MNRAS **000**, 1–3 (2017) Preprint 18 May 2017 Compiled using MNRAS LATEX style file v3.0

Monte Carlo Corrections for Bayesian Methods

Samuel R. Hinton,[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)*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
I present a rigorous treatment for truncated or biased datasets by combining Bayesian analysis methods with Monte Carlo simulations.

1 INTRODUCTION

Find some treatment of biased or truncated data before. Maybe Gull (1989).

2 THE PERFECT WORLD

In a perfect world, data is neither biased nor truncated. The data is perfect. Uncertainties are well quantified and normally distributed around true values. Presumably everything is also spherical and in a vacuum. Let us create a mock model in this perfect world. Let us observe a series of iid events $\vec{x}$, which is known perfectly and drawn from a normal distribution such that

$$\vec{x} \sim \mathcal{N}(\mu, \sigma) \tag{1}$$

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

$$P(\mu, \sigma|\vec{x}) \propto P(\vec{x}|\mu, \sigma)P(\mu, \sigma) \tag{2}$$

$$P(\mu, \sigma|\vec{x}) \propto \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma). \tag{3}$$

Generating a hundred data points with $\mu = 100,\ \sigma = 10$, we can recover our input parameters easily, as shown in Figure 1.

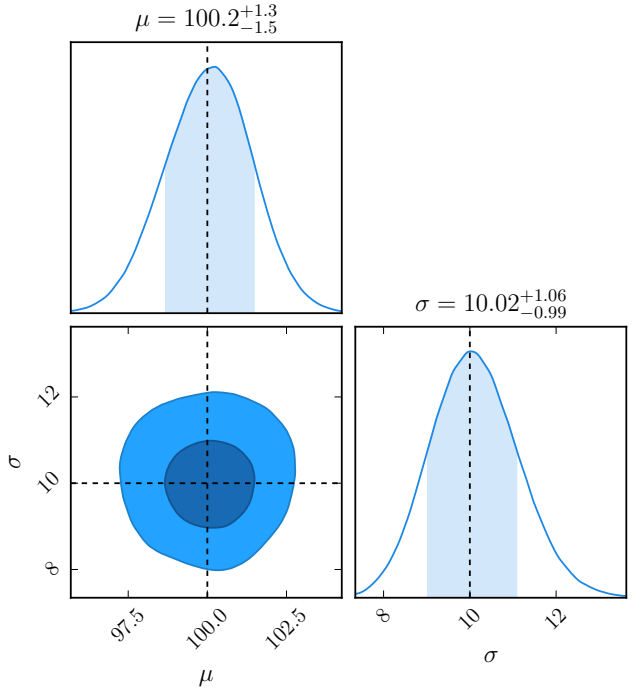

Figure 1. A systematic test of our perfect model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points. Posterior surface mapped out using emcee.

3 THE IMPERFECT WORLD

In a slightly imperfect world we may have to deal with something like truncated data. For an example, consider the previous model, but with an instrumentation deficiency such that we can only observe events above a certain threshold, such that $x > \alpha$, assigning a value $\alpha = 85$ for convenience. If we don't take this truncation into account, we will recovered biased parameter estimates, as shown in Figure 2. However, we can correct for this truncation. If restate our likelihood to take into account some selection effect S, our likelihood for a single event can be stated as

$$\mathcal{L} = P(x|\theta, S) \tag{4}$$

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

$$= \frac{P(S|x, \theta)P(x|\theta)}{\int P(S, D|\theta)\, dD} \tag{6}$$

$$= \frac{P(S|x, \theta)P(x|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{7}$$

where D is a potential observable. In our example, the selection efficiency is a step function, having observed x,

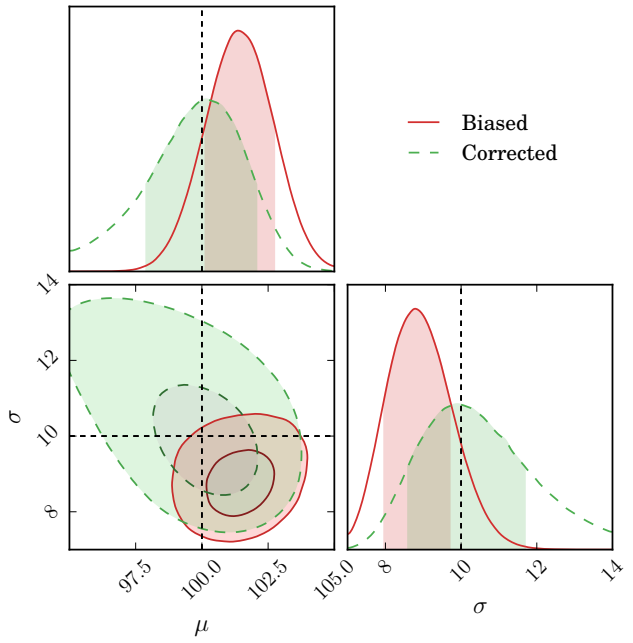

Figure 2. A systematic test of our imperfect model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points, subject to our thresholding.

$P(S|x,\theta) = 1$. To substitute in our normal model,

$$\mathcal{L} = \frac{\mathcal{N}(x|\mu,\sigma)}{\int_{-\infty}^{\infty} \mathcal{H}(D-\alpha)\mathcal{N}(D|\mu,\sigma)\,dD} \tag{8}$$

$$= \frac{\mathcal{N}(x|\mu,\sigma)}{\int_{\alpha}^{\infty} \mathcal{N}(D|\mu,\sigma)\,dD} \tag{9}$$

$$= \frac{\mathcal{N}(x|\mu,\sigma)}{\frac{1}{2}\mathrm{erfc}\left[\frac{\alpha-\mu}{\sqrt{2}\sigma}\right]}, \tag{10}$$

if $\mu > \alpha$. We can add this correction to our model, and note that we now recover unbiased parameter estimates, also shown in Figure 2.

4 THE REAL WORLD

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 scenario involves a selection efficiency which cannot be conveniently described by an analytic function. Instead, we would have a high dimensional non-analytic function. And its probably stochastic too, just to throw another wrench in the works. Provided a method of forward modelling or simulating observations, the solution is Monte Carlo integration combined with an analytic approximate correction.

So let us extend our toy model. We observe not just $x \sim \mathcal{N}(\mu,\sigma)$, but also another independent variable, $y \sim \mathcal{N}(\mu_y,\sigma_y)$. Our selection efficiency can now become a combination of x and y, such that we only observe events that satisfy $x + \beta y > \alpha$. Our likelihood for such a toy model becomes

$$\mathcal{L} = \frac{\mathcal{N}(x|\mu,\sigma)\mathcal{N}(y|\mu_y,\sigma_y)}{\iint_{-\infty}^{\infty} \mathcal{H}(x+\beta y-\alpha)\mathcal{N}(X|\mu,\sigma)\mathcal{N}(Y|\mu_y,\sigma_y)\,dX dY} \tag{11}$$

Assume that we cannot solve these 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 such as Hamiltonian MCMC that require continuous surfaces can easily fail on surfaces that, via Monte Carlo integration, fluctuate. 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. One solution is 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.

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 example. Having true values of $\mu = 100$, $\sigma = 10$, $\mu_y = 30$, $\sigma_y = 5$, and a known $\beta = 0.2$, we can give a concrete example. Assuming some prior, imperfect knowledge of mu_y we estimate that the average contribution from βy is around 4 (which is close to the correct value of 6), and from this our analytic correction to our likelihood is

$$w = \frac{1}{2}\mathrm{erfc}\left[\frac{\alpha-\mu-4}{\sqrt{2}\sigma}\right] \tag{12}$$

Further, let us explicitly break our likelihood into two parts, $\mathcal{L} = \mathcal{L}_1\mathcal{L}_2$, with the parts given by

$$\mathcal{L}_1 = \frac{\mathcal{N}(x|\mu,\sigma)\mathcal{N}(y|\mu_y,\sigma_y)}{w} \tag{13}$$

$$\mathcal{L}_2 = \frac{w}{\iint_{-\infty}^{\infty} \mathcal{H}(x+\beta y-\alpha)\mathcal{N}(X|\mu,\sigma)\mathcal{N}(Y|\mu_y,\sigma_y)\,dX dY}. \tag{14}$$

$\mathcal{L}_1$ 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. 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. Evaluating $\mathcal{L}_2$ using Monte Carlo integration of n samples, we have

$$\mathcal{L}_2 = \frac{wn}{\sum_{i=1}^{n} \mathcal{H}(x+\beta y-\alpha)\mathcal{N}(X_i|\mu,\sigma)\mathcal{N}(Y_i|\mu_y,\sigma_y)} \tag{15}$$

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 $\mathcal{N}(\mu_y,\sigma_y)$ (ie traditional importance sampling) we need only evaluate the step function for our data points. This is demonstrated in Figure 3.

Alternatively, if evaluating the probability that an event is observed is numerically expensive, it is easy to pregenerate a set of events and reuse them for all weights - provided that the number of events used when calculating the weights

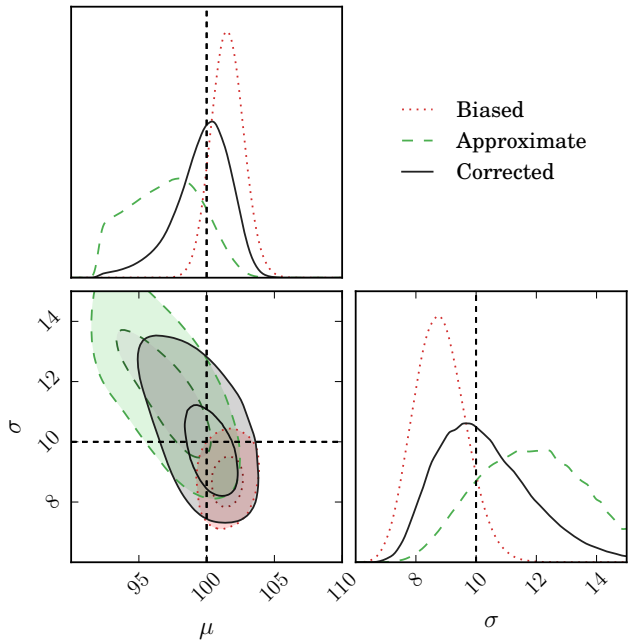

Figure 3. A systematic test of our more complicated model, done by stacking the output chains from fitting 100 independent realisations of our 100 data points, subject to our thresholding. The likelihood $\mathcal{L}_1$ was evaluated with the fitting algorithm emcee, and reweighted using Monte Carlo 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'.

is sufficient to make the statistical error of Monte Carlo integration insignificant when compared to the constraining power of your dataset. This method is however only efficient when prior knowledge of parameter values is known to allow a reasonable initial draw of events.

5 CONCLUSION

Still da bes

ACKNOWLEDGMENTS

We gratefully acknowledge the input of the many researchers that were consulted during the creation of this paper.

REFERENCES

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

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