Extracting distribution parameters from multiple uncertain observations with selection biases

Ilya Mandel and Will Farr

School of Physics and Astronomy, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK[*]

Jonathan R. Gair

School of Mathematics, University of Edinburgh and Biomathematics
and Statistics Scotland, James Clerk Maxwell Building,
Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK

(Dated: April 13, 2014; updated June 9, 2014; June 9, 2016)

Abstract

We derive a Bayesian framework for extracting the parameters of the underlying distribution based on a set of observations sampled from this distribution. We allow for both measurement uncertainty in individual measurements and, crucially, for selection biases on the measurements.

[*]Electronic address: `imandel@star.sr.bham.ac.uk`

There is a population of events with parameters $\vec{\theta}$. The distribution of events in the population is described via parameters $\vec{\lambda}$ through some functional form $p_{\rm pop}(\vec{\theta}|\vec{\lambda})$. For instance, if the underlying distribution is modeled as a multi-dimensional Gaussian, $\vec{\lambda}$ would consist of the covariance matrix; alternatively, a non-parametric distribution could be described with a (multi-dimensional) histogram, in which case $\vec{\lambda}$ represents the weights of various histogram bins.

This distribution is sampled by drawing a set of k "observed events" $\{\vec{\theta}^{(i)}\}$, for $i \in [1, k]$. The total probability of making this particular set of independent observations is

$$p(\{\vec{\theta}^{(i)}\}|\vec{\lambda}) = \prod_{i=1}^{k} \frac{p_{\rm pop}(\vec{\theta}^{(i)}|\vec{\lambda})}{\int d\vec{\theta}\, p_{\rm pop}(\vec{\theta}|\vec{\lambda})} \,. \tag{1}$$

The normalization factor here accounts for the overall probability of making an observation given a particular choice of $\vec{\lambda}$ (it will be equal to 1 if $p_{\rm pop}$ is properly normalized). Of course, we cannot uniquely reconstruct $\vec{\lambda}$ using limited observations. The best we can do is compute the posterior probability on $\vec{\lambda}$, the distribution on distributions, given the observations, which, in the usual Bayesian formalism where $\pi(\vec{\lambda})$ is the prior, is given by

$$p(\vec{\lambda}|\vec{\theta}^{(i)}) = \frac{\pi(\vec{\lambda}) p(\{\vec{\theta}^{(i)}\}|\vec{\lambda})}{p(\vec{\theta}^{(i)})}, \tag{2}$$

where the evidence $p(\vec{\theta}^{(i)})$ is the integral of the numerator over all λ. This evidence can be used to select between different models for representing the distribution, as in [1].

In practice, there is often a selection bias involved: some events are easier to observe than others. Assuming for now that there is a 1-to-1 relationship between the system parameters and the data, we can represent the selection effect with a detection probability $p_{\rm det}(\vec{\theta})$. This detection probability can be estimated empirically for a search pipeline via a large injection campaign. In some cases, it can be modeled analytically (e.g., for low-mass compact binaries, $p_{\rm det} \propto M_c^{15/6}$). With the selection effect included, equation (1) becomes [e.g., 2, 3]

$$p(\{\vec{\theta}^{(i)}\}|\vec{\lambda}) = \prod_{i=1}^{k} \frac{p(\vec{\theta}^{(i)}|\vec{\lambda}) p_{\rm det}(\vec{\theta})}{\int d(\vec{\theta}) p_{\rm pop}(\vec{\theta}^{(i)}|\vec{\lambda}) p_{\rm det}(\vec{\theta})}. \tag{3}$$

In general, we don't have the luxury of directly measuring the parameters of a given event, $\vec{\theta}^{(i)}$. Instead, we measure the data set $\vec{d}^{(i)}$ which encodes these parameters but also includes some random noise. For a given data set and search pipeline, the detectability is generally deterministic: if the data exceeds some threshold (e.g., a threshold on the signal to noise ratio), then the event is detectable; otherwise, it's not. In other words, the detection probability for a given set of parameters introduced earlier is, in fact, an integral over the possible data sets given those parameters:

$$p_{det}(\vec{\theta}) = \int_{\vec{d}>threshold} p(\vec{d}|\vec{\theta}) d\vec{d}, \tag{4}$$

where $p(\vec{d}|\vec{\theta})$ is the likelihood of obtaining the data set $\vec{d}$ given the parameters $\vec{\theta}$. We can therefore write the probability of observing a particular data set (where "observing" implies that the data are above the threshold, hence included as one of our k observations) given the assumed distribution $\vec{\lambda}$ as

$$p(\vec{d}|\vec{\lambda}) = \frac{\int d\vec{\theta} p(\vec{d}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\lambda})}{\alpha(\vec{\lambda})}, \tag{5}$$

where the normalization factor $\alpha(\vec{\lambda})$ is given by

$$\begin{aligned} \alpha(\vec{\lambda}) &\equiv \int_{\vec{d}>threshold} d\vec{d} \int d\vec{\theta} p(\vec{d}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\lambda}) \\ &= \int d\vec{\theta} \left[\int_{\vec{d}>threshold} d\vec{d} p(\vec{d}|\vec{\theta}) \right] p_{\rm pop}(\vec{\theta}|\vec{\lambda}) \equiv \int d\vec{\theta} p_{\rm det}(\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\lambda}) \,. \end{aligned} \tag{6}$$

Thus, in the presence of both measurement uncertainty and selection effects, equations (1) and (3) become:

$$p(\{\vec{d^{(i)}}\}|\vec{\lambda}) = \prod_{i=1}^{k} \frac{\int d\vec{\theta} p(\vec{d^{(i)}}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\lambda})}{\int d\vec{\theta} p_{\rm det}(\theta) p_{\rm pop}(\vec{\theta}|\vec{\lambda})} \,. \tag{7}$$

The presence of the likelihood $p(\vec{d^{(i)}}|\vec{\theta})$ in this equation is a reminder that we do not have a perfect measurement of the parameters of a given event. The likelihood can be rewritten in terms of the posterior probability density function (PDF) $p(\vec{\theta}^{(i)}|\vec{d^{(i)}})$ that is computed in the course of single-event parameter estimation using some assumed prior $\pi(\vec{\theta})$:

$$p(\vec{d^{(i)}}|\vec{\theta}^{(i)}) = \frac{p(\vec{\theta}^{(i)}|\vec{d^{(i)}}) p(\vec{d^{(i)}})}{\pi(\vec{\theta})} \,. \tag{8}$$

Thus, each term of the product in Eq. (7) is a convolution integral of the draw probability with the PDF [4]; see also the "extreme deconvolution" approach of Bovy et al. [5] and applications in Hogg et al. [6].

In practice, the PDF $p(\vec{\theta}^{(i)}|\vec{d^{(i)}})$ is often discretely sampled with N samples from the posterior, $\{{}^{j}\vec{\theta}^{(i)}\}$, for $j \in [1, N]$. Because the samples are drawn according to the posterior, the parameter space volume associated with each sample is inversely proportional to the local PDF, $d^j\vec{\theta}^{(i)} \propto \left[p(^j\vec{\theta}^{(i)}|\vec{d}^{(i)})\right]^{-1}$. This allows us to easily replace the integral in the Eq. (7) with a discrete sum over PDF samples:

$$p(\{\vec{d}^{(i)}\}|\vec{\lambda}) = \prod_{i=1}^{k} \frac{\frac{1}{N_i}\sum_{j=1}^{N_i} p_{\mathrm{pop}}(^j\vec{\theta}^{(i)}|\vec{\lambda})\frac{p(\vec{d}^{(i)})}{\pi(\vec{\theta})}}{\int d\vec{\theta} p_{\mathrm{det}}(\theta) p_{\mathrm{pop}}(\vec{\theta}|\vec{\lambda})} \,. \tag{9}$$

Finally, the posterior on the underlying population parameters $\vec{\lambda}$ is given by substituting equation (9) into equation (2):

$$p(\vec{\lambda}|\{\vec{d}^{(i)}\}) = \frac{\pi(\vec{\lambda})}{p(\{\vec{d}^{(i)}\})} \prod_{i=1}^{k} \frac{\frac{1}{N_i}\sum_{j=1}^{N_i} \frac{p_{\mathrm{pop}}(^j\vec{\theta}^{(i)}|\vec{\lambda})}{\pi(\vec{\theta})} p(\vec{d}^{(i)})}{\int d\vec{\theta} p_{\mathrm{det}}(\theta) p_{\mathrm{pop}}(\vec{\theta}|\vec{\lambda})} \,. \tag{10}$$

This expression contains the evidence factors $p(\{\vec{d}^{(i)}\})$ and $\prod_i p(\vec{d}^{(i)})$ which, in general, are not easy to calculate. However, since the evidence terms are independent of $\vec{\lambda}$, they can be factored out and subsumed into the overall normalization constant $p(\vec{\lambda})$.

Of course, if interested in the distribution of a single parameter, we can marginalize over Eq. (10) in the usual way, by integrating over the remaining parameters.

---

[1] W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera, ApJ **741**, 103 (2011), 1011.1459.

[2] J. Chennamangalam, D. R. Lorimer, I. Mandel, and M. Bagchi, MNRAS **431**, 874 (2013), 1207.5732.

[3] W. M. Farr, J. R. Gair, I. Mandel, and C. Cutler, Phys. Rev. D **91**, 023005 (2015), 1302.5341.

[4] I. Mandel, Phys. Rev. D **81**, 084029 (2010), 0912.5531.

[5] J. Bovy, D. W. Hogg, and S. T. Roweis, Annals of Applied Statistics **5** (2011), 0905.2979.

[6] D. W. Hogg, A. D. Myers, and J. Bovy, ApJ **725**, 2166 (2010), 1008.4146.