EE2211 Tutorial 3

(Probability Mass Function)

Question 1:

The random variable N has probability mass function (PMF)

$$P_N(n) = \begin{cases} c\left({}^{1}/_{2}\right)^n, & n = 0,1,2 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $Pr[N \leq 1]$?

Answer:

(a) We wish to find the value of c that makes the PMF sum up to one.

$$\sum_{n=0}^{2} P_N(n) = c + \frac{c}{2} + \frac{c}{4} = 1 \quad \text{implying } c = \frac{4}{7}$$

(b) The probability that $N \leq 1$ is

$Pr\,[N \leq 1] = Pr\,[N = 0] + Pr\,[N = 1] = 4/7 + 2/7 = 6/7$.

(Probability Density Function)

Question 2:

The random variable X has probability density function (PDF)

$$f_X(x) = \begin{cases} cx, & 0 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

Use the PDF to find

(a) the constant c,

(b) $Pr[0 \leq X \leq 1]$,

(c) $Pr[-1/2 \leq X \leq 1/2]$.

Answer:

(a) From the above PDF we can determine the value of c by integrating the PDF and setting it equal to 1.

$$\int_0^2 cx\,dx = c\left[\frac{x^2}{2}\right]_0^2 = c\left[\frac{4}{2}\right] = 2c = 1$$

Therefore $c = 1/2$.

(b) $Pr[0 \leq X \leq 1] = \int_0^1 x/2\,dx = 1/4$

(c) $Pr[-1/2 \leq X \leq 1/2] = \int_0^{1/2} x/2\,dx = 1/16$

(Bayes' rule)

Question 3:

Let A = {resistor is within 50Ω of the nominal value}. The probability that a resistor is from machine B is $Pr[B] = 0.3$. The probability that a resistor is acceptable, i.e., within 50 Ω of the nominal value, is $Pr[A] = 0.78$. Given that a resistor is from machine B, the conditional probability that it is acceptable is $Pr[A|B] = 0.6$. What is the probability that an acceptable resistor comes from machine B?

Answer:

We are given the event A that a resistor is within 50 Ω of the nominal value, and we need to find $Pr[B|A]$. Using Bayes' theorem, we have $Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A)}$.

Since all of the quantities we need are given in the problem description, our answer is
$Pr(B|A) = (0.6)(0.3)/(0.78) \approx 0.23$.

(Discrete random variable in Python)

Question 4:

Consider tossing a fair six-sided die. There are only six outcomes possible, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Suppose we toss two dice and assume that each throw is *independent.*

(a) What is the probability that the sum of the dice equals seven?
   i. List out all pairs of possible outcomes together with their sums from the two throws. (hint: enumerate all the items in `range(1,7)`)
   ii. Collect all of the *(a, b)* pairs that sum to each of the possible values from two to twelve (including the sum equals seven). (hint: use dictionary `from collections import defaultdict` to collect all of the *(a, b)* pairs that sum to each of the possible values from two to twelve)

(b) What is the probability that half the product of three dice will exceed their sum?

Answer:

Ref: Python for Probability, Statistics, and Machine Learning, Unpingco, José (pp.37-42).

(a) The first thing to do is characterize the measurable function for this as $X : (a, b) \rightarrow (a + b)$. Next, we associate all of the *(a, b)* pairs with their sum. A Python dictionary can be created like this:

```
#(i)
d={(i,j):i+j for i in range(1,7) for j in range(1,7)}
```

The next step is to collect all of the *(a, b)* pairs that sum to each of the possible values from two to twelve.

```
# (ii) collect all of the (a, b) pairs that sum to each of the possible values
#     from two to twelve
from collections import defaultdict
dinv = defaultdict(list)
for i,j in d.items(): dinv[j].append(i)
```

For example, `dinv[7]` contains the following list of pairs that sum to seven, `[(1, 6), (2, 5), (5, 2), (6, 1), (4, 3), (3, 4)]`. The next step is to compute the probability measured for each of these items. Using the independence assumption, this means we have to compute the sum of the products of the individual item probabilities in `dinv`. Because we know that each outcome is equally likely, the probability of every term in the sum equals 1/36.

```
#     Compute the probability measured for each of these items
#     including the sum equals seven
```

```
X={i:len(j)/36. for i,j in dinv.items() }
print(X)
```

```
{2: 0.027777777777777776, 3: 0.05555555555555555, 4: 0.08333333333333333, 5: 0.1111111111111111, 6: 0.1388888888888889, 7: 0.16666666666666666, 8: 0.1388888888888889, 9: 0.1111111111111111, 10: 0.08333333333333333, 11: 0.05555555555555555, 12: 0.027777777777777776}
```

(b) What is the probability that half the product of three dice will exceed their sum?
Using the same method above, we create the first mapping as follows:

```
d={(i,j,k):((i*j*k)/2>i+j+k) for i in range(1,7)
                              for j in range(1,7)
                                  for k in range(1,7)}
```

The keys of this dictionary are the triples and the values are the logical values of whether or not half the product of three dice exceeds their sum. Now, we do the inverse mapping to collect the corresponding lists,

```
dinv = defaultdict(list)
for i,j in d.items(): dinv[j].append(i)
```

Note that dinv contains only two keys, True and False. Again, because the dice are independent, the probability of any triple is $1/(6^3)$. Finally, we collect this for each outcome as in the following,

```
X={i:len(j)/6.0**3 for i,j in dinv.items() }
print(X)
```

```
{False: 0.37037037037037035, True: 0.6296296296296297}
```

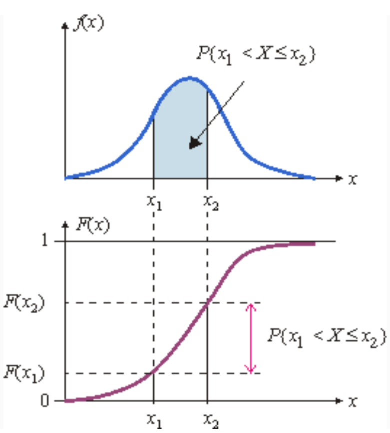
(Continuous random variable in Python)
Question 5:
Assuming a normal (Gaussian) distribution with mean 30 Ω and standard deviation of 1.8 Ω, determine the probability that a resistor coming off the production line will be within the range of 28 Ω to 33 Ω. (Hint: use `stats.norm.cdf` function `from scipy import stats`)

Answer:
Wiki: The cumulative distribution function of a real-valued random variable *X* is the function given by

$$F_X(x) = Pr(X \leq x), \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

where the right-hand side represents the probability that the random variable *X* takes on a value less than or equal to *x*. The probability that *X* lies in the semi-closed interval (*a, b*], where *a* < *b*, is therefore

$$Pr(a < X \leq b) = F_X(b) - F_X(a). \cdots\cdots\cdots\cdots\cdots (2)$$



```
from scipy import stats

# define constants mu =
30    # mean = 30Ω
sigma = 1.8  # standard deviation = 1.8Ω
x1 = 28      # lower bound = 28Ω x2 = 33
# upper bound = 33Ω

## calculate probabilities
# probability from Z=0 to lower bound
p_lower = stats.norm.cdf(x1,mu,sigma) #
probability from Z=0 to upper bound
p_upper = stats.norm.cdf(x2,mu,sigma)
# probability of the interval
Prob = (p_upper) - (p_lower)

# print the results print('\n')
print(f'Normal distribution: mean = {mu}, std dev = {sigma} \n')
print(f'Probability of occurring between {x1} and {x2}: ')
print(f'--> inside interval Pin = {round(Prob*100,1)}%') print(f'-
-> outside interval Pout = {round((1-Prob)*100,1)}% \n')
print('\n')
```

(Correlation versus Causation)

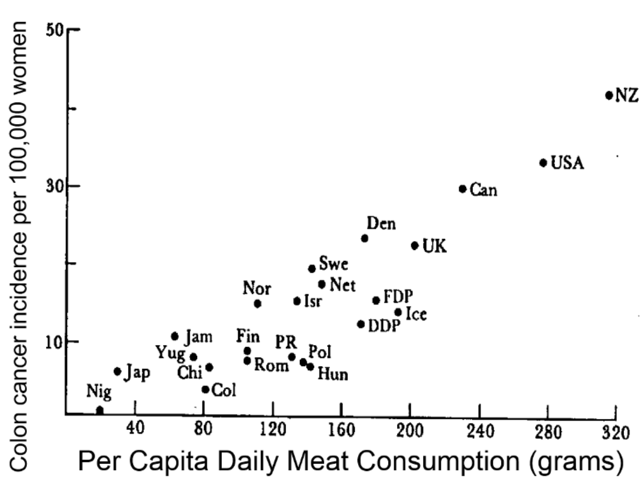
Question 6:

For each of the following graphs,

(i) State what you think the evidence is trying to suggest.

(ii) Give a reason why you agree or disagree with what the evidence is suggesting.

(iii) Identify whether the variable of the y-axis and the variable of the x-axis are correlated and/or causal?



http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1BDescriptiveStudies_and_Statistics/PH717-Module1B-DescriptiveStudies_and_Statistics6.html

Suggested Discussion:

(i) Colon cancer is correlated to the amount of daily meat consumption.

(ii) There is a clear linear trend; countries with the lowest meat consumption have the lowest rates of colon cancer, and the colon cancer rate among these countries progressively increases as meat consumption increases.

(iii) Probably causal.

Question 7: (Multiple responses – one or more answers are correct)
If A and B are correlated, but they're actually caused by C, which of the following statements are correct?

Ans:

a) A and C are correlated (Yes, A and C are correlated because A is caused by C)
b) B and C are correlated (Yes, B and C are correlated because B is caused by C)
c) A causes B to happen (No, A and B share the confounding factor C, but A and B don't have causal relationship)
d) A causes C to happen (No, C causes A to happen, however, we are not sure if A causes C to happen).

Question 8: (Multiple responses – one or more answers are correct)

We toss a coin and observe which side is facing up. Which of the following statements represent valid probability assignments for observing head P['H'] and tail P['T']?

Ans:

a) P['H']=0.2, P['T']=0.9 (Invalid because they sum to 1.1, which doesn't conform to the Axioms of probability)
b) P['H']=0.0, P['T']=1.0 (Valid because it conforms to the Axioms of probability)
c) P['H']=-0.1, P['T']=1.1 (Invalid because probability cannot have negative value)
d) P['H']=P['T']=0.5 (Valid because it conforms to the Axioms of probability)

Question 9: (True/False)

Two vectors $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ are linearly dependent?

Ans: False (because $\boldsymbol{a}$ is not a multiply of $\boldsymbol{b}$.)

Question 10: (Fill-in-blank)

The rank of the matrix $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ is 2. (try row echelon form)

Ans: $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ => $\begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix}$

Question 11: (Fill-in-blank)

The rank of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ is 2. (try row echelon form)

Ans: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ => $\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix}$=> $\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$ => $\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$