MAT231CT: LINEAR ALGEBRA AND PROBABILITY THEORY
Unit 4: Probability distributions and sampling theory

TUTORIAL SHEET-1 : Binomial Distribution

1. For each scenario described below, state whether or not the binomial distribution is a reasonable model for the random variable and why. State any assumptions you make.
   a. A production process produces thousands of temperature transducers. Let X denote the number of nonconforming transducers in a sample of size 30 selected at random from the process.
   b. From a batch of 50 temperature transducers, a sample of size 30 is selected without replacement. Let X denote the number of nonconforming transducers in the sample.
   c. Four identical electronic components are wired to a controller that can switch from a failed component to one of the remaining spares. Let X denote the number of components that have failed after a specified period of operation.
   d. Let X denote the number of accidents that occur along the federal highways in Arizona during a one-month period.

2. Let X be a binomial random variable with $p = 0.1$ and $n = 10$. Calculate the following probabilities from the binomial probability mass function
   a. $P(X \leq 2)$
   b. $P(X > 8)$
   c. $P(X = 4)$
   d. $p(5 \leq X \leq 7)$
   (Ans: a) 0.9298 b) 0 c) 0.0112 d) 0.0016)

3. The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline.
   a. What is the probability that for exactly three calls the lines are occupied?
   b. What is the probability that for at least one call the lines are not occupied?
   c. What is the expected number of calls in which the lines are all occupied?
   (Ans: a) 0.215 b) 0.9999 c) 4)

4. A particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.
   a. Over five mornings, what is the probability that the light is green on exactly one day?
   b. Over 20 mornings, what is the probability that the light is green on exactly four days?
   c. Over 20 mornings, what is the probability that the light is green on more than four days?
   (Ans: a) 0.410 b) 0.218 c) 0.37)

5. An article in Information Security Technical Report ["Malicious Software—Past, Present and Future" (2004, Vol. 9, pp. 6–18)] provided the following data on the top ten malicious software instances for 2002. The clear leader in the number of registered incidences for the year 2002 was the Internet worm "Klez," and it is still one of the most widespread threats. This virus was first detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology.

Place	Name	% instances

1	I-Worm.Klez	61.22%
2	I-Worm.Lentin	20.52%
3	I-Worm.Tanatos	2.09%
4	I-Worm.BadtransII	1.31%
5	Macro.Word97.Thus	1.19%
6	I-Worm.Hybris	0.60%
7	I-Worm.Bridex	0.32%
8	I-Worm.Magistr	0.30%
9	Win95.CIH	0.27%
10	I-Worm.Sircam	0.24%

The 10 most widespread malicious programs for 2002 (Source—Kaspersky Labs)

Suppose that 20 malicious software instances are reported. Assume that the malicious sources can be assumed to be independent.

a. What is the probability that at least one instance is "Klez"?
b. What is the probability that three or more instances are "Klez"?
c. What are the mean and standard deviation of the number of "Klez" instances among the 20 reported?
(Ans: a) 1 b) 0.99997 c) 12.244 and 2.179)

6. Because not all airline passengers show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. The probability that a passenger does not show up is 0.10, and the passengers behave independently.
   a. What is the probability that every passenger who shows up can take the flight?
   b. What is the probability that the flight departs with empty seats?
   (Ans: a) 0.9961 b) 0.9886)

7. According to Chemical Engineering Progress (November 1990), approximately 30% of all pipework failures in chemical plants are caused by operator error.
   a. What is the probability that out of the next 20 pipework failures at least 10 are due to operator error?
   b. What is the probability that no more than 4 out of 20 such failures are due to operator error?
   c. Suppose, for a particular plant, that out of the random sample of 20 such failures, exactly 5 are due to operator error. Do you feel that the 30% figure stated above applies to this plant? Comment.
   (Ans: a) 0.0480 b) 0.2375 c) $P(X = 5 \mid p = 0.3) = 0.1789$, $P = 0.3$ is reasonable)

8. One prominent physician claims that 70% of those with lung cancer are chain smokers. If his assertion is correct,
   a. find the probability that of 10 such patients recently admitted to a hospital, fewer than half are chain smokers;
   b. find the probability that of 20 such patients recently admitted to a hospital, fewer than half are chain smokers. (Ans: a) 0.0474 b) 0.2375)

9. If the probability that a fluorescent light has a useful life of at least 800 hours is 0.9, find the probabilities that among 20 such lights
   a. exactly 18 will have a useful life of at least 800 hours;
   b. at least 15 will have a useful life of at least 800 hours;
   c. at least 2 will not have a useful life of at least 800 hours.
   (Ans: a) 0.2852 b) 0.9887 c) 0.6083)

TUTORIAL SHEET-2 : Poisson Distribution

1. Suppose that the number of customers who enter a bank in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$. Determine the mean and variance of X.
   (Ans: $E[X] = V(X) = 2.996$)

2. The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour.
   a. What is the probability that there are exactly five calls in one hour? (Ans : 0.0378)
   b. What is the probability that there are three or fewer calls in one hour? (Ans: 0.0103)
   c. What is the probability that there are exactly 15 calls in two hours? (Ans : 0.051)
   d. What is the probability that there are exactly five calls in 30 minutes? (Ans : 0.175)

3. Astronomers treat the number of stars in a given volume of space as a Poisson random variable. The density in the Milky Way Galaxy in the vicinity of our solar system is one star per 16 cubic light years.
   a. What is the probability that two or more stars in 16 cubic light years? [Ans: 0.264]
   b. How many cubic light years of space must be studied so that the probability of one or more stars exceed 0.95? [Ans: 48]

4. In 1898 L. J. Bortkiewicz published a book entitled The Law of Small Numbers. He used data collected over 20 years to show that the number of soldiers killed by horse kicks each year in each corps in the Prussian cavalry followed a Poisson distribution with a mean of 0.61.
   a. What is the probability of more than one death in a corps in a year? [Ans : 0.4566]
   b. What is the probability of no deaths in a corps over five years? [Ans : 0.047]

5. The number of surface flaws in plastic panels used in the interior of automobiles has a Poisson distribution with a mean of 0.05 flaw per square foot of plastic panel. Assume an automobile interior contains 10 square feet of plastic panel.
   a. What is the probability that there are no surface flaws in an auto's interior?[Ans: 0.6065]
   b. If 10 cars are sold to a rental company, what is the probability that none of the 10 cars has any surface flaws?[Ans: 0.0067]
   c. If 10 cars are sold to a rental company, what is the probability that at most one car has any surface flaws?
      [Ans: $P(W \leq 1) = P(W = 0) + P(W = 1) = 0.0067 + 0.0437 = 0.0504$]

6. The number of content changes to a Web site follows a Poisson distribution with a mean of 0.25 per day.
   a. What is the probability of two or more changes in a day? [Ans: 0.026]
   b. What is the probability of no content changes in five days? [Ans: 0.287]
   c. What is the probability of two or fewer changes in five days?[Ans:8.868]

TUTORIAL SHEET-3 : Exponential Distribution

1. Suppose the counts recorded by a Geiger counter follow a Poisson process with an average of two counts per minute.
   a. What is the probability that there are no counts in a 30-second interval? [Ans: 0.3679]
   b. What is the probability that the first count occurs in less than 10 seconds? [Ans:0.1353]
   c. What is the probability that the first count occurs between 1 and 2 minutes after start-up?[Ans:0.1170]

2. The time between calls to a plumbing supply business is exponentially distributed with a mean time between calls of 15 minutes.
   a. What is the probability that there are no calls within a 30-minute interval?
   b. What is the probability that at least one call arrives within a 10-minute interval?
   c. What is the probability that the first call arrives within 5 and 10 minutes after opening?
   d. Determine the length of an interval of time such that the probability of at least one call in the interval is 0.90
   [Ans: a) 0.1353 b) 0.4866 c) 0.2031 d) 34.54]

3. Suppose that the time to failure (in hours) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$
   a. What proportion of the fans will last at least 10,000 hours? [Ans: 0.0498]
   b. What proportion of the fans will last at most 7000 hours? [Ans: 0.8775]

4. The time between arrivals of taxis at a busy intersection is exponentially distributed with a mean of 10 minutes.
   a. What is the probability that you wait longer than one hour for a taxi? [Ans: .00025]
   b. Suppose you have already been waiting for one hour for a taxi. What is the probability that one arrives within the next 10 minutes? [Ans:0.6321]
   c. Determine x such that the probability that you wait more than x minutes is 0.10. [Ans:23.03]
   d. Determine x such that the probability that you wait less than x minutes is 0.90. [Ans:23.03]
   e. Determine x such that the probability that you wait less than x minutes is 0.50.[Ans:6.93]

5. The lifetime of a mechanical assembly in a vibration test is exponentially distributed with a mean of 400 hours.
   a. What is the probability that an assembly on test fails in less than 100 hours?
   b. What is the probability that an assembly operates for more than 500 hours before failure?
   c. If an assembly has been on test for 400 hours without a failure, what is the probability of a failure in the next 100 hours?
   d. If 10 assemblies are tested, what is the probability that at least one fails in less than 100 hours? Assume that the assemblies fail independently.
   e. If 10 assemblies are tested, what is the probability that all have failed by 800 hours? Assume the assemblies fail independently.
   [Ans: a) 0.2212 b) 0.2865 c) 0.2212 d) 0.9179 e) 0.2337]

6. The time between calls to a corporate office is exponentially distributed with a mean of 10 minutes.
   a. What is the probability that there are more than three calls in one-half hour?

b. What is the probability that there are no calls within one-half hour?
c. Determine x such that the probability that there are no calls within x hours is 0.01.
d. What is the probability that there are no calls within a two-hour interval?
e. If four nonoverlapping one-half-hour intervals are selected, what is the probability that none of these intervals contains any call?
[Ans: a) 0.3528 b) 0.04979 c) 46.05 d) 6.14×10^{-6} e) e^{-12}]

7. Web crawlers need to estimate the frequency of changes to Web sites to maintain a current index for Web searches. Assume that the changes to a Web site follow a Poisson process with a mean of 3.5 days.
   a. What is the probability that the next change occurs in less than two days? [Ans: 0.435]
   b. What is the probability that the next change occurs in greater than seven days? [Ans:0.135]
   c. What is the time of the next change that is exceeded with probability 90%? [Ans:0.369]
   d. What is the probability that the next change occurs in less than 10 days, given that it has not yet occurred after three days? [Ans: 0.865]

TUTORIAL SHEET- 4 : Normal Distribution

1. Assume X is normally distributed with a mean of 5 and a standard deviation of 4. Determine the following:
   a. $P(X < 11)$ [Ans: 0.93319]
   b. $P(X > 0)$ [Ans: 0.89435]
   c. $P(3 < X < 7)$ [Ans: 0.38292]
   d. $P(-2 < X < 9)$ [Ans: 0.80128]
   e. $P(2 < X < 8)$ {Ans: 0.54674]

2. The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.
   a. What is the probability that a sample's strength is less than 6250 Kg/cm^2?
   b. What is the probability that a sample's strength is between 5800 and 5900 Kg/cm^2?
   c. What strength is exceeded by 95% of the samples?
   [Ans: a) 0.99379 b) 0.13591 c) 5835]

3. The line width for semiconductor manufacturing is assumed to be normally distributed with a mean of 0.5 micrometer and a standard deviation of 0.05 micrometer.
   a. What is the probability that a line width is greater than 0.62 micrometer? [Ans; 0.0082]
   b. What is the probability that a line width is between 0.47 and 0.63 micrometer? [Ans: 0.72109]
   c. The line width of 90% of samples is below what value? [Ans: 0.564]

4. The speed of a file transfer from a server on campus to a personal computer at a student's home on a weekday evening is normally distributed with a mean of 60 kilobits per second and a standard deviation of 4 kilobits per second.
   a. What is the probability that the file will transfer at a speed of 70 kilobits per second or more? [Ans: 0.00621]
   b. What is the probability that the file will transfer at a speed of less than 58 kilobits per second? [Ans:0.308538]

c. If the file is 1 megabyte, what is the average time it will take to transfer the file? [Ans: 133.3]

5. The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.
   a. What is the probability that a laser fails before 5000 hours? [Ans: 0.00043]
   b. What is the life in hours that 95% of the lasers exceed? [Ans: 6016]
   c. If three lasers are used in a product and they are assumed to fail independently, what is the probability that all three are still operating after 7000 hours? [Ans; 1/8]

6. The weight of a sophisticated running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.5 ounce.
   a. What is the probability that a shoe weighs more than 13 ounces? [Ans: 0.02275]
   b. What must the standard deviation of weight be in order for the company to state that 99.9% of its shoes are less than 13 ounces? [Ans: 0.324]
   c. If the standard deviation remains at 0.5 ounce, what must the mean weight be in order for the company to state that 99.9% of its shoes are less than 13 ounces? [Ans: 11.455]

7. The serum cholesterol level X in 14-year-old boys has approximately a normal distribution with mean 170 and standard deviation 30.
   a. Find the probability that the serum cholesterol level of a randomly chosen 14-year-old boy exceeds 230. [Ans: 0.0228]
   b. In a middle school there are 300 14-year-old boys. Find the probability that at least 8 boys have a serum cholesterol level that exceeds 230. [Ans: 0.3974]

8. A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.
   a. What is the probability that a trip will take at least 1/2 hour? [Ans: 0.0571]
   b. If the office opens at 9:00 A.M. and the lawyer leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work? [Ans: 99.11%]
   c. If he leaves the house at 8:35 A.M. and coffee is served at the office from 8:50 A.M. until 9:00 A.M., what is the probability that he misses coffee? [Ans: 0.3974]
   d. Find the length of time above which we find the slowest 15% of the trips. [Ans: 27.95 min]
   e. Find the probability that 2 of the next 3 trips will take at least 1/2 hour.[Ans: 0.0092]

TUTORIAL SHEET- 5 : SAMPLING DISTRIBUTION OF MEANS

1. The average life of a bread-making machine is 7 years, with a standard deviation of 1 year. Assuming that the lives of these machines follow approximately a normal distribution, find
   a) The probability that the life of a machine is less that 6 years [Ans: 0.158655]
   b) the probability that the mean life of a random sample of 9 such machines falls between 6.4 and 7.2 years; [Ans : 0.6898]
   c) the value of x to the right of which 15% of the means computed from random samples of size 9 would fall.[Ans : 7.35]

2. PVC pipe is manufactured with a mean diameter of 1.01 inch and a standard deviation of 0.003 inch. Find the probability that a random sample of $n = 9$ sections of pipe will have a sample mean diameter greater than 1.009 inch and less than 1.012 inch. [Ans : 08186]

3. If all possible samples of size 16 are drawn from a normal population with mean equal to 50 and standard deviation equal to 5, what is the probability that a sample mean $\bar{X}$ will fall in the interval from $\mu_{\bar{X}} - 1.9\sigma_{\bar{X}}$ to $\mu_{\bar{X}} - 1.9\sigma_{\bar{X}}$? Assume that the sample means can be measured to any degree of accuracy. [Ans: 0.3159]

4. If the standard deviation of the mean for the sampling distribution of random samples of size 36 from a large or infinite population is 2, how large must the sample size become if the standard deviation is to be reduced to 1.2? [Ans = 100]

5. A certain type of thread is manufactured with a mean tensile strength of 78.3 kilograms and a standard deviation of 5.6 kilograms. How is the variance of the sample mean changed when the sample size is
   a) increased from 64 to 196? [Ans : Var is reduced from 0.49 to 0.16
   b) decreased from 784 to 49? [Ans : Variance is increased from 0.04 to 0.64]

6. A soft-drink machine is regulated so that the amount of drink dispensed averages 240 milliliters with standard deviation of 15 milliliters. Periodically, the machine is checked by taking a sample of 40 drinks and computing the average content. If the mean of the 40 drinks is a value within the interval $\mu_{\bar{X}} \pm 2\sigma_{\bar{X}}$, the machine is thought to be operating satisfactorily; otherwise, adjustments are made. In Section 8.3, the company official found the mean of 40 drinks to be $\bar{x}$ = 236 milliliters and concluded that the machine needed no adjustment. Was this a reasonable decision? [Ans : Yes]

7. Suppose that X has a discrete uniform distribution

$$f(x) = \begin{cases} 1/3, & x = 1, 2, 3 \\ 0, & otherwise \end{cases}$$

   A random sample of $n = 36$ is selected from this population. Find the probability that the sample mean is greater than 2.1 but less than 2.5, assuming that the sample mean would be measured to the nearest tenth. [Ans : 0.2312]

TUTORIAL SHEET- 6 : SAMPLING DISTRIBUTION OF DIFFERENCES OF MEANS

1. The distribution of heights of a certain breed of terrier has a mean of 72 centimeters and a standard deviation of 10 centimeters, whereas the distribution of heights of a certain breed of poodle has a mean of 28 centimeters with a standard deviation of 5 centimeters. Assuming that the sample means can be measured to any degree of accuracy, find the probability that the sample mean for a random sample of heights of 64 terriers exceeds the sample mean for a random sample of heights of 100 poodles by at most 44.2 centimeters. [Ans : 0.5596]

2. A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size $n_2 = 9$ is taken from another normal population with mean 70 and standard deviation 12. Let $\bar{X}_1$ and $\bar{X}_2$ be the two sample means. Find:
   a) The probability that $\bar{X}_1 - \bar{X}_2$ exceeds 4 [Ans : 0.5885]
   b) The probability that $3.5 \leq \bar{X}_1 - \bar{X}_2 \leq 5.5$ [Ans : 0.1759]

3. The mean score for freshmen on an aptitude test at a certain college is 540, with a standard deviation of 50. Assume the means to be measured to any degree of accuracy. Determine the probability that two groups selected at random, consisting of 32 and 50 students, respectively, will differ in their mean scores by
   a) more than 20 points;
   b) an amount between 5 and 10 points.

4. For boys, the average number of absences in the first grade is 15 with a standard deviation of 7; for girls, the average number of absences is 10 with a standard deviation of 6. In a nationwide survey, suppose 100 boys and 50 girls are sampled. What is the probability that the male sample will have *at most* three more days of absences than the female sample? [Ans : 0.03452]

5. Assume there are two species of green beings on Mars. The mean height of Species 1 is 32 while the mean height of Species 2 is 22. The variances of the two species are 60 and 70, respectively and the heights of both species are normally distributed. You randomly sample 10 members of Species 1 and 14 members of Species 2. What is the probability that the mean of the 10 members of Species 1 will exceed the mean of the 14 members of Species 2 by 5 or more? Without doing any calculations, you probably know that the probability is pretty high since the difference in population means is 10. But what exactly is the probability? [Ans : 0.934]

6. The mean height of 15-year-old boys (in cm) is 175 and the variance is 64. For girls, the mean is 165 and the variance is 64. If eight boys and eight girls were sampled, what is the probability that the mean height of the sample of girls would be higher than the mean height of the sample of boys? In other words, what is the probability that the mean height of girls minus the mean height of boys is greater than 0? [Ans : 0.0062]

TUTORIAL SHEET- 7 : SAMPLING DISTRIBUTION PROPORTIONS

1. Suppose it is known that 43% of Americans own an iPhone. If a random sample of 50 Americans were surveyed,
   a) What is the probability that the proportion of the sample who owned an iPhone is between 45% and 50%? [Ans: 0.2287]
   b) If a random sample of size of seventy-five was surveyed, what is the probability we would find more than 50% of Americans with an iPhone? [Ans: 0.1112]

2. According to the US Census Bureau's American Community Survey, 87% of Americans over the age of 25 have earned a high school diploma. Suppose we are going to take a random sample of Americans in this age group and calculate what proportion of the sample has a high school diploma.What is the probability that the proportion of people in the sample with a high school diploma is less than 85% ? [Ans: 0.20]

3. Suppose that in a population of voters in a certain region 38% are in favour of particular bond issue. Nine hundred randomly selected voters are asked if they favour the bond issue.
   a) Verify that the sample proportion $\hat{P}$ computed from samples of size 900 meets the condition that its sampling distribution be approximately normal. [Ans: yes]
   b) Find the probability that the sample proportion computed from a sample of size 900 will be within 5 percentage points of the true population proportion. [Ans: 0.9980]

4. An online retailer claims that 90% of all orders are shipped within 12 hours of being received. A consumer group placed 121 orders of different sizes and at different times of day; 102 orders were shipped within 12 hours.
   a) Compute the sample proportion of items shipped within 12 hours. [Ans: 0.84]
   b) Confirm that the sample is large enough to assume that the sample proportion is normally distributed. Use $p = 0.90$, corresponding to the assumption that the retailer's claim is valid. [Ans: Yes]
   c) Assuming the retailer's claim is true, find the probability that a sample of size 121 would produce a sample proportion so low as was observed in this sample. [Ans: 0.0139]
   d) Based on the answer to part (c), draw a conclusion about the retailer's claim [Ans: actual value of p is less than 90%]