Assignment 3

Problem 1: Following problems are from Jain's textbook, page 201: Problems 12.5, 12.6, 12.7, 12.8, 12.9, 12.10, and 12.11.

(1) The response time of a computer system has an Erlang distribution with the following cumulative distribution function (CDF):

$$F(x) = 1 - e^{-x/a} \left\{ \sum_{i=0}^{m-1} \frac{(x/a)^i}{i!} \right\}, \quad x \geq 0$$

Find expressions for the pdf, mean, variance, mode, and coefficient of variation of the response time.

Answer:

$$f(x) = \frac{x^{m-1} e^{-x/a}}{(m-1)! a^m}, \quad x \geq 0$$

mean $= am$, variance $= a^2 m$, mode $= a(m-1)$, coefficient of variance $= 1/\sqrt{m}$.

(2) The cumulative distribution function (CDF) of a Pareto variate is given by

$$F(x) = 1 - x^{-a}, \quad 1 \leq x < \infty$$

Find expressions for pdf, mean, variance, mode, and coefficient of variation.

Answers: $f(x) = ax^{-(a+1)}$; mean $= a/(a-1)$ provided $a > 1$, variance $= a/\left\{(a-1)^2 (a-2)\right\}$ provided $a > 2$, mode $= 1$, and coefficient of variation $= \{a(a-2)\}^{-1/2}$.

(3) The execution times of queries on a database is normally distributed with a mean of 5 seconds and a standard deviation of one second. Determine the following:

(a) What is the probability of the execution time being more than 8 seconds?

(b) What is the probability of the execution time being less than 6 seconds?

(c) What percentage of responses will take between 4 and 7 seconds?

(d) What is the 95-percentile execution time?

Answers: (a) 0.0013 (b) 0.8413 (c) 0.8185 (d) 6.644 seconds.

(4) What index of central tendency should be used to report?

(a) Response time (symmetrical pdf).

(b) Number of packets per day (symmetrical pdf).

(c) Number of packets per second (skewed pdf).

(d) Frequency of keywords in a language.

Answers: (a) Mean (b) Mean (c) Median (d) Mode.

(5) How would you summarize an average personal computer configuration: (a) CPU type, (b) Memory size (c) Disk type (d) Number of peripherals (e) Cost.

Answers: (a) Mode (b) Median (c) Mode (d) Median (e) Median.

(6) The CPU times in milliseconds for 11 workloads on a processor are: 0.74, 0.43, 0.24, 2.24, 262.08, 8960, 4720, 19740, 7360, 22440, and 28560. Which index of central tendency would you choose and why?

Answer: Since the ratio of maximum to minimum is very high, use the median.

(7) The number of disk I/O's performed by a number of programs were measured as follows: {23,33, 14, 15, 42, 28, 33, 45, 23, 34, 39, 21, 36, 23, 34, 36, 25, 9, 11, 19, 35, 24, 31, 29, 16, 23, 34, 24, 38, 15, 13, 35, 28}. Which index of central tendency would you choose and why?

Answer: Arithmetic mean since the data is very clustered together (not skewed) and $y_{\max}/y_{\min}$ ratio is small.

Problem 2: You are given an undirected graph $G = (V, E)$, where elements of the vertex set V are $1, 2, 3$, and 4. The edge set E is given by $\{v, w, x, z\}$, where $v = (1, 3), w = (2, 3), x = (2, 4)$, and $z = (1, 4)$.

- Using the matrix-tree theorem prove that the number of spanning trees is 4.
- Using the generalized form of matrix tree theorem, list all the spanning trees.

Problem 3: Consider a three node fully connected undirected graph. The degree of each node in this graph is two. The probability that each link is operational is p.

- List all the 3 distinct spanning trees of this graph.
- Using the principle of inclusion and exclusion theorem of probability theory, prove that the reliability of this network is given by $p^2(1 + 2q)$, where $q = (1 - p)$.

Solution:

(Note: In this outline, for the sake of clarity, we will indicate the trees by their edge-sets.)

This graph looks like a triangle. Denote its edges (links) by a, b, and c. It has three spanning trees. Denote the edge sets of the spanning trees by $\tau_i, 1 \leq i \leq 3$, where $\tau_1 = \{a, b\}, \tau_2 = \{b, c\}$, and $\tau_3 = \{c, a\}$. Note that T_i, is a random variable associated with the tree τ_i, where once again $i = 1, 2$, and 3. The random variable T_i takes two values. These are either 0 or 1. More specifically

$$T_i = \begin{cases} 1 & \text{if tree } \tau_i \text{ is working (connected)} \\ 0 & \text{if tree } \tau_i \text{ is not working (disconnected connected)} \end{cases}$$

The tree τ_1 is connected, if both the links a and b are working. Therefore the probability that this tree is connected, is p^2. Recall that $P(T_1)$ is the probability that the tree τ_1 is connected. Therefore $P(T_1) = p^2$. Similarly $P(T_2) = P(T_3) = p^2$.

Now, $T_1 \cap T_2$ implies that both the trees τ_1 and τ_2 have to be connected. This will occur if all the links in the sets τ_1 and τ_2 are operational, that is we want all the links

in the set $\{a,b\} \cup \{b,c\}$ to be operational. This set is $\{a,b,c\}$. Therefore if all the links a, b, and c work, then the trees τ_1 and τ_2 are operating simultaneously. Consequently $P(T_1 \cap T_2) = p^3$. Similarly $P(T_2 \cap T_3) = P(T_1 \cap T_3) = p^3$.

Once again, for all the three trees $\tau_i, 1 \leq i \leq 3$ to be functioning simultaneously, the links in the set $\{a,b\} \cup \{b,c\} \cup \{c,a\} = \{a,b,c\}$ have to function simultaneously. The probability of this event is equal to p^3. That is $P(T_1 \cap T_2 \cap T_3) = p^3$.

Summarizing, we have

$$\begin{aligned} P(T_1) &= P(T_2) = P(T_3) = p^2 \\ P(T_1 \cap T_2) &= P(T_2 \cap T_3) = P(T_3 \cap T_1) = p^3 \\ P(T_1 \cap T_2 \cap T_3) &= p^3 \end{aligned}$$

Using the inclusion-exclusion principle we obtain $\mathcal{R}(G)$ to be

$$\begin{aligned} \mathcal{R}(G) &= P(T_1 \cup T_2 \cup T_3) \\ &= P(T_1) + P(T_2) + P(T_3) - P(T_1 \cap T_2) - P(T_1 \cap T_3) - P(T_2 \cap T_3) \\ &\quad + P(T_1 \cap T_2 \cap T_3) \\ &= 3p^2 - 3p^3 + p^3 \\ &= 3p^2 - 2p^3 \\ &= p^2(3 - 2p) \end{aligned}$$

Substituting $p = (1-q)$ in $(3-2p)$, we get

$$\mathcal{R}(G) = p^2(1+2q)$$

□