Correcting for sample selection in Bayesian analyses

Samuel R. Hinton,[1,2]★ Alex Kim,[3] Tamara M. Davis,[1,2]
[1]*School of Mathematics and Physics, The University of Queensland, Brisbane, QLD 4072, Australia*
[2]*ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)*
[3]*Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

ABSTRACT
Samples of astronomical sources are drawn from underlying distributions. The samples we observe generally do not necessarily represent the full population of sources in the Universe, but are preferentially biased toward those sources that are more likely to be detected. Malmquist bias is a well-known example. Sample selection is not unique to astronomy - noisy data is pruned in almost all fields, and physical limits (such as age only being positive) are inescapable. These sample selection effects need to be accounted for in statistical analyses. In this paper we present a simple overview of a Bayesian consideration of sample selection, giving solutions to both analytically tractable and intractable models. This is accomplished via a combination of analytic approximations and Monte Carlo integration, in which dataset simulation is efficiently used to correct for issues in the observed dataset. Toy models are included for demonstration, along with discussions of numerical considerations and how to optimise for implementation.

1 INTRODUCTION

Sample selection is a problem in many areas of scientific inquiry. For example, it is one of the primary difficulties when performing supernova cosmology analysis, as telescopes more easily detect bright objects than faint ones, causing our observed supernova distribution to deviate from the actual underlying distribution. This "Malmquist bias", has been a subject of much investigation (Butkevich et al. 2005). It is taken into account during analyses by either modifying the observed data to remove the expected bias (Betoule et al. 2014; Conley et al. 2011), or by incorporating the expected bias into the underlying model (Rubin et al. 2015). Truncated data are also commonly encountered in biological fields, where data such as mortality rates are left-truncated (Colchero & Clark 2012). Simplified and generalised examples have been investigated in numerous fashions (Woodroofe 1985; Gull 1989; Grogger & Carson 1991; O'Neill & Barry 1995) and with application to specific fitting algorithms (Gelfand et al. 1992). Whilst generalised resources exist that provide a comprehensive overview of sample selection and analysis techniques in a similar fashion to this work (Klein & Moeschberger 2005), these sources are often opaque due to volume and mathematical complexity.

This work provides a simple treatment of sample selection using a common Bayesian technique. The general theory for considering selection effects is discussed in Section 2. Section 3 provides three examples of increasing complexity with sample selection. Section 4 details numeric concerns and tricks to be aware of for effective implementation of Monte Carlo corrections applied to analytic approximations.

2 THEORY

When formulating and fitting a model using a constraining dataset, we wish to resolve the posterior surface defined by

$$P(\theta|\text{data}) \propto P(\text{data}|\theta)P(\theta), \tag{1}$$

which gives the probability of the model parameter values (θ) given the data. Prior knowledge of the allowed values of the model parameters is encapsulated in the prior probability $P(\theta)$. Of primary interest to us is the likelihood of observing the data given our parametrised model, $\mathcal{L} \equiv P(\text{data}|\theta)$. When dealing with experiments which have imperfect selection efficiency, our likelihood must take that efficiency into account. We need to describe the probability that the events we observe were both drawn from the distribution predicted by the underlying theoretical model *and* that those events, given they happened, were subsequently successfully observed. To make this extra conditional explicit, we write the likelihood of the data given an underlying model, θ, *and* selection efficiency, S, as:

$$\mathcal{L} = P(\text{data}|\theta, S). \tag{2}$$

If we seek to describe the chance of our observation passing the sample selection as a function of the observation itself and model parameters, we need to rearrange the above equation to include $P(S|\text{data}, \theta)$, which colloquially can be stated as *the probability of selection given our observation and model*. This can be done in a few lines:

$$P(\text{data}, S|\theta) = P(\text{data}, S|\theta) \tag{3}$$

$$P(\text{data}|\theta, S)P(S|\theta) = P(S|\text{data}, \theta)P(\text{data}|\theta) \tag{4}$$

$$P(\text{data}|\theta, S) = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{P(S|\theta)}, \tag{5}$$

thus allowing us to explicitly write our likelihood as

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{P(S|\theta)}. \tag{6}$$

Introducing an integral over all possible data D, so we can evaluate $P(S|\theta)$,

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S, D|\theta)\,dD} \tag{7}$$

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\,dD}, \tag{8}$$

where the integral in the denominator has the same dimensionality as the experimental data. **Equation 8 is the generalised likelihood of experiments with sample selection.**

3   SAMPLE SELECTION

In this section we present sample selections of increasing complexity. Each sample selection is accompanied by an illustrative example inspired by Type Ia supernova cosmology, where we characterize the properties of a standard or standardizable candle. Python code implementations of all examples, including plot generation, can be found at the Github repository `SampleSelection`[1].

3.1   Complete Selection

In a perfect world, data is neither biased nor truncated. Data is perfect. Uncertainties are well quantified and normally distributed around true values. Presumably everything is also spherical and in a vacuum.

We thus begin by considering an ideal situation where the sample is complete. All events are included in the sample such that $P(S|\text{data}, \theta) = 1$. Trivially this expression is independent of our data and model parameters θ, and our likelihood from Equation (8) reduces down to Equation (2). As a concrete example of this case, let us consider a model for a population of objects whose brightnesses form a normal distribution with average μ and standard deviation σ. Let us also assume that our experiment produces data x_i, which are independent measurements of unique objects' brightnesses (with negligible measurement uncertainty), which we can formally state as

$$\vec{x} \sim \mathcal{N}(\mu, \sigma). \tag{9}$$

If, having collected our observations $\vec{x}$, we wanted to constrain μ and σ, this would be a simple task of modelling the posterior surface. We simply wish to map the surface

$$P(\theta|\text{data}) \propto P(\text{data}|\theta)P(\theta), \tag{10}$$

where our model parameters $\theta = \{\mu, \sigma\}$ and our data is given by $\vec{x}$.

$$P(\mu, \sigma|\vec{x}) \propto P(\vec{x}|\mu, \sigma)P(\mu, \sigma) \tag{11}$$

With uniform priors, $P(\mu, \sigma) = $ constant, and can be absorbed into the constant of proportionality. Expanding our
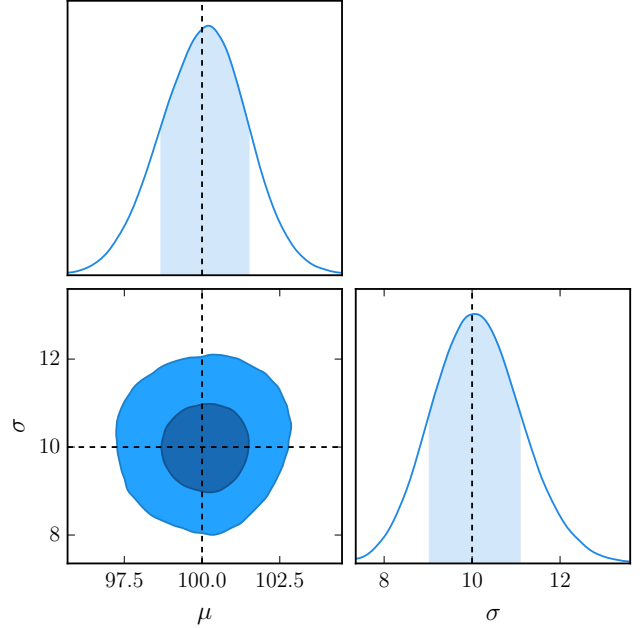
Figure 1. A systematic test of our perfect model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points. Any systematic offset in our model would be revealed by a shift in the stacked results away from the true parameter values. This plot can be generated by executing `0_perfect.py`.

observation vector to take into account all observations and not just one, the posterior surface is given by

$$P(\mu, \sigma|\vec{x}) \propto \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma), \tag{12}$$

where $\mathcal{N}(x_i|\mu, \sigma)$ represents a the probability of drawing sample x_i from the Normal distribution with mean μ and standard deviation σ. Generating a hundred data points with $\mu = 100$, $\sigma = 10$, we can recover our input parameters easily, as shown in Figure 1.

3.2   Analytic Sample Selection

We now consider a case with selection bias that allows us to form an analytic expression for the likelihood. This illustrative case is useful for the reader as the influence of selection bias is simple and intuitive. The case we consider is identical to the one in the previous section, *except* that only the subset of objects brighter than a threshold α can be observed.

With this sample selection, all events satisfy $x_i > \alpha$, giving $P(S|x, \theta) = \mathcal{H}(x - \alpha)$, where $\mathcal{H}$ is the Heaviside step function

$$\mathcal{H}(y) \equiv \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

We assign an arbitrary value $\alpha = 85$ for convenience. If we do not take this truncation into account, we will recover biased parameter estimates. However, we can correct for this truncation using Equation (8), as the integral in the denominator has an analytic solution. Having successfully observed

x_i, it follows that $x_i > \alpha$ and so $P(S|x_i, \theta) = 1$. To substitute in our normal model,

$$\mathcal{L}_i = \frac{P(S|x_i, \theta)P(x_i|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD} \tag{14}$$

$$= \frac{\mathcal{N}(x_i|\mu, \sigma)}{\int_{-\infty}^{\infty} \mathcal{H}(D - \alpha)\mathcal{N}(D|\mu, \sigma)\, dD} \tag{15}$$

$$= \frac{\mathcal{N}(x_i|\mu, \sigma)}{\int_{\alpha}^{\infty} \mathcal{N}(D|\mu, \sigma)\, dD} \tag{16}$$

$$= \frac{\mathcal{N}(x_i|\mu, \sigma)}{\frac{1}{2}\mathrm{erfc}\left[\frac{\alpha - \mu}{\sqrt{2}\sigma}\right]}, \tag{17}$$

where in the last line we have evaluated the integral in the case $\mu > \alpha$. Note that this is for a single observation, and so for a set of independent observations we need to introduce the product found in Equation (12).

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i = \prod_{i=1}^{N} \frac{\mathcal{N}(x|\mu, \sigma)}{\frac{1}{2}\mathrm{erfc}\left[\frac{\alpha - \mu}{\sqrt{2}\sigma}\right]}, \tag{18}$$

However, as our selection efficiency correction is independent of our observations $\vec{x}$, we can take it outside the product.

$$\mathcal{L} = 2\left(\mathrm{erfc}\left[\frac{\alpha - \mu}{\sqrt{2}\sigma}\right]\right)^{-N} \prod_{i=1}^{N} \mathcal{N}(x|\mu, \sigma), \tag{19}$$

We can add this correction to our model, and note that we now recover unbiased parameter estimates. This is demonstrated in Figure 2, which shows the posterior surfaces for when you take sample selection into account and when you do not. The selection bias preferentially selects intrinsically brighter objects and, by cutting out some of the distribution, narrows the observed distribution. As such, we note that the bias correction correctly increases the weight of low μ and high σ parametrisations, as those models would be subject to the most sample bias. We also note that not only does the best fit location for each parameter shift, but the shape of the posterior surface itself is significantly modified.

3.3 Analytically Intractable Sample Selection

Unfortunately it is a rare scenario when dealing with nature and all her faults for us to have an analytic selection function, let alone a function encapsulated by a single parameter. A more realistic selection efficiency would take the form of non-analytic function of many model parameters. And the function would probably be stochastic too, just to make things more difficult. Provided a method of forward modelling or simulating observations, the solution is to combine an analytic approximate correction with Monte Carlo integration. Whilst other integration algorithms can be used, Monte Carlo integration is an appropriate choice, as it is based on drawing random numbers from underlying distributions and is efficient for evaluating high dimensional integrals.

In this subsection we introduce an example that presents computational challenges in the evaluation of the posterior, and which better reflects the model complexity in supernova cosmology analysis. The specific challenge is that the denominator in Equation (8) is dependent on θ and does not have an analytic solution (i.e. has to be solved numerically). When using Monte Carlo sampling methods to
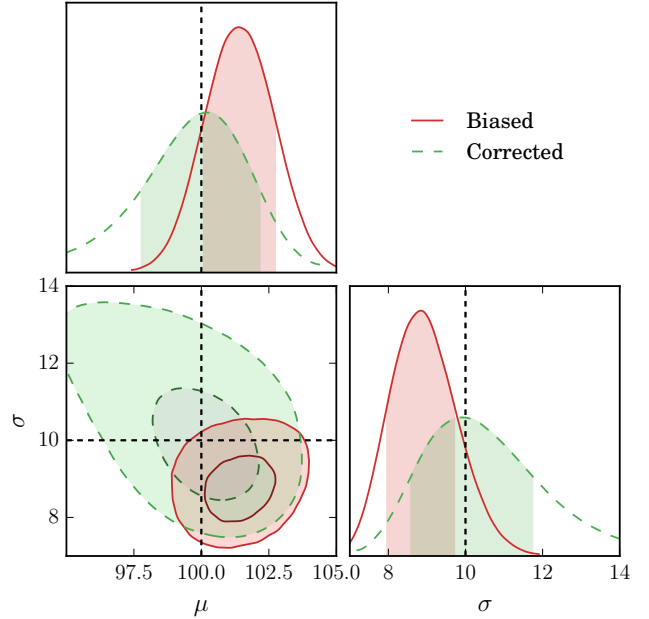
Figure 2. A systematic test of our imperfect model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points, subject to our thresholding. The bias shown in the red 'Biased' contour can be corrected to via the techniques shown in Section 3.2 to recover unbiased surfaces. This plot can be generated by executing `1_imperfect.py`.

probe the posterior with this kind of likelihood, the integral in the denominator must be calculated with each draw. We therefore present importance sampling as an approach that makes these posteriors more computationally tractable.

Let us modify our imperfect toy model from the previous section. Instead of observing just one variable, x, we also observe a new independent variable, y, which is drawn from its own distribution $y \sim \mathcal{N}(\mu_y, \sigma_y)$, and has no measurement uncertainty (like x). Our selection efficiency can now become a combination of x and y, such that we only observe events that satisfy $x + \beta y > \alpha$, giving $P(S|x, y, \theta) = \mathcal{H}(x + \beta y - \alpha)$, giving a relationship reminiscent of the stretch and colour corrections for standardising supernovae. Our likelihood for such a toy model becomes now the combination of probabilities for observing both x and y, with the denominator becoming an integral over all possible X and Y observations subject to our selection effects.

$$\mathcal{L}_i = \frac{P(S|x_i, y_i, \theta)P(x_i, y_i|\theta)}{\iint P(S|X, Y, \theta)P(X, Y|\theta)\, dX\, dY}, \tag{20}$$

where $\theta = \{\mu, \sigma, \mu_y, \sigma_y\}$. As before, given a successful observation and a deterministic selection efficiency, $P(S|x_i, y_i, \theta) = 1$. We continue to use uniform priors to keep the example simple. Substituting in the step function selection efficiency for $P(S|X, Y, \theta)$ and the normal distributions for $P(x_i, y_i|\theta)$, we have

$$\mathcal{L}_i = \frac{\mathcal{N}(x_i|\mu, \sigma)\mathcal{N}(y_i|\mu_y, \sigma_y)}{\iint_{-\infty}^{\infty} \mathcal{H}(X + \beta Y - \alpha)\mathcal{N}(X|\mu, \sigma)\mathcal{N}(Y|\mu_y, \sigma_y)\, dX dY} \tag{21}$$

Assume that we cannot solve this integral analytically, and must resort to numeric solutions. These often clash with

sampling methods, especially for high dimensional integrals. Inserting Monte Carlo integration into fitting algorithms can drastically slow them down, and algorithms such as Hamiltonian MCMC that require continuous surfaces can easily fail on surfaces that fluctuate due to the stochastic nature of Monte Carlo integration. Even by fixing the samples used in MC integration (thereby giving a continuous surface), the complexity of the surface derivatives will pose almost insurmountable problems for any algorithms that utilise surface gradients.

It is tempting to separate out the numerator and denominator of Equation (21) and only sample from the numerator, adding corrections from the denominator after the fact. However, this strategy rarely works, as ignoring the denominator completely can lead to fits in completely wrong areas of parameter space. The solution to find an approximate, analytic correction we can utilise in our fitting algorithm which seeks to shift the region of parameter space sampled by the sampler closer to the correct area, and then importance sample our MC chains to provide a fully corrected surface. Given an approximate analytic correction w_{approx}, we explicitly break our likelihood into two parts, $\mathcal{L}_i = \mathcal{L}_{i1}\mathcal{L}_{i2}$, with the parts given in general by

$$\mathcal{L}_{i1} = \frac{P(S|\mathrm{data},\theta)P(\mathrm{data}|\theta)}{w_{\mathrm{approx}}} \tag{22}$$

$$\mathcal{L}_{i2} = \frac{w_{\mathrm{approx}}}{\int P(S,D|\theta)\,dD}. \tag{23}$$

In our example, if $\beta \ll 1$, such that the majority of selection effect is encapsulated by x and not y, our approximate correction can take the form found in the previous correction from Section 3.2. Having true values of $\mu = 100$, $\sigma = 10$, $\mu_y = 30$, $\sigma_y = 5$, $\alpha = 92$ and a known $\beta = 0.2$, we can give a concrete example. Assuming some prior, imperfect knowledge of μ_y (perhaps we believe it is approximately 20) we estimate that the average contribution from βy is approximately $20\beta = 4$ (which is close to the correct value of 6), and from this we add in a small adjustment to the analytic correction from Section 3.2:

$$w_{\mathrm{approx}} = \frac{1}{2}\mathrm{erfc}\left[\frac{\alpha - \mu - 4}{\sqrt{2}\sigma}\right]. \tag{24}$$

This gives our likelihood parts as

$$\mathcal{L}_{i1} = \frac{\mathcal{N}(x_i|\mu,\sigma)\mathcal{N}(y_i|\mu_y,\sigma_y)}{w_{\mathrm{approx}}} \tag{25}$$

$$\mathcal{L}_{i2} = \frac{w_{\mathrm{approx}}}{\iint_{-\infty}^{\infty}\mathcal{H}(x+\beta y-\alpha)\mathcal{N}(X|\mu,\sigma)\mathcal{N}(Y|\mu_y,\sigma_y)\,dX\,dY}. \tag{26}$$

Evaluating $\mathcal{L}_{i2}$ using Monte Carlo integration of n samples, we have for a single observation that

$$\mathcal{L}_{i2} = \frac{w_{\mathrm{approx}}n}{\sum_{j=1}^{n}\mathcal{H}(X_j+\beta Y_j-\alpha)\mathcal{N}(X_j|\mu,\sigma)\mathcal{N}(Y_j|\mu_y,\sigma_y)}. \tag{27}$$

We now wish to move to a set of observations. As $\mathcal{L}_{i2}$ is independent of the specific observation i, this is a simple
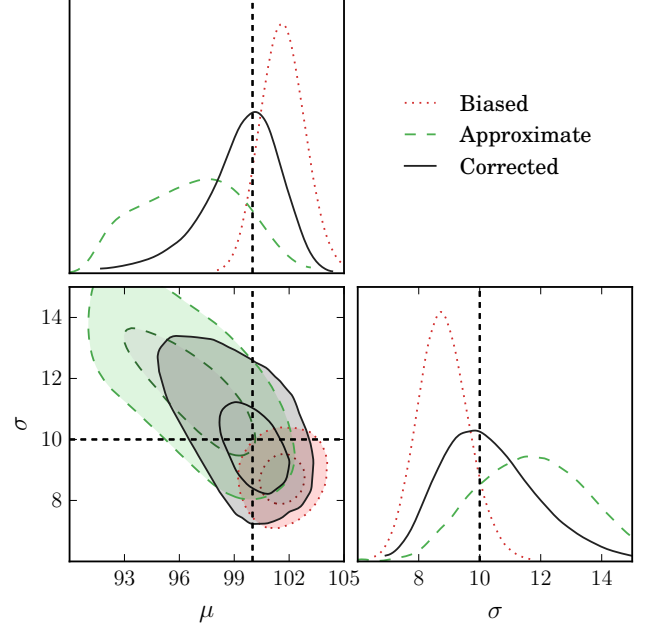

Figure 3. A systematic test of our more complicated model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points, subject to our thresholding. The likelihood $\mathcal{L}_1$ was evaluated, and reweighted using Monte Carlo integration of a hundred thousand possible events as per $\mathcal{L}_2$. The truncated data with no correction is shown as 'Biased' in dotted red, the 'Approximate' only correction ($\mathcal{L}_1$) shown in dashed green, and the final reweighted chain shown in solid black as 'Corrected'. A large number of samples of $\mathcal{L}_1$ had to be generated to ensure sufficient sampling of our 'Corrected' posterior. This plot can be generated by executing `2_real.py`.

step:

$$\mathcal{L}_1 = w_{\mathrm{approx}}^{-N}\prod_{i=1}^{N}\mathcal{N}(x_i|\mu,\sigma)\mathcal{N}(y_i|\mu_y,\sigma_y) \tag{28}$$

$$\mathcal{L}_2 = \left(\frac{w_{\mathrm{approx}}n}{\sum_{j=1}^{n}\mathcal{H}(X_j+\beta Y_j-\alpha)\mathcal{N}(X_j|\mu,\sigma)\mathcal{N}(Y_j|\mu_y,\sigma_y)}\right)^N \tag{29}$$

$\mathcal{L}_1$ can thus be fitted with a traditional sampler without numeric difficulty or slowdown, and $\mathcal{L}_2$ allows us to calculate the weight of each sample given by fitting $\mathcal{L}_1$. We are effectively importance sampling our likelihood evaluations. The computational benefits of this should not be understated either — each sample in our chains can be reweighted independently, providing a task that is trivially parallelisable.

Taking our concrete example stated previously, we end up with a corrected posterior surface as shown in Figure 3. It is important to note that the point of maximum likelihood is biased by roughly the same amount in the biased and approximate posterior surfaces in Figure 3. However, as the approximately correctly posterior is broader than the biased surface, it has far more samples in the region of parameter space mapped by the fully corrected posterior. This is the entire purpose of the approximate correction — to maximise the number of samples in the correct region of parameter space, so that we can importance sample our chains efficiently.

4 NUMERICAL TECHNIQUES

4.1 Importance Sampling

Further tricks can be used to increase the efficiency with which the samples are reweighted. Firstly, the overarching analytic model often provides functions which can be drawn from efficiently. In the case of our example, by drawing the random numbers X and Y respectively from the normal distributions $\mathcal{N}(\mu,\sigma)$ and $\mathcal{N}(\mu_y,\sigma_y)$ (i.e. traditional importance sampling) we need only evaluate the step function for our data points. That is, for a simplified 1D example only in X, we replace

$$\int f(X)\mathcal{N}(X|\mu,\sigma)\,dX = \frac{1}{N}\sum_{i=1}^{N} f(X_i)\mathcal{N}(X_i|\mu,\sigma), \qquad (30)$$

where X is drawn from a uniform distribution over all space, with

$$\int f(X)\mathcal{N}(X|\mu,\sigma)\,dX = \frac{1}{N}\sum_{i=1}^{N} f(X_i), \qquad (31)$$

where X drawn from $\mathcal{N}(X_i|\mu,\sigma)$.

4.2 Precomputing selection

The integral in the denominator of Equation (8) must be recalculated for each set of θ parameter-values that are considered. A significant cost saving is possible when Monte Carlo integration is done, if the same Monte Carlo realizations are used for all sets of θ. This is especially true if evaluating the sample selection $P(S|\text{data},\theta)$ is numerically expensive, but can only be used when the sample selection is independent of θ, so that $P(S|\text{data},\theta) = P(S|\text{data})$, and when there is some prior knowledge of θ such that we can estimate a reasonable θ_{approx}. Considering the integral from Equation (8), we have

$$\int P(S|D,\theta)P(D|\theta)\,dD =$$
$$\int P(S|D)\frac{P(D|\theta)}{P(D|\theta_{\text{approx}})}\left[P(D|\theta_{\text{approx}})\,dD\right] \qquad (32)$$

With the mathematics laid as so, it is easy to pregenerate a set of events D_i drawn from distribution $P(D|\theta_{\text{approx}})$, evaluate $P(S|D_i)$, and reuse them for all samples. Formulated using Monte Carlo integration,

$$\int P(S|D,\theta)P(D|\theta)\,dD = \frac{1}{N}\sum_{i=1}^{N} \frac{P(S|D_i)}{P(D_i|\theta_{\text{approx}})}P(D_i|\theta). \qquad (33)$$

The *only* term that needs to be evaluated for each sample is $P(D_i|\theta)$, as all other terms can be pre-computed and stored. When using this technique, take care that the number of events used when calculating the weights is sufficient to make the statistical error of Monte Carlo integration insignificant when compared to the constraining power of your dataset.

Consider the imperfect example — where we observe x drawn from an underlying normal distribution as was done in Section 3.2, but utilise the Monte Carlo integration technique from Section 3.3, and work without an analytic approximation (i.e. we set $w_{\text{approx}} = 1$). We could estimate, given some prior knowledge, that variable $x \approx$

$\mathcal{N}(\mu_{\text{approx}},\sigma_{\text{approx}})$. In this example, it would be better to overestimate σ_{approx} rather than underestimate it, so that we still effectively sample the correct area of parameter space. We then draw samples of x from this approximate distribution, recording the probability of each draw and then calculating whether our potential observation of x would be observed in the experiment or not. That is, we assign $P(S|x,\theta) = P(S|x) = 1$ or 0 given it passed cuts or not. We discard all events with 0 weight (as they have 0 weight), and only track those events which pass. Then, when calculating the sample reweighting after running chains, $\mathcal{L}_{i2}$ becomes

$$\mathcal{L}_{i2} \propto w_{\text{approx}}\left[\sum_{j=1}^{n} \frac{\mathcal{N}(X_j|\mu,\sigma)}{\mathcal{N}(X_j|\mu_{\text{approx}},\sigma_{\text{approx}})}\right]^{-1}, \qquad (34)$$

where you can see that we discard the constant n from Equation (27) as we only care about likelihood proportionality. Provided our parameter estimate is reasonably well informed, the computation benefit this precomputation provides is enormous for any nontrivial selection function. Not only do we now waste no time when calculating $\mathcal{L}_{i2}$ determining $P(S|\text{data},\theta)$, as we only save results that pass the cuts, we have no wasted evaluations of $\mathcal{N}(X_j|\mu,\sigma)$.

As stated previously, this technique requires that selection efficiency of an observation is independent of model parameters θ. For many experimental cases this may hold, however if it does not this method cannot be used to increase efficiency. Gridding or interpolating the parameter space is strongly not recommended due the required accuracy of $\mathcal{L}_{i2}$. Even a small error when raised to the power of N can spiral out of control.

4.3 Log-space

Following from the previous section, as our reweighting $\mathcal{L}_2$ is raised to the power of the number of our observations, they should definitely be computed in log-space, which turns the power into a linear factor. As most probabilistic work is already computed in log-space, this subsection barely needs to be stated. However, whilst working in log-space an efficient way of increasing the accuracy of the approximate analytic correction is to fit the correction such that the spread of the distribution $\log(\mathcal{L}_2)$ is minimised.

5 CONCLUSION

Sample selection is a pervasive issue in many scientific domains. For simple cases of sample selection which can be encapsulated with analytic functions, it is possible to analytically correct likelihood surfaces by introducing selection efficiency into the likelihood formulation. When analytic corrections fail to provide an adequate description of selection effects, Monte Carlo integration can be used on top of analytic approximations to further correct the likelihood surface, provided the experiment can be effectively simulated.

REFERENCES

Betoule M., et al., 2014, A&A, 568, A22

Butkevich A. G., Berdyugin A. V., Teerikorpi P., 2005, Monthly Notices of the Royal Astronomical Society, 362, 321

Colchero F., Clark J. S., 2012, Journal of Animal Ecology, 81, 139

Conley A., et al., 2011, ApJS, 192, 1

Gelfand A. E., Smith A. F. M., Lee T.-M., 1992, Journal of the American Statistical Association, 87, 523

Grogger J. T., Carson R. T., 1991, Journal of applied econometrics, 6, 225

Gull S. F., 1989, in , Maximum Entropy and Bayesian Methods. Springer, pp 511–518

Klein J. P., Moeschberger M. L., 2005, Survival analysis: techniques for censored and truncated data. Springer Science & Business Media

O'Neill T. J., Barry S. C., 1995, Biometrics, pp 533–541

Rubin D., et al., 2015, ApJ, 813, 137

Woodroofe M., 1985, The Annals of Statistics, pp 163–177

This paper has been typeset from a TEX/LATEX file prepared by the author.