Assignment - 1

Siddik Ayyappa
2020101089
CSE, UG2k20

SMAI Homework Assignment

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

August 4, 2022

Problem 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.

Solution. Consider the example of the following PMF n: Poisson Random Variable

$$p(n) = \frac{\lambda^n}{n!}e^{-\lambda} \qquad [\lambda > 0]$$

This is an example of a probability mass function with an infinite range. As
n $\to\infty$
p(n) $\to 0$
Hence there are infinitely many values for which p(n) $\neq 0$

$$\begin{aligned}
p(n) &\geq 0 && \text{[all the terms are positive]}\\
& && \text{[condition (1) satisfied]}\\
\sum_{n=0}^{\infty} p(n) &= \sum_{n=0}^{\infty} \frac{\frac{\lambda^n}{n!}}{e^{\lambda}}\\
&= \frac{\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}}{e^{\lambda}} && [\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}]\\
&= \frac{e^{\lambda}}{e^{\lambda}} = 1\\
\therefore \sum_{n=0}^{\infty} p(n) &= 1 && \text{[condition (2) satisfied]}
\end{aligned}$$

Now, consider Another random variable, m, which is the number of students in a lab in IIITH. Clearly the range of the random variable is finite since there is a limit to the upper bound u_b = sup(M) and a lower bound l_b = inf(M). The values of the random variable, beyond the region [l_b, u_b] do not exist.

$$\therefore p(x) = 0 \qquad \forall x \in (-\infty, l_b) \cup (u_b, \infty)$$

$$\begin{aligned}
p(x) &= \frac{\text{Number of Labs with x students}}{\text{Total number of labs}} && \text{[all the terms are positive]}\\
p(x) &\geq 0 && \text{[condition (1) satisfied]}\\
\sum_{i=l_b}^{u_b} p(i) &= \sum_{i=l_b}^{u_b} \frac{\text{Number of Labs with i students}}{\text{Total Number of Labs}}\\
&= \frac{\sum_{i=l_b}^{u_b} \text{Number of Labs with i students}}{\text{Total Number of Labs}}\\
&= \frac{\text{Total Number of Labs}}{\text{Total Number of Labs}}\\
\sum_{i=l_b}^{u_b} p(i) &= 1 && \text{[condition (2) satisfied]}
\end{aligned}$$

Problem 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. We know that the variance of a density is given by $Var(X) = E((x-\mu)^2)$ where μ is the mean of the density.

$$
\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty}((x-\mu^2)).p(x)dx \\
&= \int_{-\infty}^{\infty}(x^2-2\mu x+\mu^2).p(x)dx \\
&= \int_{-\infty}^{a}(x^2-2\mu x+\mu^2).p(x)dx + \int_{a}^{b}(x^2-2\mu x+\mu^2).p(x)dx + \int_{b}^{\infty}(x^2-2\mu x+\mu^2).p(x)dx \\
&= 0 + \int_{a}^{b}(x^2-2\mu x+\mu^2).p(x)dx + 0 && [\because eq(9)] \\
&= \int_{a}^{b}(x^2-2\mu x+\mu^2).p(x)dx \\
&= \int_{a}^{b}(x^2).p(x)dx - 2\mu\int_{a}^{b}x.p(x)dx + \mu^2 && [\because eq(3)] \\
&= \int_{a}^{b}(x^2).p(x)dx - 2\mu * \mu + \mu^2 && [\because eq(9)] \\
&= \int_{a}^{b}x^2.p(x) - (\frac{a+b}{2})^2 \\
&= \int_{a}^{b}x^2.\frac{1}{b-a}.p(x)dx - \frac{a^2+2ab+b^2}{4} \\
&= \left[\frac{x^3}{3} * \frac{1}{b-a}\right]_{b}^{a} - \frac{a^2+2ab+b^2}{4} \\
&= \left[\frac{b^3-a^3}{3} * \frac{1}{b-a}\right] - \frac{a^2+2ab+b^2}{4} \\
&= \left[\frac{(b-a)(b^2+ab+a^2)}{3} * \frac{1}{b-a}\right] - \frac{a^2+2ab+b^2}{4} \\
&= \left[\frac{(b^2+ab+a^2)}{3}\right] - \frac{a^2+2ab+b^2}{4} \\
&= \left[\frac{(4b^2+4ab+4a^2)}{12}\right] - \frac{3a^2+6ab+3b^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} \\
\sigma^2 &= \frac{(b-a)^2}{12}
\end{aligned}
$$

Hence the variance of uniform density is $\sigma^2 = \frac{(b-a)^2}{12}$

Problem 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. Consider the following case. We have 2 different distributions with the same mean and variance.

Let us take a normal distribution with mean and variance 0 and 1 respectively.

$$
\begin{aligned}
p(x) &= N(\mu, \sigma) \\
&= N(0,1) \qquad \text{[This gives us a normal distribution with given parameters]}
\end{aligned}
$$

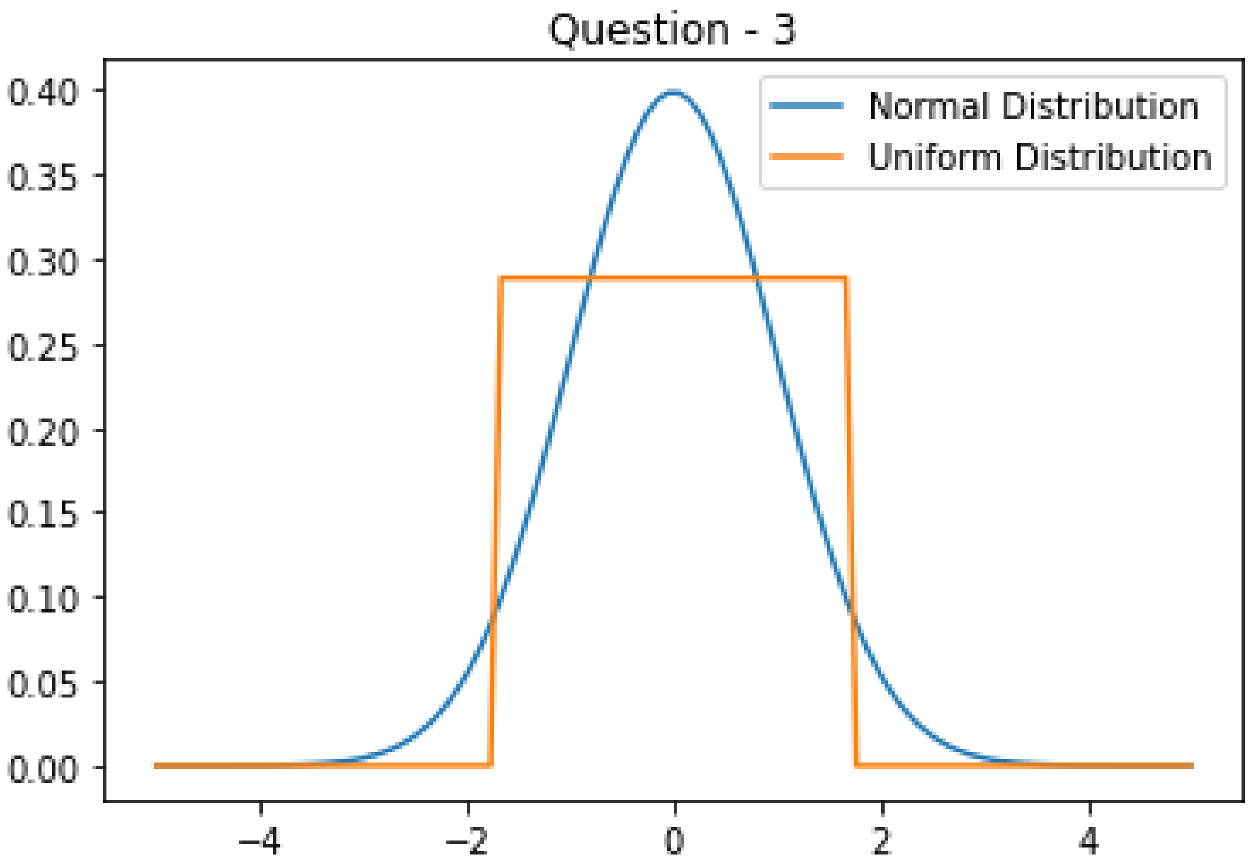
Now, consider the case of a Uniform Distribution U[a, b].

$$
\begin{aligned}
\mu &= \frac{a+b}{2} \implies a+b=2 \\
\sigma^2 &= \frac{(b-a)^2}{12} \implies b-a = \pm 2\sqrt{3}\sigma
\end{aligned}
$$

Solving the system of linear equations for a, b, we get ...

$$
\begin{aligned}
a &= \mu \mp \sqrt{3}\sigma \\
b &= \mu \pm \sqrt{3}\sigma \qquad [\because b > a] \\
a &= \mu - \sqrt{3}\sigma \\
b &= \mu + \sqrt{3}\sigma
\end{aligned}
$$

Hence, with the help of that, we would be able to plot uniform and normal distributions with mean and variance values. Upon plotting the Uniform and Normal distributions with $\mu = 0$ and $\sigma = 1$, we get ...



Problem 4

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

Solution. The equation for the variance of a discrete random variable is given by eq.4, which is

$$\begin{aligned}
Var(x) \equiv \sigma^2 = E[(x-\mu)^2] &= \sum_{i=1}^{n} (v_i - \mu)^2 p(v_i) \\
&= \sum_{i=1}^{n} (v_i^2 - 2\mu v_i + \mu^2).p(v_i) \\
&= \sum_{i=1}^{n} v_i^2 p(v_i) - 2\mu \sum_{i=1}^{n} v_i p(v_i) + \mu^2 \sum_{i=1}^{n} p(v_i) \\
& \qquad \because \sum_{i=1}^{n} v_i p(v_i) = \mu \text{ and } \because \sum_{i=1}^{n} p(v_i) = 1 \\
&= \sum_{i=1}^{n} v_i^2 p(v_i) - 2\mu * \mu + \mu^2 \qquad \because \sum_{i=1}^{n} f(v_i) p(v_i) = E(f(x)) \\
&= E[x^2] - \mu^2 \qquad \because E[x] = \mu \\
\sigma^2 &= E[x^2] - (E[x])^2
\end{aligned}$$

Hence the expression for variance given in equation 5 holds for discrete random variables as well.

Problem 5

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

Solution. We know that the Normal Function is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Where μ and σ^2 are the mean and variance of the distribution respectively.

Proof of Mean.

$$\begin{aligned} Mean(X) \equiv E(X) &= \int_{-\infty}^{\infty} x.p(x).dx \\ &= \int_{-\infty}^{\infty} x.\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}.dx && \text{substituting } \frac{x-\mu}{\sigma} = t \\ & && x = \sigma t + \mu \\ & && dx = \sigma dt \\ &= \int_{-\infty}^{\infty} \sigma t \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{\frac{-t^2}{2}}.\cancel{\sigma} dt + \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{\frac{-t^2}{2}}.\cancel{\sigma} dt \end{aligned}$$

since the 1^{st} term is an odd function with integral limits of -a to a and it equals to 0. The 2^{nd} term is an even function, with similar integral limits, it equals double of the same integral, with limits 0 to a. source link (equations 5.53 and 5.5.4)

$$E(X) = 0 + 2.\int_{0}^{\infty} \mu \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{\frac{-t^2}{2}}.\cancel{\sigma} dt$$

$$\text{substituting } \frac{t^2}{2} = y \implies t.dt = dy \implies dt = \frac{dy}{t} = \frac{dy}{\sqrt{2y}} \text{ and } t = \sqrt{2y}$$

$$E(X) = 2.\int_{0}^{\infty} \mu \frac{1}{\sqrt{2\pi}} e^{-y}.\frac{dy}{\sqrt{2y}}$$

$$E(X) = \frac{\cancel{2}\mu}{\cancel{2}.\sqrt{\pi}} \int_{0}^{\infty} y^{-1/2} e^{-y}.dy \qquad \left[\Gamma(a+1) = \int_{0}^{\infty} t^a e^{-t}\, dt\right]$$

[where Γ is the Gamma function]

[source link]

$$E(X) = \frac{\mu}{\sqrt{\pi}} \Gamma(-1/2+1) = \frac{\mu}{\sqrt{\pi}} \Gamma(1/2) \qquad [\because \Gamma(1/2) = \sqrt{\pi}]$$

[source link]

$$E(X) = \frac{\mu}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}}$$

$Mean(X) = \mu$

Hence Proved.

Proof of Variance. We know that the variance of a distribution is given by $Var(X) = E[X^2] - E[X]^2$, hence we shall begin by calculating E[X^2].

$$E[X^2] = \int_{-\infty}^{\infty} x^2.p(x).dx$$

$$= \int_{-\infty}^{\infty} x^2.\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}.dx \qquad \text{substituting } \frac{x-\mu}{\sigma} = t$$

$$x = \sigma t + \mu$$

$$dx = \sigma dt$$

$$= \int_{-\infty}^{\infty} (\sigma t + \mu)^2.\frac{1}{\sqrt{2\pi}\cancel{\sigma}}.e^{-t^2/2}.\cancel{\sigma}.dt$$

$$= \frac{1}{\sqrt{2\pi}}(\int_{-\infty}^{\infty} \mu^2.e^{-t^2/2}.dt + \int_{-\infty}^{\infty} \sigma^2 t^2.e^{-t^2/2}.dt - \int_{-\infty}^{\infty} 2\sigma\mu t.e^{-t^2/2}.dt)$$

The 3^{rd} term in the equation's RHS, equals to 0, since it is the integral of an odd function, with limits -a to a. The 1^{st} and 2^{nd} terms are even functions with limits -a to a, hence are equal to double of the same integral with limits 0 to a. source link (equations 5.5.3 and 5.5.4)

$$E[X^2] = \frac{2}{\sqrt{2\pi}}(\int_{0}^{\infty} \mu^2.e^{-t^2/2}.dt + \int_{0}^{\infty} \sigma^2 t^2.e^{-t^2/2}.dt)$$

substituting y = $t^2/2 \implies t = \sqrt{2y}$ and $dy = t.dt$

$$dt = \frac{dy}{\sqrt{2y}}$$

$$t = \sqrt{2y}$$

$$= \frac{2}{\sqrt{2\pi}}(\int_{0}^{\infty} \mu^2.e^{-y}.\frac{dy}{\sqrt{2y}} + \int_{0}^{\infty} \sigma^2.2y.e^{-y}.\frac{dy}{\sqrt{2y}})$$

$$= \frac{\cancel{2}}{\cancel{2}\sqrt{\pi}}(\mu^2.\int_{0}^{\infty} e^{-y}.\frac{dy}{\sqrt{y}} + 2\sigma^2.\int_{0}^{\infty} y.e^{-y}.\frac{dy}{\sqrt{y}})$$

$$= \frac{1}{\sqrt{\pi}}(\mu^2.\int_{0}^{\infty} y^{-1/2}.e^{-y}dy + 2\sigma^2.\int_{0}^{\infty} y^{1/2}.e^{-y}.dy) \qquad [\Gamma(a+1) = \int_{0}^{\infty} t^a e^{-t}\,dt]$$

[where Γ is the Gamma function]

$$= \frac{1}{\sqrt{\pi}}(\mu^2.\Gamma(1/2) + 2.\sigma^2.\Gamma(3/2)) \qquad [\because \Gamma(1/2) = \sqrt{\pi}]$$

$$[\because \Gamma(3/2) = \sqrt{\pi}/2]$$

[source link]

$$= \frac{1}{\sqrt{\pi}}(\mu^2.\sqrt{\pi} + \cancel{2}.\sigma^2.\frac{\sqrt{\pi}}{\cancel{2}})$$

$$= \frac{\cancel{\sqrt{\pi}}}{\cancel{\sqrt{\pi}}}(\sigma^2 + \mu^2)$$

$$E[X^2] = \sigma^2 + \mu^2$$

$$\therefore E[X^2] = \sigma^2 + \mu^2$$

Substituting in Var(X) = $E[X^2] - E[X]^2$
We know that $E[X] = \mu$

$$Var(X) = E[X^2] - E[X]^2$$

$$= \sigma^2 + \cancel{\mu^2} - \cancel{(\mu^2)}$$

$$Var(X) = \sigma^2$$

$Var(X) = \sigma^2$
Hence Proved.

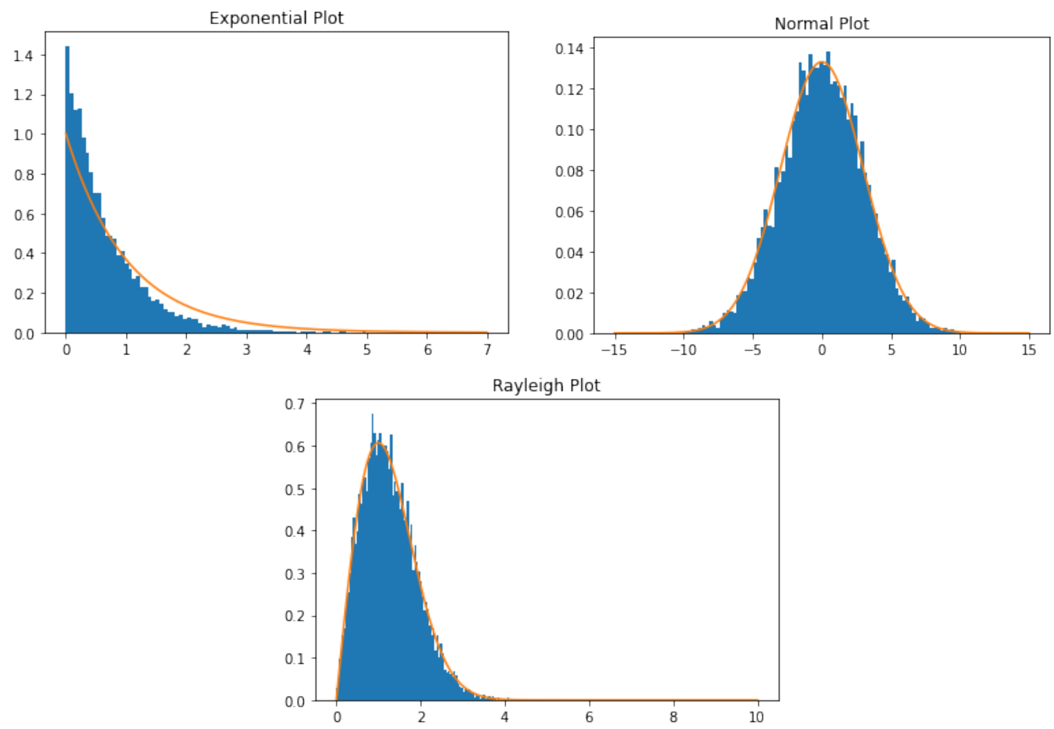
Problem 6

Using the inverse of CDFs, map a set of 10, 000 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.

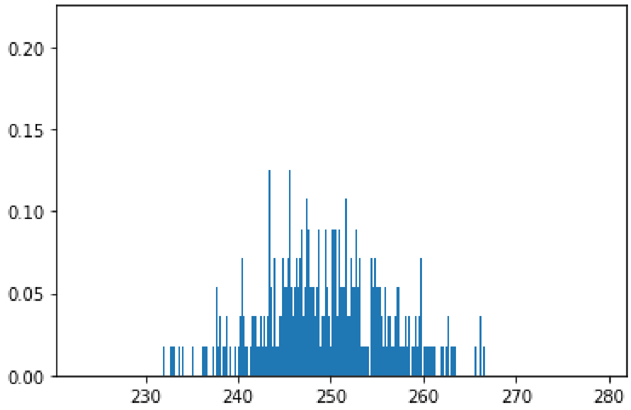


From the above plot we can infer that the plots are resembling, i.e. the blue and orange parts. Judging from the method we have used to generate the histograms, we can say that given a set of random numbers y_i with uniform density U[0, 1], one can map it to set of random variables x_i with any desired PDF using the inverse CDF fucntion !!!.

Problem 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 50,000 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 histogram resembles that of a normal distribution, with a mean of 250, approx.. The following plot shows the plot for truly random numbers, and to achieve the constraint of bin-size = 1, the plot has been made into 50000 bins, resulting in the following plot.



Upon making some tweaks, and generating random integers in range [1, 10] and using 5000 bins, hence reaching constraint (bin-size=1) in the plot, we can get a clearer picture of the plot being a normal distribution. To support the argument, we can see that the mean of the data is always close to 250. [Refer to the point in The python Notebook]