PSA64 Signal Loss: Effect of Prior

Carina Cheng

May 18, 2018

1 Overview

The revised PAPER-64 analysis computes an upper limit on EoR using a Bayesian approach, namely:

$$p(P_{\rm in}|\widehat{P}_{\rm out}) \propto p(\widehat{P}_{\rm out}|P_{\rm in})\, p(P_{\rm in}), \tag{1}$$

where $P_{\rm in}$ is the EoR injection amplitude used in our signal loss simulations and $P_{\rm out}$ is the output power spectrum values (of data plus EoR) that result from bootstrapping. Our simulations produce a 2D transfer function ($P_{\rm in}$ vs. $\widehat{P}_{\rm out}$) over a range of injection amplitudes (the likelihood function), and we simply read off the likelihood function for a $P_{\rm out}$ level that matches our data-only power spectrum value (the posterior). See Cheng et al. (2018) for more details.

Our final posteriors (functions of $P_{\rm in}$ are dependent on our choice of *prior*. In our simulations, we populate our transfer function by injecting EoR signals uniformly in log-space. A common prior, however, is often a uniform prior, because it is often an uninformative prior (i.e. assuming all EoR amplitudes are equally likely). A unifoJeffreysrm prior would be flat vs. $P_{\rm in}$.

But is a uniform prior actually uninformative for the PSA64 signal loss analysis? We can check this by computing the Jeffreys prior, which is always a true uninformative prior (one of its key features is that it is invariant under reparameterization).

2 Jeffreys Prior Derivation

For simplicity in notation, let's rename our variables: $x = P_{\rm in}$ and $y = \widehat{P}_{\rm out}$. The Jeffreys prior is defined as:

$$p(x) \propto \sqrt{\left\langle \left(\frac{\partial \mathcal{L}}{\partial x} \right)^2 \right\rangle}, \tag{2}$$

where

$$\mathcal{L} = log\, p(y|x), \tag{3}$$

or the log of the probability of obtaining y given x (obtaining $P_{\rm out}$ given a $P_{\rm in}$). We model this probability as a Gaussian, namely:

$$p(y|x) = \frac{1}{\sigma(x)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}(x)}{\sigma}\right)^2} \tag{4}$$

where σ is the standard deviation of the y-values and $\bar{y}$ is the mean of the y-values. Both are functions of x. Now we do the math:

$$\begin{aligned}
\mathcal{L} &= log\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}}{\sigma}\right)^2}\right] &(5)\\
&= log\left[\frac{1}{\sigma\sqrt{2\pi}}\right] - \frac{1}{2}\Big(\frac{y-\bar{y}}{\sigma}\Big)^2 &(6)\\
&= -log\,\sqrt{2\pi} - log\,\sigma - \frac{1}{2}\Big(\frac{y-\bar{y}}{\sigma}\Big)^2 &(7)
\end{aligned}$$

We take the partial derivative with respect to x, keeping in mind that both σ and $\bar{y}$ are functions of x. Then we square the entire quantity.

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial x} &= -\frac{1}{\sigma}\frac{\partial\sigma}{\partial x} + \Big(\frac{y-\bar{y}}{\sigma^2}\Big)\frac{\partial\bar{y}}{\partial x} + \Big(\frac{(y-\bar{y})^2}{\sigma^3}\Big)\frac{\partial\sigma}{\partial x} &(8)\\
\Big(\frac{\partial\mathcal{L}}{\partial x}\Big)^2 &= \frac{1}{\sigma^2}\Big(\frac{\partial\sigma}{\partial x}\Big)^2 - \Big(\frac{2(y-\bar{y})}{\sigma^3}\Big)\frac{\partial\sigma}{\partial x}\frac{\partial\bar{y}}{\partial x} - \Big(\frac{2(y-\bar{y})^2}{\sigma^4}\Big)\Big(\frac{\partial\sigma}{\partial x}\Big)^2\\
&+ \Big(\frac{(y-\bar{y})^2}{\sigma^4}\Big)\Big(\frac{\partial\bar{y}}{\partial x}\Big)^2 + \Big(\frac{2(y-\bar{y})^3}{\sigma^5}\Big)\frac{\partial\sigma}{\partial x}\frac{\partial\bar{y}}{\partial x} + \Big(\frac{(y-\bar{y})^4}{\sigma^6}\Big)\Big(\frac{\partial\sigma}{\partial x}\Big)^2 &(9)
\end{aligned}$$

Next, taking the expectation value removes all terms with odd powers of $(y-\bar{y})$ because those moments of a Gaussian are 0. Additionally, the second moment is simplified as $\langle(y-\bar{y})^2\rangle = \sigma^2$ and the fourth moment is simplified as $\langle(y-\bar{y})^4\rangle = 3\sigma^4$.

$$\begin{aligned}
\Big\langle\Big(\frac{\partial\mathcal{L}}{\partial x}\Big)^2\Big\rangle &= \frac{1}{\sigma^2}\Big(\frac{\partial\sigma}{\partial x}\Big)^2 - \frac{2}{\sigma^2}\Big(\frac{\partial\sigma}{\partial x}\Big)^2 + \frac{1}{\sigma^2}\Big(\frac{\partial\bar{y}}{\partial x}\Big)^2 + \frac{3}{\sigma^2}\Big(\frac{\partial\sigma}{\partial x}\Big)^2 &(10)\\
&= \frac{1}{\sigma^2}\Big(2\Big(\frac{\partial\sigma}{\partial x}\Big)^2 + \Big(\frac{\partial\bar{y}}{\partial x}\Big)^2\Big) &(11)
\end{aligned}$$

Finally, the Jeffreys prior becomes:

$$p(x) \propto \sqrt{\frac{1}{\sigma^2}\Big(2\Big(\frac{\partial\sigma}{\partial x}\Big)^2 + \Big(\frac{\partial\bar{y}}{\partial x}\Big)^2\Big)} \tag{12}$$

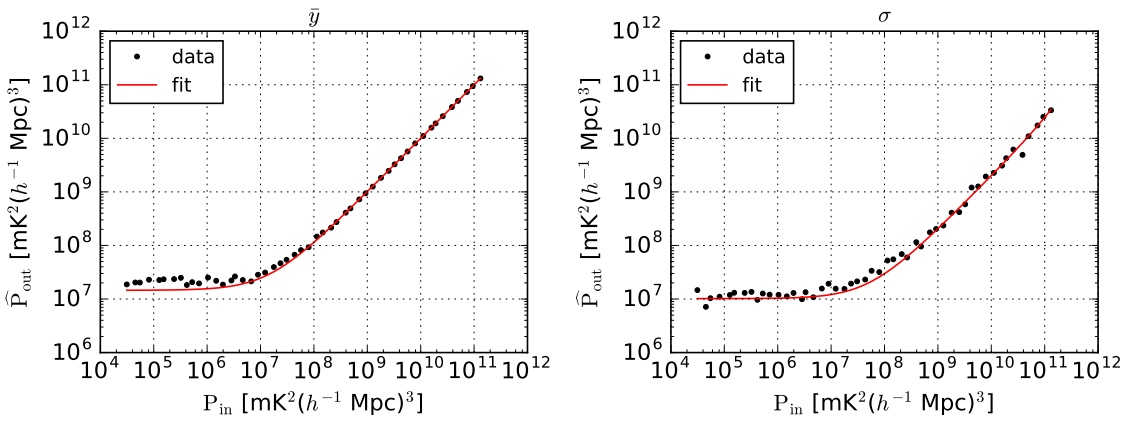

Figure 1: For each injection level, the mean of P_{out} (left plot) and standard deviation of P_{out} (right plot) are shown as the black points. Polynomials are fit to each (red) to describe how $\bar{y}$ and σ evolve with x (injection level), respectively, for the computation of the Jeffreys prior.

3 Jeffreys Prior for PSA64

As a simple test case and to get a general picture of what the Jeffreys prior looks like for PSA64, we fit analytical functions to σ (standard deviation of P_{out}) and $\bar{y}$ (mean of P_{out}) as a function of x (injection level) for one k-value and "I" weighted PSA64 data. For example, Figure 1 shows these quantities (black points) as well as the polynomial fits (red). We note that their general shapes are expected to be similar across all PSA64 analysis, even though we are only showing one specific example.

With functional forms for σ and $\bar{y}$ in hand, we can compute Equation 12 easily. The final Jeffreys prior is shown in Figure 2. It is interesting that the prior is *not flat* in shape, meaning that **the uniform prior is informative**.

4 Effect of Different Priors on PSA64

In this section we briefly show how three different choices of prior affect PSA64 posteriors and power spectra. The three priors are:

- uniform prior
- Jeffreys prior
- log-uniform prior

In Figure 3, we show the posterior distributions obtained for one k-value using the three different priors. Most notably, the uniform prior (left column) up-weights high injections

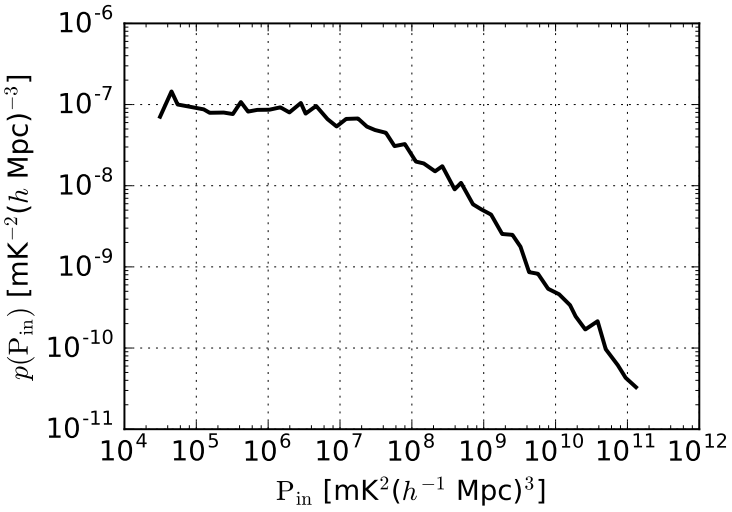

Figure 2: The computed Jeffreys prior shape for PSA64.

greatly (because to enforce a uniform prior, we multiply each 'column' of transfer function (each P_{in}) by the effective bin width, and the bins are huge moving to the right), overshadowing the fact that the Gaussian tails (per injection) drop off at low P_{out} values (i.e. the prior increases the amplitude of these tails). The Jeffreys prior suppresses this behavior significantly, but still gives quite a bit of weight to the high injections.

In Figure 4, we show the PSA64 power spectrum results (for empirical inverse covariance weighting) for the three cases. The thermal noise prediction (green), data points and errors (black/gray), and thermal noise errors on the data points (gray shaded) are identical (i.e. prior to signal loss estimation). The solid red and solid blue are the upper limits on EoR using the full injection framework, for the weighted and unweighted cases respectively. It is interesting that they change significantly depending on the choice of prior.

5 Takeaways

My personal takeaways and opinions based on this investigation into priors are:

- The uniform prior does not seem to be a good choice, as it substantially alters results and now we know it's not uninformative.
- Investigating noise simulations (Figure 5) suggests that a log-uniform prior should be used over the other two options, as it produces power spectrum results for the uniform weighted case that agree extremely well with the analytic prediction for noise. However, the fact that there seems to be a discrepancy between the $p(k)$ and $\Delta^2(k)$

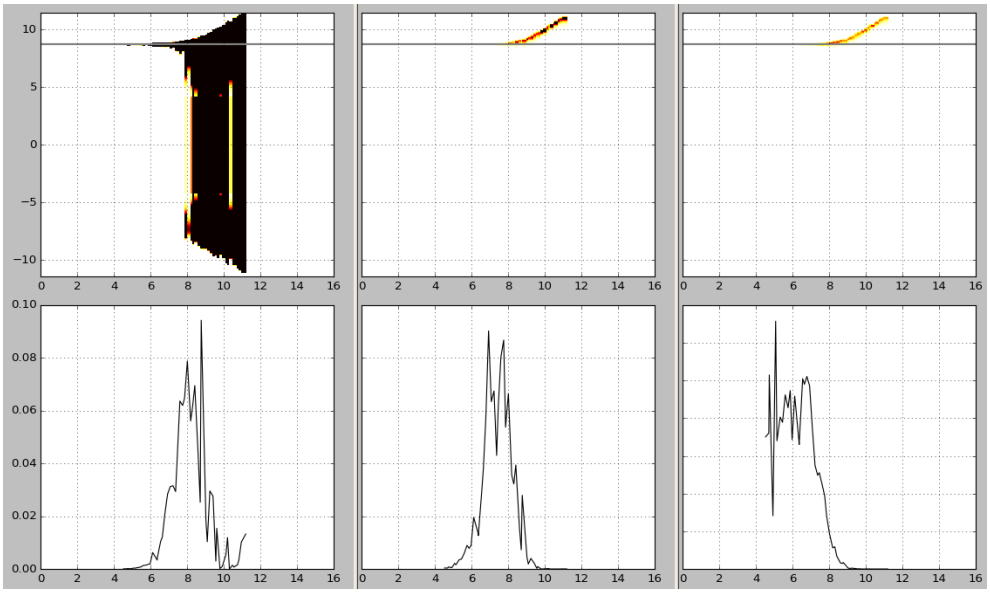

Figure 3: PSA64 likelihood functions (top) and posteriors (bottom) for $k = 0$, using a uniform prior (left), Jeffreys prior (middle), and log-uniform prior (right). Axes are in log-units and the x-axis is $P_{\rm in}$. The gray horizontal lines denote the data level.

versions (for all cases) suggests that folding power spectra values has a significant effect and that we may not be fully converged (i.e. more bootstraps might be needed to make a more robust decision).

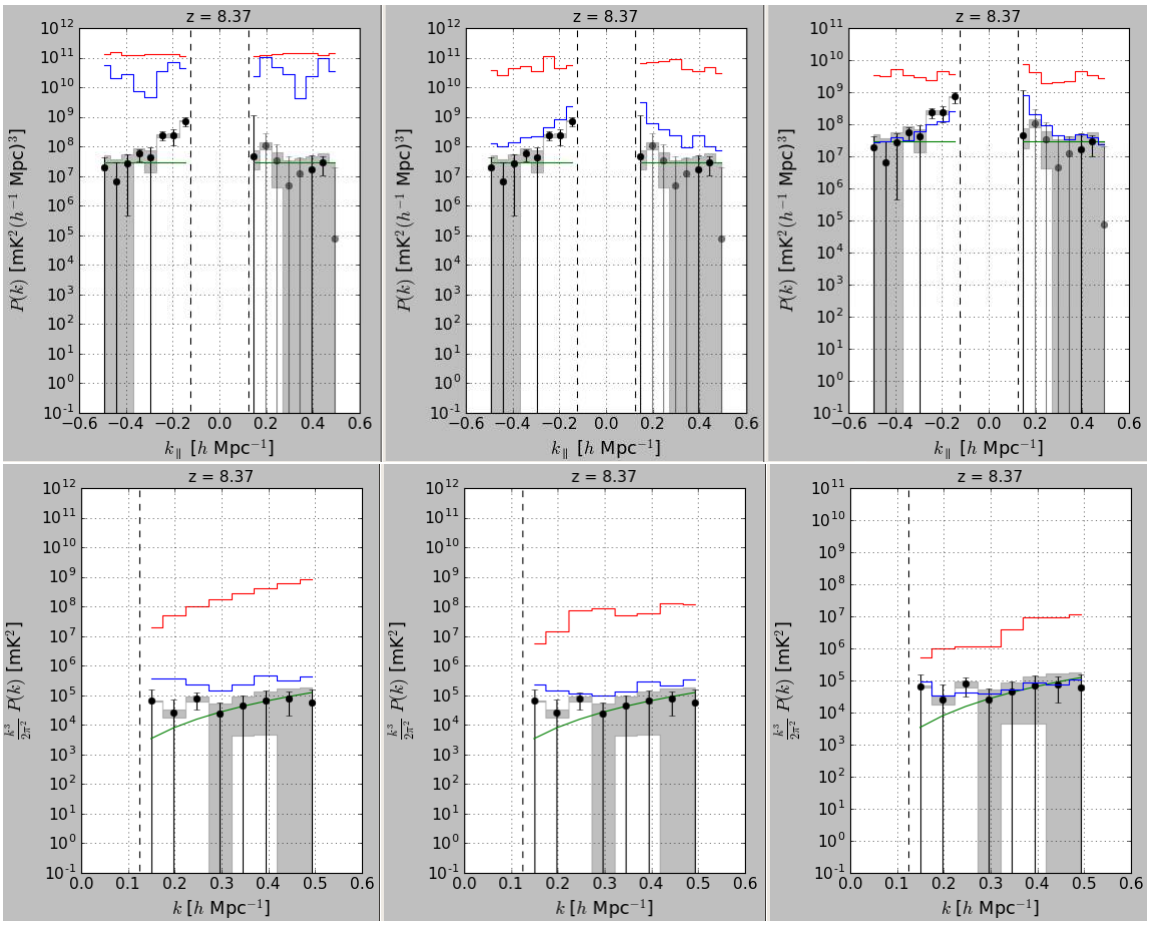

Figure 4: PSA64 power spectra using a uniform prior (left), Jeffreys prior (middle), and log-uniform prior (right). Empirical inverse covariance weighting is used. The 2σ upper limit on EoR is in red, and our 2σ thermal noise prediction is in green. The black/gray data points are positive/negative power spectrum values, with 2σ error bars, prior to estimating signal loss. The gray shaded regions are thermal noise errors plotted on top of the measured power spectrum values. Finally, the blue is the uniform-weighted upper limit on EoR using the full signal loss injection framework — it is interesting that this changes significantly depending on the choice of prior, especially for $p(k)$, which is not as tightly constrained as $\Delta^2(k)$ (i.e. the posteriors have longer tails to high injections).

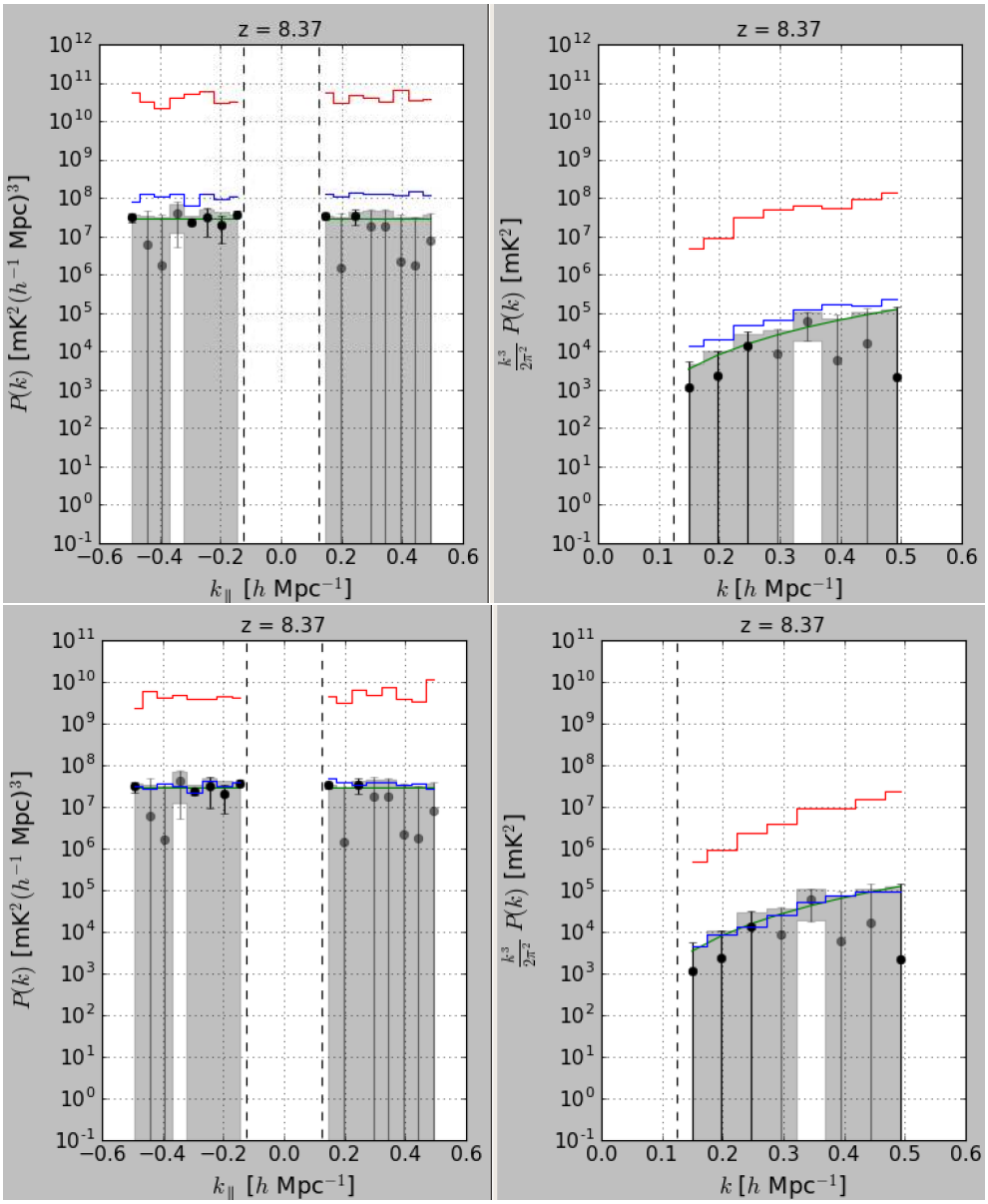

Figure 5: Pure noise power spectra using a Jeffreys prior (top), and log-uniform prior (bottom). Empirical inverse covariance weighting is used. The 2σ upper limit on EoR is in red, and our 2σ thermal noise prediction is in green. The black/gray data points are positive/negative power spectrum values, with 2σ error bars, prior to estimating signal loss. The gray shaded regions are thermal noise errors plotted on top of the measured power spectrum values. Finally, the blue is the uniform-weighted upper limit on EoR using the full signal loss injection framework — it seems to agree very well with the thermal noise prediction for the log-uniform prior case.