SMAI Assignment I

Alapan Chaudhuri, 2019111023

September 2021

Question 1

Give an example each of probability mass functions with finite and infinite ranges. Show that the conditions on PMF are satisfied by your example.

PMF with finite range

Consider a Bernoulli distribution, we can represent its probability mass function as follows (with $0 \leq p \leq 1$).

$$p_X(x) = \begin{cases} p, & \text{for x = 1} \\ 1 - p, & \text{for x = 0} \\ 0, & \text{otherwise} \end{cases}$$

As is evident in the above PMF, the range of X, $R_X = \{0, 1\}$ which is finite. Hence, we can have the PMF corresponding to Bernoulli distribution as a PMF with finite range.

Moreover, the necessary PMF conditions are satisfied, namely

1. $\forall_{x \in R_X} 0 \leq p_X(x) \leq 1$ holds true

   This does hold true in our case since we have,

   $$p_X(x = 1) = p, \; p_X(x = 0) = 1 - p$$

   and given that $0 \leq p \leq 1$,

   $$0 \leq p_X(x = 0), p_X(x = 1) \leq 1$$

2. $\sum_{x \in R_X} p_X(x) = 1$

   We have $p_X(x = 0) + p_X(x = 1) = p + 1 - p = 1$ which confirms this condition.

PMF with infinite range

Consider a Poisson distribution given by the PMF as follows.

$$p_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!} & \text{for } k \in 0, 1, 2, 3, ... \\ 0 & \text{otherwise} \end{cases}$$

The range of X, $R_X = \{0, 1, 2, 3, ...\}$ which is countably infinite. Hence the PMF of a Poisson distribution can be considered as a PMF with an infinite range.

Moreover, the necessary PMF conditions are satisfied, namely

1. $\sum_{k \in R_X} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$
2. $\forall_{x \in R_X} 0 \leq p_X(x)$ holds true since $\frac{e^{-\lambda}\lambda^k}{k!} \geq 0$ for all k.

Question 2

Show with complete steps that the variance of uniform density is given by equation 10. (Hint: use the expression for variance in equation 5.)

Solution

The uniform density function is characterized by the following function.

$$U(a, b) = \begin{cases} \frac{1}{b-a}, \text{ if } a \leq x \leq b \\ 0, \text{ otherwise} \end{cases}$$

Then the equation of variance is given by,

$$\sigma^2 = E[x^2] - E[x]^2$$

where,

$$E[x] = \int_a^b x \frac{1}{b-a} dx = \frac{b-a}{2}$$

and,

$$E[x^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{x^3}{3(b-a)} |_a^b = \frac{a^2 + ab + b^2}{3}$$

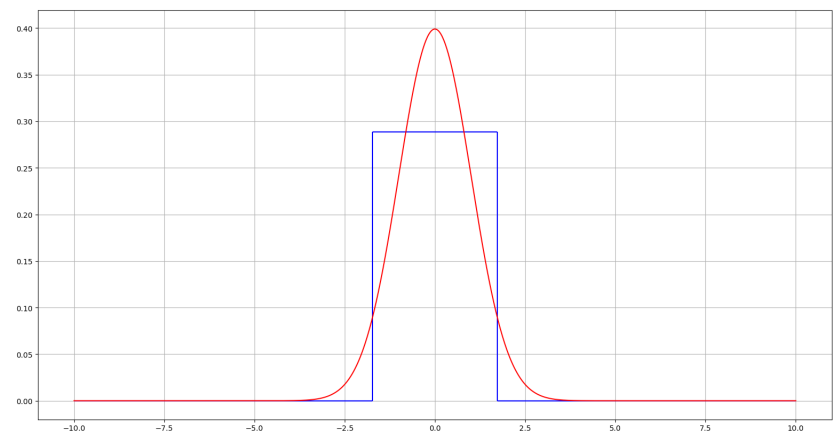
Thus, the required variance is,

$$\begin{aligned}\sigma^2 &= E[x^2] - E[x]^2 \\ &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{a^2 - 2ab + b^2}{3} \\ &= \frac{(b-a)^2}{12}\end{aligned}$$

Question 3

Show examples of two density functions (draw the function plots) that have the same mean and variance, but clearly different distributions. Plot both functions in the same graph with different colours.

Solution



In the above graph, the blue line represent a normal distribution where as the red line represents an uniform distribution which shares the same mean and variance as that of the normal distribution.

Uniform distribution function: $a = -\sqrt{3},\ b = \sqrt{3},\ \mu = 0,\ \sigma^2 = 1$
Gaussian distribution function: $\mu = 0,\ \sigma^2 = 1$

Clearly even though both of the distributions have the same mean and variance, their distributions are different.

Question 4

Show that the alternate expression for variance given in equation 5 holds for discrete random variables as well.

Solution

For a discrete random variable, we have

$$E[x] = \mu = \sum_{x \in \chi} x p_X(x)$$

$$\sum_{x \in \chi} p_X(x) = 1$$

Thus, the variance is given by

$$\begin{aligned}
\sigma^2 &= E[(x-\mu)^2] \\
&= \sum_{x \in \chi} (x-\mu)^2 p_X(x) \\
&= \sum_{x \in \chi} (x^2 - 2x\mu + \mu^2) p_X(x) \\
&= \sum_{x \in \chi} x^2 p_X(x) - 2\mu \sum_{x \in \chi} x p_X(x) + \mu^2 \sum_{x \in \chi} p_X(x) \\
&= E[x^2] - 2\mu.\mu + \mu^2.1 \\
&= E[x^2] - \mu^2 \\
&= E[x^2] - E[x]^2
\end{aligned}$$

Question 5

Prove that the mean and variance of a normal density, $N(\mu, \sigma^2)$ are indeed its parameters, μ and σ^2.

Solution

PDF of a normal distribution is given by

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

where μ and σ are two parameters with $\sigma > 0$.

Now, the mean of the distribution is given by

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} (x+\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= 0 + \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx
\end{aligned}
$$

The first term goes to 0 since, the function is a odd function. Now, solving the second term we have

$$
E[X] = \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx
$$

We substitute x with $\sqrt{2}\sigma x$ and get the following

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{\pi}} e^{-x^2} dx \\
&= \int_{0}^{\infty} \mu \frac{2}{\sqrt{\pi}} e^{-x^2} dx \\
&= \frac{2\mu}{\sqrt{\pi}} \int_{0}^{\infty} e^{-x^2} dx \\
&= \frac{2\mu}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} \\
&= \mu
\end{aligned}
$$

Hence, $E[X] = \mu$.

Now, the variance is given by

$$
Var(X) = E[X^2] - E[X]^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2
$$

By substituting x as $\dfrac{x-\mu}{\sqrt{2}\sigma}$ we get,

$$
E[X^2] = \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma x + \mu)^2 e^{-x^2} dt
$$

$$
\begin{aligned}
E[X^2] &= \frac{1}{\sqrt{\pi}}(2\sigma^2\int_{-\infty}^{\infty}x^2e^{-x^2}dx + 2\sqrt{2}\sigma\mu\int_{-\infty}^{\infty}xe^{-x^2}dx + \mu^2\int_{-\infty}^{\infty}e^{-x^2}dx)\\
&= \frac{1}{\sqrt{\pi}}(2\sigma^2\int_{-\infty}^{\infty}x^2e^{-x^2}dx + 2\sqrt{2}\sigma\mu[-\frac{1}{2}e^{-x^2}]_{-\infty}^{\infty} + \mu^2\sqrt{\pi})\\
&= \frac{1}{\sqrt{\pi}}(2\sigma^2\int_{-\infty}^{\infty}x^2e^{-x^2}dx + 2\sqrt{2}\sigma\mu\cdot 0) + \mu^2\\
&= \frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}x^2e^{-x^2}dx + \mu^2\\
&= \frac{2\sigma^2}{\sqrt{\pi}}([-\frac{x}{2}e^{-x^2}]_{-\infty}^{\infty} + \frac{1}{2}\int_{-\infty}^{\infty}e^{-x^2}dx) + \mu^2\\
&= \frac{2\sigma^2}{\sqrt{\pi}}\cdot\frac{1}{2}\int_{-\infty}^{\infty}e^{-x^2}dt + \mu^2\\
&= \frac{\sigma^2}{\sqrt{\pi}}\sqrt{\pi} + \mu^2\\
&= \sigma^2 + \mu^2
\end{aligned}
$$

Thus, we have the required value of variance as

$$Var(X) = E[X^2] - E[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Hence, we have correctly proved that the mean and variance of a normal density, $N(\mu, \sigma^2)$ are indeed its parameters, μ and σ^2.

Question 6

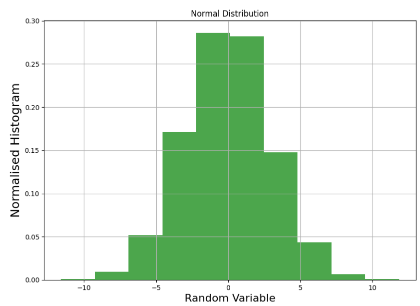
Using the inverse of CDFs, map a set of 10000 random numbers from $U[0, 1]$ to follow the following pdfs:

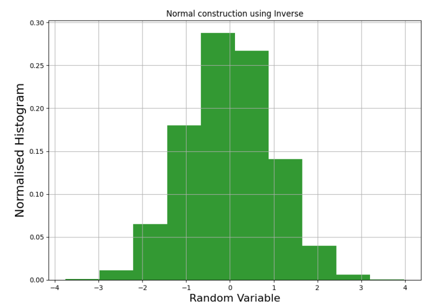
- Normal density with $\mu = 0, \sigma = 3.0$.
- Rayleigh density with $\sigma = 1.0$.
- Exponential density with $\lambda = 1.5$. Once the numbers are generated, plot the normalized histograms (the values in the bins should add up to 1) of the new random numbers with appropriate bin sizes in each case; along with their pdfs.

What do you infer from the plots?

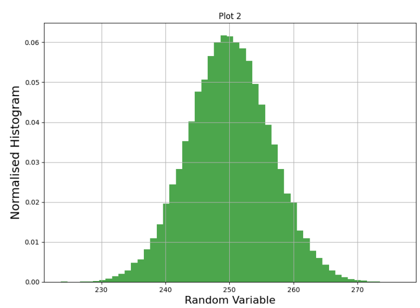
Solution

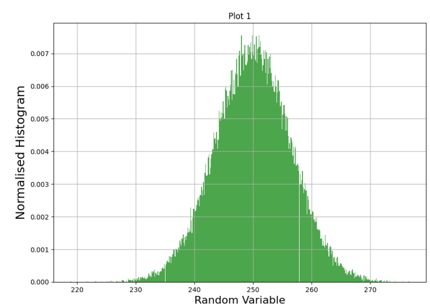
1. Normal Density Function:



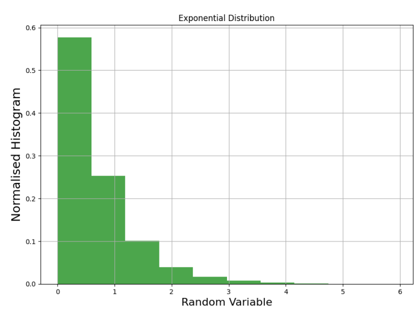


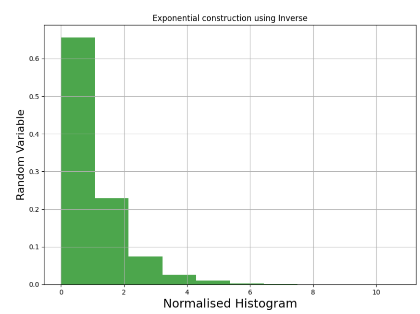
2. Rayleigh Density Function:





3. Exponential Density Function:



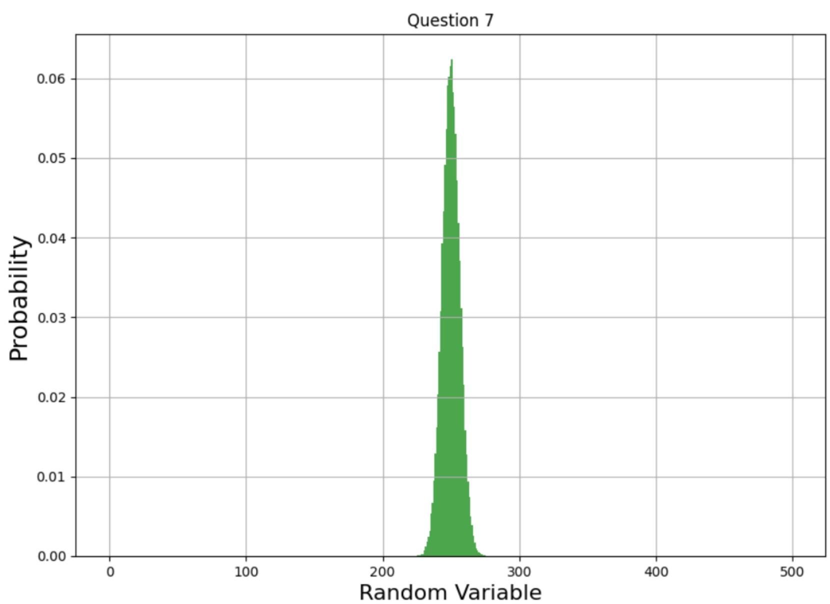


Thus, we can infer that the inverse of the cumulative density function, which is the sum or integral of all values of the distribution up to a point, can be used for approximating the probability density function.

Question 7

Write a function to generate a random number as follows: Every time the function is called, it generates 500 new random numbers from $U[0,1]$ and outputs their sum. Generate 50000 random numbers by repeatedly calling the above function, and plot their normalized histogram (with bin-size = 1). What do you find about the shape of the resulting histogram?

Solution



The shape of the resulting histogram resembles that of a Gaussian or Normal distribution.

This can be considered a direct implication of the Central Limit Theorem, stating that the sum of independent random variables tends towards a normal distribution.

We can moreover represent the density of the sum of n uniform random variables as follows.

$$\frac{1}{(n-1)!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-k)_+^{n-1}$$

Here the expression $(x-k)_+$ equals $(x-k)$ if $(x-k)$ is positive and equals 0 otherwise. Ideally the peak should occur at $\frac{n}{2}$ which is 250 in our case as $\binom{n}{k}$ is maximum at $\frac{n}{2}$.