1) Give Example of PMF with finite and infinite Range.

a) Finite Range :-

<u>Toss of coin</u> :- Consider a fair coin, with random variable X defined over its state space S (heads / Tails). The category tails is given 0, and heads is 1.

Since the coin is fair the PMF is

$$P(x) = \begin{cases} 1/2 & x = \text{heads} \\ 1/2 & x = \text{tails} \\ 0 & x \notin \{\text{heads, tails}\}. \end{cases}$$

We can observe $P(x = \text{heads}) \geq 0$
and $P(x = \text{tails}) \geq 0$

[Condition ① of PMF $P(x = x_i) \geq 0$ $\forall x_i \in X$ is obeyed].

$$P(x = \text{heads}) + P(x = \text{tails}) = 1/2 + 1/2 = 1$$

[Condition ② of PMF $\sum_{i=1}^{N} P(x_i) = 1$ is obeyed].

b) Infinite Range :-

Consider an infinite set of Natural numbers

$N = \{1, 2, 3 \ldots\ldots\}$

and let the probability that every number occurs be defined as

$$P(X = n_i) = \frac{1}{2^{n_i}} \quad \forall n_i \in N$$

we observe that $\frac{1}{2^{n_i}} \geq 0 \quad \forall n_i \in N$

[condition ① of PMF is obeyed
$p(n_i) \geq 0 \quad \forall n_i \in X$]

and

$$\frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots\ldots$$

This is an infinite Geometric progression

$$S_\infty = \frac{a}{1-r} = \frac{1/2}{1-1/2} = \underline{\underline{1}}$$

$$\Rightarrow \sum_{i=1}^{\infty} 1/2^i = 1 \qquad \left[\sum_{i=1}^{N(=\infty)} P(x_i) = 1 \right]$$

[Condition ② Sum of all probabilities is = 1
the second condition is obeyed].

2) Variance of Uniform density function

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

Variance is Given by

$$\sigma^2 = E[x^2] - [E(x)]^2 \rightarrow ①$$

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

$$= \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{b-a} \frac{1}{3} [x^3]_a^b$$

$$= \frac{1}{b-a} \frac{1}{3} (b^3 - a^3)$$

$$= \frac{1}{3} \frac{(b^3 - a^3)}{(b-a)} \rightarrow ②$$

$$E(x) = \int_{-\infty}^{\infty} x\, p(x)\, dx = \int_a^b x \frac{1}{b-a} \cdot dx$$

$$= \frac{1}{b-a} \cdot \frac{1}{2} [x^2]_a^b$$

$$= \frac{1}{2} \frac{(b^2 - a^2)}{b-a}$$

$$E(x) = \frac{b+a}{2} \leftarrow \frac{(b+a)}{2} \frac{\cancel{(b-a)}}{\cancel{b-a}}$$

$\hookrightarrow ③$

Substitute results of (2) & (3) in (1)

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

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

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

$\sigma^2 = \frac{(b-a)^2}{12}$ Hence proved that

Variance of uniform density function is

$$\sigma^2 = \frac{(b-a)^2}{12}$$

4) Consider a discrete random variable ~~to~~ defined over a ~~n~~n output state space $\{v_1, v_2, \ldots, v_n\}$.

The Expected value is

$$E[v_i] = \mu = \sum_{i=1}^{n} v_i P(v_i)$$

The variance is given by

$$\sigma^2 = E\left[(v_i - \mu)^2\right]$$

$$= \sum_{i=1}^{N} (v_i - \mu)^2 P(v_i)$$

$$= \sum_{i=1}^{N} (v_i^2 + \mu^2 - 2 v_i \mu) \, P(v_i)$$

$$= \sum_{i=1}^{N} v_i^2 P(v_i) + \mu^2 \sum_{i=1}^{N} P(v_i) - 2\mu \sum_{i=1}^{N} v P(v_i)$$

$$= E[v_i^2] + \mu^2 - 2\mu \underbrace{\sum_{i=1}^{N} v P(v_i)}_{(\mu)}$$

$$= E[v_i^2] + \mu^2 - 2\mu^2$$

$$= E[v_i^2] - \mu^2 = \cancel{E[v_i^2] - (E[v_i])^2}$$

$$= \underline{E[v_i^2] - \left(E[v_i]\right)^2}$$

5) Consider ~~the~~ the Normal distribution

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

Mean = $\cancel{E[x] = \int_{-\infty}^{\infty} N(\mu, \sigma^2)}$

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

Substitute $x = x + \mu$

$dx = dx$

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

$$= \underbrace{\int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx}_{I_1} + \underbrace{\int_{-\infty}^{\infty} \frac{\mu}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx}_{I_2}$$

$$I_1 = \underbrace{\int_{-\infty}^{0} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx}_{I_1'} + \int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

Date
Page

$$I_1' = \int_{-\infty}^{0} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

invert the limits

$$= -\int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

Substitute $x = -x$, $dx = -dx$

$\therefore$ limits $= [0, \infty)$

$$-\int_{0}^{\infty} \frac{(-x)}{\sigma\sqrt{2\pi}} e^{-(-x)^2/2\sigma^2} (-dx)$$

$$= -\int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

$$\Rightarrow I_1 = I_1' + \int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

$$= -\int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} + \int_{0}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

$\underline{\underline{I_1 = 0}}$

$$I_2 = \int_{-\infty}^{\infty} \frac{M}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

Substitute $\quad x = \sigma\sqrt{2}\, x$

$\qquad dx = \sigma\sqrt{2}\, dx$

$$= \int_{-\infty}^{\infty} \frac{M}{\sigma\sqrt{2\pi}} \cdot \sigma\sqrt{2}\; e^{-\left(\frac{\cancel{\sigma^2}\cancel{2}\, x^2}{\cancel{\sigma^2}\cancel{2}}\right)} dx$$

$$= \int_{-\infty}^{\infty} \frac{M}{\sqrt{\pi}} e^{-x^2} dx$$

$$= 2\int_{0}^{\infty} \frac{M}{\sqrt{\pi}} e^{-x^2} dx$$

$$M \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} dx = M \lim_{t\to\infty} \frac{2}{\sqrt{\pi}} \int_0^{t} e^{-x^2} dx$$

$$\boxed{= M \lim_{t\to\infty} \operatorname{erf}(t) = \underline{\underline{M}}}$$

(1)

[Result]

$\therefore$ the mean of normal distribution is the parameter μ

$$E[X] = \mu$$

→ Variance of distribution :-

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

Substitute $x = x + \mu \quad dx = dx$

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

Substitute $x = \sigma\sqrt{2}\,x$

$dx = \sigma\sqrt{2}\,dx$

$$\int_{-\infty}^{\infty} \sigma^2 2 x^2 \frac{1}{\sigma\sqrt{2\pi}} \sigma\sqrt{2}\, e^{-\left(\frac{(\sigma\sqrt{2}x)^2}{2\sigma^2}\right)} dx$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2} dx = \frac{4\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} x^2 e^{-x^2} dx$$

Substitute $t = x^2 = x = \sqrt{t}$

$$dt = 2x\,dx = 2\sqrt{t}\,dx$$

$$dx = \underline{(2\sqrt{t})^{-1}\,dx}$$

$$V(x) = \sigma^2 \frac{4}{\sqrt{\pi}} \int_0^\infty (\sqrt{t})^2 (2\sqrt{t})^{-1} e^{-t}\,dt$$

$$\cancel{= \sigma^2 \frac{4}{\sqrt{\pi}} \frac{1}{2} \int_0^\infty t^{3/2-1}}$$

$$= \sigma^2 \frac{4}{\sqrt{\pi}} \frac{1}{2} \underbrace{\int_0^\infty t^{\frac{3}{2}-1} e^{-t}\,dt}_{} \rightarrow (a)$$

$$\left[\Gamma(3/2) = \int_0^\infty e^{-t}\, t^{\frac{3}{2}-1}\,dt\right]$$

We know that $\Gamma(3/2) = \frac{\sqrt{\pi}}{2}$

Substitute result in equation (a).

$$= \sigma^2 \frac{4}{\cancel{\sqrt{\pi}}} \cdot \frac{1}{2} \cdot \frac{\cancel{\sqrt{\pi}}}{\cancel{2}} = \boxed{\underline{\sigma^2} \rightarrow (2)}$$

[Result]

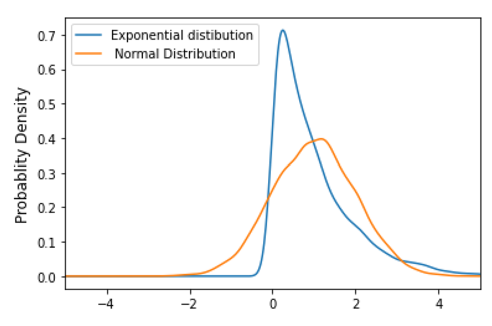
$\therefore$ It is proved that ~~the~~ from Result (1) and Result (2), the mean and variance of normal distribution are its parameters ~~~~ μ and σ^2.

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

We shall consider the example of a Normal Distribution and an exponential distribution each of them having a mean and variance of 1.0

In [56]:
```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Draw 10000 samples from an exponential distribution with
# scale parameter =1
exponential_dist1 = np.random.exponential(1, 10000)

# Draw 10000 samples from a normal distribution with
#mean and variance equal to 1
normal_dist = np.random.normal(1, 1, 10000)

# We use the inbuilt KDE (kernel density estimate)
#to plot the density functions

s_normal = pd.Series(normal_dist)
s_exp = pd.Series(exponential_dist1)

print("----------------")


variance_exp = np.var( exponential_dist1)
mean_exp = np.mean(exponential_dist1)


print( "variance of exponential distribution = " + str( variance_exp))
print( "mean of exponential distribution = " +  str(mean_exp))


print("----------------")

variance_normal = np.var( normal_dist)
mean_normal = np.mean(normal_dist)

print( "variance of normal distribution = " +  str( variance_normal))
print(  "mean of normal distribution = " +  str(mean_normal))


print("----------------")


ax1 = s_exp.plot.kde(label ='Exponential distibution' )
ax2 = s_normal.plot.kde(label =' Normal Distribution')

plt.xlim([-5, 5.0])
plt.ylabel("Probablity Density" , fontsize=12)
plt.legend()

plt.show()
```

```
----------------
variance of exponential distribution = 1.0387532539610607
mean of exponential distribution = 1.0214694002873796
----------------
variance of normal distribution = 0.9946853632262337
mean of normal distribution = 1.0045287923846868
----------------
```



It can be observed from the above graphs that both the density functions have the same mean and variance, but are completely different distributions.

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

The Following procedure is followed to map a set of random numbers from U[0,1] to any distribution.

Step1 Determine the CDF of the given distribution \ **Step2** Take the inverse of the given CDF \ **Step3** Substitute values from U[0,1] into the iverse CDF function to obtain the PDF of the corresponding function.

a) Normal density with μ = 0, σ = 3.0.

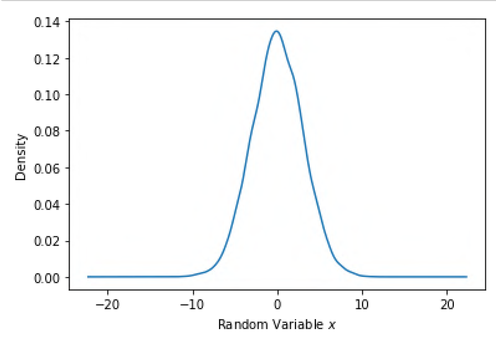
The CDF of a normal density function with mean μ and variance σ^2 is given by

$$cdf(x) = (1/2)(1 + erf((x_{pdf} - \mu)/\sigma(2)^{1/2}))$$

This implies

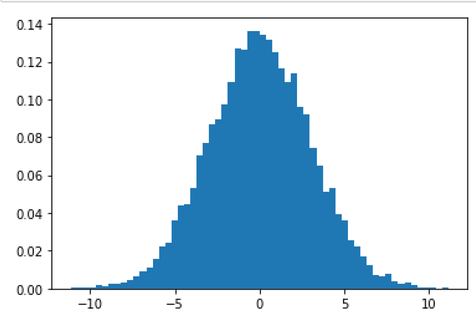
$$x_{pdf} = \sigma(2)^{1/2} erf^{-1}(2cdf(x) - 1) + \mu$$

In [ ]:
```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from scipy import special
import math

std =3
temp_val = std*math.sqrt(2)
U = np.arange(  0.00000001, 1 , 1/30000 )
U = np.random.choice(U, replace=False,  size=10000 ) ## U is the cdf distribution


temp1 = 2*U - 1
inv_err = special.erfinv(temp1)
normal_dist = temp_val*inv_err

s_normal = pd.Series(normal_dist)


ax1 = s_normal.plot.kde(label ='Normal distibution' )
plt.xlabel("Random Variable $x$")

plt.show()

mean= np.mean( normal_dist)
variance = np.var(normal_dist)

print(" Mean of Distribution = " +str(mean) )
print(" standard deviation of Distribution  = " + str( np.sqrt( variance )))
```




```
 Mean of Distribution = 0.021155687173977412
 standard deviation of Distribution  = 2.9992663557287482
```

In [ ]:
```python
## Plot the Normalized Histogram
plt.hist(normal_dist,
         bins= 'auto',   density='True')  ## Plot the normaized histogram
plt.show()
```



b) Rayleigh density with $\sigma = 1.0$

Given the CDF of Rayleigh desnity is F and σ is the scale parameter

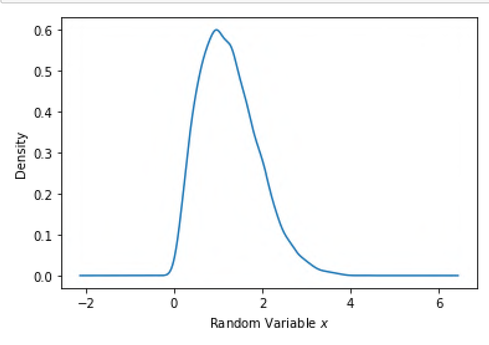
$F = 1 - e^{(-x_{pdf}^2/2\sigma^2)}$ \ for $x \in [0, \infty)$

This implies

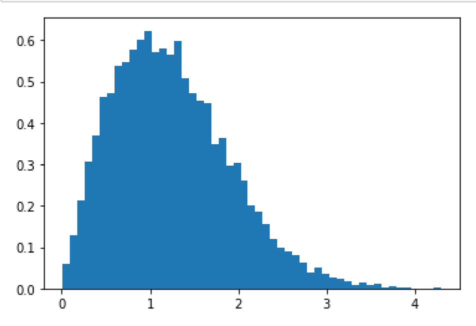
$x_{pdf}^2 = -2\sigma^2 \ln(1 - F)$

$x_{pdf} = (-2\sigma^2 \ln(1 - F))^{1/2}$

In [ ]:
```python
U = np.arange(  0, 1 , 1/30000 )
U = np.random.choice(U, replace=False,  size=10000 ) ## U is the cdf distribution
sigma = 1.0
temp_rayleigh =  np.sqrt( -2*(sigma**2)*(np.log( 1- U) ))
s_rayleigh = pd.Series(temp_rayleigh)

ax1 = s_rayleigh.plot.kde(label ='Rayleigh distibution' )
plt.xlabel("Random Variable $x$")
plt.show()
```



In [ ]:
```python
## Plot the Normalized Histogram

plt.hist(temp_rayleigh,
         bins= 'auto',   density='True')  ## Plot the normaized histogram
plt.show()
```



c) Exponential density with $\lambda = 1.5$

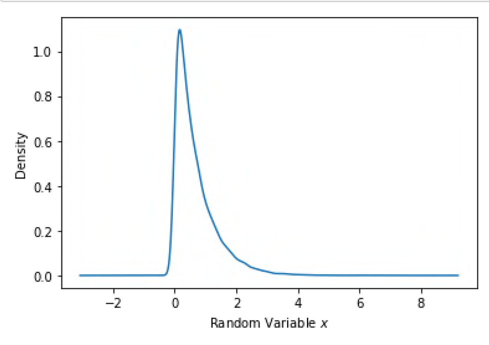
Given the CDF F with λ as the rate parameter

$F = 1 - e^{(-\lambda x_{pdf})}$

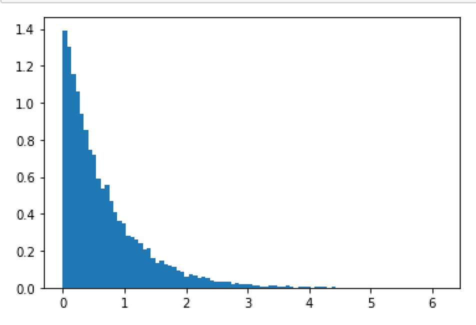
This implies

$x_{pdf} = -(1/\lambda)\ln(1 - F)$

In [ ]:
```python
U = np.arange(  0, 1 , 1/30000 )
U = np.random.choice(U, replace=False,  size=10000 ) ## U is the cdf distribution
lambda_scale = 1.5
temp_exp = -(1/lambda_scale)*(np.log( 1- U) )
s_exp = pd.Series(temp_exp)

ax1 = s_exp.plot.kde(label ='Exponential distibution' )
plt.xlabel("Random Variable $x$")
plt.show()
```



In [ ]:
```python
## Plot the Normalized Histogram

plt.hist(temp_exp,
         bins= 'auto',   density='True')  ## Plot the normaized histogram
plt.show()
```



Inference: We can observe that any desired distribution can be generated given a random variable that follows the uniform distribution, the PDF of the corresponding distribution can be determined by the inverse of the CDF function.

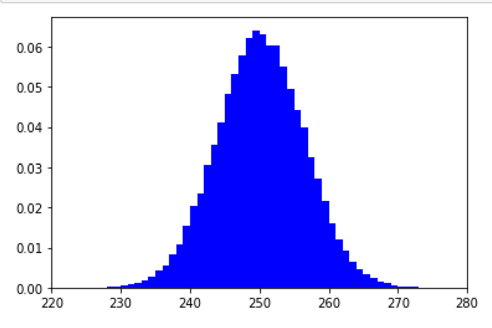
7) Write a function to generate a random number as follows: Every time the function is called, it generates 500 new random numbers from U [0, 1] and outputs their sum.

In [ ]:
```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from scipy import special
import math
import random


U = np.arange(  0.001, 1 , 1/10000 )

def generate_500_random_pts( dist ):

        # take random samples in the distribution array
        sampled_values = np.random.choice(dist, replace=False,  size=500 )
  # sum all elements in the distrin=bution
        sum_val = np.sum(sampled_values)
        return sampled_values , sum_val


for i in range(5):
        ## Call the function multiple times to demonstrate that sum of random numbers are generated
  dist, sum_res = generate_500_random_pts(U)

  print(sum_res)
```

```
254.55580000000012
247.28780000000012
243.9150000000001
255.69610000000011
262.6934000000001
```

Generate 50, 000 random numbers by repeatedly calling the above function, and plot their normalized histogram (with bin-size = 1).

In [ ]:
```python
union_sample_list = np.array([0])

for i in range(50000):

        temp , sum_val= generate_500_random_pts(U)

        union_sample_list  = np.append( union_sample_list , [sum_val] , axis =0)

plt.hist(union_sample_list[1:],
         bins=[k for k in range(round (min(union_sample_list)-1),
                                round(max(union_sample_list)+1))],
                                    color ="blue" , density='True')  ## Plot the normaized histogram
plt.xlim([220 , 280])
plt.show()
```



Inference: We observe that the result is a gaussian distribution