Better understanding of Random Variables

Objective of this notebook

I was going over the required reading for "Statistical Concepts Useful in Describing Risky Events and Risky Outcomes" and I found the explanation to be a little definition based and not intuition based. And so I thought of developing a notebook where I show how these concepts can be practically used to solve a certain problem at hand.

I came up with this problem to help me understand the concept and I do hope that this will also help you all as well :)

A little background

Assume you have a fair coin at hand. If you toss the coin, it has a chance of either going up heads (H) or tails (T). Let us set the stage that you are making a bet with your friend that it will be heads on the next toss. The probability of the experiment of being heads (H) is `0.5`. By definition, probability is just the ratio of occurence for a favorable event occuring against the entire possible events that can occur. So, out of 2 possible occurances `{ H , T }`, we are interested in how many of these occurances comes out as heads `H`. The result becomes:

$$P(H) = \frac{number\ of\ favorable\ occurences}{total\ possible\ occurences} = \frac{1}{2} = 0.5$$

The Problem

So far, so good. And for simple cases like this, the need for defining random variables, pdf's and other variables might not be as interesting and not even neccesary and you would either win/lose with a fifty-fifty chance. But let us extend the problem and see why wee need to define and understand these terms mentioned in the required reading. Specifically, we will intuitively and interactively see what random variables are and what pdf's might look like graphically

Defining Random Variables

In the background, we bet that it will be heads, (hereafter as `H`) in the next toss. Now, if we extend the game for two coins. Here, `H` can occur once, twice, or will not appear at all (when both the coins turns out to be tails). and now the game becomes 'Guess how many heads will appear in two coin tosses'

Here are the outcomes for the game:

$$possible\ outcomes = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$$

So, if we are **interested** in "number of heads occuring in the game", we can see that it can have the values $\{0, 1, 2\}$. It would appear that the number of heads occuring in this game is a variable and we will represent it as X from now on.

Normally, if we imagine a typical variable called y for example, we know from linear algebra that it can have any value within a specified domain. But that is preety much it. So our variable X does satisfy this definition but this variable has another property to it. For every value, there is a corresponding probability assigned to it, which makes it a little different from an ordinary variable. For example, the probability that no heads occuring is

$p(0) = \frac{1}{4} = 0.25$; the probability that one head will occur is $p(1) = \frac{2}{4} = 0.5$ and the probability that two heads will occur is $p(2) = \frac{1}{4} = 0.25$.

Secondly, the variable is generated by first definining what we are interested in (in our case, it is the number of heads occuring by tossing two coins), run the experiment and log all the outcomes and finally, tally all distinct values of the variable.

so, in simple terms, the variable is definied by the outcomes of a random experiment being run, hence "random variable". So, when you think of random variable, associate all these values are possible outcomes of our interest parameter and has a corresponding probability value to it.

As a result, by looking at the random variable, if we want to have better odds to win the game, it would be in our best interest to guess `1` of the coins will turn out heads.

Visualizing random variables

Hopefully, we have some understanding of random variables so far. Now let us extend the game to more coins. Meaning,"Guess how many heads will turn out for three, four, ten or a hundred coins?" And our problem is to pick an optimal number to have better odds to win this game!

Let us first generalize by saying that the number of coins in the game will be y.

We still are interested in the number of heads occuring in this game. This is a random variable and we will still denote it as X. As it is a random variable, to completely understand this variable, we need to know:

1. what the possible values are
2. What the corresponding probabilities for these values are

For two coins: we already have solved it: $X = \{0, 1, 2\}$ and $p(X) = \{0.25, 0.5, 0.25\}$ and it would be best to answer "one coin will turn out heads".

Now, For three or more coins, we can generalize by using principles of counting (such as combinations and permutations). I will not go in detail as to how this results came about, as they are not the point of this topic.(If you want me to go deeper in this topics, you can reply to the forum so that I could explain them in detail)

So to answer for (1) what the possible values are: it can be any value from 0 to y.

To answer for (2) what the corresponding probabilities for these values are: We need two values; a) the total possible occurences in the game and b) the number of occurences X heads will turn out

For a) we have 2^y. For example, if the coins are two, we can see that the total possible outcomes are $possible\ outcomes = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$ which means there are four possible outcomes or $2^2 = 4$

For b) we have $\frac{y!}{(y-X)X!}$. For example, the number of occurences 1 coin will turn out out of two coins is $\frac{2!}{(2-1)1!} = 2$, which makes sense as $\{\{H, T\}, \{T, H\}\}$ are the only possible results. By putting it all together, we have:

$$(1)\ X = \{x \mid 0 \leq x \leq y\}$$

$$(2)\ p(X) = \frac{y!}{(y-X)X!2^y}$$

(2) is simply the probability density function. It is a function which outputs the probability of a random variable.

So, to visualize this, let us get coding!

In [38]:

```
# Import important libraries
import math as m
import matplotlib.pyplot as plt
```

In [39]:

```
# Define inputs to set the game, please feel free to change this variable

#number of coins - defined as y
y = 20
```

In [40]:

```
# Random variable X - number of heads to turn out
X = [x for x in range(y+1)]
if len(X)<=100:
    print("possible random variables:", X)
else:
    print("Kindly input smaller values to allow printing of the random variables")
```

```
possible random variables: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20]
```

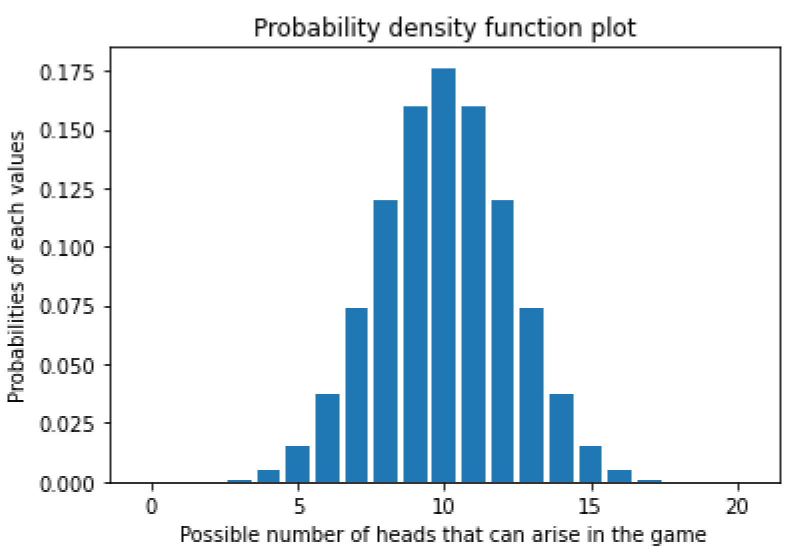
In [48]:

```
# Probability density function p(x) labeled as p_X
p_X = [ (m.factorial(y))/(math.factorial(y-x)*m.factorial(x)*m.pow(2,y)) for x in X ]
if len(X)<=100:
    print("probaility distributions of the variables:",[round(x,3) for x in p_X])
else:
    print("Kindly input smaller values to allow printing of the probability distributions'
```

```
probaility distributions of the variables: [0.0, 0.0, 0.0, 0.001, 0.005, 0.015, 0.037, 0.0
74, 0.12, 0.16, 0.176, 0.16, 0.12, 0.074, 0.037, 0.015, 0.005, 0.001, 0.0, 0.0, 0.0]
```

In [42]:

```
# Plot the distribution function
plt.bar(X,p_X);
plt.xlabel("Possible number of heads that can arise in the game")
plt.ylabel("Probabilities of each values")
plt.title("Probability density function plot");
```



In [47]:

```
# Expected value labelled as Ev - Use to estimate the best answer to have better chances t
```

```
Ev = sum([px * x for px,x in zip(p_X,X)])
Ev
```

Out[47]:

```
10.0
```

Conclusion

So, we can see how defining random variables and plotting corresponding density functions might help us to solve a specific problem. If we want to win the game "Guess how many heads will occur in y coin tosses?". It would be in our best interest to choose the value where the probability is maximum. Better yet, it would be good to calculate the expected value by multiplying each random variable with its corresponding probability and summing the result. This can be extended if the coins are not fair for example to introduce skewness or the outcomes becoming more than two, for example a dice toss instead of a coin toss. Even though the problems will be much more complex, the concepts are basically similar.

Going through this notebook, I hope you find the examples interesting and understand how to use these concepts to solve problems related to analyzing different outcomes.

In [ ]:

```

```