HW3 - Jacob Shkrob

Jacob Shkrob

October 14, 2022

1 Exercise 28

In this exercise, we use a one-step resampling importance sampling methods (multinomial, Bernoulli, and systematic) to generate samples approximately from $\mathcal{N}(0,\sigma^2)$ using $\mathcal{N}(0,1)$ random variables. In the following description, we go through the details of each method below, with their corresponding implementations in `python`.

(i) **Multinomial resampling:** After generating $\{X_j\}_{j=1}^N$ i.i.d. $\mathcal{N}(0,1)$ random variables, select $M \sim \text{Multinom}(N, \{\omega\}_{j=1}^N)$, where each ω_i is a normalized importance weight i.e.

$$\omega_i = \frac{\frac{\phi_\sigma(X_i)}{\phi(X_i)}}{\sum_{j=1}^N \frac{\phi_\sigma(X_j)}{\phi(X_j)}},$$

where ϕ_σ and ϕ are the probability densities of $\mathcal{N}(0,\sigma^2)$ and $\mathcal{N}(0,1)$ respectively. The final output sample is then $\{Y_j\}_{j=1}^N$, where for $j = 1,...,N-1$,

$$Y_j = X_i \text{ for } \sum_{k=1}^{i} M_k \leq j \leq \sum_{k=1}^{i+1} M_k,$$

and $Y_N = X_N$ for $N - \sum_{k=1}^{N-1} M_k \leq j \leq N$.

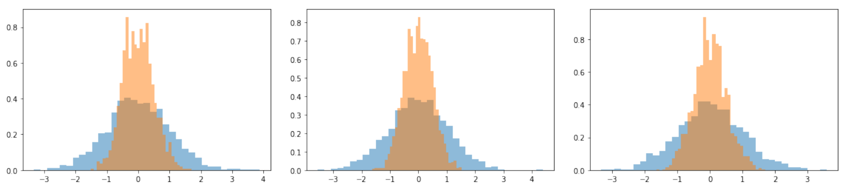
(ii) **Bernoulli resampling** After generating $\{X_j\}_{j=1}^N$ i.i.d. $\mathcal{N}(0,1)$ random variables, select floored weights $Y = \{\lfloor N \cdot \omega_i \rfloor\}_{i=1}^N$. Next, we select a random uniform random variables $U_i \sim \text{Unif}[0,1]$ for each $i = 1,...,N$ and similarly sample as above with $\{M_j\}_{j=1}^N$ defined to be

$$M_j = Y_j + \mathbb{1}_{U_i < N\omega_j - Y_j}\}.$$

(iii) **Systematic sampling:** After generating $\{X_j\}_{j=1}^N$ i.i.d. $\mathcal{N}(0,1)$ random variables, we generate $U \sim \text{Unif}[0,1]$ and write $Y_j = \frac{1}{N}(j-U)$ for $j = 1,...,N$. Our M_j as described above will now be determined by

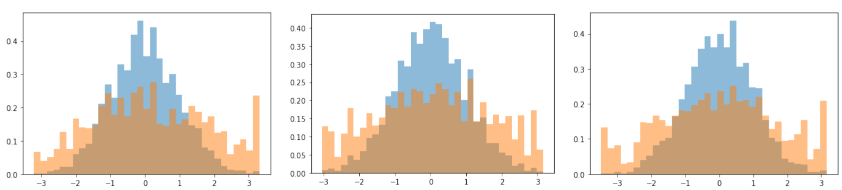
$$M_j = \left|\left\{k \leq N : Y_k \in \left[\sum_{l=1}^{k} \omega_l, \sum_{l=1}^{k+1} \omega_l\right]\right\}\right|.$$

To demonstrate each of the sampling outputs, below are three histograms demonstraing multinomial, Bernoulli, and systemtaic resampling methods for the $\mathcal{N}(0,2)$ target density on $N = 2000$ examples. The orange histograms in Figures 1, 2, and 3 clearly show that the samples of $\mathcal{N}(0,1)$ are resampled with higher probability in different regions, depedning on the choice of σ^2, allowing for the new sample to appear sparser/denser accordingly. This can also be verified by the distribution of importance weights relative to the original sample (shown in Figure 4). we also include plots for various levels of σ^2, such as $\sigma^2 = 0.5, 2, 5$, to demonstrate issues arising in IS methods when the distance between the reference and target distributions increase (in total-variation distance for example).



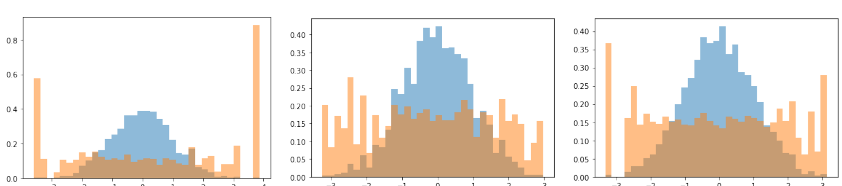
(a) Multinomial method (b) Bernoulli method (c) Systematic method

Figure 1: Histogram of values of $\mathcal{N}(0,1)$ (blue) and $\mathcal{N}(0,0.5)$ (orange) using resampled IS methods



(a) Multinomial method (b) Bernoulli method (c) Systematic method

Figure 2: Histogram of values of $\mathcal{N}(0,1)$ (blue) and $\mathcal{N}(0,2)$ (orange) using resampled IS methods



(a) Multinomial method (b) Bernoulli method (c) Systematic method

Figure 3: Histogram of values of $\mathcal{N}(0,1)$ (blue) and $\mathcal{N}(0,5)$ (orange) using resampled IS methods

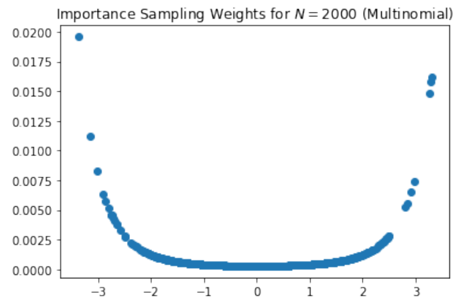


Figure 4: Importance weights v.s. original $\mathcal{N}(0,1)$ sample

To study the robustness of the the resampling method, we study what happens to $\bar{M}$, the emperical average of resampled copies of each $\{X_j\}_{j=1}^{N}$, as the variance σ^2 increases. As we can see, the number of points that are resampled on average increases, as more extreneous copying is needed to provide a more "uniform" distribution (Figure 5). Therefore, it appears that the methods are not robust to increasing σ^2, since the fraction of resampled values needs to increase closer to $\left(\frac{1}{10}\right)^{th}$ of the original sample size ($N = 100$). As we have already demonstrated in Figure 3, importance resampling will resample values closer to the tails of the distribution.

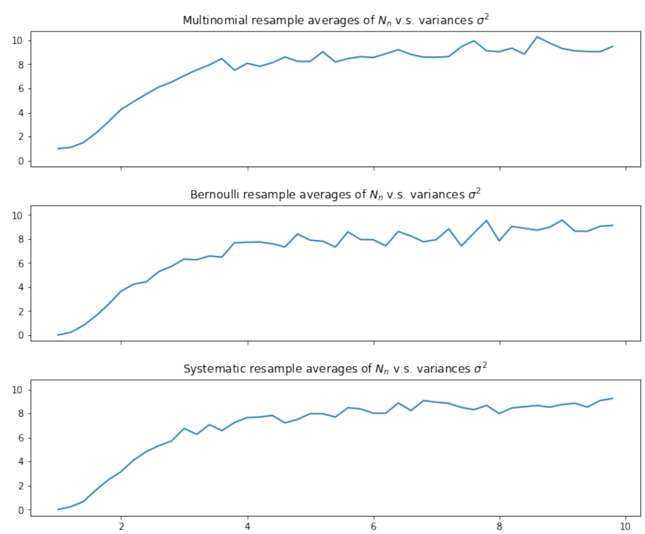
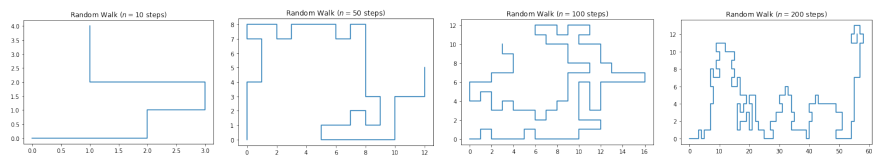


Figure 5: Averaged resampling numbers $\{M_j\}$ for Multinomial, Bernoulli, and systematic methods for $\sigma^2 \in [1, 10]$

2 Exercise 29

When performing averages over the uniform distribution over self-avoiding random walks on $\mathbb{Z}_d^2 = \{0, 1, ..., d-1\}^2$, otherwise known as SAW(d), it is difficult to naturally sample without the help of sequential importance sampling techniques which make use of the most natural form of sampling technique: for a SAW(d-1) $X_{1:d-1}$, samples each *acceptable* neighbor of X_{d-1} uniformly at random (where *acceptable* means that $X_{1:d} \in SAW(d)$). Below are a selected group of such self-avoiding random walks of size $n = 10, 20, 100$, and 200.



(a) Example of SAW(10)

(b) Example of SAW(20)

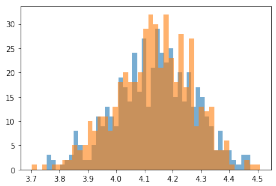
(c) Example of SAW(100)

(d) Example of SAW(200)

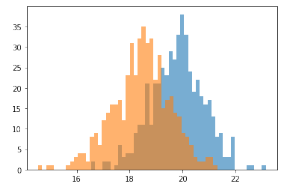
Using the (resampling) sequential importance sampling method to estimate averages against SAW(d), by the property of the recursive importance weight $w_n(X_{1:n}) = \frac{m(X_{1:n-1})\mathcal{Z}_n}{\mathcal{Z}_{n-1}}$,where $m(X_{1:n})$ is the number of *acceptable* neighbors of $X_{1:n}$, and the normalization of the importance weights, we can write the recursive formula for the importance weight at instance $(n+1)$ of sample (k) as

$$W_{n+1}^{(k)} = \frac{W_n^{(k)} m(X_{1:d}^{(k)})}{\sum_{j=1}^{K} W_n^{(j)} m(X_{1:d}^{(j)})}.$$

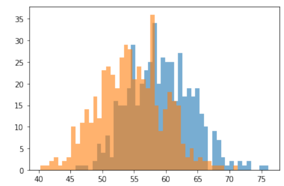
Note that using normalization removes the partition function constants $\mathcal{Z}_n$ through cancelation. The resampling sequential importance sampling method is also implemented as well. We use this technique in our implementation. In our analysis of both methods, three functions were tested in each IS method to see whether their statistics match what is expected of the uniform SAW(d) measure. They are (i) a lattice-site visiting function for a specific lattice site $(i, j) \in \mathbb{Z}_d^2$, (ii) a crashing function which is an indicator for whether the SAW "crashes" into itself, and (iii) a mean-square distance function which is just the L_2 norm of the last point in SAW(d). Below in the following figures are the SIS estimates of (i), (ii), and (iii), where the lattice site is $(5, 5)$. Below we show the SIS method and the resampling SIS method with the multinomial resampling feature turned off, noting that they behave quite similarly. However, when we do turn on multinomial resampling, the estimates differ (here we are estimated $\langle M_d^2 \rangle$, the mean-square displacement from the origin). We show the SIS estimates for $d = 8$ and for $d = 15$ dimensional SAW(d) paths averaged over $N = 100$ samples (we repeated the experiment 500 to create a histogram of IS estimates, as done previously throughout the paper).



(a) SIS v.s. Resampled SIS w/ resampling turned off, $d = 8$



(b) SIS (orange) v.s. Resampled SIS (blue), $d = 8$



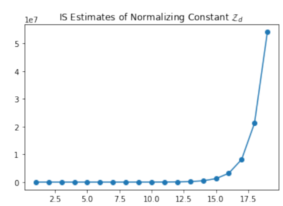
(c) SIS (orange) v.s. Resampled SIS (blue), $d = 15$

Figure 7: Estimates of $\langle M_d^2 \rangle$ (mean-squared displacement from origin) for $d = 8, 15$

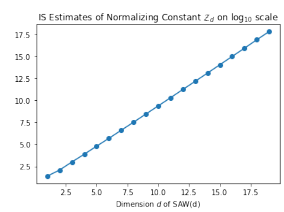
To compute the normalizing constant, notice that

$$\mathbb{E}\left[\prod_{j=1}^{n} w_j\right] = \mathbb{E}\left[\frac{m(X_{1:n-1})\mathcal{Z}_n}{\mathcal{Z}_{n-1}} \cdot \frac{m(X_{1:n-2})\mathcal{Z}_{n-1}}{\mathcal{Z}_{n-2}} \cdot \ldots \cdot \frac{m(X_2)\mathcal{Z}_2}{\mathcal{Z}_1} \cdot \frac{m(X_2)\mathcal{Z}_1}{\mathcal{Z}_0}\right] = \mathcal{Z}_n,$$

i.e. all we need to compute is the average of the products of importance weights over many instances of IS. Below in Figure is the scatter plot of $\mathcal{Z}_d$ against d, the dimension (in both the regular and $\log_{10}$ scale), so $\mathcal{Z}_d \sim q\alpha^N$, for some constant α. See Figure 8a and Figure 8b for more emperical evidence of the scaling law.



(a) Plot of IS estimates of normalization constant



(b) Plot of IS estimates of $\log_1 0$ normalization constant