Draft version March 30, 2022
Typeset using LaTeX/show your work! modern style in AASTeX631

A Note About Marginalizing Over Gaussian Populations With KDE Likelihood Representations

Will M. Farr

ABSTRACT

I work through the marginalization over parameters whose population is Gaussian using a kernel density estimate (KDE) approximation to the likelihood function. The result is an analytic KDE representation of the marginal likelihood for the Gaussian population parameters.

1. MARGINALIZATION

Suppose we are conducting a population analysis with a Gaussian population (e.g. Isi et al. 2019; Miller et al. 2020). The hierarchical posterior for the unknown parameter x_i with the Gaussian population from each of the $i = 1, \ldots, N$ events and the population parameters μ and σ is

$$p\left(\left\{x_i\right\}, \mu, \sigma \mid \left\{d_i\right\}\right) \propto p\left(\mu, \sigma\right) \prod_{i=1}^{N} p\left(d_i \mid x_i\right) p\left(x_i \mid \mu, \sigma\right) \tag{1}$$

where $p\left(\mu, \sigma\right)$ is the prior we impose on the population parameters, and d_i is the data for each observation. The population is assumed to be Gaussian with mean μ and standard deviation σ:

$$p\left(x \mid \mu, \sigma\right) = N\left(x \mid \mu, \sigma\right). \tag{2}$$

In problems of interest, the likelihood function is often not simple to write down, and does not take a straightforward analytic form. In such situations, we usually draw *samples* from the likelihood (or a posterior which is the likelihood times some simple prior) for each event. Let there be $j = 1, \ldots, M_i$ samples x_{ij} from (a density proportional to) the likelihood (i.e. from a posterior with a flat prior):

$$x_{ij} \sim p\left(d \mid x_i\right). \tag{3}$$

A useful representation of the likelihood function (up to a proportionality constant that we usually ignore—unless we are computing the Bayesian evidence) then is a KDE with a Gaussian kernel over the samples from the likelihood

$$p\left(d \mid x_i\right) \simeq \frac{1}{M_i} \sum_{j} N\left(x_i \mid x_{ij}, \sigma_i\right) \tag{4}$$

where σ_i is a reasonable *bandwidth* for the KDE[1].

Because we can do Gaussian integrals (though the discussion here is for a one-dimensional parameter x, the same trick works is multiple dimensions; Hogg et al. (2020) is helpful here), we can analytically integrate out the true parameter values x_i from the posterior in Eq. (1) using the KDE approximation to the likelihood. The result is

$$p\left(\mu, \sigma \mid \{d_i\}\right) \propto p\left(\mu, \sigma\right) \prod_{i=1}^{N} \frac{1}{M_i} \sum_{j=1}^{M_i} N\left(x_{ij} \mid \mu, \sqrt{\sigma_i^2 + \sigma^2}\right) \tag{6}$$

In this special case of Gaussian populations and Gaussian KDE approximations to the likelihood function, the above expression is considerably more robust to differences in scale between the population and measurement (i.e. $\sigma_i \ll \sigma$ or $\sigma \ll \sigma_i$) than the usual Monte-Carlo approximation to the integral of the likelihood (e.g. Miller et al. 2020).

REFERENCES

Hogg, D. W., Price-Whelan, A. M., & Leistedt, B. 2020, arXiv e-prints, arXiv:2005.14199. https://arxiv.org/abs/2005.14199

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

Miller, S., Callister, T. A., & Farr, W. M. 2020, ApJ, 895, 128, doi: 10.3847/1538-4357/ab80c0

Scott, D. W. 1992, Multivariate Density Estimation

[1] When in doubt, I usually follow Scott (1992), with

$$\sigma_i^2 = \frac{\operatorname{Var} x_{ij}}{M_i^{2/5}}. \tag{5}$$