show your work! REPRODUCIBLE ARTICLE
/farr/HowManySamples

Draft version March 21, 2025
Typeset using LaTeX **modern** style in AASTeX631

How Many Samples Do We Need?

Will M. Farr [1,2]

[1] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[2] *Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA*

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque, morbi tristique senectus et netus et. Mauris ut leo, cras viverra metus rhoncus sem, nulla et lectus vestibulum. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus, morbi dolor nulla, malesuada eu, pulvinar at, mollis ac. Curabitur auctor semper nulla donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1. INTRODUCTION

The form of the hierarchical marginal likelihood is

$$\log \mathcal{L}\left(\mathbf{d} \mid \lambda\right) = \sum_{i=1}^{N} \log \int \mathrm{d}\theta_i \, p\left(d_i \mid \theta_i\right) p\left(\theta_i \mid \lambda\right). \tag{1}$$

We often approximate the integrals in this expression using Monte Carlo integration over samples for each event i, $\theta_i^{(s)}$, $s = 1, \ldots, S_i$, drawn from a posterior density using some fiducial prior, $p\left(\theta\right)$:

$$\log \mathcal{L}\left(\mathbf{d} \mid \lambda\right) \approx \sum_{i=1}^{N} \log \frac{1}{S_i} \sum_{s=1}^{S_i} \frac{p\left(\theta_i^{(s)} \mid \lambda\right)}{p\left(\theta_i^{(s)}\right)}. \tag{2}$$

Each term in the sum takes the form of

$$\log \sum_{s=1}^{S_i} w_i^{(s)} + \mathrm{const}, \tag{3}$$

will.farr@stonybrook.edu
wfarr@flatironinstitute.org

with

$$w_i^{(s)} = \frac{p\left(\theta_i^{(s)} \mid \lambda\right)}{p\left(\theta_i^{(s)}\right)}. \tag{4}$$

The Monte-Carlo sum introduces some variance into the estimate of the log likelihood. The variance of each term is given by

$$\operatorname{var}\left(\log \sum_{s=1}^{S_i} w_i^{(s)}\right) \simeq \frac{\operatorname{var}\left(\sum_{s=1}^{S_i} w_i^{(s)}\right)}{\left(\sum_{s=1}^{S_i} w_i^{(s)}\right)^2} = \frac{S_i \operatorname{var}\left(w_i\right)}{\left(\sum_{s=1}^{S_i} w_i^{(s)}\right)^2}. \tag{5}$$

The total variance in the log-likelihood is a sum of the variances for each term.

If we define the "effective number of samples" N_{eff} as the inverse of the variance of the log of the sum of the weights, then we can estimate N_{eff} using the empirical variance and the sum of the weights w_i via[1]

$$N_{\text{eff}} = \frac{\left(\sum_{s=1}^{S_i} w_i^{(s)}\right)^2}{S_i \operatorname{var}\left(w_i\right)}. \tag{6}$$

Note that in the limit that the sum of the weights is dominated by a single weight, say (without loss of generality) the $s = 1$ weight, then we have

$$\sum_{s=1}^{S_i} w_i^{(s)} \simeq w_i^{(1)}, \tag{7}$$

and (assuming $S_i \gg 1$)

$$S_i \operatorname{var}\left(w_i\right) \simeq \left(w_i^{(1)}\right)^2, \tag{8}$$

so that $N_{\text{eff}} \simeq 1$. There is no upper bound on N_{eff} (equal weights would have $N_{\text{eff}} \to \infty$); but suppose that some small subset of the weights, $s = 1, \ldots, N_i$ with $1 \ll N_i \ll S_i$ contribute about equally to the sum of the weights, and all other contributions are negligable. Then

$$\sum_{s=1}^{S_i} w_i^{(s)} \simeq N_s w_i^{(1)}, \tag{9}$$

and

$$S_i \operatorname{var}\left(w_i\right) \simeq N_s \left(w_i^{(1)}\right)^2, \tag{10}$$

and $N_{\text{eff}} \simeq N_s$, as would be expected.

[1] Equation (6) is invariant under rescaling of the weights by a constant factor; a convenient choice of rescaling can help with numerical stability in the estimate of N_{eff}. Usually we have access to the log of the weights (to avoid under- or overflow). Let $l_s = \log w_i^{(s)}$ (suppressing the i index for clarity). Then let $\mu_l = \log \sum_s \exp l_s$ be the log of the sum of the weights; μ_l is computable without over- or underflow via the well-known `logsumexp` special function. Then let $\tilde{w}_i^{(s)} = \exp\left(w_i^{(s)} - \mu_l\right)$; the $\tilde{w}_i^{(s)}$ sum to 1, and therefore their variance can be computed without undue roundoff error. With this rescaling, $N_{\text{eff}} = 1/\left(S \operatorname{var}\left(\tilde{w}_i\right)\right)$.

We demonstrate here with a simple hierarchical statistical model that a sufficient condition on the accuracy of the Monte-Carlo estimate of the the marginal log likelihood is that each term in the sum has $N_{\text{eff}} \gg 1$ (here we use a threshold of 10), even if the total variance of the marginal log likelihood is larger than 1. We argue that this condition should be sufficient for other, more complex hierarchical models because it does not depend on the *structure* of the likelihood function; in some sense, as long as many samples contribute to the Monte-Carlo estimate of the log-likelihood for each observation, the model "knows" about the degree of uncertainty in that observation, and can "take it into account."

Our hierarchical model is a simple "normal-normal" two-level hierarchy. Suppose we have a population of N objects with parameter x, and that the population distribution is normal with mean μ and s.d. σ:

$$x_i \sim \mathcal{N}\left(\mu, \sigma\right) \quad i = 1, \ldots, N. \tag{11}$$

Each object is observed and its x parameter is measured with some (known) normal uncertainty σ_i and no bias to be $x_{\text{obs},i}$:

$$x_{\text{obs},i} \sim \mathcal{N}\left(x_i, \sigma_i\right) \quad i = 1, \ldots, N. \tag{12}$$

A hierarchical model for estimating the population parameters μ and σ would have a likelihood of the form

$$\log \mathcal{L}\left(\mathbf{x}_{\text{obs}} \mid \mu, \sigma, \boldsymbol{\sigma}\right) = \sum_{i=1}^{N} \log \int \mathrm{d}x_i \, \mathcal{N}\left(x_{\text{obs},i} \mid x_i, \sigma_i\right) \mathcal{N}\left(x_i \mid \mu, \sigma\right), \tag{13}$$

where we are using $\mathbf{x}_{\text{obs}}$ and $\boldsymbol{\sigma}$ to represent the complete set of observations and their uncertainties. This model is simple enough that we can compute the exact marginal likelihood via a Gaussian integral; it is

$$\log \mathcal{L}\left(\mathbf{x}_{\text{obs}} \mid \mu, \sigma, \boldsymbol{\sigma}\right) = \sum_{i=1}^{N} \log \mathcal{N}\left(x_{\text{obs},i} \mid \mu, \sqrt{\sigma^2 + \sigma_i^2}\right). \tag{14}$$

But we can also draw S_i samples from a fiducial "posterior" for x_i, $p_i\left(x_i \mid x_{\text{obs},i}, \sigma_i\right)$, and use Monte Carlo integration to estimate the marginal likelihood:

$$\log \mathcal{L}\left(\mathbf{x}_{\text{obs}} \mid \mu, \sigma, \boldsymbol{\sigma}\right) \simeq \sum_{i=1}^{N} \log \frac{1}{S_i} \sum_{s=1}^{S_i} \frac{\mathcal{N}\left(x_i^{(s)} \mid \mu, \sigma\right)}{p_i\left(x_i^{(s)} \mid x_{\text{obs},i}, \sigma_i\right)}. \tag{15}$$

We have simulated $N = 128$ observations from this model with population parameters $\mu = 0$, $\sigma = 1$. Each observation has a randomly-chosen $\sigma_i \sim \mathcal{U}(1, 2)$, with $x_{\text{obs},i}$ drawn accordingly. We impose broad priors, $\mathcal{N}(0, 1)$ on μ and $\mathrm{Exp}(1)$ on σ. We then use the No-U-Turn Sampler (NUTS) (Hoffman & Gelman 2011) variant of

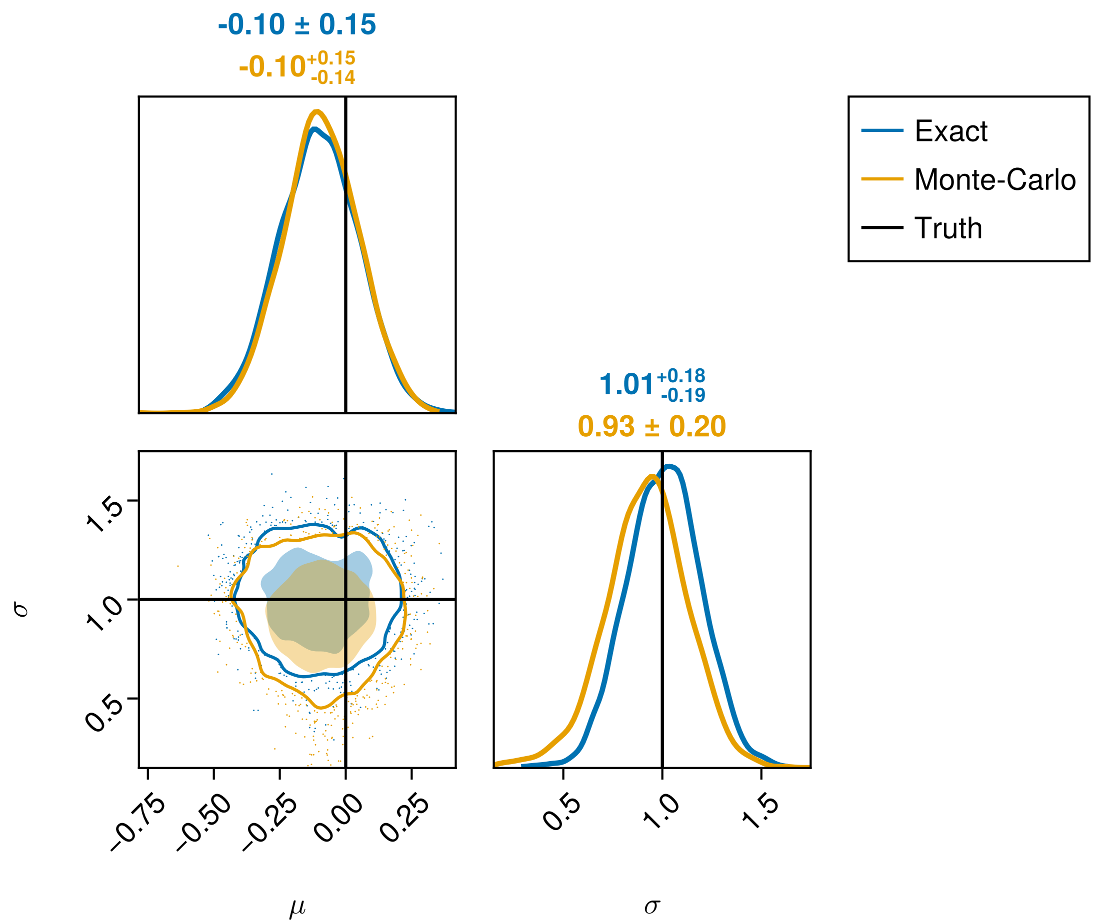

Figure 1. The posterior density for the parameters μ and σ in the hierarchical model. The sampling using the exact marginal likelihood is shown in blue; the sampling from the approximate Monte-Carlo likelihood once all observations have $N_{\text{eff}} > 10$ (the minimum $N_{\text{eff}} = 13.0$ for this sampling) is shown in orange. Over all posterior samples in the converged Monte-Carlo sampling, the largest total variance for the marginal log likelihood was 5.8. The sampling from the approximate likelihood has converged to that of the exact likelihood even though the total variance of the log-likelihood is considerably larger than 1.

Hamiltonian Monte-Carlo (Neal 2011) as implemented in the `Turing.jl` package (Ge et al. 2018) to draw samples of μ and σ from the exact and approximate posteriors. After sampling, we compute the effective number of samples for each observation's likelihood using Eq. (6), and re-run the MCMC after drawing additional samples for any observation that has $N_{\text{eff}} < 10$. Ultimately, our final MCMC sampling has $N_{\text{eff}} = 13.0$ for the observation with the smallest N_{eff} and total log-likelihood variance $\text{var}\left(\log \mathcal{L}\right) = 5.8$. The results are shown in Figure 1; it is apparent that the sampling from the approximate likelihood has converged to that of the exact likelihood *even though the total variance of the log-likelihood is larger than 1.*

REFERENCES

Ge, H., Xu, K., & Ghahramani, Z. 2018, in International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain, 1682–1690. http://proceedings.mlr.press/v84/ge18b.html

Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints, arXiv:1111.4246, doi: 10.48550/arXiv.1111.4246

Neal, R. 2011, in Handbook of Markov Chain Monte Carlo, 113–162, doi: 10.1201/b10905