Reweighting and Bias

Mark Dewing

(Dated: August 14, 2007)

Abstract

Standard reweighting exhibits a bias with finite sample sizes. We derive expressions to demonstrate the bias, and explore it with a simple example.

I. REWEIGHTING

Reweighting is used in the evaluation of integrals by Monte Carlo when the samples are drawn from one probability distribution, but the desired average is with respect to a different distribution.

The standard formula for evaluating averages of some estimator, $\mathcal{O}(R)$, over the probability distribution Q is written as

$$\langle \mathcal{O} \rangle = \frac{\int dR \mathcal{O}(R) Q(R)}{\int dR Q(R)} \approx \sum O(R_i) \tag{1}$$

Now let P be the distribution the samples are actually drawn from. The following transformation gives the standard derivation of the reweighting equations.

$$\langle \mathcal{O} \rangle = \frac{\int dR \mathcal{O}(R) \frac{Q(R)}{P(R)} P(R)}{\int dR \frac{Q(R)}{P(R)} P(R)} \approx \frac{\sum w(R_i) \mathcal{O}(R_i)}{\sum w(R_i)} \tag{2}$$

where $w = Q/P$

In the final step, the integral is approxmiated by a sum. Let us examine the consequences of this approximation in more detail.

II. FINITE SAMPLE SIZE BIAS

The general approach for this investigation is to compute the expectation value of the expectation value. We take the expression for the sum (Eqn. 2) for a fixed number of points and compute its expected value by integrating over the probability distribution of the samples.

Consider the $N = 1$ case.

$$\langle \mathcal{O} \rangle \approx \frac{\mathcal{O}(x_1) w(x_1)}{w(x_1)} = \mathcal{O}(x_1) \tag{3}$$

Integrate x_1 over $P(x)$, and we get

$$\int dx_1 P(x_1) \mathcal{O}(x_1) \tag{4}$$

which is clearly biased. (The desired, unbiased result is $\int dx_1 Q(x_1) \mathcal{O}(x_1)$.)

Now consider the $N = 2$ case.

$$\langle \mathcal{O} \rangle \approx \frac{\mathcal{O}(x_1) w(x_1) + \mathcal{O}(x_2) w(x_2)}{w(x_1) + w(x_2)} \tag{5}$$

Average over $P(x_1)$ and $P(x_2)$.

$$= \int dx_1 dx_2 P(x_1)P(x_2)\left[\frac{\mathcal{O}(x_1)w(x_1)+\mathcal{O}(x_2)w(x_2)}{w(x_1)+w(x_2)}\right] \quad (6)$$

$$= \int dx_1 dx_2 \frac{1}{w(x_1)+w(x_2)}\left[\mathcal{O}(x_1)Q(x_1)P(x_2)+\mathcal{O}(x_2)P(x_1)Q(x_2)\right] \quad (7)$$

$$= 2\int dx_1 \mathcal{O}(x_1)Q(x_1)\int dx_2 \frac{P(x_2)}{w(x_1)+w(x_2)} \quad (8)$$

$$= \int dx_1 \mathcal{O}(x_1)Q(x_1)F_2(x_1) \quad (9)$$

where

$$F_2(x) = 2\int dx_2 \frac{P(x_2)}{w(x)+w(x_2)} \quad (10)$$

The bias term is $F_2(x)$, and the rest of the expression is the unbiased result.

The general expression for $F_N(x)$ is

$$F_N(x_1) = N\int \frac{\prod_{i=2}^{N} dx_i P(x_i)}{\sum_{j=1}^{N} w(x_j)} \quad (11)$$

This expression is expected to become constant and have a value of 1 as $N \to \infty$ (that is, we expect the bias to vanish in this limit).

III. SIMPLE EXAMPLE

As an example, we use two gaussians in one dimension with the same width ($\sigma = 1$), but with different centers. Take P to have a center at one ($\mu = 1$) and Q has a center ranging from $2 - 4$.

In Figure 1 we see the computed value of $\langle x \rangle$ as a function of the number of samples used in the reweighting estimate. Each reweighting estimate was computed 2000 times and the average of those estimates is shown in the graph. This graph shows the bias depends on the overlap between the distributions and the number of samples used.

The bias function F_N is shown in Figure 2. As expected, the shape of the curve gets flatter and closer to 1 with increasing N.

A. Further analysis

Consider the numerator and denominator separately. The denominator is the ratio of normalizations of P and Q. In this example they are the same, so D should be 1.

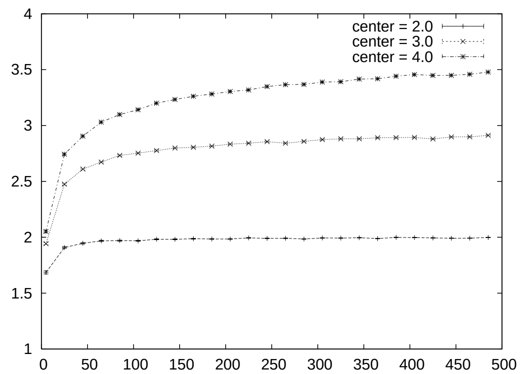

FIG. 1: $\langle x \rangle$ versus number of configurations for various values of μ, with $\mu_1 = 1.0$ and $\mu_2 = 2.0, 3.0$ and 4.0

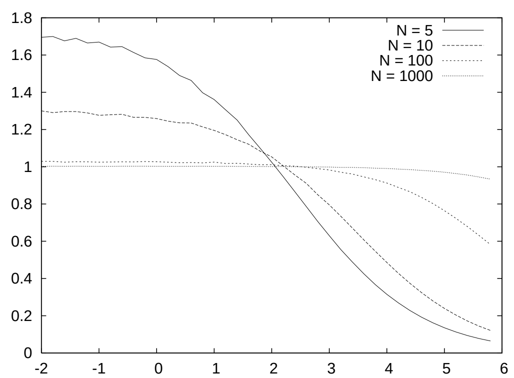

FIG. 2: $F_N(x)$ for various values of N, with $\mu_1 = 1.0$ and $\mu_2 = 2.0$

It appears the $1/D$ part is primarily responsible for the bias (if the observable is also non-linear, it may have a bias as well)

The various elements are show in Figure 3. The total reweighted value ($\langle x * w \rangle / \langle w \rangle$) is as we saw previously. The numerator ($\langle w * x \rangle$) is noisy, but shows little bias. The denominator ($\langle w \rangle$) is close to the expected value of 1. The reciprocal of the denominator ($1/\langle w \rangle$), however, shows a large bias.

IV. POSSIBLE FIXES

The obvious way to create an unbiased estimator is to define a new estimator

$$\tilde{\mathcal{O}}(x) = \mathcal{O}(x)/F_N(x) \tag{12}$$

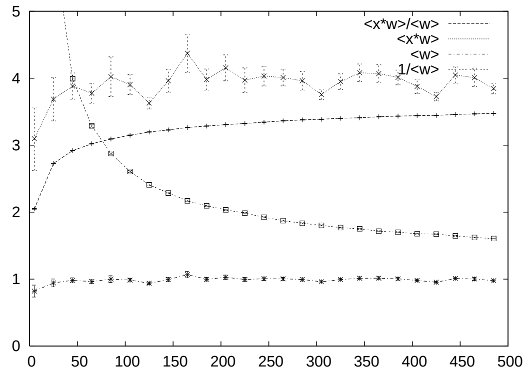

FIG. 3: Various elements the calculation versus number of configurations for $\mu = 4.0$

Unfortunately, the computation of $F_N(x)$ is as least as difficult as solving the original problem, if not more so.

Other possible solutions:

- If the integrals for F_N could be done in the large N limit, the correction factor could be computed as an expansion in $1/N$.

- There are ways to evaluate series expansions stochastically. Perhaps the denominator could be expanded in a series and evaluated this way.

- A resampling method (such as the bootstrap) could be used to estimate and correct the bias (see [1] for such a method). I have explored this approach a little. It can improve the estimate of $1/D$, but this does not translate into an improvment for the full expression. This is likely due to correlations because the same weights are used in the numerator and denominator.

- An alternate use for a resampling method could be to compute the estimator vs. N, and extract the answer from extrapolation to infinite N.

- Does a techinque like black-box reweighting ([3]) suffer from this problem? The theoretical distribution used to generate the sample points is ignored, and the sampled points themselves are used to reconstruct an estimate of the probability distribution for the weights.

V. DIAGNOSTICS

If the problem can't be solved, can it at least be detected?

The overlap between two distributions is often measured by the effective number of points

$$N_{\mathrm{e}ff} = \frac{(\sum w_i)^2}{\sum w_i^2} \tag{13}$$

The Kullback-Leibler number is also used [2].

The bias also depends on the number of points sampled, in addition to the amount of overlap, so these diagnostics provide an incomplete measure of the amount of bias present.

Perhaps a resampling method can be used to compute the average value for $N/2$ and compare to the average at N, and see if a bias is detectable?

[1] Javier Cabrera and Peter Meer. Unbiased estimation of ellipses by bootstraping. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18, 1996.

[2] Martin Hohenadler, Hans Gerd Evertz, and Wolfgang von der Linden. Quantum Monte Carlo and variational approaches to the Holstein model. *arXiv:cond-mat/0305387*, 2003.

[3] F. Marty Ytreberg and Daniel M. Zuckerman. Black-box re-weighting. *arXiv:physics/0609194*, 2006.