The Hong Kong University of Science and Technology
Department of Electronic and Computer Engineering
ELEC2600 Fall 2019 Homework-2
Please submit the soft copy of your homework solutions to Canvas
Due at 17:00 on Mar 21, 2019

1. (Lec 6 & 7 discrete R.V., 16 pts)
The pmf (probability mass function) of a random variable X is shown below:

X	-2	-1	0	1	2	otherwise
$p_X(x)$	0.2	0.4	a	0.1	0.1	0

Let A be the event that X is less than 0.
(a) Find the value of the constant a.
(b) Find $E[X]$ and $Var[X]$.
(c) Find the conditional pmf of X given A.
(d) Find $E[X|A]$ and $Var[X|A]$.
(e) Let $Y = 2X + 3$. Find the pmf of Y.

2. (Lec 8 & 11 discrete R.V. & conditional moments, 7 pts)
Consider a communication system with two transmitters and one receiver. Each millisecond, only one transmitter sends packets to the receiver. Transmitter 1 transmits packets with probability 0.7. Otherwise, transmitter 2 will transmit packets.

Assume the numbers of packets sent by transmitter 1 and transmitter 2 in one millisecond are geometric random variables ranging from 1 to infinity with mean 3 and 2, respectively.

Let X be the number of packets sent to the receiver in one millisecond.

(a) Find the pmf of X.
(b) Find the expected value of X.
(c) Find the variance of X.

3. (Lec 9 & 10 & 11 continuous R.V. & conditional moments, 11 pts)
The pdf (probability density function) of a continuous random variable X is given by
$$f_X(x) = \begin{cases} cx(2+x), & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$
(a) Find the value of the constant c.
(b) Find the expression for the cdf (cumulative distribution function) of X.
(c) Find the expected value and variance of X.
(d) Find the probability of event, $A = \{0.4 < X < 0.8\}$.
(e) Find the conditional pdf and conditional expected value of X given A, $f_{X|A}(x|A)$ and $E[X|A]$.

4. (Lec 10 & 11 continuous R.V. & conditional moment, 12 pts)
Suppose you are waiting for the mini-bus11 M at the bus station at the north gate of HKUST. Let T be the waiting time (measured in minutes) of the mini-bus 11M at this station, where T follows an exponential distribution with parameter $\lambda = 0.5$.
(a) Plot the pdf of T from 0 to 10. If you plot by hand, make sure that your plot is clear. Label the y-intercept and make sure values at 2, 4, 6, 8, 10 are close to the actual values.
(b) Compute the expected value of T.
(c) Find the conditional pdf of T given that the waiting time is less than 2 minutes. Plot it from 0 to 10. If you plot by hand, make sure that your plot is clear. Label the y-intercept and the values at any discontinuities in the function.
(d) Compute conditional expected value of T given that the waiting time is less than 2 minutes.
(e) Find the probability that the waiting time exceeds 4 minutes.
(f) Find the probability that you will continue to wait for at least another 1 minute, given that you have been waiting for 4 minutes already.

5. (Lec 12 function of R.V., 12 pts)
Suppose the charge remaining in the battery of an iPhone, X, is continuous and uniformly distributed in the interval $[150, 2600]$ mAh (milliamp-hours). Let Y be the estimated remaining battery life of the phone in minutes, where $Y = 0.25X - 37.5$.
(a) Plot the pdf of Y.
(b) Find the expected value and variance of Y.
(c) Find the pdf of Z, where $Z = 4X^2$.

6. (Python, 10 pts) Use Python to complete this question. Please submit a pdf file of your completed notebook on Canvas.

   The file “iris data.xlsx”, stores the petal lengths of two types of iris (Setosa and Versicolor).

   (a) Plot the normalized histogram of the petal length for all of the data without regard to iris type from 0 to 6 with bin size 0.2. The normalized histogram can be computed by dividing the count in each bin by the number of observations times the bin width. In python, this can be selected by passing “density=True” into the **matplotlib.pyplot.hist** or **pandas.DataFrame.hist** functions.

   (b) Calculate empirical mean *m* and standard deviation σ of the petal length. For a set of data $\{l_1,\ l_2, \cdots, l_n\}$, these can be computed by:
   $$m = \frac{1}{n}\sum_{i=1}^{n} l_i \text{ and } \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(l_i - m)^2}.$$
   In python, these can be computed by **pandas.DataFrame.mean** and **pandas.DataFrame.std** functions.

   (c) Assume that the probability density function of the petal length, $f_L(l)$, is given by a Gaussian distribution with mean and standard deviation computed above. You can compute values of the Gaussian distribution in python using the **scyipy.stats.norm.pdf** function. Note that the Gaussian distribution is also called the normal distribution.

   Compare the normalized histogram with the plot of $f_L(l)$ by plotting both together in the same figure for *l* from 0 to 6. Does this look like a good model of the data?

   (d) Now generate plots of normalized histogram of petal length for each iris type separately.

   (e) Compute the empirical conditional mean and conditional variance of the petal length for each iris type using the same equation as in (b) except the summations only contain data for each iris type. In Python, you can do this automatically using “**groupby**”.

   (f) Assume that the conditional probability density functions of the petal length given Setosa and Virginica, $f_{L|\text{type}}(l|\text{Setosa})$ and $f_{L|\text{type}}(l|\text{Versicolor})$, are given by Gaussian distributions with means and variances computed in (e), respectively.

   For each iris type, compare the normalized histogram with the conditional density by plotting them both in the same figure for *l* from 0 to 6. Does the Gaussian assumption look like a good assumption when each iris type is considered in isolation?

   (g) Assume that P[Setosa] = P[Versicolor] = 0.5. Use the total probability theorem to combine the two conditional densities to obtain a new model distribution for the petal length, $g_L(l)$. Compare the plot of $g_L(l)$ with the normalized histogram in (a) by plotting both in the same figure for *l* from 0 to 6. How does this compare with the single Gaussian assumption in part (c)?