IE 415/515 Probability and Statistics Review Questions

Random variables and functions characterizing random variables

a) What are the two main types of random variables? Give an example of each type.

b) What is the notational convention for random variables and realizations of random variables?

c) Name the three different probability distribution functions and the type of random variables to which they apply.

d) What are the notational conventions used for the functions in the prior question?

e) What are the mathematical properties that the functions in c must satisfy?

f) How are the two functions used to characterize discrete random variables mathematically related? How are the two functions used to characterize continuous random variables mathematically related?

g) For a random variable X what is the definition of the expected value ($E[X]$) and variance ($Var[X]$) when X is discrete and continuous?

h) What do probability distributions with names (e.g., normal, exponential, Poisson, …) represent?

T F i) Random variables only take on non-negative values.

T F j) If a set of data represent realizations of some random variable, the "shape" of a histogram constructed from the data is an estimate of the shape of the cumulative distribution function of the random variable.

T F k) If $x_1, x_2, \ldots, x_n$ represent n independent observations of some random variable X, then as n gets larger $\bar{x}$ gets "closer" to the true mean $E[X]$.

Statistics and statistical inference

a) What is a statistic?

b) What is a sampling distribution?

c) For a normally distributed random variable X, with mean μ and standard deviation σ, what are the mean and standard deviation of the sample average of n observations of X? What is the distribution of the sample average of n observations?

d) What are two graphical methods for validating that observed data are from a normal distribution?

e) Given a set of numerical data, name two measures (computed from the data) of central tendency, and two measures of unpredictability.

f) What is the definition of the coefficient of variation for a random variable X? For n observations of X, how would you estimate the coefficient of variation of X?

g) In statistical inference what does the null hypothesis represent?

h) What are the two types of errors in statistical inference – explain what they mean?

i) In a hypothesis test what are the three items (parameters) that affect type II error, and how does type II error change as each parameter gets larger.

j) What is the name of the set of graphs that show the relationship between type II error, sample size, and normalized difference to detect for a specific hypothesis test?

T F k) The information used to complete a hypothesis test (at significance level α), can also be used to construct a (1-α)*100% confidence interval.

T F l) When conducting a hypothesis test on the mean of a normally distributed random variable where the standard deviation is estimated from the sample data, the sampling distribution of the test statistics is a standard normal distribution.

T F m) The central limit theorem implies that the sampling distribution of the sample standard deviation of a large (≥30) number of independent observations (of the same random variable) is approximately normal.

T F n) In a hypothesis test Type II error is always larger than type I error.

T F o) If $x_1, x_2, \ldots, x_n$ represent *n* independent observations from some distribution and $n > 30$, then the sample average $\bar{x}$ is an observation from a (approximately) normal distribution.

T F p) The sample average as a measure of central tendency is less sensitive to presence of extreme or "outlier" values than the sample median.

T F q) Statistics computed from realizations of random variables are also realizations of a random variable.

Problems

Problem I.

10, 10, 0, 5, 0 are realizations of a discrete random variable X with the following probability mass function.

$$p(i) = \begin{cases} 1/4 \ \mathit{for} & i = 0 \\ y \ \mathit{for} & \mathrm{i} = 5 \\ 1/2 \ \mathit{for} & i = 10 \\ 0 & \mathit{otherwise} \end{cases}$$

a. What is the value of y?
b. Compute the expected value ($\mu = E[X]$) and variance ($\sigma^2 = Var[X]$) of the distribution.
c. Compute the sample mean, median, mode, variance, range, and coefficient of variation from the realizations.
d. Write out the probability distribution function *F(x)*, and draw its graph

Problem II

Individual customers arrive to a coffee shop after it opens at 6AM. The number of customers arriving by 7AM has a Poisson distribution with parameter $\lambda = 10$. 25% of the customers buy a single donut and coffee, and 75% just buy coffee. Answer the following.

a. What are the two random variables described above and what are their probability mass functions?
b. What is the probability that exactly five customers arrive by 7AM?
c. If 10 customers enter by 7AM, what is the probability that five donuts are purchased?
d. What is the expected number (true mean) of donuts a single customer will purchase, and what is the expected number of customers that arrive by 10AM?

Problem III

A normally distributed random variable X is assumed to have a mean of 15. The standard deviation of this random variable is 2.5. A random sample of five observations are obtained and the sample average = 12.5.

a. Test the hypothesis that μ (the mean of X) equals 15 (the alternative hypothesis is that the mean does not equal 15). Use $\alpha = 0.05$.

b. What is the p-value for the test statistic computed in part a?

c. Assume the true mean of X is 13 (with $\sigma = 2.5$) and five new observations of X will be obtained. If a test that $\mu = 15$ will be conducted with the five new observations, what is the probability of Type II error (β)?

d. The sample standard deviation obtained from yet another five observations of X obtained at a different time (with sample average = 11.9) is 5. Because of this value the assumption that the standard deviation of X is 2.5 is no longer made. Using these most recent five observations conduct a test of the hypothesis that μ (the mean of X) equals 15. Use $\alpha = 0.05$.