Simple Parameter Estimation with PolyChord and GetDist

Raul A. Monsalve: raul.monsalve@physics.mcgill.ca

August 29, 2018

1 Description

Here we show a simple exercise of parameter estimation using the Python versions of PolyChord and GetDist.

PolyChord [1][2][1, 2] is *"a Bayesian inference tool for the simultaneous calculation of evidences and sampling of posterior distributions"*. It uses the Nested Sampling technique [3, 4] and can robustly and efficiently explore multi-mode posteriors. Due to all these properties, this algorithm and code are becoming popular statistical tools in physics [5].

GetDist [3] is *"a Python package for analyzing Monte Carlo samples"*. This package is a natural extension to PolyChord for presenting the results and elegantly comparing those from different analyses. It has been widely used in recent CMB analyses in conjunction with the CosmoMC and CAMB packages [6].

We fit non-linear parameters to a simulated power law measurement that resembles the signal from diffuse synchrotron emission. We use three models for the measurement noise. This simple exercise is intended to be an end-to-end demonstration of parameter estimation using the Nested Sampling technique and popular modern tools. It can be easily adapted to analyze real data and evaluate complex linear and non-linear models.

2 Model

We simulate the noiseless synchrotron power law with the equation:

$$\text{noiseless simulation} = T_{75}\left(\frac{\nu}{75 \ \text{MHz}}\right)^{\beta} \quad [\text{K}] \tag{1}$$

We use the values $T_{75} = 1500$ K for the brightness temperature at 75 MHz, and $\beta = -2.5$ for the spectral index, which broadly represent a measurement off the Galactic plane. Our frequency vector spans $\nu = 50 - 100$ MHz with a resolution of 1 MHz. We then add Gaussian noise to this simulation. We test three cases: 1) Noise with spectrally-flat standard deviation $\sigma = 2$ K, 2) Noise with spectrally-flat standard deviation $\sigma = 1$ K, and 3) Noise with frequency-dependent standard deviation given by:

$$\sigma = \left(\frac{\nu}{75 \ \text{MHz}}\right)^{-2.5} \quad [\text{K}]. \tag{2}$$

[1] https://ccpforge.cse.rl.ac.uk/gf/project/polychord/
[2] https://polychord.co.uk/
[3] https://getdist.readthedocs.io/en/latest/index.html

In this last model, the standard deviation also follows a power law and has a value of 1 K at 75 MHz. This is intended to mimic more closely real measurement noise that follows the radiometer equation.

Figure 1 shows the simulated brightness temperature power law and the three models of noise standard deviation.

3 Parameter Estimation

We fit Equation 1 to the noisy simulated data, where the non-linear fit parameters are T_{75} and β.

To compute the model evidence and (our objective here) map the parameter posterior distributions, PolyChord maximizes the log-likelihood function. For Gaussian frequency-independent noise with standard deviation σ (such as our first two noise models), the log-likelihood function is given by:

$$logL = -\frac{1}{2}\sum_{i=1}^{N}\left(\frac{d_i - m_i}{\sigma}\right)^2 - \frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2). \tag{3}$$

where d is the measured (simulated) spectrum, m is the model (Equation 1), and N is the number of data points.

In general, frequency-dependent Gaussian noise with non-zero channel-to-channel correlations is characterized in terms of its covariance matrix C. In this general case (which represents our third noise model), the log-likelihood function is given by:

$$logL = -\frac{1}{2}(d-m)^T C^{-1}(d-m) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|C|. \tag{4}$$

4 Results

Figure 2 presents the results. It was produced using GetDist from the samples obtained by exploring the parameter space and maximizing the log-likelihood function with PolyChord. The bottom-left panel shows the 68% and 95% joint probability contours for T_{75} and β, while the other panels show the 1D marginalized distributions, for each of the noise models evaluated.

PolyChord provides functions to quickly determine parameter limits from their probability density. 68% limits are presented in Table 1.

We see in the figure and table that, as expected, when the (frequency-independent) noise standard deviation is reduced from 2 K to 1 K, the limits become tighter. However, the improvement is not linear with respect to the noise level. Some of the non-linearity could be due to the imperfect accuracy of the limits themselves, rather than a systematic effect.

We also see in the figure that in the cases of frequency-independent noise (which is modeled correctly in the log-likelihood function as frequency-independent), there is a correlation in the estimates of T_{75} and β (elliptical contours orientated in diagonal). This can be intuitively understood by noticing that, for a fixed measured spectrum, if the value of T_{75} is increased, the best-fit spectrum will try to compesate by preferring a smaller (in absolute value) spectral index. In this way, the model will match better the data at low frequencies, which is where the dominant (largest) residuals tend to occur.

On the other hand, in the case where the measurement noise follows a power law (which is modeled correctly in the log-likelihood function as a power law), there is no obvious correlation between the estimates of T_{75} and β (blue contour not orientated in diagonal). This can be understood by realizing that an increase in T_{75} does not necessarily require a smaller (in absolute value) β, because at low frequencies the data have higher noise than in the previous cases, and the model can still be consistent with the data without needing a systematic decrease (in absolute value) in β.

In Figure 2 we also notice that in the cases with spectrally-flat noise, the estimates seem to be biased low, i.e., lower values of T_{75} and β are preferred relative to the true input values. On the other hand, the joint PDF of T_{75} and β for power-law noise seems to be centered close to the input values, i.e., the input values are well within the 68% limits. Since the noise levels considered are small, the potential biases in the cases of flat noise are very small ($\Delta\beta \approx 0.001$) and, moreover, from only two cases ($\sigma = 1$ K and $\sigma = 2$ K) we cannot confidently conclude that this is a real effect.

Finally, Figure 3 shows the chains of posterior samples accepted during the Nested Sampling computations. The chains require few samples compared with other algorithms, and converge exponentially to the best fit due to a step in the calculation *that compreses the prior exponentially* [2].

Table 1: Means and 68% limits obtained from the parameter PDFs.

	T_{75}			β		
	-68%	mean	$+68\%$	-68%	mean	$+68\%$
$\sigma = 2$ K	1495.81	1499.67	1503.31	-2.510	-2.501	-2.492
$\sigma = 1$ K	1498.04	1499.59	1501.60	-2.506	-2.501	-2.497
$\sigma = (\nu/75\ \ \mathrm{MHz})^{\beta}$ K	1498.25	1499.90	1501.27	-2.504	-2.500	-2.494

References

[1] Handley, W. J., Hobson, M. P., & Lasenby, A. N., ArXiv: 1502.01856.

[2] Handley, W. J., Hobson, M. P., & Lasenby, A. N., ArXiv: 1506.00171.

[3] Skilling, J., AIP Conference Proceedings 2004, 735, 395, doi:10.1063/1.1835238.

[4] Feroz, F., & Skilling, J., ArXiv: 1312.5638.

[5] AbdusSalam, S. S., ArXiv: 1710.10785

[6] Planck collaboration, ArXiv: 1807.06209.

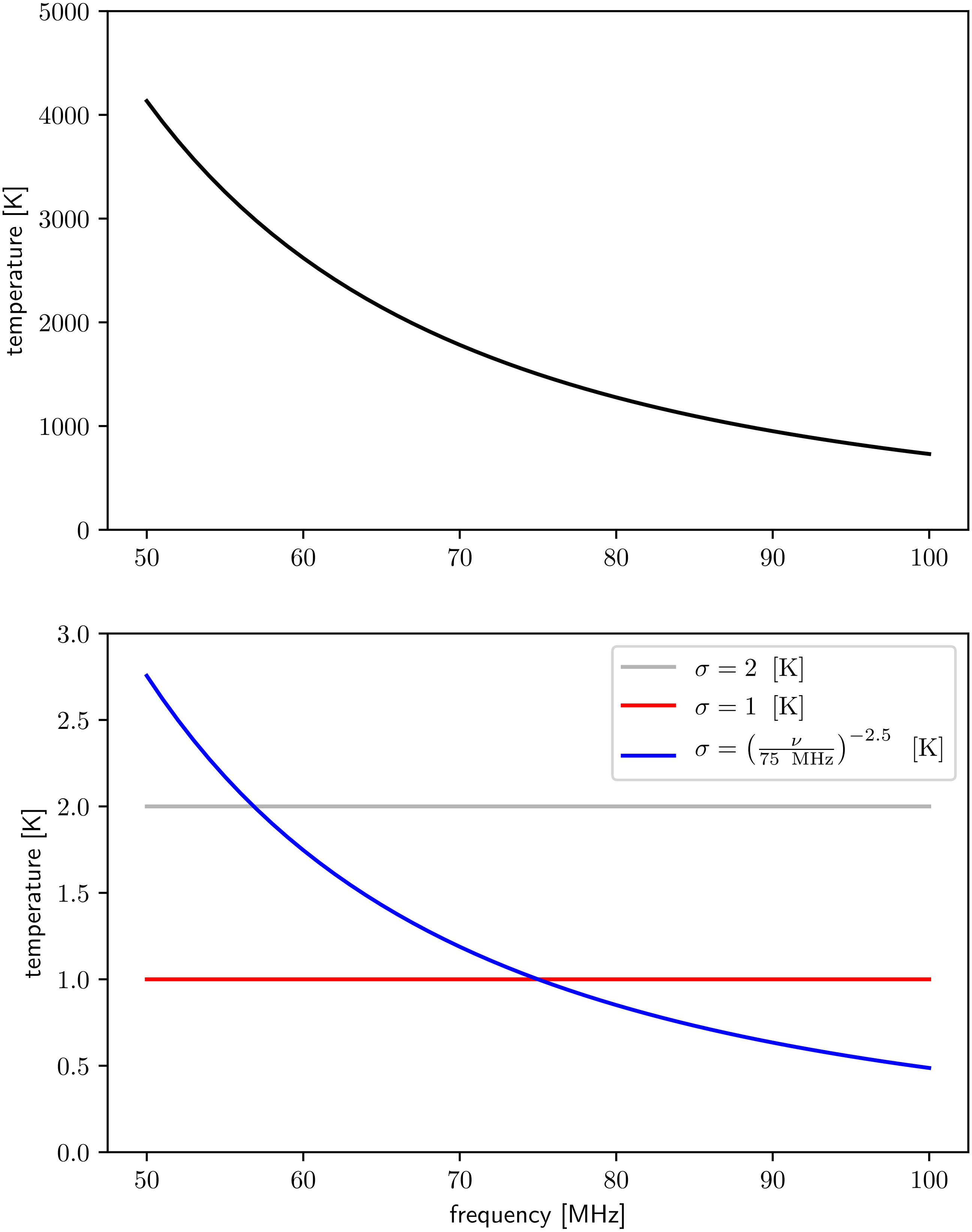

Figure 1: Top: Simulated brightness temperature that follows a power law and resembles the signal from synchrotron radiation when observing off the Galactic plane. Bottom: The three cases of noise standard deviation considered in the simulations.

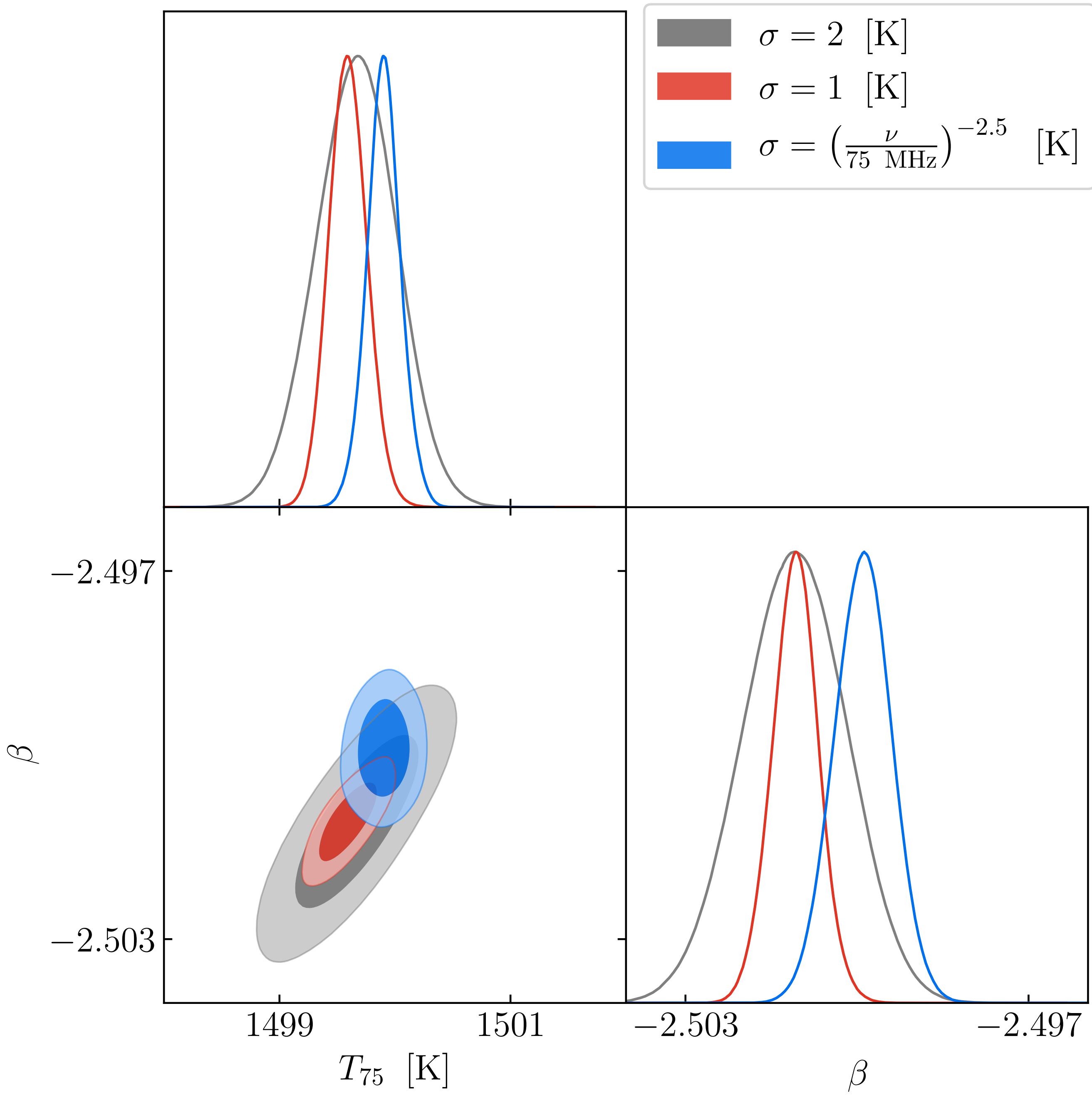

Figure 2: Results of the parameter estimation exercises, for the three models of noise standard deviation. For the calculation we used PolyChord. For the presentation we used GetDist.

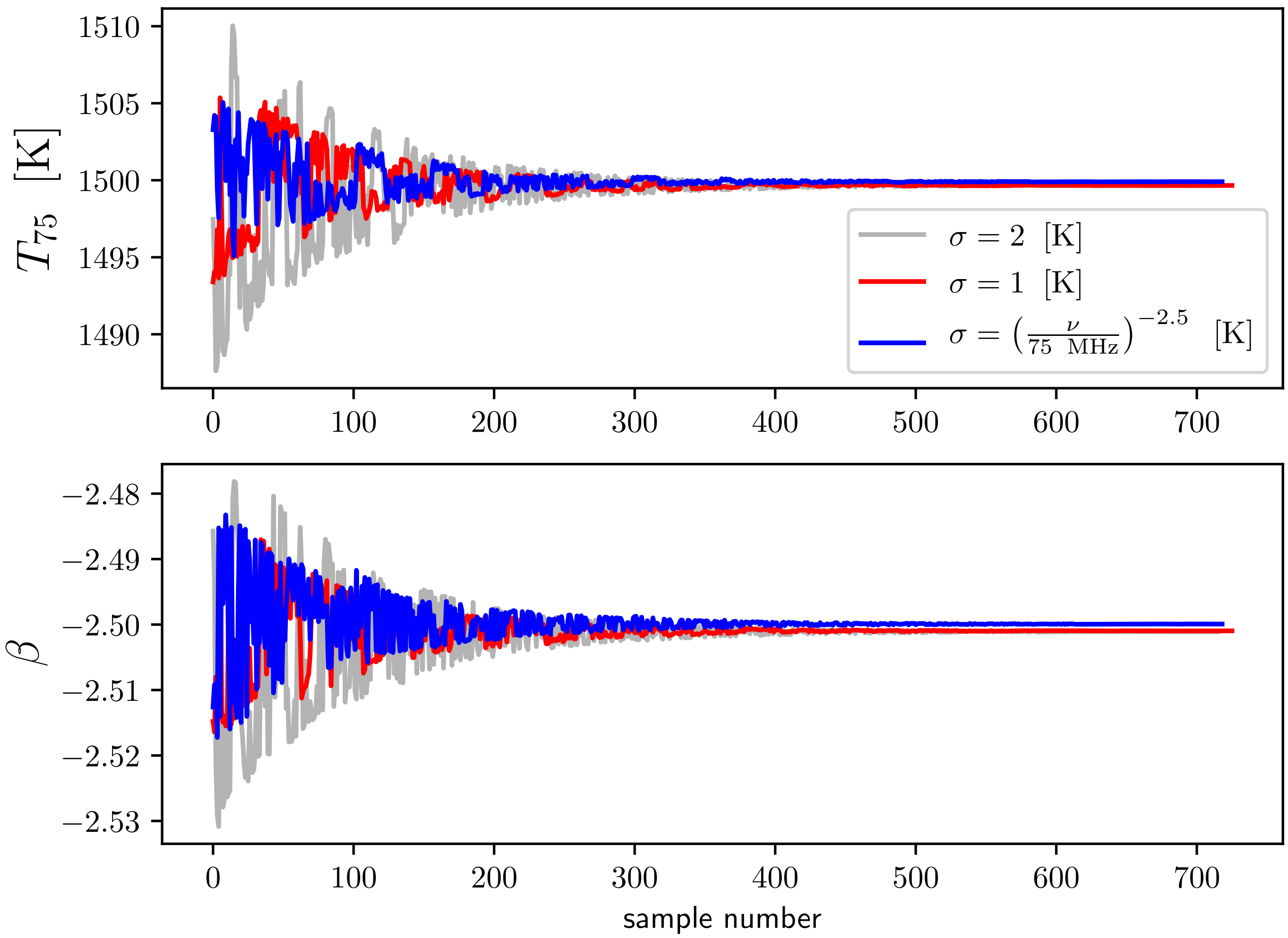

Figure 3: Chains of accepted samples in the Nested Sampling exploration of our parameter space using PolyChord. The x-axis also represents time, evolving toward the right.