HW2 - Jacob Shkrob

Jacob Shkrob

September 28, 2022

1 Exercise 16

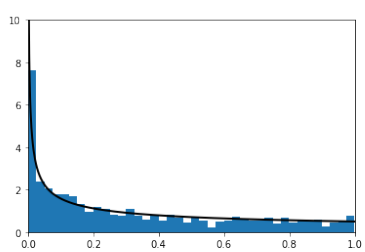
In this exercise, our goal is to write code which can generate samples from $\pi(x)$, where

$$\pi(x) = \frac{1}{2\sqrt{x}},$$

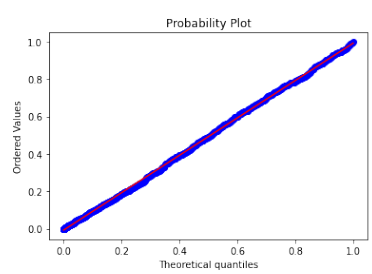
using the inversion method. In the inversion method, we use $U_i \sim \mathrm{Unif}[0,1]$ iid random variables and compute $F_\pi^{-1}(U_i)$, where $F_\pi(x)$ is the cummulative distribution function for the measure $\pi(x)\,dx$. In this case, it is very straightforward to compute the CDF of π:

$$\int_0^x \frac{1}{2\sqrt{y}}\,dy = \sqrt{x} = F_\pi(x). \tag{1.1}$$

Therefore, the inverse $F_\pi^{-1}(x)$ must trivially be written as $\frac{1}{2\sqrt{x}}$. The following diagram shows a histogram of $N = 1000$ simulated $\mathrm{Unif}[0,1]$ random variables after mapping them with $F_\pi^{-1}(x)$. The black graph in Figure 1a is the density $\pi(x)\,dx$, which aligns with the formed histogram as expected. To compute the QQ-plot for our samples, we used the `scipy.stats.probplot` API to generate the correct quantiles for our distribution and align our data. Figure 1b shows that the fit of the generated data seems to be accurate.



(a) Histogram of $F_\pi^{-1}(U_i)$, where $i \in \{1,\ldots,N\}$



(b) QQ-plot w.r.t π from $N = 1000$ samples $F_\pi^{-1}(U_i)$

Figure 1: Exercise 16 Figures

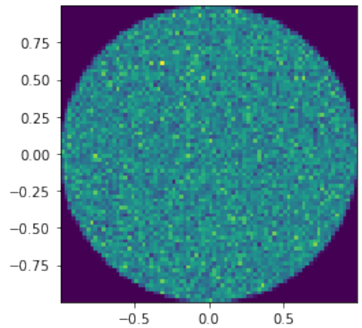


Figure 2: Histogram for change of variable method to generate $\mathrm{Unif}(S^1)$

2 Exercise 18

Given two uniform random variables $U_i \sim \mathrm{Unif}[0,1]$ where $i = 1, 2$, our goal is to randomly generate a uniformly distributed point in the unit disc. Suppose $\mathrm{Unif}(S^1)$ denotes the uniform distribution on the unit disk. Then the density of $X \sim \mathrm{Unif}(S^1)$ is $\frac{1}{\pi}\mathbb{1}_{\|x\|\le 1}$. To generate such a density, we need to use the transformation formula for functions of random variables. From our class, we know that if $Y \sim \tilde{\pi}$, then $X = \pi(Y)$ has the density

$$\frac{\tilde{\pi}(\phi^{-1}(x))}{|J\phi(\phi^{-1}(x))|},$$

where J is the Jacobian i.e. $J\phi = \det(D\phi)$. Define the transformations $\phi(u_0, u_1) = (\phi_0(u_0, u_1), \phi_1(u_0, u_1))$ to be

$$\phi_0(u_0, u_1) = \sqrt{u_0}\cos(2\pi u_1) \tag{2.1}$$
$$\phi_1(u_0, u_1) = \sqrt{u_0}\sin(2\pi u_1). \tag{2.2}$$

Then the Jacobian $J\phi$ is

$$\begin{vmatrix} \frac{1}{2\sqrt{u_0}}\cos(2\pi u_1) & -2\pi\sqrt{u_0}\sin(2\pi u_1) \\ \frac{1}{2\sqrt{u_0}}\sin(2\pi u_1) & 2\pi\sqrt{u_0}cos(2\pi u_1) \end{vmatrix} = \pi,$$

and therefore the distribution of $\phi(U_1, U_2)$ must be $\frac{1}{\pi}\mathbb{1}_{\|x\|\le 1}$. To verify that this transformation is correct, we plot a 2-dimensional histogram of samples generated by $\phi(U_0, U_1)$ in Figure 2. Note that our histogram uses $N = 100,000$ samples and appears to uniformly space the unit disc.

3 Exercise 19

We impement the rejection sampling method where the target density is the uniform measure on S^1 and the reference measure is $U[0,1]$. Since the only

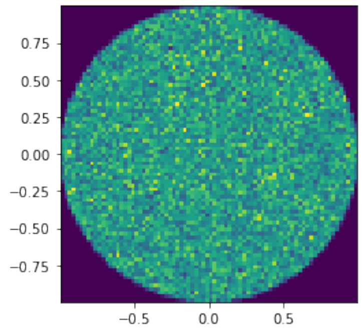


Figure 3: Histogram for rejection sampling method to generate $\mathrm{Unif}(S^1)$

constant we have to worry about is $\frac{1}{\pi}$, the most natural constant K to select for the algorithm to work (i.e. for $\tilde{\pi}(x) \leq K\pi(x)$) is $K = 1$. One slight caviat for rejection sampling using the standard method is that our choice of representative $Y^{(\tau)}$, after the rejection sampling condition, may not be fully supported in the original domain of the target measure π. For example, $Y^{(\tau)}$ is this case will only be a sample from $\mathrm{Unif}(S^1)$ conditioned on staying in the top-right quadrant. To get a truely uniform distribution on the sphere, another random variable $W \sim \mathrm{Unif}[0,1]$ is drawn in order to determine which quadrant the eventual drawn rejection sample goes to (this is done by multiplying 1 or -1 to one or both coordinates in $Y = (Y_1, Y_2)$). As shown in Figure 3, our samples appear to be uniformly distributed in S^1.

4 Exercise 20

Our goal for this exercise is to estimate the probability $\mathbb{P}(X > 2)$, where $X \sim \mathcal{N}(0,1)$, using importance sampling techniques. In this case, our estimator $\tilde{f}_N$ is equal to

$$\tilde{f}_N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{(Y_i > 2)} \frac{\phi(Y_i)}{\phi_{m,s}(Y_i)},$$

where ϕ is the probability density function (pdf) for $\mathcal{N}(0,1)$, $\phi_{m,s}$ is the pdf of $\mathcal{N}(m,s)$, and $Y_i \sim \mathcal{N}(m,s)$. Figure 4 shows a histogram of 1000 simulations of the estimator $\tilde{f}_N$, where $N = 1000$. As we can see, the estimator is cenetered at approximately 0.023, so we can safely say that the probability $\mathbb{P}(X > 2) \approx 0.023$ (the true answer is around 0.02275013).

What happens when instead of using a fixed m and s, we vary m and s to achieve a bester estimate? To do so, we run two different experiments: the first experiment varies the mean levels (i.e. m) while fixing the variance $s = 1$ and the second experiment varies the standard deviation levels (i.e. s) while fixing the mean $m = 0$ In both experiment, the metric we use is the sample variance for each instance of samples generated using IS. These two experiments will

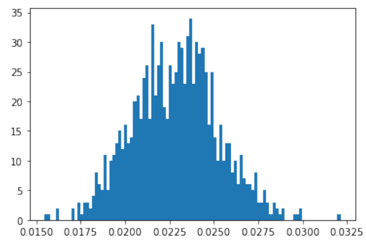
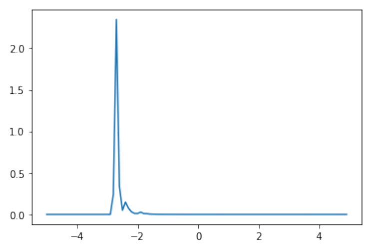
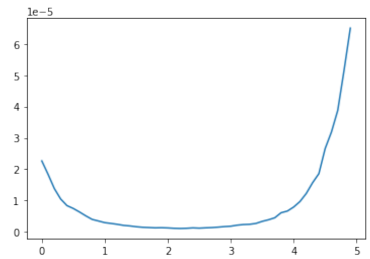


Figure 4: Histogram of importance sampling estimates using $\mathcal{N}(0,2)$ to estimate $\mathcal{N}(0,1)$



(a) Sample $\mathbf{Var}\left(\tilde{f_N}\right)$ estimates for $m \in [-5.0, 5.0]$, $s = 1$



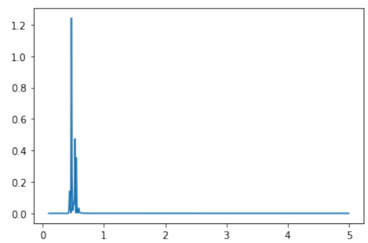
(b) Sample $\mathbf{Var}\left(\tilde{f_N}\right)$ estimates for $m \in [0, 5.0]$, $s = 1$

Figure 5: Graphs illustrating sample variance of $\tilde{f}_N$ at various levels of m mean

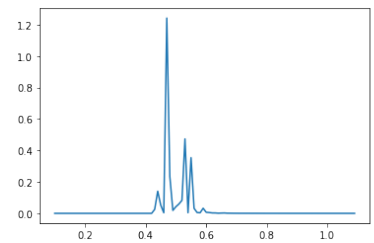
show how $\tilde{f}_N$ is affected by changes in m and s, as well as the proper reference distribution to select.

For the first experiment, we considered mean values between -5.0 and 5.0, with increments of 0.1. We fixed $N = 1000$ (the average in the estimate) and $M = 1000$ (the number of importance samples taken) for both exepriments. The first plot shows the sample variances within each of the importance sample groups for each $m = -5.0, ..., 5.0$. In Figure 5a, the effect of the mean m is still unclear, as most cases appear to have little effect on the overall spread of the random variable. Only when the mean is near $[-2.5, -2]$ does the actual variance of the importance samples become unstable. However, once we zoom into an appropriate part of the graph, as in Figure 5b, we can see that the optimal mean m is closer to 2 than it is to 0, by a relatively small factor of $2 \cdot 10^{-5}$.

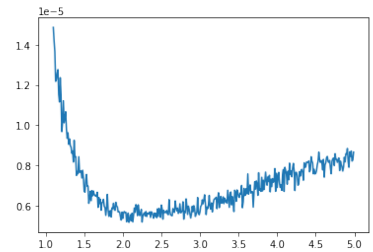
For the second experiment, we considered variance values between 0.1 and 5, with increments of 0.01, fixing N and M exactly as in the first experiment (we fix $m = 1$ in this setting as well). Again, as we can see in Figure 6, the stochasticity lies within a narrow region between $[0, 1]$, where the variance becomes unstable. If we examine the estimates in the first half, the tale is that the variance explodes near 0.5. In the second half, the variance minimizes when $s \approx 2$ and slowly climbs upwards as $s \to 5^+$. This suggests that a reference distribution with thicker tails may actually help with estimation.



(a) Sample $\mathbf{Var}(\tilde{f_N})$ estimates for $s \in [0.1, 5.0]$, $m = 1$



(b) Sample $\mathbf{Var}(\tilde{f_N})$ estimates for $s \in [0.1, 2.5]$, $m = 1$



(c) Sample $\mathbf{Var}(\tilde{f_N})$ estimates for $m \in [2.5, 5.0]$, $m = 1$

Figure 6: Graphs illustrating sample variance of $\tilde{f}_N$ at various levels of m mean

5 Exercise 21

This exercise is virtually identical to Exercise 20, only instead we need to estimate $\mathcal{Z}$, which is the normalizing constant $\int_{\mathbb{R}} e^{-|x|^3}\,dx$. To estimate this, we use the fact that if a target distribution p has normalizing constant $\mathcal{Z}_p$ and a reference distribution q has a normalizing constant $\mathcal{Z}_q$, then

$$\frac{1}{N}\sum_{i=1}^{N}\frac{p(Y_i)}{q(Y_i)} \xrightarrow{\text{Pr}} \frac{\mathcal{Z}_p}{\mathcal{Z}_q}, \text{ as } N \to \infty.$$

In our case, we have that $p(x) = e^{-|x|^3}$ and $\frac{q(x)}{\mathcal{Z}_q} = \phi(x)$. In Figure 7, we show the histogram of the estimates of $\mathcal{Z}$.

6 Exercise 22

We now look at the estimator $\frac{\tilde{f}_N}{\tilde{1}_N}$ which is biased estimator of $\pi[f]$, but asymptotically unbiased as $N \to \infty$. In Figure 8, we compare different levels of centrality m of the reference distribution when running IS. Notice that as we increase the centrality from $m = -1$ to $m = 3$, the IS method $\tilde{f}_N$ concentrates around the correct value, whereas the alternative IS method $\frac{\tilde{f}_N}{\tilde{1}_N}$ has higher variance. For this reason, it is wiser in this case to consider using the estimator $\tilde{f}_N$ for the better control in the estimates.

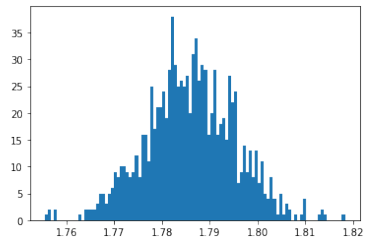
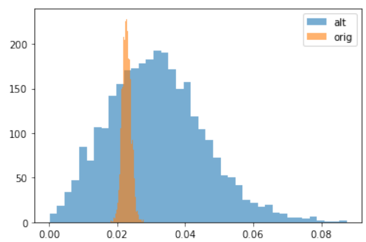
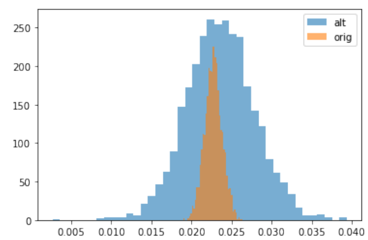


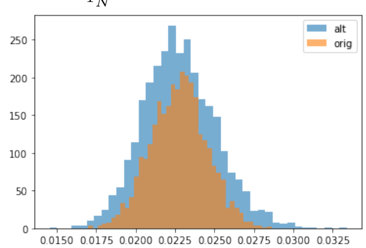
Figure 7: Histogram of importance sampling estimates using $\mathcal{N}(0,1)$ to estimate $\mathcal{Z} = \int_{\mathbb{R}} e^{-|x|^3}\, dx$



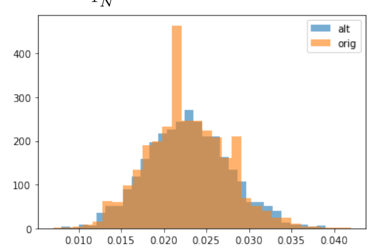
(a) $\frac{\tilde{f}_N}{\tilde{1}_N}$ v.s. $\tilde{f}_N$: $\mathcal{N}(3,1)$ case



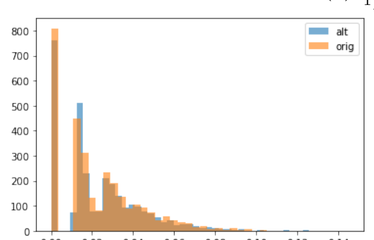
(b) $\frac{\tilde{f}_N}{\tilde{1}_N}$ v.s. $\tilde{f}_N$: $\mathcal{N}(2,1)$ case



(c) $\frac{\tilde{f}_N}{\tilde{1}_N}$ v.s. $\tilde{f}_N$: $\mathcal{N}(1,1)$ case



(d) $\frac{\tilde{f}_N}{\tilde{1}_N}$ v.s. $\tilde{f}_N$: $\mathcal{N}(0,1)$ case



(e) $\frac{\tilde{f}_N}{\tilde{1}_N}$ v.s. $\tilde{f}_N$: $\mathcal{N}(-1,1)$ case

Figure 8: Histograms for importance sampling estimates using $\frac{\tilde{f}_N}{N}$ and $\tilde{f}_N$, where $M = 3000, N = 1000$. In the key, "alt" means estimator is $\frac{\tilde{f}_N}{\tilde{1}_N}$ and "orig" means estimator is $\tilde{f}_N$.