Helwan University
Faculty of Computers & Information
Computer Science Department

Machine Learning - Sheet # 3:

Continuous Random Variables, their Probability Distributions, and Multivariate Probability Distributions

Reference Book: Dennis D. Wackerly, William Mendenhall III, and Richard L. Scheaffer, *Mathematical Statistics with Applications (6th Edition)*

1) Suppose that Y has density function

$$f(y) = \begin{cases} ky(1-y), & 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

- **a.** Find the value of k that makes $f(y)$ a probability density function.
- **b.** Find $P(.4 \le Y \le 1)$.
- **c.** Find $P(.4 \le Y < 1)$.
- **d.** Find $P(Y \le .4 \mid Y \le .8)$.
- **e.** Find $P(Y < .4 \mid Y < .8)$.

THEOREM 4.5 Let c be a constant, and let $g(Y)$, $g_1(Y)$, $g_2(Y)$, …, $g_k(Y)$ be functions of a continuous random variable Y. Then the following results hold:

1. $E(c) = c$.
2. $E[cg(Y)] = cE[g(Y)]$.
3. $E[g_1(Y) + g_2(Y) + \ldots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \ldots + E[g_k(Y)]$.

2) If Y is a continuous random variable with mean μ and variance σ^2 and a and b are constants, use theorem 4.5 to prove the following:

- **a.** $E(aY + b) = aE(Y) + b = a\mu + b$.
- **b.** $V(aY + b) = a^2V(Y) = a^2\sigma^2$.

3) The proportion of time Y that an industrial robot is in operation during a 40-hour week is a random variable with probability density function

$$f(y) = \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

- **a.** Find $E(Y)$ and $V(Y)$.
- **b.** For the robot under study, the profit X for a week is given by $X = 200Y - 60$. Find $E(X)$ and $V(X)$.
- **c.** Find an interval in which the profit should lie for at least 75% of the weeks that the robot is in use.

4) Weekly CPU time used by an accounting firm has probability density function (measured in hours) given by

$$f(y) = \begin{cases} (3/64)y^2(4-y), & 0 \le y \le 4 \\ 0, & elsewhere \end{cases}$$

- **a.** Find the expected value and variance of weekly CPU time.
- **b.** The CPU time costs the firm \$200 per hour. Find the expected value and variance of the weekly cost for CPU time.
- **c.** Would you expect the weekly cost to exceed \$600 very often? Why?

5) A company that manufactures and bottles apple juice uses a machine that automatically fills 16-ounce bottles. There is some variation, however, in the amounts of liquid dispensed into the bottles that are filled. The amount dispensed has been observed to be approximately normally distributed with mean 16 ounces and standard deviation 1 ounce. What proportion of bottles will have more than 17 ounces dispensed into them?

<u>6)</u> Assume that Y is normally distributed with mean μ and standard deviation σ. After observing a value of Y, a mathematician constructs a rectangle with length $L = |Y|$ and width $W = 3|Y|$. Let A denote the area of the resulting rectangle. What is $E(A)$?

<u>7)</u> An environmental engineer measures the amount (by weight) of particulate pollution in air samples of a certain volume collected over two smokestacks at a coal-operated power plant. One of the stacks is equipped with a cleaning device. Let Y_1 denote the amount of pollutant per sample collected above the stack that has no cleaning device, and let Y_2 denote the amount of pollutant per sample collected above the stack that is equipped with the cleaning device. Suppose that the relative frequency behavior of Y_1 and Y_2 can be modeled by

$$f(y_1, y_2) = \begin{cases} k, & 0 \le y_1 \le 2, \quad 0 \le y_2 \le 1, \quad 2y_2 \le y_1, \\ 0, & elsewhere \end{cases}$$

That is, Y_1 and Y_2 are uniformly distributed over the region inside the triangle bounded by $y_1 = 2$, $y_2 = 0$, and $2y_2 = y_1$.

- **a.** Find the value of k that makes this function a probability density function.
- **b.** Find $P(Y_1 \ge 3Y_2)$. That is, find the probability that the cleaning device reduces the amount of pollutant by one-third or more.

<u>8)</u> Suppose that the random variables Y_1 and Y_2 have joint probability density function $f(y_1,y_2)$ given by

$$f(y_1, y_2) = \begin{cases} 6y_1^2 y^2, & 0 \le y_1 \le y_2, \quad y_1 + y_2 \le 2 \\ 0, & elsewhere. \end{cases}$$

- **a.** Verify that this is a valid joint density function.
- **b.** What is the probability that $Y_1 + Y_2$ is less than 1?

<u>9)</u> Let Y_1 and Y_2 be independent exponentially distributed random variables, each with mean 1. Find $P(Y_1 > Y_2 | Y_1 < 2Y_2)$.

<u>10)</u> Assume that:

$$f(y_1, y_2) = \begin{cases} 4y_1 y_2, & 0 \le y_1 \le 1, \quad 0 \le y_2 \le 1 \\ 0, & elsewhere \end{cases}$$

- **a.** Find $E(Y_1)$.
- **b.** Find $V(Y_1)$.
- **c.** Find $E(Y_1 - Y_2)$.

<u>11)</u> Let Y_1 and Y_2 be uncorrelated random variables. Find the covariance and correlation between $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 - Y_2$ in terms of the variances of Y_1 and Y_2.

<u>12)</u> If Y_1 is the total time between a customer's arrival in the store and departure from the service window, and if Y_2 is the time spent in line before reaching the window, the joint density of these variables is given as

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \le y_2 \le y_1 \le \infty \\ 0, & elsewhere \end{cases}$$

- **a.** If 2 minutes elapse between a customer's arrival at the store and his departure from the service window, find the probability that he waited in line less than 1 minute to reach the window.
- **b.** Are Y_1 and Y_2 independent?
- **c.** The random variable $Y_1 - Y_2$ represents the time spent at the service window. Find $E(Y_1 - Y_2)$ and $V(Y_1 - Y_2)$. Is it highly likely that a randomly selected customer would spend more than 4 minutes at the service window?