynamic probability $P(c, t)$ at time g such that $P(c, t) \rightarrow \alpha e^{-\beta E(c)}$ for each configuration. In particular, at each time step, the probability distribution over the next configuration sampled depends only on the current configuration, endowing the solution with the Markov property. Then at equilibrium,

$$P(c, \infty) = \sum_{c'} T(c' \rightarrow c) P(c', \infty) \quad (9)$$

where T is the weight matrix. This condition imposes the detailed balance condition

$$P(c', \infty) T(c' \rightarrow c) = P(c, \infty) T(c \rightarrow c') \quad (10)$$

$$\frac{T(c' \rightarrow c)}{T(c \rightarrow c')} = e^{-\beta \Delta E} \quad (11)$$

Thus, the probability of transition from a given configuration to the next is dependent only upon the ratio of the probabilities of transition. Equivalently, this means that at a given timestep, a new configuration should be automatically accepted when $\Delta E < 0$ and accepted with a finite probability when $\Delta E > 0$, giving rise to the Metropolis-Hastings algorithm for sampling (Sec. III). Assuming such an equilibrium state exists, then regardless of initial starting point in the parameter space, after a sufficiently long "burn-in" time, data aggregation should yield a reasonable estimate for the underlying parameter [2].

In a nutshell, E , the log probability is analogous to the "energy" of a probability landscape to be explored by the solver. Because all sampling occurs locally (only two configurations are calculated) rather than globally, this approach avoids the problem of enumerating all configurations in order to obtain a faithful estimate of the target distribution.

C. Bayesian Methods

Ultimately, the goal of the MCMC Solver in this project is to estimate the joint probability distributions of the six parameters given the observed data in C_ℓ 's. This is equivalent to finding a posterior using Bayes' rule

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \quad (12)$$

where θ is the parameters, \mathbf{x} is the data, $p(\mathbf{x}|\theta)$ is the likelihood of seeing the data given a particular choice of parameters, and $p(\theta)$ is the prior: the probability of a particular choice of parameters based on some previous insight. As the data do not change and their underlying probabilities are unknown, the relation is often written $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$ [3].

Assuming Gaussian error in the measurement, the likelihood for a given set of parameters is then calculated on the predicted \hat{C}_ℓ 's against the measured C_ℓ 's:

$$p(\{C_\ell\}|\{\hat{C}_\ell\}, \theta) = \prod_\ell \frac{1}{\sigma_{C_\ell} \sqrt{2\pi}} \exp\left(-\frac{(\hat{C}_\ell - C_\ell)^2}{2\sigma_{C_\ell}^2}\right) \quad (13)$$

III. METHODS

A. Metropolis-Hastings Algorithm

As mentioned in Sec. II, the asymptotic behavior of the Metropolis-Hastings algorithm depends on the random choice of configuration at timestep $t + 1$ given only the current configuration t . In particular, if $E(C)$ is the log likelihood of a configuration C (a particular set of values for the six cosmological parameters), then after samples C at time t , the algorithm samples C' with probability

$$\min\left(e^{-E(C') - E(C)}, 1\right)$$

In the event C' is not sampled, the algorithm then resamples C at timestep $t + 1$. The following describes the overall scheme of the algorithm[2,4]:

1. Pick some initial configuration (set of values) for the cosmological parameters ($H_0, \Omega_b h^2, \Omega_{CDM} h^2, \tau, n_s, A_s$) in allowed parameter space.
2. Calculate the posterior of the configuration using the priors on the cosmological parameters and the likelihood on C_ℓ .
3. Choose a configuration local to the current configuration with the probability scheme detailed above.
3. Cycle between 2 and 3 until the algorithm reaches equilibrium at which point the sampling should stop drifting.
4. Discard samples from earlier timesteps in which non-equilibrium occurs and calculate statistical quantities for the remaining data.

B. Stretch-Move Optimization

While the Metropolis-Hastings algorithm provides significant increases in computational efficiency over total enumeration of the parameter space, this project used a modification by Mackey et al. known as the "stretch-move" to further optimize the runtime efficiency [4]

To start, computation of the posterior at each configuration involves not only the likelihood, but also the prior. Although sufficiently long trials would asymptotically remove bias in the posterior from incorrect information in the prior, for N parameters, even a simple multivariate Gaussian prior has N^2 degrees of freedom in its covariance matrix. By using an affine invariant algorithm, an algorithm that is invariant under linear transformation of the data, the MCMC solver becomes sensitive only to the N diagonal terms of the covariance matrix as the system can be rotated into one for which variation occurs only along principal axes. This then reduces the degrees of freedom (and thus guesswork) needed for initializing a prior.[4]

Second, the Metropolis-Hastings algorithm examines a single configuration at a time. Imagining the algorithm as "walker" between points of parameter space, the locality of MCMC methods means that a single "walker" will need quite a bit of time to sufficiently explore parameter space while multiple walkers under the Metropolis-Hastings algorithm are independent, so their data cannot be aggregated. Under the stretch-move algorithm proposed by Mackey et al., this idea is better optimized by having an ensemble of walkers, leading to parallelizability. Therefore, although the number of samples and computations remains approximately the same, this algorithm allows shorter wait-times when parallelized.

The algorithm is as follows[4]:

1. Pick several initial configurations to serve as starting points for different walkers in two ensembles of K walkers, S_K and S'_K .
2. At time t , randomly pick a walkers X_k, X'_k from the ensembles S_K, S'_K , a random number z from the unit interval.
3. Move X_k $z(X_k - X'_k)$ along the line segment between X_k and X'_k in parameter space to a new configuration Y
4. Accept and sample the new configuration with the probability scheme $P(\text{accept}) = z^{N-1} \frac{p(Y)}{p(X_k)}$ where $p(X)$ is the prior on that configuration. In the event of rejection, return X_k to its original configuration at time t and sample that configuration.
5. Repeat steps 2-4 to move walker in X'_k .
6. Repeat steps 2-5 until the position of the walkers stops drifting.

C. Other Implementation Specifics

The Boltzmann solver CAMB (Code for Anisotropies in the Microwave Background) was used to calculate predicted \hat{C}_ℓ 's for each configuration of $(H_0, \Omega_b h^2, \Omega_{CDM} h^2, \tau, n_s, A_s)$, with all other values fixed to their best-fit values in the Planck collaboration (I) and assuming an infinitesimal reionization depth. For the measured C_ℓ 's, high- ℓ data and uncertainties were taken from the Planck Legacy Archive[5,6]. Following the Planck collaboration, a Gaussian prior was used for τ while the tophat prior was used for the other five

TABLE I: Priors and fixed values for various cosmological parameters used by CAMB

Parameter	Prior
H_0	[40, 100]
$\Omega_b h^2$	[0.005, 0.1]
$\Omega_{CDM} h^2$	[.001, .99]
τ	$\mathcal{N}(.07, .02)$
n_s	[0.9, 1.1]
$\ln(10^{10} A_s)$	[2.7, 4.0]

cosmological parameters studied[5]. Finally, the project implemented a multi-threaded version with a small number of walkers intended for a quad-core laptop and an MPI-based version with a large number of walkers intended for the Harvard ODYSSEY cluster.

IV. RESULTS AND DISCUSSION

A. High- ℓ C_ℓ Spectrum

Two extended trials were run, one with 12 walkers over 700 steps (single node) and one with 16 walkers over 1600 steps (cluster). Results can be found in the Appendix.

Overall, both runs had significant overlap on values for the cosmological parameters. Examining the run data (FIG. 1, FIG. 2), however, showed that fewer walkers generally converged faster than more walkers. In particular, burn-in seems to occur at around 200 steps for the 12-walker trial and 600-steps for the 16-walker trial. This is most likely due to overexploration of the parameter space by the trial with more walkers. In particular, a higher density of walkers implies a smaller distance between them. As the stretch-move causes walkers to jump configurations proportional to the distance between walkers, this causes the low-walker trial to have more jumps, thereby smoothing local behavior in the parameter space. As the acceptance fraction, the fraction of times a new configuration is sampled over the current one, was .767 for the 16-walker trial and .344 for the 12-walker trial, this supports the idea that a 16-walker trial might be too large a number of walkers for a six-parameter model.

Examining the corner plots (FIG. 3, FIG. 4), the 16-walker model showed much more ovoid features around the data peaks while the 12-walker model not only had greater spreads, but also managed to capture very elongated relationships. This again supports the idea that the 16-walker model might not be appropriate, as a high acceptance fraction generally means walkers are already concentrated near the maximum likelihood estimate, whereas the goal of the algorithm is to discover the distribution as well.

V. CONCLUSION

MCMC Solvers offer a way to faithfully estimate an underlying target distribution. From the Monte Carlo property, the solvers avoid the problem of having to enumerate every possible point in the target distribution. From the Markov property, the solvers further reduce computation demand by strategically sampling local configurations, allowing for relatively small calculations. Combined with a Boltzmann solver, MCMC solvers are useful in cosmology as a way to explore the target distribution over various cosmological parameters without having to directly and simultaneously optimize the many parameters and equations governing cosmological evolution. Future steps for this project might involve implementing

a Hamiltonian MC solver. Such a solver uses Hamiltonian equations with occasional kicks (ergodic evolution), rather than a Markov process (stochastic evolution) to decide amongst local configurations.

VI. APPENDIX

Acknowledgments

The authors are grateful Prof. Cora Dvorkin and Tansu Daylan for both their assistance on the project and their wonderful teaching.

-
- [1] Tojeiro, R. 2006. *Understanding the Cosmic Microwave Background Temperature Spectrum*
 - [2] Levine, E. 2015. *Numerical Simulations*
 - [3] Mackey et al., 2013. *emcee: The MCMC Hammer*
 - [4] Heavens, A. 2001. *Statistics in Cosmology*
 - [5] Planck Collaboration: *Planck* 2015 results. XIII
 - [6] www.pla.esac.esa.int

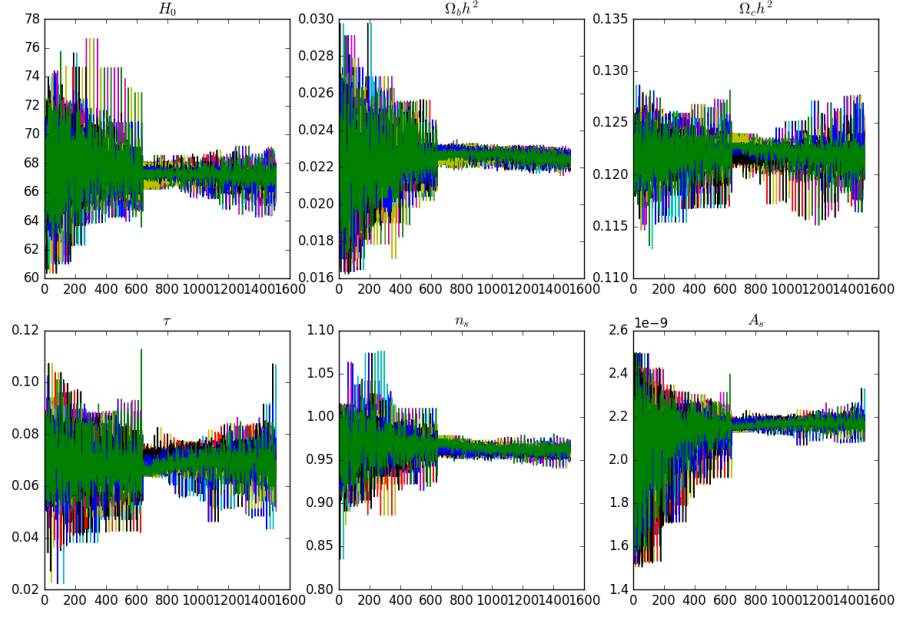


FIG. 1: Walker Plot for the 16-walker, 1600-step trial

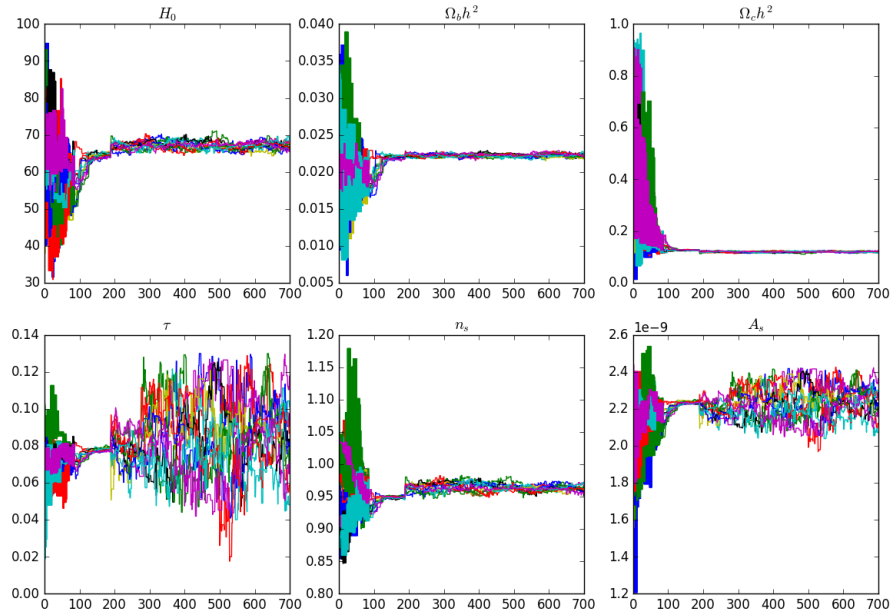


FIG. 2: Walker Plot for the 12-walker, 700-step trial

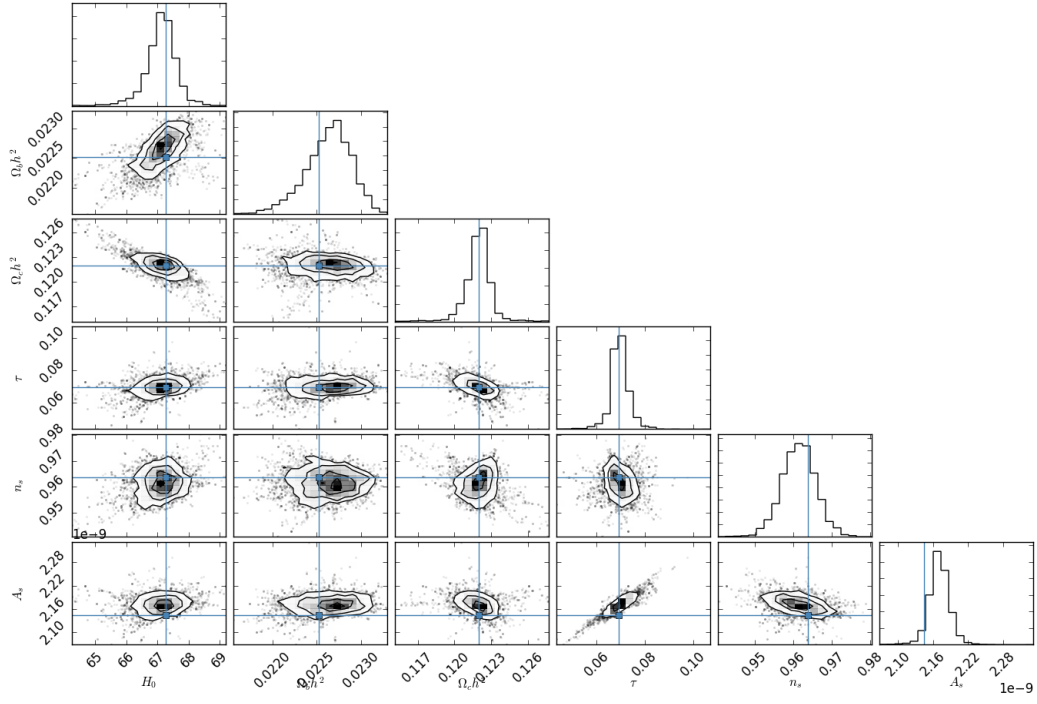


FIG. 3: Corner Plot for the 16-walker, 1600-step trial

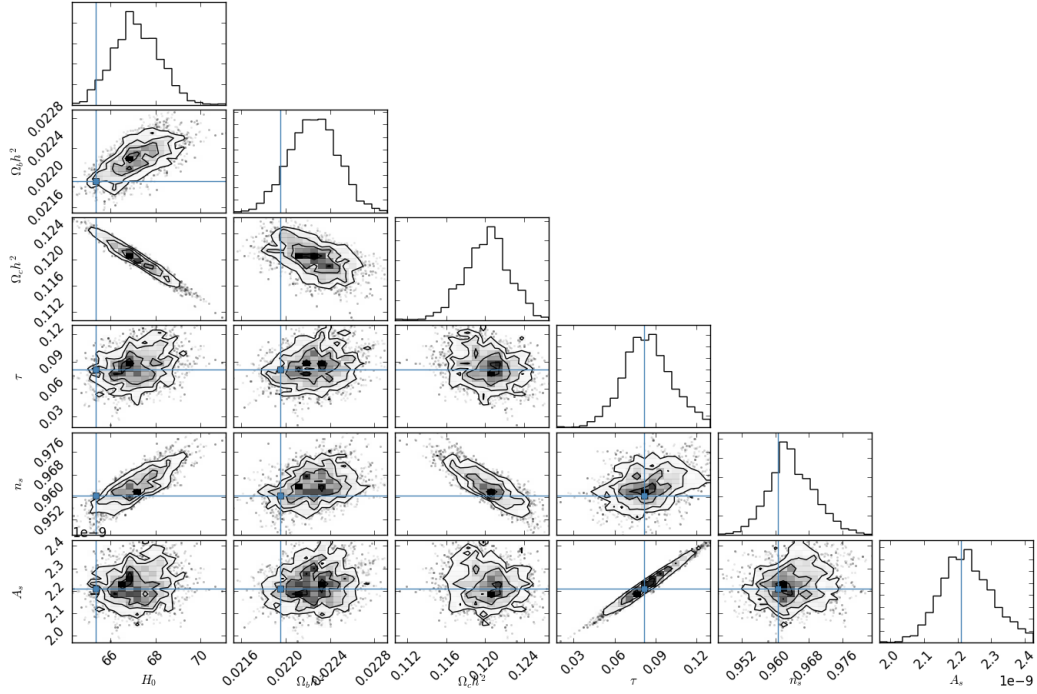


FIG. 4: Corner Plot for the 12-walker, 700-step trial

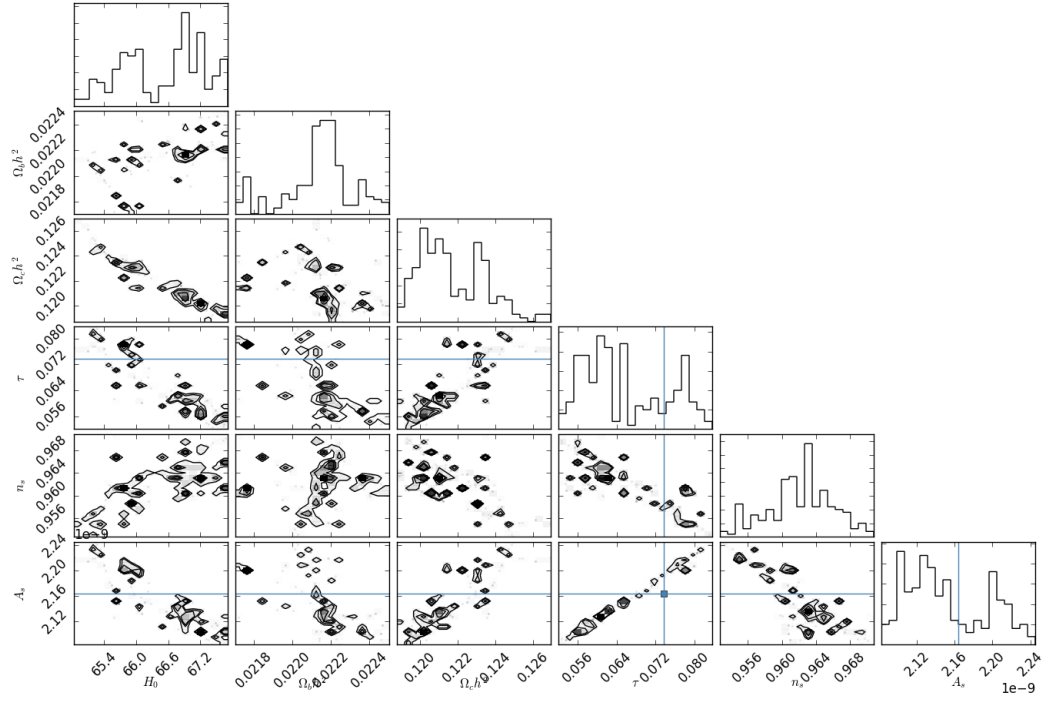


FIG. 5: A Bad Run: Corner Plot for the 12-walker, 200-step trial. Note the disjoint regions, which indicate the walkers have not had sufficient time to explore parameter space and settle near an optima.

TABLE II: Estimates of Cosmological Parameters from a 12-walker, 700-step and 16-walker, 1600-step MCMC trial. Uncertainties are given at the 68% level.

Parameter	12-walker run		16-walker run	
	Max Likelihood Estimate	68% <i>bounds</i>	Max Likelihood Estimate	68% <i>bounds</i>
H_0	65.22	54.14 – 76.30	67.27	64.67 – 69.86
Ω_b^2	.02193	.01782 – .02604	.02252	.02033 – .02471
$\Omega_c h^2$.1389	0 – .3053	.1220	.1193 – .1247
τ	.08178	.04887 – .14469	.06919	.05557 – .08281
n_s	.9602	.9210 – .9995	.9637	.9382 – .9894
A_s	2.212×10^{-9}	$2.008 \times 10^{-9} - 2.415 \times 10^{-9}$	2.144×10^{-9}	$1.954 \times 10^{-9} - 2.334 \times 10^{-9}$