Statistical Methods in AI: Assignment1

Soumya Jahagirdar

September 2021

1 Introduction

This document contains the answers for the questions present in Assignment 1.

2 Questions

2.1 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.

Solution

Probability mass function of a random variable is the measure of probabilities of the possible set of values for a random variable, i.e

$$P_X(x_k) = P(X = x_k), k = 1, 2, 3, ..., \tag{1}$$

An example of **Probability mass function** with **finite range** is an experiment of drawing a spade from a deck of cards. In this case, the task is to pick a card from a deck of cards and find the probability that a randomly picked card is a spade is

$$P(x = spade) = N_{spade}/N_{total} \tag{2}$$

where $P(x = spade)$ is the random experiment to pick a spade, and N_{spade} is the total number of spade cards in the deck and, N_{total} is the total number of cards present in the deck.

Similarly, along with $P(x = spade)$, if we consider $P(x = diamond)$, $P(x = club)$, $P(x = heart)$, then in general we can write the above equation as

$$P(x = v_i) = N_{v_i}/N_{total} \tag{3}$$

where v_i belongs to X and $X = P(x = diamond)$, $P(x = club)$, $P(x = heart)$, $P(x = spade)$.

Here we call **P(.)** as the probability mass function of the experiment where a card is drawn from a deck of cards with finite range, i.e x takes a set of finite countable values.

In the above case probability of each case considered will be 0.25 as there are 13 cards of each type amongst a set of 52 cards. (assuming that there is replacement procedure followed).

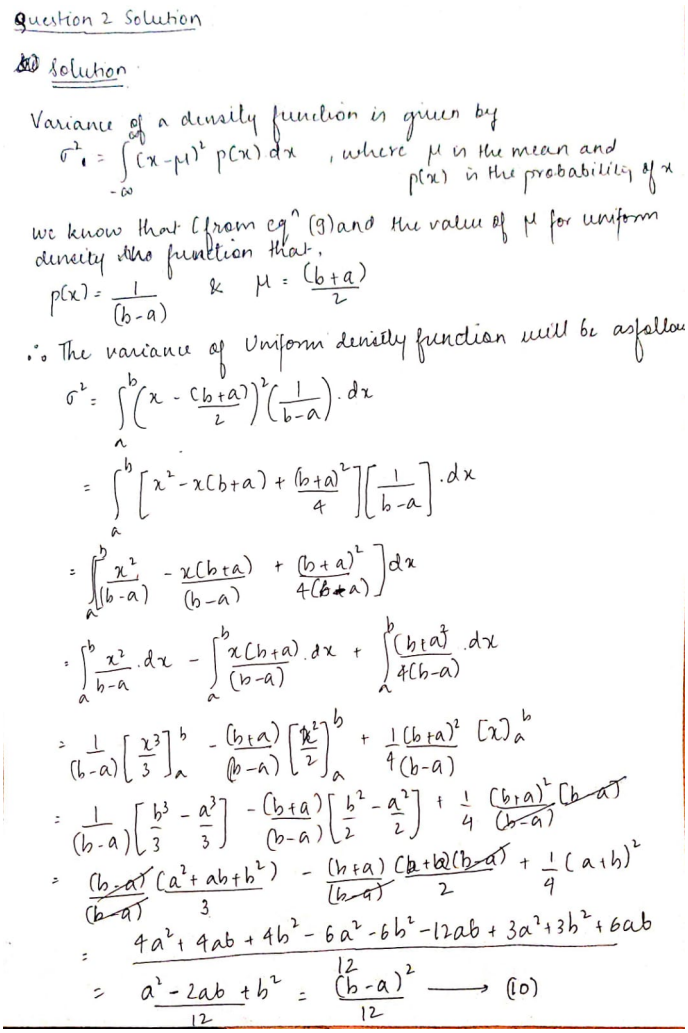
An example of **Probability mass function** with **infinite range** is the experiment of obtaining a number divisible by 3 from a set of positive integers, i.e

Where $X = 3, 6, 9, 12, ...,$

2.2 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



Question 2 Solution

Solution

Variance of a density function is given by

$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\, dx$, where μ is the mean and $p(x)$ is the probability of x

we know that (from eqn (9) and the value of μ for uniform density) the function that,

$p(x) = \frac{1}{(b-a)}$ & $\mu = \frac{(b+a)}{2}$

∴ The variance of Uniform density function will be as follow

$$\sigma^2 = \int_a^b \left(x - \frac{(b+a)}{2}\right)^2 \left(\frac{1}{b-a}\right) dx$$

$$= \int_a^b \left[x^2 - x(b+a) + \frac{(b+a)^2}{4}\right]\left[\frac{1}{b-a}\right] dx$$

$$= \int_a^b \left[\frac{x^2}{(b-a)} - \frac{x(b+a)}{(b-a)} + \frac{(b+a)^2}{4(b-a)}\right] dx$$

$$= \int_a^b \frac{x^2}{b-a}\, dx - \int_a^b \frac{x(b+a)}{(b-a)}\, dx + \int_a^b \frac{(b+a)^2}{4(b-a)}\, dx$$

$$= \frac{1}{(b-a)}\left[\frac{x^3}{3}\right]_a^b - \frac{(b+a)}{(b-a)}\left[\frac{x^2}{2}\right]_a^b + \frac{1}{4}\frac{(b+a)^2}{(b-a)}[x]_a^b$$

$$= \frac{1}{(b-a)}\left[\frac{b^3}{3} - \frac{a^3}{3}\right] - \frac{(b+a)}{(b-a)}\left[\frac{b^2}{2} - \frac{a^2}{2}\right] + \frac{1}{4}\frac{(b+a)^2}{(b-a)}(b-a)$$

$$= \frac{(b-a)}{(b-a)}\frac{(a^2+ab+b^2)}{3} - \frac{(b+a)}{(b-a)}\frac{(b+a)(b-a)}{2} + \frac{1}{4}(a+b)^2$$

$$= \frac{4a^2+4ab+4b^2-6a^2-6b^2-12ab+3a^2+3b^2+6ab}{12}$$

$$= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12} \longrightarrow (10)$$

Figure 1: Solution for second question

2.3 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** The following are the examples of the two density functions

1. A single normal distribution *N(0, 5)*
2. A combination of two normal distributions, where each distribution has standard deviation-1, and respective means are 2 and -2. The resulting combination has a mean of 0 and standard variance of 5.

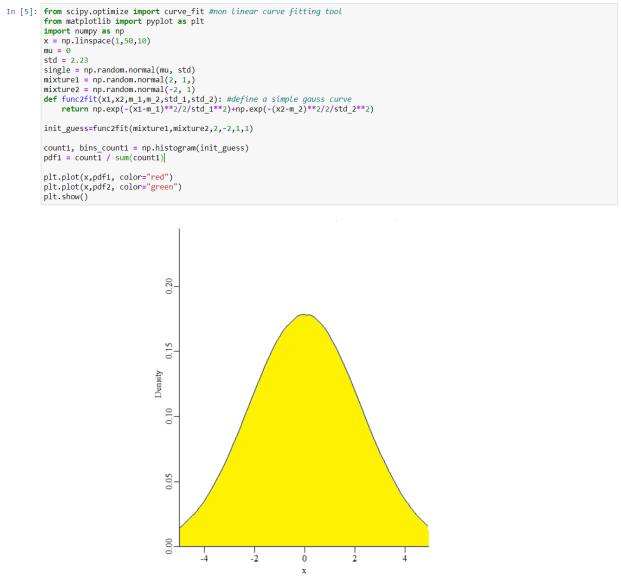


Figure 2: Solution for third question: example of one normal distribution function

These are clearly different distributions but they have same mean and variance.

In [5]:

```
from scipy.optimize import curve_fit #non linear curve fitting tool
from matplotlib import pyplot as plt
import numpy as np
x = np.linspace(1,50,10)
mu = 0
std = 2.23
single = np.random.normal(mu, std)
mixture1 = np.random.normal(2, 1,)
mixture2 = np.random.normal(-2, 1)
def func2fit(x1,x2,m_1,m_2,std_1,std_2): #define a simple gauss curve
    return np.exp(-(x1-m_1)**2/2/std_1**2)+np.exp(-(x2-m_2)**2/2/std_2**2)

init_guess=func2fit(mixture1,mixture2,2,-2,1,1)

count1, bins_count1 = np.histogram(init_guess)
pdf1 = count1 / sum(count1)

plt.plot(x,pdf1, color="red")
plt.plot(x,pdf2, color="green")
plt.show()
```

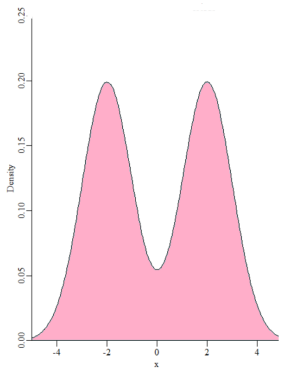


Figure 3: Solution for third question: example of two normal distribution function

2.4 Question 4

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

Solution

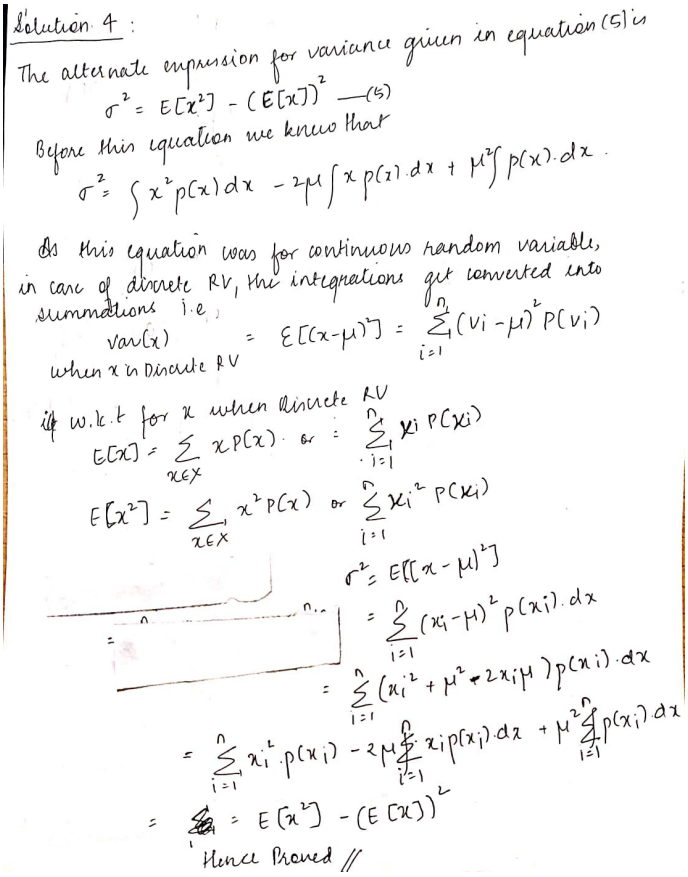


Figure 4: Solution for fourth question

2.5 Question 5

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

Solution

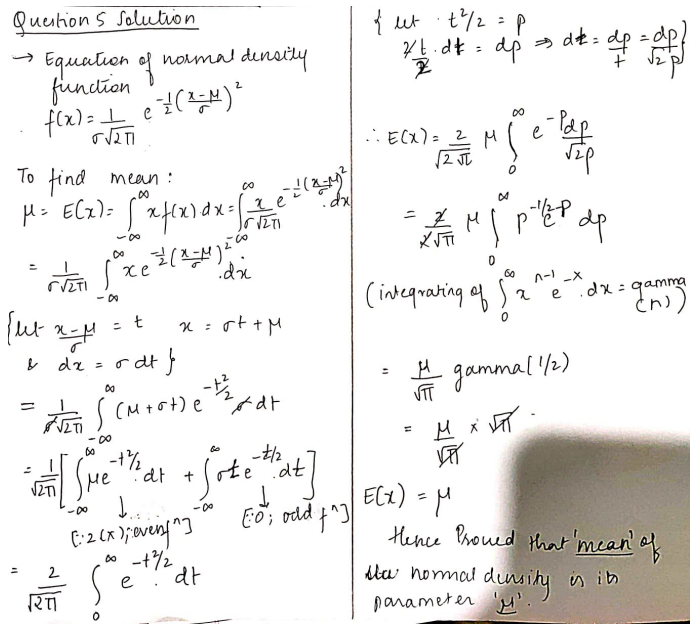


Figure 5: Solution for fifth question "mean"

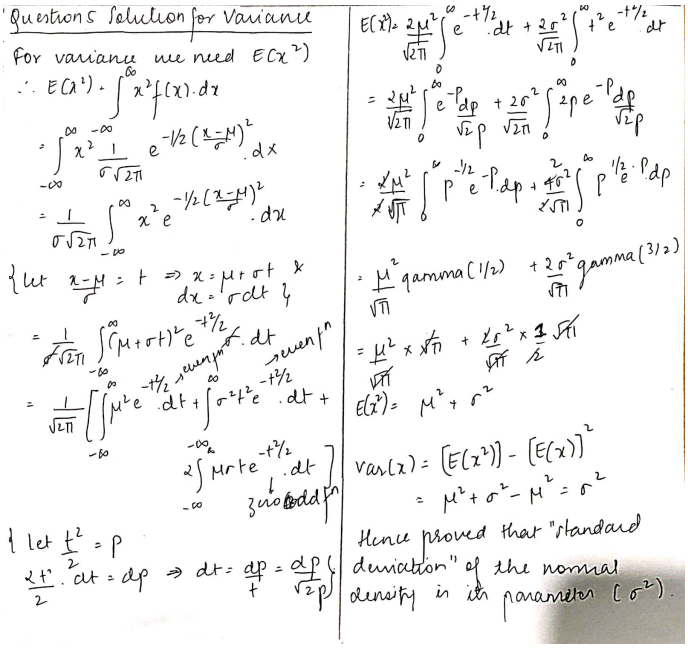


Figure 6: Solution for fifth question "variance"

2.6 Question 6

Using the inverse of CDFs, map a set of 10,000 random numbers from U[0; 1] to follow the following pdfs:

1. Normal density with $\mu = 0$, $\sigma = 3.0$.
2. Rayleigh density with $\sigma = 1.0$.
3. 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? Note: see rand() function in C for U[0, INT_MAX]. **Solution**

```
In [125]: #6th question

          #Plotting uniform distribution
          s = np.random.uniform(0,1,10000)
          plt.hist(s, bins=100, alpha=0.5, label="Generated r.v.")
          plt.title("Uniform distribution")
          plt.show()
```

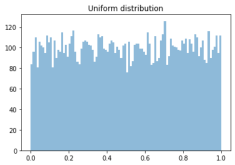


Figure 7: Solution for sixth question uniform distribution

```
In [128]: #For Normal density with mu = 0, sigma = 3.0
          from scipy.stats import norm
          from scipy.special import ndtri


          mu, sigma = 0, 3.0
          s1 = np.random.normal(mu, sigma, 10000)

          count, bins_count = np.histogram(s1, bins=10000)
          pdf = count / sum(count)
          cdf = np.cumsum(pdf)
          norm = ndtri(cdf)*s
          plt.hist(norm, bins=50, alpha=0.5, label="Generated r.v.")
          plt.title("normal")

          plt.show()
```

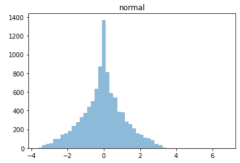


Figure 8: Solution for sixth question normal distribution

```
In [91]: #For Exponential density with lambda = 1.5

         from scipy.stats import expon

         lambda1=1.5
         X=-lambda1*np.log(1-s)

         plt.hist(X, bins=50, alpha=0.5, label="Generated r.v.")
         plt.title("Exponential density with lambda = 1.5")
         plt.legend()
         plt.show()
```

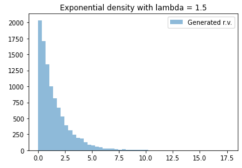


Figure 9: Solution for sixth question exponential distribution

2.7 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 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

```
In [9]: import numpy as np
        import matplotlib.pyplot as plt

        lis= []

        def gen_random_number():
            s = np.random.uniform(0,1,500)
        #     print(np.sum(s))

            lis.append(np.sum(s))
            return np.sum(s)

        for i in range(100):
            (gen_random_number())

        # count, bins = plt.hist(np.array(lis)//50000, 1)
        plt.hist(np.array(lis),density=True)
        plt.title("normalized histogram of sums of random numbers")
        plt.show()
```

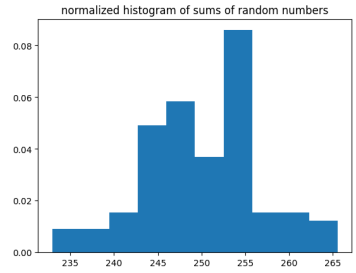


Figure 10: Solution for seventh question

After adding the numbers sampled from randomly distribution, the resultant of the sum of the numbers obtained is not uniformly distribution. It is seen that the numbers are normally distributed.