Assignment1

TASK1

the probability that the urn Fred is using is urn u :

$$
\begin{aligned}
p(u|D) &= \frac{p(D,u)}{p(D)} \\
&= \frac{p(D,u)}{\sum_u p(D,u)} \\
&= \frac{p(u)p(D|u)}{\sum_u p(D,u)}
\end{aligned} \tag{1}
$$

supposing he randomly select a urn, then $p(u) = \frac{1}{11}$

and

$$
p(D|u) = C_{10}^{3}(\frac{u}{10})^3(1-\frac{u}{10})^7 \tag{2}
$$

so we could compute the $p(u|D)$ for each urn:

the label of urn	$p(u\|D)$
0	0.0
1	0.06307262464852743
2	0.22123978796753002
3	0.29321985989557925
4	0.2362555743579042
5	0.12877850558581408
6	0.04666776777440084
7	0.009892048584565579
8	0.0008642179217481626
9	9.613263930578772e-06
10	0.0
SUM	**1.0000000000**

so we can see that when $u = 3$, the post probability is the maximum.

TASK2

part A

because for every $t \in [0,1]$

$$F(G(t)) = t \tag{3}$$

we could get:

$$F(G(x)) = x \tag{4}$$

because $G(X) = Y$, for each $x \in [0,1]$, there has a $y = G(x)$ corresponding to x

from equation 2, we have :

$$F(y) = x, for\ any\ x \in [0,1] \tag{5}$$

because x is a real-value random varibles, and $x \in [0,1]$, so our cumulative distribution function $F(y)$ is meaningful, so that we could write it as the form:

$$F(y) = \int_{-\infty}^{y} p(v)dv \tag{6}$$

Hint: we have been told that $F(u) = \int_{-\infty}^{u} p(v)dv$

then we could compute the PDF $p_Y(y)$ by using the derivate formula:

$$p_Y(y) = p(y) \tag{7}$$

part B

we could use the previous conclusions, and use Y to represent the random variable, we want to make:

$$p_Y(y) = p(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

the cumulative distribution is:

$$F_Y(y) = \int_{-\infty}^{y} p(v)dv \tag{9}$$

the result is:

$$F_Y(y) = \begin{cases} 1 & y > 1 \\ y^2 & 0 \le y \le 1 \\ 0 & y < 0 \end{cases} \tag{10}$$

we only focus on the $y \in [0,1]$,

$$F_Y(y) = y^2 \tag{11}$$

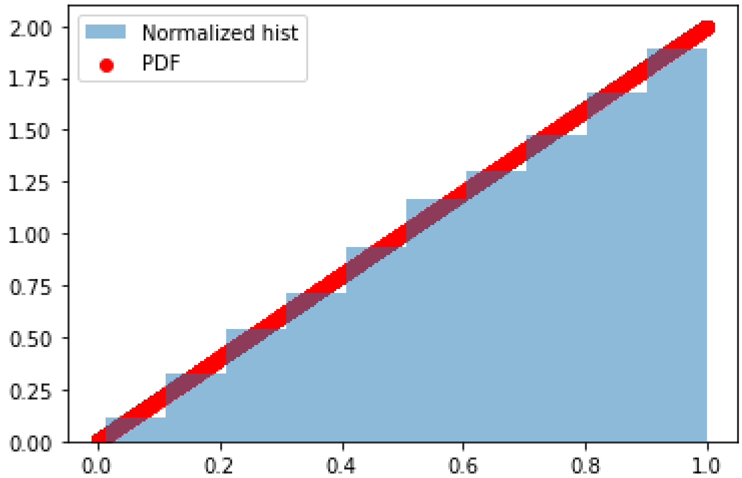
so that the $G(y) = \sqrt{y}$, and

$$y = G(x) = \sqrt{x} \tag{12}$$

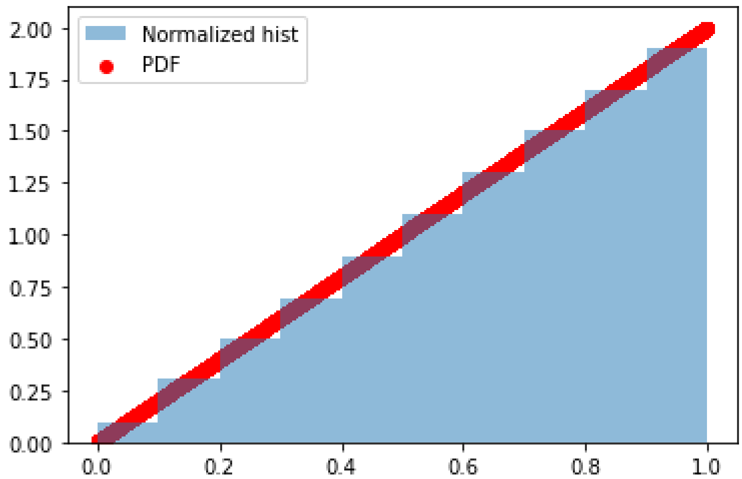
now we generate points and draw image:

```python
import numpy as np
import matplotlib.pyplot as plt
from numpy import random
fig,ax = plt.subplots()
x = random.uniform(size=(10000,))
y = np.sqrt(x)
y2 = 2*x
ax.hist(y,density=True, label="Normalized hist", alpha = 0.5)
ax.scatter(x,y2,color = "red", label = "PDF")
ax.legend()
plt.show()
```

$N = 10^4$



$N = 10^6$



we can clearly see that our generated points' histogram is very similar to the true PDF, this shows that **our method to "generate random variable with given PDF" is useful.**

as for the normalized procedure, it is to make the measure of "area of histogram" equals to 1.