A SIMPLE BAYESIAN MODEL FOR GLOBAL 21CM DATA

EDGES MEMO

Steven G. Murray (on behalf of the EDGES collaboration)
School of Earth and Space Exploration
Arizona State University
Tempe, AZ
steven.g.murray@asu.edu

August 15, 2019

ABSTRACT

This memo seeks to develop a simple Bayesian model (or set thereof) for estimating phenomenological parameters of the 21 cm signal during Cosmic Dawn/EoR. From the set of possible particular models, it discusses benefits and trade-offs concerning computational efficiency and accuracy.

1 Introduction

The purpose of this memo is to specify an exact Bayesian model for parameter estimation in the case of EDGES. In fact, it will present a *family* of such models, with some specifications drawing particular attention. The most general model is really rather too general for our purposes here (eg. Liu2012). Instead, we assume that the data we have on-hand has already been averaged over LST, and is thus purely a function of frequency.

In §2 we will introduce the general form for such a likelihood, and our nomenclature and assumptions. Following this, we will attempt to draw conclusions.

2 Overview of Bayesian Parameter Estimation

Bayes' theorem is

$$\mathcal{P}(\vec{p}|D) = \frac{\pi(\vec{p})\mathcal{L}(D|\vec{p})}{\mathcal{P}(D)}, \tag{1}$$

where the LHS is called the "posterior" and denotes the probability of the statistical variables (or "parameters") $\vec{p}$ being a given value, given the data D. On the RHS, π is the called the "prior", and $\mathcal{P}(D)$ is called the "evidence", which we shall largely ignore for the rest of the memo, since MCMC is not sensitive to it. Finally, $\mathcal{L}$ is the "likelihood" of the data given specific values of the parameters.

This equation holds all the information we can ever use. Nevertheless, two cases are of potential interest to us. The first is the case in which some of the $\vec{p}$ is dependent on other elements of $\vec{p}$. For example, perhaps we are interested in soil temperature at a range of times. At each time, the temperature is a random variable, so $\vec{p}$ contains N_t parameters, labelled, say, T_i. However, we may know that these temperatures are drawn from a Gaussian distribution, but we are unsure of its variance. The variance then is another parameter in $\vec{p}$, upon which the $\vec{T_i}$ depend. This case is often called a "Bayesian graph" or "hierarchical model". In this case, letting the dependent parameters be $\vec{q}$, the prior can be written more specifically as

$$\pi(\vec{p}) = \pi'(\vec{p}')\pi''(\vec{q}|\vec{p}''), \tag{2}$$

where we have used a single prime to indicate the sub-vector of $\vec{p}$ which do not have any dependent parameters, and double prime to indicate its complement.

The other case that is of interest is that in which there is some subset of parameters, let us call them $\vec{\varphi}$, which we have no interest in. For example, the T_i from above would probably fall into this category. In fact, so would the precise values of the thermal noise[1].

The process of properly "ignoring" a subset of $\vec{p}$ is called "marginalization", which in detail is performed by integrating the posterior over the marginalized parameters. In practice it can be performed in two ways. Typically the simplest is to obtain the posterior via MCMC for all parameters, in the form of a *sample* of points from the posterior. Marginalization then amounts to projecting that sample onto the dimensions which are sought after. Let's call these $\vec{p}_\varphi$. In reality, this is a (potentially) numerically-efficient algorithm for computing

$$\mathcal{P}(\vec{p}_\varphi|D) = \int d\vec{\varphi}\, \frac{\pi(\vec{p}_\varphi, \vec{\varphi})\mathcal{L}(D|\vec{p}_\varphi, \vec{\varphi})}{\mathcal{P}(D)} \tag{3}$$

If the parameters $\vec{p}_\varphi$ are *a priori* independent of φ (which is the only case we will consider in this memo), we can re-write this as

$$\begin{aligned} \mathcal{P}(\vec{p}|D) &= \frac{\pi(\vec{p})\mathcal{L}'(D|\vec{p}, \vec{\varphi})}{\mathcal{P}(D)}, \quad \text{with} \\ \mathcal{L}'(D|\vec{p}, \vec{\varphi}) &= \int d\vec{\varphi}\, \pi(\vec{\varphi})\mathcal{L}(D|\vec{p}_\varphi, \vec{\varphi}) \end{aligned} \tag{4}$$

where for ease of notation we have understood that $\vec{p} = \vec{p}_\varphi$. While often it will be the case that the first approach will be more efficient (especially if the integral in Eq. 4 is not separable), there are cases in which performing the integral directly is more efficient. One such case is that of the exact values of the thermal "noise". Instead of deriving estimates of each of them, one typically pre-marginalizes by defining $\mathcal{L} \equiv \mathcal{L}'(D|\vec{p}, \vec{n})$. This is suitable since it is typically assumed that thermal noise is normally-distributed and independent, rendering Eq. 4 quite manageable (in fact, much *more* manageable than the likelihood of delta-functions that would ensue otherwise).

One of the primary questions that this memo seeks to answer is what difference some of the nuisance parameters make to the "covariance function" of the data. This is an ill-defined question to begin with. However, perhaps it is more suitably stated as: "assuming the likelihood is a multivariate normal distribution (or well approximated by one), and that the process of marginalization over some nuisance parameters $\vec{\varphi}$ does not change the Gaussianity of the effective likelihood ($\mathcal{L}'$), what are the covariance matrix Σ, and mean vector μ, that uniquely characterize the likelihood after marginalisation?". In particular, if these assumptions are true, *and the covariance function does not depend on any of* $\vec{p}$, due to the symmetrical properties of the normal distribution, the covariance function that defines the likelihood may then be interchangeably interpreted as the covariance "of the data". This would allow storing a single pre-computed covariance matrix and using it exlusively to perform parameter inference, according to the likelihood:

$$\ln \mathcal{L}(D|\vec{p}) = (D - \vec{\mu})\Sigma_\varphi^{-1}(D - \vec{\mu})^\dagger, \tag{5}$$

which has well-studied solutions.

Despite none of these assumptions being particularly likely, it is not unlikely that within a suitable small ball around the maximum likelihood estimate, the likelihood (even the effective likelihood) could be well-approximated as multivariate normal, and that in this regime it may also be roughly independent of $\vec{p}$. Given the simplicity of the resulting problem if these approximations hold, it is worth attempting to determine if they are valid for our particular problem.

3 Likelihood Construction

In this memo, as previously stated, we will exclusively consider data that is pre-averaged over LST and is thus a function of frequency only. Henceforth, we let x_j be the j_{th} frequency-component of our *model*. This is understood to be a random variable – it depends deterministically on the set of parameters $\vec{p}$ that we wish to constrain, and non-deterministically on the set of nuisance parameters $\vec{\varphi}$[2].

[1] This observation leads one to ponder the extreme limit of this line of thinking: if *everything* "unknown" is considered as a parameter in $\vec{p}$, then $\mathcal{L}$ is a set of delta-functions at the set of precise values of $\vec{p}$ which can reproduce D. However, at the same time, π is modified to include the probability of obtaining those sets of parameters, ultimately maintaining the equality of Bayes' theorem (after marginalising, as we shall see).

[2] Though post-marginalization is equivalent to pre-marginalization (i.e. the "direct" integration given by Eq. 4), for the remainder of the memo we consider $\vec{\varphi}$ to consist only of those parameters that we wish to *pre*-marginalize. That is, they will not enter into any MCMC sampling routine.

We split both $\vec{p}$ and $\vec{\varphi}$ into several subsets (and potentially, further subdivisions are possible), explicitly

$$\vec{p} = \{\vec{p}_{21}, \vec{p}_{\rm fg}, \vec{p}_{\rm beam}\} \tag{6}$$
$$\vec{\varphi} = \{\vec{\varphi}_{\rm fg}, \vec{\varphi}_{\rm beam}, \vec{n}\}, \tag{7}$$

where n is the thermal noise, which we exclusively consider to be nuisance parameters, and we have omitted 21 cm signal parameters from $\vec{\varphi}$ as ostensibly these will never be nuisance parameters. It is clear that each of the sets (21 cm, fg and beam) will be independent of each other. However, it is not clear whether n will be independent of any of the other sets, as the level of thermal noise depends on both the instrument and sky.

The model can then be parameterized by

$$x_j(\vec{p}) = \frac{1}{N_t}\sum_{i=1}^{N_t}\int d^2\theta A(\theta, t_i, \nu_j, \vec{p}_{\rm beam}, \vec{\varphi}_{\rm beam}) I(\theta, \nu_j, \vec{p}_{\rm sky}, \vec{\varphi}_{\rm sky}) + \vec{n}_{ij}(\vec{p}, \vec{\varphi}_{\rm sky}, \vec{\varphi}_{\rm beam}), \tag{8}$$

where θ is the angular location on the sky (assumed to be in equatorial co-ordinates, eg. RA and DEC), A is the primary beam of the antenna, and I is the intensity of the sky.

The likelihood, as a function of $\vec{p}$, is equivalently the probability density function of $\vec{x}$ (which in general is not trivial to determine). Note that if φ solely consists of $\vec{n}$, i.e. we only consider the thermal noise to be nuisance, and if that thermal noise is Gaussian-distributed for any given choice of $\vec{p}$, then the likelihood is multivariate normal. We will consider this simplest of cases in detail in the next subsection, and then move to more involved cases.

3.1 No nuisance parameters

Appendix A.1 gives the derivation of the full joint-model for x_j from first-principles. Summarily, it is equivalent to Eq. 8 with

$$n_{ij} \sim \mathcal{N}\left[0, \frac{\kappa}{\sqrt{\Delta\nu}}\int d^2\theta\ A(\theta, t_i, \nu_j) I(\theta, \nu_j)\right]. \tag{9}$$

(i.e. a normally-distributed variable with mean zero and the given standard deviation). Importantly, there are no correlations between frequencies, and therefore the covariance is diagonal.

Since the noise component is mean-zero, the expectation is simply[3]

$$\mu(\vec{p}) = \frac{1}{N_t}\sum_{i=1}^{N_t}\int d\alpha d\delta A(ALT(\alpha, \delta, t_i), AZ(\alpha, \delta, t_i), \nu_j, \vec{p}_{\rm beam}) I(\alpha, \delta, \nu_j, \vec{p}_{\rm sky}). \tag{13}$$

This is a function only of ν, and we denote it

$$\mu(\vec{p}) = \mu_{21}(\vec{p}_{21}, \vec{p}_{\rm beam}) + \mu_{\rm fg}(\vec{p}_{\rm fg}, \vec{p}_{\rm beam}). \tag{14}$$

These components correspond to the usual phenomenological forms for the foregrounds and 21 cm absorption trough.

As previously noted, since the noise is independent in frequency and time, the covariance is diagonal, and the standard deviation is given by

$$\begin{aligned}\sigma_j &= \frac{\kappa}{N_t\sqrt{\Delta\nu}}\sum_{i=1}^{N_t}\int d\alpha d\delta\ \ I(\alpha, \delta, \nu_j) A(ALT_i, AZ_i, \nu_j) \\ &= \frac{\kappa}{\sqrt{\Delta\nu}}\mu(\vec{p}).\end{aligned} \tag{15}$$

[3] We can be a little more explicit about the time-dependence of the model. In particular, the sky is (for our purposes) constant in time when expressed in (RA, DEC) co-ordinates (denoted α, δ). Altitude (ALT) and azimuth (AZ), for an observing site at latitude (γ) and LST (t) are given by

$$ALT = \sin^{-1}\left[\sin\delta\sin\gamma + \cos\delta\cos\gamma\cos(t-\alpha)\right] \tag{10}$$
$$AZ = -\tan^{-1}\left[-\frac{\sin(t-\alpha)}{\sin\gamma\cos(t-\alpha) - \cos\gamma\tan\delta}\right], \tag{11}$$

where AZ is measured Westward of South[4]. Thus in detail we have

$$A(\theta, t) = A(ALT(\alpha, \delta, t), AZ(\alpha, \delta, t)). \tag{12}$$

The likelihood in such a model is simply given by

$$\ln \mathcal{L}(\vec{p}, \kappa) = \sum_j -\ln \sigma(\vec{p})_j - \frac{(\vec{x}_j - \mu_j(\vec{p}))^2}{2\sigma_j^2(\vec{p})}$$
$$= -\sum_j \ln \kappa + \ln \mu_j(\vec{p}) + \frac{\Delta\nu(\vec{x}_j - \mu_j(\vec{p}))^2}{2\kappa^2\mu_j^2(\vec{p})}. \quad (16)$$

In this model, κ is best treated a free parameter to be fit.

Note that this model is the simplest possible – by assuming that simple 1D models for μ_{21} and $\mu_{\rm fg}$ can fully capture the averaged form of the observed sky, *there is no correlation between frequency bins*. This is exactly the fitting procedure currently used. There are however two caveats to this approach:

1. This formula assumes that we have accounted for *all* possible variable parameters in $\vec{p}$. If other variables in truth exist, but are not included in $\vec{p}$, then the model is *wrong*. The *correct* model in this case is that in which the neglected parameters form part of φ and are marginalized over (either during MCMC or beforehand). This is almost always most *easily* accomplished by actually running the MCMC again with all the parameters that are considered unknown. In the next section, however, we consider how we might estimate what the appropriate modified likelihood should be given that we would prefer to completely neglect the parameters.
2. The second caveat is that while under the stated assumptions Eq. 16 is a *valid* likelihood (if all unknown parameters are accounted for in $\vec{p}$), it is not necessarily the *best* one. That is, it does not necessarily make use of all available information optimally. One clear way in which this is true is that the time-average on the data removes information. Nevertheless, in this memo we are restricting ourselves to time-averaged data, so we will not yet consider that particular generalization. The second way in which information is potentially lost is in the priors. If the model for μ chosen is highly flexible, and not strongly founded on physics, then the priors on its parameters may be unnecessarily wide. Furthermore, it may not be clear whether there are *a priori* correlations between those parameters which may restrict the posterior. However, moving to a less-flexible model means that μ will be less likely to be able to match reality if any parameters are ignored. In this case, the considerations of the following section become important.

3.2 Nuisance Parameters

As in the previous section, in this section we will continue to assume that a purely phenomenological model for the sky temperature, as a function of only ν, provides a good-enough fit to the data. However, we will assume also that there are known random variables that are ignored, i.e. they form $\vec{\varphi}$. We can view these simply as altering the proper likelihood, given the sought-after $\vec{p}$.

The most general solution is to use Eq. 4. Then we have

$$\mathcal{L}' \propto \int d\vec{\varphi}\, \pi(\vec{\varphi}) \kappa^{-N_t} \prod_j^{N_t} \frac{1}{\mu_j(\vec{p}, \vec{\varphi})} \exp\left(-\frac{\Delta\nu(\vec{D}_j - \mu_j(\vec{p}, \vec{\varphi}))^2}{2\kappa^2\mu_j^2(\vec{p}, \vec{\varphi})}\right). \quad (17)$$

Remember that the likelihood is a pdf as a function of the *data*, D, not the parameters. Even so, it is difficult to see how one would evaluate this integral in many cases. The likelihood is *not* in general even Gaussian any longer.

Nevertheless, if we assume that the likelihood remains approximately Gaussian (at least close to the maximum likelihood estimate) we may characterize the likelihood solely by the mean and covariance. There are two ways to think about this. If we are certain that all physical effects have been accounted for (and thus our physical model of the sky reduces to a simple function $\mu(\vec{p})$), but we are not sure that our model for $\mu(\vec{p})$ adequate, then we modify the likelihood as follows. The model is given by

$$x_j = \mu_j(\vec{p}, \vec{\varphi}) + n_j. \quad (18)$$

The expectation of x_j is

$$E[x_j] = \int d\varphi\, \mu_j(\vec{p}, \vec{\varphi}) \pi(\vec{\varphi}), \quad (19)$$

but the variance is quite difficult to determine. Needless to say, it is almost certainly *not* proportional to $E[x_j]$, which makes the simple model invalid. Furthermore, it would not be enough to calculate the variance once – it will depend on the parameters $\vec{p}$ – unless one approximates the variance as constant around the maximum likelihood estimate, $\hat{p}$.

What happens if instead we understand that there is, say, some physical parameter that affects the beam, but which is not in our model explicitly? This is conceptually tricky, because to arrive at the flexible polynomial model, we

have already made the assumption that there are no other physical parameters that aren't being modeled explicitly. Essentially, the flexible polynomial model is a reparameterization of a theoretically-obtained model. What if one of the physical parameters were neglected (i.e. assumed to be known when it was uncertain)? Presumably, whatever model *would* have been produced by including that parameter is a model that is able to be fit by the polynomial model. So long as this is the case, then the flexible polynomial model with likelihood given by Eq. 16 is still valid – we are just not using the full information, so the posterior may be wider than we could achieve with physical parameters. What changes here is the physical interpretation (if we had made one) of the polynomial coefficients. Using Eq. 16, the interpretation is that the coefficients are a mapping from *the complete* set of possible physical parameters (instead of a limited subset in which some have been marginalised over). Given that this interpretation is unlikely to ever occur, we can probably be content with the values for the coefficients themselves.

In the case that we were interested in more physical models of the foregrounds and signal that gave detailed predictions of the beam as a function of position on the sky, Appendix A.2 gives the mean and covariance of the model, if some parameters (only of the beam at this point) were to be marginalized.

4 Conclusion

In this memo we have examined the ramifications of neglecting the uncertainty of some physical parameters in a Bayesian analysis of global spectra. We have shown that, as long as the flexible models for the spectra are *able* to model the realistic spectra, then a basic Gaussian model with diagonal covariance is valid. This validity is robust to uncertainties in the physical model, so long as the ability to map from whatever the physical parameters might be to the flexible parameters is maintained.

What exactly renders a flexible model "able" to model the realistic spectra is not entirely clear. It *is* clear that not every flexible model will do – the simple case of a polynomial model illustrates this: adding an extra polynomial coefficient creates a model which is in general incompatible with its simpler precedent. We showed that determining the effective likelihood (or even its Gaussianized mean and covariance) which produces the correct posterior, assuming that this single extra coefficient's uncertainty has been ignored, is highly non-trivial; it is far better to use the MCMC integration to determine the posterior directly in each case (perhaps comparing Bayesian evidence to decide whether the increase in complexity is warranted).

All of this points to ultimately using more physical models, which can be constrained more tightly by physical priors, and which use the full gamut of time and frequency information at our disposal.

A Derivation of Autocorrelation

A.1 Deterministic sky and beam

Let the electric field for the sky be labelled $\mathcal{E}$ and given by

$$\mathcal{E}(\theta) \sim \mathcal{N}(0, I(\theta, \nu, t)). \tag{20}$$

with the distribution being independent in θ, ν and t. Then the autocorrelation at time t (this is pre-FFT to frequency domain) is

$$\int d^2\theta d^2\theta' \mathcal{A}(\theta, \nu, t)\mathcal{A}^\dagger(\theta', \nu, t)\mathcal{E}(\theta, \nu, t)\mathcal{E}^\dagger(\theta', \nu, t), \tag{21}$$

with $\mathcal{A}$ the far-field pattern of the antenna and θ the position on the sky. We assume that the distribution of $\mathcal{E}$ is stationary w.r.t t over the timescales of the FFT to frequency.

Given that here $\mathcal{A}$ and I are considered deterministic, the expectation is simply

$$\begin{aligned} E[V] &= \int d^2\theta d^2\theta' \mathcal{A}(\theta)\mathcal{A}^\dagger(\theta') \cdot \mathrm{Cov}[\mathcal{E}(\theta), \mathcal{E}^\dagger(\theta')] \\ &= \int d^2\theta \, |\mathcal{A}|^2 \mathrm{Var}(\mathcal{E}(\theta)) \\ &= \int d^2\theta \, A(\theta) I(\theta) \end{aligned} \tag{22}$$

The variance is

$$\begin{aligned}\mathrm{Var}(V) &= \int d^2\theta d^2\theta' d^2\theta'' d^2\theta''' \mathcal{A}(\theta)\mathcal{A}^\dagger(\theta')\mathcal{A}(\theta''')\mathcal{A}^\dagger(\theta'')\mathrm{Cov}[\mathcal{E}\mathcal{E}^{\dagger\prime}, \mathcal{E}'''\mathcal{E}^{\dagger\prime\prime}] \\ &= \int d^2\theta d^2\theta' d^2\theta'' d^2\theta''' \mathcal{A}(\theta)\mathcal{A}^\dagger(\theta')\mathcal{A}(\theta''')\mathcal{A}^\dagger(\theta'') \left[\langle\mathcal{E}\mathcal{E}^{\dagger\prime\prime}\rangle\langle\mathcal{E}'''\mathcal{E}^{\dagger\prime}\rangle + \langle\mathcal{E}\mathcal{E}'''\rangle\langle\mathcal{E}^{\dagger\prime}\mathcal{E}^{\dagger\prime\prime}\rangle\right] \\ &= \left|\int d\theta A(\theta) I(\theta)\right|^2, \end{aligned} \tag{23}$$

where the second equality follows from Isserlis' theorem, and the final line uses the diagonality of the covariance of $\mathcal{E}(\theta)$ and separates integrals.

Note that the resulting distribution of V is *not* Gaussian (it is strictly positive). Indeed, its variance is large compared to its mean so it is quite far from Gaussianity. Nevertheless, for a typical visibility, $\sim 10^6$ of these samples are summed (as part of a Fourier Transform), so that according to the law of large numbers, the result is approximately Gaussian, with variance

$$\mathrm{Var}(V) \approx \left|\int d\theta A(\theta) I(\theta)\right|^2 / n_t. \tag{24}$$

A.2 Random beam

Let's now dispense with the assumption that the beam is deterministic, though we will maintain that the beam and sky are independent (and that the sky intensity is deterministic). Then the expectation is

$$\begin{aligned} E[V] &= \int d^2\theta d^2\theta' \langle\mathcal{A}(\theta)\mathcal{A}^\dagger(\theta')\rangle \mathrm{Cov}[\mathcal{E}(\theta), \mathcal{E}^\dagger(\theta')] \\ &= \int d^2\theta \, \langle A(\theta)\rangle I(\theta) \end{aligned} \tag{25}$$

And the variance is

$$\begin{aligned}\mathrm{Var}(V) &= \int d^2\theta d^2\theta' d^2\theta'' d^2\theta''' \mathrm{Cov}[\mathcal{A}\mathcal{A}^{\dagger\prime}\mathcal{E}\mathcal{E}^{\dagger\prime}, \mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\mathcal{E}'''\mathcal{E}^{\dagger\prime\prime}] \\ &= \int d^2\theta d^2\theta' d^2\theta'' d^2\theta''' \langle\mathcal{A}\mathcal{A}^{\dagger\prime}\mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\rangle\langle\mathcal{E}\mathcal{E}^{\dagger\prime}\mathcal{E}'''\mathcal{E}^{\dagger\prime\prime}\rangle - \langle\mathcal{A}\mathcal{A}^{\dagger\prime}\rangle\langle\mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\rangle\langle\mathcal{E}\mathcal{E}^{\dagger\prime}\rangle\langle\mathcal{E}'''\mathcal{E}^{\dagger\prime\prime}\rangle \\ &= \int d^2\theta d^2\theta' d^2\theta'' d^2\theta''' \langle\mathcal{E}\mathcal{E}^{\dagger\prime}\rangle\langle\mathcal{E}'''\mathcal{E}^{\dagger\prime\prime}\rangle \left[\langle\mathcal{A}\mathcal{A}^{\dagger\prime}\mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\rangle - \langle\mathcal{A}\mathcal{A}^{\dagger\prime}\rangle\langle\mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\rangle\right] + \langle\mathcal{A}\mathcal{A}^{\dagger\prime}\mathcal{A}'''\mathcal{A}^{\dagger\prime\prime}\rangle\langle\mathcal{E}\mathcal{E}^{\dagger\prime\prime}\rangle\langle\mathcal{E}'''\mathcal{E}^{\dagger\prime}\rangle \\ &= \int d^2\theta d^2\theta' \mathrm{Var}(\mathcal{E})\mathrm{Var}(\mathcal{E}') \left[\langle A\rangle\langle A'\rangle + \mathrm{Cov}[A, A']\right] \\ &= \int d^2\theta d^2\theta' I(\theta) I(\theta') \left[\langle A\rangle\langle A'\rangle + \mathrm{Cov}[A, A']\right] \end{aligned} \tag{26}$$

As expected, if A is deterministic, this reduces to the previous answer. Furthermore, if the beam is independent for different angles, it reduces to the standard result (with the exception that A is replaced by $\langle A\rangle$). Nevertheless, it is not clear that either of these will be true *a priori*.

A.3 Random beam and sky

`NotImplementedError`