Reinforcement Learning Assignment-1

Utkarsh Prakash
180030042

February 4, 2022

1. Histogram obtained after sampling $N = 100$ samples:

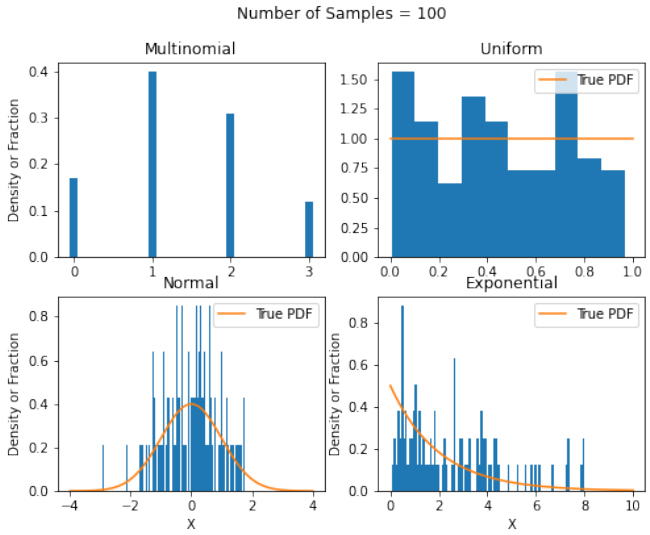


Figure 1: Histogram obtained after sampling $N = 100$ samples.

Here, since the number of samples is less hence we see a lot deviation between the plotted histogram and the true PDF.

Histogram obtained after sampling $N = 1000$ samples:

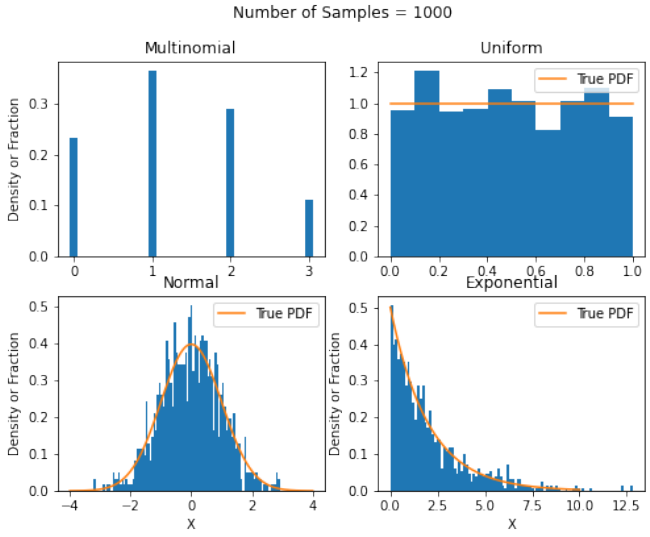


Figure 2: Histogram obtained after sampling $N = 1000$ samples.

Histogram obtained after sampling $N = 10000$ samples:

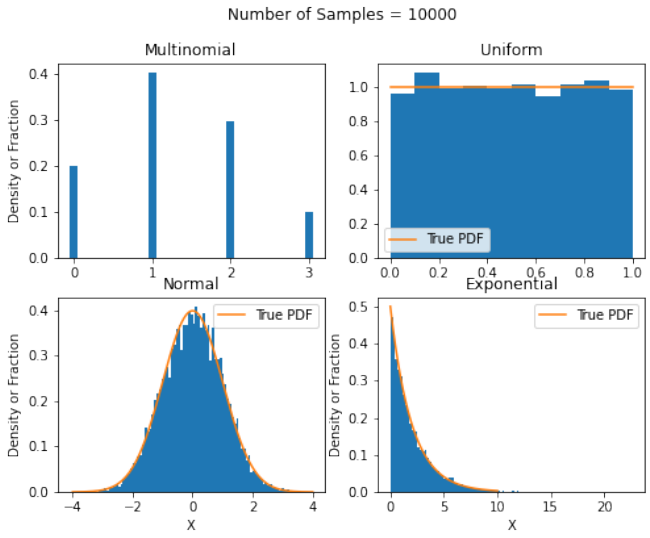


Figure 3: Histogram obtained after sampling $N = 10000$ samples.

We observe that as we increase the number of samples the plotted histogram and the true PDF tend to match. This way we are guaranteed that the samples were drawn from the same distribution.

2. **Solution 1:** Using Central Limit Theorem, we know that if $X_1, X_2, ..., X_n$ are i.i.d. random variable with mean μ and variance σ^2, then

$$\sqrt(n)\frac{\bar{X_n} - \mu}{\sigma} \xrightarrow{n\to\infty} \mathcal{N}(0, 1)$$

where $\bar{X_n}$ is the sample mean. Now, we have samples from uniform distribution between 0 and 1 i.e., $\mu = 1/2$ and $\sigma^2 = 1/12$. The sample mean of these samples will be a sample from

standard normal distribution. We can generate samples from a normal distribution with mean μ and σ, using the transformation $y = \sigma z + \mu$, where z is the sample from standard normal distribution.

The following histogram plots the samples drawn from normal distribution with $\mu = 2$ and $\sigma = 2$ using the method described above. The graph also shows the true PDF for the normal distribution. Since, the two match, we are guaranteed that the samples were drawn from the normal distribution with $\mu = 2$ and $\sigma = 2$.

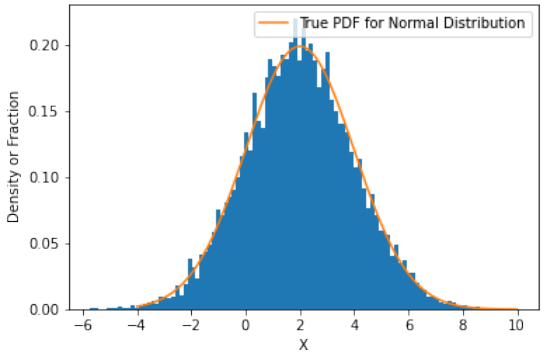


Figure 4: Histogram obtained after sampling $N = 10000$ samples.

Solution 2: Let $U \sim Unif(0, 1)$ and $\Phi(x)$ denote the standard normal CDF. Then, we can show that if $X = \Phi^{-1}(U)$, then $X \sim \mathcal{N}(0, 1)$. This is because $P(X \leq t) = P(\Phi^{-1}(U) \leq t) = P(U \leq \Phi(t)) = \Phi(t)$. Now, since the CDF X is $\Phi(x)$, hence, $X \sim \mathcal{N}(0, 1)$.

Let u be a sample from $Unif(0, 1)$. Then using the above fact, we can generate samples from standard normal random samples as $x = \Phi^{-1}(u)$. Then, to generate normal samples with mean μ and variance σ^2 we can simply transform as $y = \sigma x + \mu$, where y is the sample from normal random variable with mean μ and variance σ^2.

The plotted histogram are the samples drawn using the method described above using $\mu = 1$ and $\sigma^2 = 1$. The graph also shows the true PDF for the standard normal distribution. Since, the two match, we are guaranteed that the samples were drawn from the standard normal distribution.

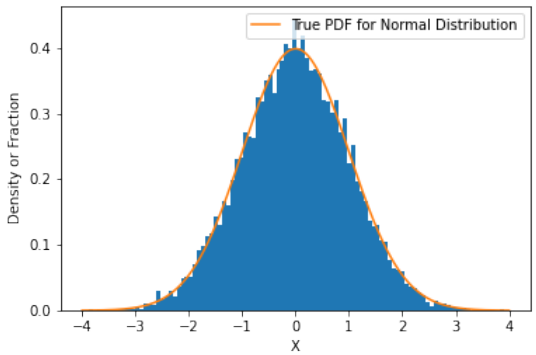


Figure 5: Histogram obtained after sampling $N = 10000$ samples.

3. In order to compute the integral, we can represent the integration as an expectation of a random variable. Then using Law of Large Numbers, we can compute the expectation, by sampling from the distribution of that random variable.

   Let's $X \sim Unif(0, \pi)$ and $f(x)$ be it's PDF i.e.,

$$f(x) = \begin{cases} \frac{1}{\pi} & if 0 \leq x \leq \pi \\ 0 & otherwise \end{cases}$$

   (a) The graph for $\sqrt{(}sin(x))$ is plotted as below:

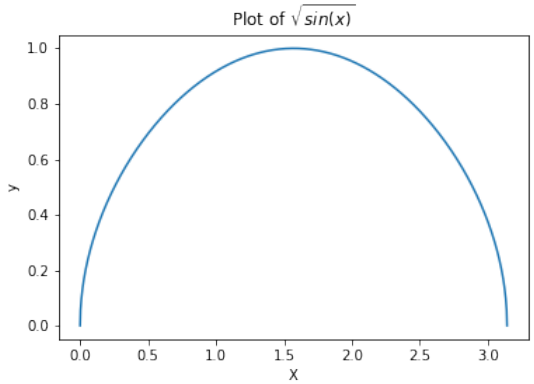


Figure 6: Graph for $\sqrt{(}sin(x))$

   Let $g(X) = \sqrt{(}sin(X))$, where X is as defined above. Then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx = \frac{1}{\pi} \int_{0}^{\pi} g(x)\, dx \tag{1}$$

   Let $X_1, X_2, ..., X_N \overset{i.i.d.}{\sim} \text{Unif}(0, \pi)$ and $Y_i = g(X_i)$ for $i \in 1, 2, ...N$. Then according to Law of Large Numbers, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Y_i = \mathbb{E}[g(X)]$$

   Therefore,

$$\int_{0}^{\pi} \sqrt{(}sin(x))\, dx = \pi \left(\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Y_i \right) \tag{2}$$

   The value of the integral obtained by this method is 2.4061907079036167.

   (b) The graph for $\sqrt{(}sin(x))exp(-x^2)$ is plotted as below:

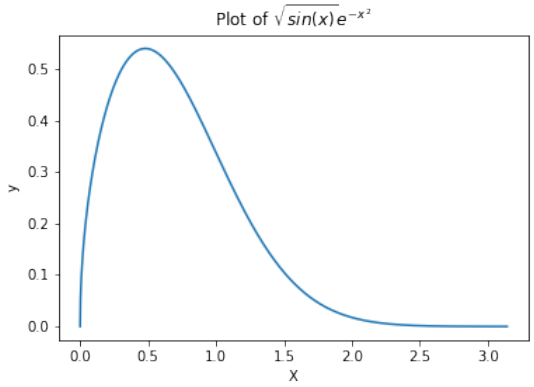


Figure 7: Graph for $\sqrt{(}sin(x))exp(-x^2)$

Using Normal Random Variable

Let $X \sim \mathcal{N}(0, \frac{1}{2})$ where $\frac{1}{2}$ is the variance X. Now, to apply the above trick of converting integral into expectation of a random variable, we need to confine the density of the above random variable between $[0, \pi]$. To do this we construct a new random variable Y such that we sample from the distribution of $|X|$ and discard the samples which are greater than π. Once, we do this we can write the integral as:

$$\int_0^\pi \sqrt{(}sin(x))exp(-x^2)\,dx = \lambda\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sqrt{(}sin(Y_i))\right)$$

where $Y_i \overset{i.i.d.}{\sim} Y$ Now, all that remains is to determine the value of λ. This can be determined if we compute the PDF for Y. We know that the PDF of $|X|$ is defined as follows:

$$f_x(|X|) = \frac{2\sqrt{2}}{\pi}exp(-x^2)$$

We know that the PDF of Y is:

$$f_y(Y) = \begin{cases} Cf_x(|X|) & if 0 \le X \le \pi \\ 0 & otherwise \end{cases}$$

where C is a normalizing constant such that

$$\int_0^\pi f_y(Y)\,dy = 1$$
$$\implies \int_0^\pi Cf_x(|X|)\,dx = 1$$
$$\implies C = \frac{1}{\int_0^\pi f_x(|X|)\,dx}$$

We can numerically estimate the value of $\int_0^\pi Cf_x(|X|)\,dx$ by using samples from Uniform distribution and applying the procedure as in part(a). We found this value to be 0.7956. Therefore, $C = 1.2569$. Therefore,

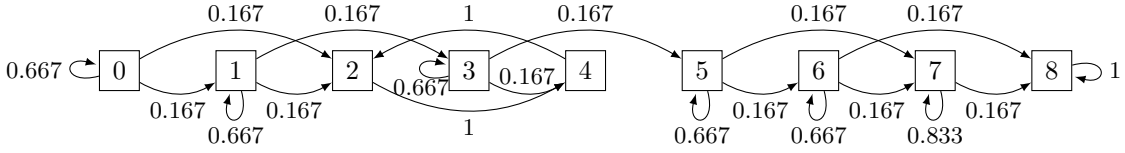
$$\lambda = \frac{2\sqrt{2}C}{\pi}$$

Using, the technique above we found the value of integral $\int_0^\pi \sqrt{(}sin(x))exp(-x^2)\,dx$ to be 0.570899 with just 10 samples from Y.

Using Uniform Random Variable

We can follow the same procedure as in part (a) but by defining $g(X) = \sqrt{(sin(X))exp(-X^2)}$. The value of the integral obtained by this method is 0.501713 with 10 samples from the uniform distribution.
Clearly, we can see the value obtained from uniform distribution is way far-off from the value obtained from normal distribution. Therefore, we can say that using Normal Distribution is sample-efficient.

4. **Assumption:** If the die shows up 2 when in the state 7, the player remains in the same state. The Markov Chain for the game can be drawn as follows:



The transition probability matrix can be represented as follows:

$$Q = \begin{bmatrix} 2/3 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2/3 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & 1/6 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/3 & 1/6 & 1/6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The initial distribution of states can be represented as follows:

$$\pi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Simulation

We can consider an indicator random variable I such that it takes value of 1 when the player reaches the end state else it takes the value of 0. Now, the expectation of this random variable is the probability of the player ever reaching the end state. Using, Law of Large Numbers we can calculate this probability by simulating this Markov Chain (i.e. playing this game) multiple times and calculating the average number of times the player reaches the end state.

The question now that arises is that how long should we simulate each game? For this we calculate the value of $\pi_0 Q^T$ for large values of T. We observe that the values of $\pi_0 Q^T$ for large values of T alternate between $\begin{bmatrix} 0 & 0 & 0.482 & 0 & 0.393 & 0 & 0 & 0 & 0.125 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 & 0.393 & 0 & 0.482 & 0 & 0 & 0 & 0.125 \end{bmatrix}$. We know from this that in the long run following things happen:

- The probability of player being in states 0, 1, 3, 5, 6 and 7 is almost 0.
- The probability of player being in states 2 and 4 alternates between 0.393 and 0.482.
- The probability of player being in state 8 is 0.125.

In other words, in the long run the player can only be in states 2, 4 or 8. Therefore, while simulating the Chain, we need to only simulate to the point where the player either reaches the state 2 or 4 (then the player cannot reach end state) or the point where the player reaches the state 5 (because then the player is guaranteed in the long run will reach the end state).

On simulating the Chain for 10,000 times the probability of ever reaching the end state is found to be 0.12555.

Analytic Solution

Let α_i denote the probability of ever-reaching the end state from state i. We can create the following equations recursively:

$$\alpha_0 = \frac{1}{6}\alpha_1 + \frac{1}{6}\alpha_2 + \frac{2}{3}\alpha_0$$
$$\alpha_1 = \frac{1}{6}\alpha_2 + \frac{1}{6}\alpha_3 + \frac{2}{3}\alpha_1$$
$$\alpha_2 = \frac{1}{6}\alpha_3 + \frac{1}{6}\alpha_4 + \frac{2}{3}\alpha_2$$
$$\alpha_3 = \frac{1}{6}\alpha_4 + \frac{1}{6}\alpha_5 + \frac{2}{3}\alpha_3$$
$$\alpha_4 = \frac{1}{6}\alpha_5 + \frac{1}{6}\alpha_6 + \frac{2}{3}\alpha_4$$
$$\alpha_5 = \frac{1}{6}\alpha_6 + \frac{1}{6}\alpha_7 + \frac{2}{3}\alpha_5$$
$$\alpha_6 = \frac{1}{6}\alpha_7 + \frac{1}{6}\alpha_8 + \frac{2}{3}\alpha_6$$
$$\alpha_7 = \frac{1}{6}\alpha_8 + \frac{5}{6}\alpha_7$$

We know that $\alpha_2 = \alpha_4 = 0$ and $\alpha_8 = 1$. On solving the above set of equations we get $\alpha_5 = \alpha_6 = \alpha_7 = 1$, $\alpha_3 = 1/2$, $\alpha_1 = 1/4$ and $\alpha_0 = 1/8$. Hence, the required probability is $\alpha_0 = 0.125$.

Another Solution

Let us consider a random variable X_n denoting the state of the player at timestamp n. Let I be an indicator random variable such that it takes value of 1 when the player reaches the end state else it takes the value of 0. Now, the probability of the player ever reaching the end state is sum of probabilities of it reaching the end state at timestamp $T = 5, 6, ...$, where the probability of a player reaching the end state at time T is probability that the player is in state 8 at time T but was not in state 8 at $T-1$ since 8 is an absorbing state. We start the sum from 5 because the earliest the player can reach the end state is at timestamp 5. This can be written down as follows:

$$P(I=1) = \sum_{t=5}^{\infty} P(X_t = 8, X_{t-1} \neq 8) = \sum_{t=5}^{\infty} P\left(X_t = 8, \bigcup_{i=0}^{7}(X_{t-1} = i)\right)$$

The last equality holds because if the player is not in state 8 at time $T-1$ implies that the player was in one of the states from 0 to 7. We can further simplify the above equation as

$$P(I=1) = \sum_{t=5}^{\infty}\sum_{i=0}^{7} P(X_t = 8, X_{t-1} = i)$$
$$= \sum_{t=5}^{\infty}\sum_{i=0}^{7} P(X_t = 8 | X_{t-1} = i) P(X_{t-1} = i)$$
$$= \sum_{t=5}^{\infty}(P(X_t = 8|X_{t-1} = 6)P(X_{t-1} = 6) + P(X_t = 8|X_{t-1} = 7)P(X_{t-1} = 7))$$

The last equality holds because the transition probabilities $P(X_t = 8|X_{t-1} = i) = 0 \forall i \in 0, 1, ..., 5$. Now, we know that $P(X_{t-1} = j) = (\pi_0 Q^{t-1})_j$, where $(\cdot)_j$ represents the jth element of the matrix. Using this we can simplify the above equation as

$$P(I=1) = \sum_{t=5}^{\infty} \frac{1}{6}((\pi_0 Q^{t-1})_6) + \frac{1}{6}((\pi_0 Q^{t-1})_7)$$

Now, we know that $\sum_{t=0}^{\infty} A^t$ for any matrix A converges only when the largest eigenvalue of $A < 1$. However, we find that the largest eigenvalue of Q is 1. So, in compute the above sum

approximately, we compute it till timestamp $t = 10000$ i.e.,

$$P(I = 1) \approx \sum_{t=5}^{10000} \frac{1}{6}((\pi_0 Q^{t-1})_6) + \frac{1}{6}((\pi_0 Q^{t-1})_7)$$

The probability thus obtained is 0.1249, which conforms to the probability found earlier using simulation.