Reweighting Single Event Posteriors with Hyperparameter Marginalization
LIGO DCC No. T2100301

Tom Callister
(Dated: July 1, 2021)

Given data d for a single event, let's assume that we've done standard parameter estimation to obtain a posterior $p_\mathrm{pe}(\theta|d)$ on a parameter θ of this particular event. To do parameter estimation, we've assumed a default prior $p_\mathrm{pe}(\theta)$. Additionally, given a set of data $\{d_i\}$ for a number of events (including our single event of interest), we've obtained posteriors $p(\Lambda|\{d\})$ for a collection of hyperparameters Λ parametrizing the ensemble distribution $p(\theta|\Lambda)$. Using this newfound information about the ensemble distribution of θ, our goal is to obtain a reweighted posterior $p(\theta|\{d\})$ for our single event, marginalized over possible values of Λ.

Let's begin with the joint posterior on the parameters $\{\theta\}$ of *all* our events and Λ:

$$p(\{\theta\},\Lambda|\{d\}) = p(\Lambda)\prod_i \frac{p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}, \tag{1}$$

where $\xi(\Lambda)$ is the population-averaged detection efficiency. If we were interested only in Λ, we might marginalize Eq. (1) over all θ_i, leaving

$$p(\Lambda|\{d\}) = p(\Lambda)\prod_i \frac{\int d\theta_i\, p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}, \tag{2}$$

Instead, we'll marginalize Eq. (1) over Λ and over all parameters $\theta_{i\neq j}$, leaving only a posterior on the particular θ_j of interest that has been "reweighted" from the initial prior to a new prior informed by our population fit:

$$\begin{aligned}
p(\theta_j|\{d\}) &= \int d\Lambda\, d\theta_{i\neq j}\, p(\{\theta\},\Lambda|\{d\}) \\
&= \int d\Lambda\, d\theta_{i\neq j}\, p(\Lambda)\prod_i \frac{p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)} \\
&= \int d\Lambda\, \frac{p(d_j|\theta_j)p(\theta_j|\Lambda)}{\xi(\Lambda)} p(\Lambda) \left[\prod_{i\neq j}\frac{\int d\theta_i p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}\right],
\end{aligned} \tag{3}$$

where in the last line I've pulled the factors $p(d_j|\theta_j)p(\theta_j|\Lambda)/\xi(\Lambda)$ for our event of interest out of the product. In the final line of Eq. (3), it is (hopefully) clear that no double-counting is occurring: the quantity in square brackets, which achieves the population reweighting, explicitly does *not* depend on the event j that we seek to reweight.

In practice, though, Eq. (3) is not terribly convenient to work with. Reweighting each of N events in a catalog would necessitate running our population inference N times, each time leaving out a different single event. Let's proceed instead by both multiplying and dividing the integrand of Eq. (3) by the evidence integral $\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)$:

$$\begin{aligned}
p(\theta_j|\{d\}) &= \int d\Lambda\, \frac{p(d_j|\theta_j)p(\theta_j|\Lambda)}{\xi(\Lambda)} p(\Lambda) \left[\prod_{i\neq j}\frac{\int d\theta_i p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}\right] \frac{\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)}{\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)} \\
&= \int d\Lambda\, \frac{p(d_j|\theta_j)p(\theta_j|\Lambda)}{\left[\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)\right]} p(\Lambda) \left[\prod_{i\neq j}\frac{\int d\theta_i p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}\right] \frac{\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)}{\xi(\Lambda)} \\
&= \int d\Lambda \frac{p(d_j|\theta_j)p(\theta_j|\Lambda)}{\left[\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)\right]} p(\Lambda) \left[\prod_{i}\frac{\int d\theta_i p(d_i|\theta_i)p(\theta_i|\Lambda)}{\xi(\Lambda)}\right] \\
&= \int d\Lambda \frac{p(d_j|\theta_j)p(\theta_j|\Lambda)}{\left[\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)\right]} p(\Lambda|\{d\}),
\end{aligned} \tag{4}$$

using Eq. (2) for the marginalized posterior on Λ. In contrast to Eq. (3), which depended on the posterior for Λ using all events *other* than j, we now have an expression that depends on the posterior $p(\Lambda|\{d\})$ obtained using *all* events, even the event j we are seeking to reweight. The cost of this simplification, however, is that we've picked up an additional evidence integral $\left[\int d\theta'_j p(d_j|\theta'_j)p(\theta'_j|\Lambda)\right]$ appearing in the denominator of Eq. (4). This factor is not a constant – it depends on Λ, and so we cannot factor it outside of the integral and discard it as a constant of proportionality. Its presence, though, is exactly what prevents us from double counting information when reweighting θ_j using a posterior $p(\Lambda|\{d\})$ that was *itself* informed by event j. More on this, however, below.

As usual, in practice we don't have direct access to the likelihood $p(d|\theta)$ (now dropping the subscript j for convenience) but only to the default posterior $p_{\rm pe}(\theta|d)$ mentioned above. These quantities are related via

$$p(d|\theta) = \frac{p_{\rm pe}(\theta|d)p_{\rm pe}(d)}{p_{\rm pe}(\theta)}, \tag{5}$$

where $p_{\rm pe}(d)$ is the evidence obtained using our default prior. Substituting into Eq. (4),

$$\begin{aligned} p(\theta|\{d\}) &= \int d\Lambda \frac{p_{\rm pe}(\theta|d)p_{\rm pe}(d)}{p_{\rm pe}(\theta)} \frac{p(\theta|\Lambda)}{\left[\int d\theta' p(d_j|\theta')p(\theta'|\Lambda)\right]} p(\Lambda|\{d\}) \\ &\propto \int d\Lambda \, p_{\rm pe}(\theta|d) \frac{p(\theta|\Lambda)}{p_{\rm pe}(\theta)} \frac{p(\Lambda|\{d\})}{\left[\int d\theta' p(d|\theta')p(\theta'|\Lambda)\right]}. \end{aligned} \tag{6}$$

Here, $p_{\rm pe}(d)$ *is* a proper constant, and so we've discarded it in the second line.

We still have the pesky evidence integral in Eq. (6), which currently make this expression hard to apply to the problem of reweighting a set of discrete posterior samples. Fortunately, a loophole exists whereby we can effectively ignore this problematic term. Note that the conditional probability $p(\theta|\Lambda, \{d\})$ is *defined* by

$$p(\theta|\{d\}) = \int d\Lambda \, p(\theta|\Lambda, \{d\}) \, p(\Lambda|\{d\}). \tag{7}$$

Comparing Eqs. (6) and (7), we see that

$$p(\theta|\Lambda, \{d\}) \propto p_{\rm pe}(\theta|d) \frac{p(\theta|\Lambda)}{p_{\rm pe}(\theta)} \left[\int d\theta' p(d|\theta')p(\theta'|\Lambda)\right]^{-1}. \tag{8}$$

Since we have conditioned on Λ, however, the evidence integral is now a true constant of proportionality – we can happily ignore it as long as Λ is fixed, yielding

$$p(\theta|\Lambda, \{d\}) \propto p_{\rm pe}(\theta|d) \frac{p(\theta|\Lambda)}{p_{\rm pe}(\theta)}. \tag{9}$$

This fact suggests the following algorithm for producing a reweighted set of posterior samples, giving an initial set $\{\theta\}$ from parameter estimation and a set of hyperparameter samples $\{\Lambda\}$ (inferred with a catalog that included the event of interest):

1. Randomly select a hyperparameter sample $\Lambda_i \in \{\Lambda\}$

2. Having conditioned on this Λ_i, we can ignore the evidence integral as in Eq. (9), and assign every $\theta_j \in \{\theta\}$ a draw probability

$$w_j \propto p_{\rm pe}(\theta_j|d) \frac{p(\theta_j|\Lambda_i)}{p_{\rm pe}(\theta_j)}, \tag{10}$$

   normalizing to $\sum_j w_j = 1$.

3. Finally, select and store a single such θ_j according to the probability weights w_j.

4. Repeat Steps 1-4 until the desired number of reweighted samples are obtained!