Draft version November 2, 2023
Typeset using LaTeX **modern** style in AASTeX631

Comments on the Normal-Normal Hierarchical Model

Will M. Farr[1,2]

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

Here we examine the model for the hierarchical stacking analysis suggested in Isi et al. (2019) and currently employed to combine information from multiple gravitational wave detections in tests of general relativity (The LIGO Scientific Collaboration et al. 2021). Recently Pacilio et al. (2023) suggested that the method may not converge to a consistent posterior, and pointed out that the frequentist coverage of the credible intervals from the method is not uniform in the limit of repeated ensembles of observations. The first observation is in error, while the second is a consequence of the prior-population mismatch between the hierarchical model and the underlying repeated ensemble of observations. Pacilio et al. (2023) note this latter explanation, but focus instead on bootstrap calibration of the credible level coverage; here we explore alternative prior choices that can improve the frequentist coverge of the Bayesian credible levels from this model.

Undelying the approach of Isi et al. (2019) is an assumption that the deviation parameter of interest for each event i, x_i, is drawn from a normal distribution with

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

mean μ and standard deviation σ:

$$x_i \sim N(\mu, \sigma). \tag{1}$$

This assumption can interpolate smoothly from the delta-function limit, with $\mu = \sigma \to 0$ where the parameter of interest is a universal constant that is the same for each event, to the case where the parameter of interest is dependent (in an unknown, or at least un-modeled) way on the other parameters of the event and therefore follows some (unknown, or un-modeled) population distribution. In the latter case, and in the absence of selection effects, the μ and σ population parameters will be related to the population mean and population standard deviation of the parameter of interest.

While in practice the observation of the parameter of interest (and other parameters) for each event generates a complicated likelihood function that couples the deviation parameter of interest to other parameters describing the event (see, e.g., The LIGO Scientific Collaboration et al. (2021); Payne et al. (2023)), it is illuminating to consider a simplified observational model where each x_i is observed with additive normal noise with standard deviation σ_i:

$$x_{\mathrm{obs},i} \sim N(x_i, \sigma_i). \tag{2}$$

This simplified model is analyzed in detail at the beginning of Pacilio et al. (2023). Note that we are not assuming "heteroskedastic" observations; these comprise the special case where the observational uncertainties are all equal, $\sigma_i \equiv \sigma_{\mathrm{obs}}$.

With this simplified observational model, the latent x_i parameters can be integrated out to give the marginal likelihood for μ and σ given an ensemble of observations:

$$p(\{x_{\mathrm{obs},i}\} \mid \mu, \sigma) = \prod_i N\left(x_{\mathrm{obs},i} \mid \mu, \sqrt{\sigma^2 + \sigma_i^2}\right). \tag{3}$$

Here we show, contrary to the claims in Pacilio et al. (2023), that the maximum-likelihood estimators for the parameters μ and $V \equiv \sigma^2$, which we denote $\hat{\mu}$ and $\hat{V}$, are unbiased and asymptotically unbiased, and therefore the likelihood function above converges weakly (check this, maybe converges in distribution or probability? Or almost surely?) to a delta-function at the true parameters in the limit of an infinite ensemble of observations.

The maximum likelihood estimator $\hat{\mu}$ for an ensemble of observations $\{x_{\mathrm{obs},i}\}$ is

$$\hat{\mu} = \frac{1}{W} \sum_i \frac{x_{\mathrm{obs},i}}{\sigma_i^2 + \hat{V}}, \tag{4}$$

where

$$W \equiv \sum_i \frac{1}{\sigma_i^2 + \hat{V}}, \tag{5}$$

and the maximum-likelihood estimator for the variance $\hat{V}$ is implicitly defined by

$$\sum_i \frac{\left(x_{\mathrm{obs},i} - \hat{\mu}\right)^2}{\left(\sigma_i^2 + \hat{V}\right)^2} = W. \tag{6}$$

The estimator $\hat{\mu}$ is unbiased:

$$\langle\hat{\mu}\rangle = \frac{1}{W}\sum_i \frac{\langle x_{\mathrm{obs},i}\rangle}{\sigma_i^2+\hat{V}} = \frac{1}{W}\sum_i \frac{\mu}{\sigma_i^2+\hat{V}} = \mu, \tag{7}$$

and has variance

$$\mathrm{Var}\left(\hat{\mu}\right) = \frac{1}{W^2}\sum_i \frac{\mathrm{Var}\left(x_{\mathrm{obs},i}\right)}{\left(\sigma_i^2+\hat{V}\right)^2} = \frac{1}{W^2}\sum_i \frac{\sigma_i^2+V}{\left(\sigma_i^2+\hat{V}\right)^2} = \frac{1}{W} + \mathcal{O}\left(\frac{1}{N}\right). \tag{8}$$

with the convergence in the last relation following from the convergence of $\hat{V}$ to the true variance V at leading order that we will demonstrate presently.

Taking expectations of both sides of the implicit Eq. (6) defining $\hat{V}$, we find

$$W = \sum_i \frac{\left\langle\left(x_{\mathrm{obs},i}-\hat{\mu}\right)^2\right\rangle}{\left(\sigma_i^2+\hat{V}\right)^2}. \tag{9}$$

Expanding the quantity in angle brackets, we find

$$\left\langle\left(x_{\mathrm{obs},i}-\hat{\mu}\right)^2\right\rangle = \left\langle x_{\mathrm{obs},i}^2\right\rangle - 2\left\langle x_{\mathrm{obs},i}\hat{\mu}\right\rangle + \left\langle\hat{\mu}^2\right\rangle. \tag{10}$$

The first term is

$$\left\langle x_{\mathrm{obs},i}^2\right\rangle = \mu^2+\sigma_i^2+V, \tag{11}$$

the second term becomes

$$\left\langle x_{\mathrm{obs},i}\hat{\mu}\right\rangle = \frac{1}{W}\sum_{j\neq i}\frac{\mu^2}{\sigma_j^2+\hat{V}} + \frac{1}{W}\frac{\mu^2+\sigma_i^2+V}{\sigma_i^2+\hat{V}} = \mu^2 + \frac{1}{W}\frac{\sigma_i^2+V}{\sigma_i^2+\hat{V}}, \tag{12}$$

and the third term becomes

$$\left\langle\hat{\mu}^2\right\rangle = \frac{1}{W^2}\sum_{i,j}\frac{\left\langle x_{\mathrm{obs},i}x_{\mathrm{obs},j}\right\rangle}{\left(\sigma_i^2+\hat{V}\right)\left(\sigma_j^2+\hat{V}\right)} = \mu^2 + \frac{1}{W^2}\sum_j \frac{\sigma_j^2+V}{\left(\sigma_j^2+\hat{V}\right)^2}. \tag{13}$$

Substituting in Eq. (9), we obtain

$$W = \sum_i \frac{1}{\left(\sigma_i^2+\hat{V}\right)^2}\left(\sigma_i^2+V-\frac{2}{W}\frac{\sigma_i^2+V}{\sigma_i^2+\hat{V}}+\frac{1}{W^2}\sum_j\frac{\sigma_j^2+V}{\left(\sigma_j^2+\hat{V}\right)^2}\right). \tag{14}$$

If $\hat{V}=V$, this becomes

$$W|_{\hat{V}=V} = \sum_i \frac{1}{\sigma_i^2+V} - \left.\frac{1}{W}\right|_{\hat{V}=V}\sum_i\frac{1}{\left(\sigma_i^2+V\right)^2}. \tag{15}$$

Intuition about this complicated expression can be gathered by considering the homoskedastic case, where $\sigma_i \equiv s$ for all i. Then this expression becomes

$$\frac{N}{s^2+V} = \frac{N}{s^2+V} - \frac{s^2+V}{N}\frac{N}{\left(s^2+V\right)^2} = \frac{N-1}{s^2+V}, \tag{16}$$

indicating that the relation is satisfied up to order $1/N$ when the observational uncertainties are homoskedastic and $\hat{V} = V$; equivalently, we could say $\left\langle \hat{V} \right\rangle = V + \mathcal{O}\left(1/N\right)$ in this case. Similarly, if the distribution of observational errors has enough finite moments so that we can write

$$W = \frac{N}{\left\langle \sigma_i^2 + \hat{V} \right\rangle} + \mathcal{O}\left(N^0\right), \tag{17}$$

then in the heteroskedastic case we have

$$\hat{V} = V + \mathcal{O}\left(\frac{1}{N}\right), \tag{18}$$

so the maximum likelihood variance estimator is (asymptotically) unbiased.

TODO: discussing calibration of the frequentist coverage of the Bayesian credible intervals. And then take a shower becauese that is truly a horrible sentence.

REFERENCES

Isi, M., Chatziioannou, K., & Farr, W. M. 2019, PhRvL, 123, 121101, doi: 10.1103/PhysRevLett.123.121101

Pacilio, C., Gerosa, D., & Bhagwat, S. 2023, arXiv e-prints, arXiv:2310.03811, doi: 10.48550/arXiv.2310.03811

Payne, E., Isi, M., Chatziioannou, K., & Farr, W. M. 2023, arXiv e-prints, arXiv:2309.04528, doi: 10.48550/arXiv.2309.04528

The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2021, arXiv e-prints, arXiv:2112.06861, doi: 10.48550/arXiv.2112.06861