Department of Computer Science

Computer Science 177: Modelling and Simulation

FINAL EXAMINATION — Winter 2006

1. Indicate whether each of the following statements is true or false. (2 points for each correct answer, -1 point for each wrong answer, 0 points for each blank.)

_____ a. If the $Q-Q$ plot for two cumulative distribution functions $F_X(x)$ and $F_Y(Y)$ is a *straight line*, then the random variables X and Y have the same distribution, except for a location or scale parameter.

_____ b. The *covariance* of two random variables can be used to test whether they are independent in the following way. If the covariance is close to *zero*, then we *cannot determine anything* about independence without gathering more data. Otherwise, as the covariance becomes *more positive*, we can conclude with greater confidence that the two variables are *not independent*, and as the covariance becomes *more negative* we can conclude with greater confidence that the two variables are *independent*.

_____ c. The exponential distribution represents the number of events during some fixed-length interval, if the time between events has the Poisson distribution.

_____ d. The Law of Large Numbers tells us not to add together large numbers of independent measurements in a computer program because of the risk of numerical errors due the properties of floating point arithmetic.

_____ e. According to the principle of maximum entropy, it is always better to represent each attribute in in your simulation model by a *random variable with a large variance* because Mother Nature greatly prefers chaos more than order.

_____ f. For an event-driven simulation model, you can reduce the execution time for your program by measuring time in a larger unit, i.e., hours instead of seconds.

_____ g. If a CSIM-19 process executed the statement "hold (30)" when the time on the clock was already greater than 30, then the process would keep going without stopping.

_____ h. If one long simulation run is split up to create several measurements using the regenerative method, the measurements are guaranteed to be independent of one another, but they probably won't be identically distributed.

_____ i. The Chi-Squared distribution with d degrees of freedom describes a value of a sum of squares $X_1^2, \cdots, X_d^2$ where each term X_i has an i.i.d. Standard Normal distribution.

_____ j. You should always choose the method for computing an average based on the units associated with the data. For example, the average passenger waiting time should be computed as a time average, whereas the average number of passengers standing on the moving sidewalk should be computed as a population average.

2. Consider the elevator simulation problem in your programming assignment. One of the required measurements was the *average trip length* for a passenger, since the building designers wanted to ensure that the average trip length for a passenger is less than 2 minutes. Suppose you misunderstood the requirement, and instead measured *the average time for elevator to complete a "trip"*, defined as the length of time it is going in a single direction (up or down) before it reverses direction to start the "reverse trip".

   a. Give some examples to show that these two measurements are different. Show how the average trip time for passengers might be either smaller or larger than the average trip time for an elevator.

   b. Briefly explain the difference between a time average and a population average, and this distinction means the average class size at UCR is bigger if you survey students, than if you survey professors.

   c. Use the result from part (b) to explain why the the average trip time for an elevator is likely to be significantly smaller than the average trip time for passengers.

3. Suppose you just finished a five-day data-collection study in the elevator lobby of a large building. During each day, you spent many hours counting passenger arrivals and recording *the number of passenger arrivals each 10 minute interval.* At the end of the ith day, you recorded x_i, *the maximum number of arrivals* during any 10 minute interval in that day, which gave you the following set of 5 measurements:

   38, 42, 46, 41, and 43.

   a. Let X be a random variable representing the *maximum number of arrivals* you would find in any 10 minute period during one day of studying this buidling. Find the sample mean and sample variance for X.

   b. Calculate the 95% confidence interval for the average value of X. (HINT: Since each value for X comes from a large number of 10-minute measurement periods, you may assume that X has a Normal distribution.)

   c. The building architect wants to redesign the elevator lobby so it can hold larger crowds. Analyse your data to tell him how much capacity will be needed for the new layout, by estimating:

      i. X_M, the maximum number of arrivals at the elevator lobby we should see during any 10 minute period during a typical month of normal operations. (HINT: assume that the working definition for a ‘‘month’’ consists of 4 weeks, each containing 5 days of operation.)

      ii. X_Y, the maximum number of arrivals at the elevator lobby we should see during a typical year of normal operations. (HINT: assume that the working definition for a ‘‘year’’ consists of 10 months.)

Formulas:

Poisson Distribution: Number of arrivals during time t with arrival rate $\lambda$is (mean= λt, variance= λt).

$$P[k] = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t}$$

Chi-Squared test: $\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$

Confidence Intervals: $\bar{X}(n) = \frac{1}{n} \sum_{i=1}^{n} X_i$ $\qquad S^2(n) = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X}(n))^2$

$Prob[abs(\bar{X}(n) - \mu)/\sqrt{\sigma^2/n} < Z_{(1-\alpha/2)}] = 1 - \alpha$ $\qquad Prob[abs(\bar{X}(n) - \mu)/\sqrt{S^2(n)/n} < t_{(n-1),(1-\alpha/2)}] = 1 - \alpha$

 END OF EXAMINATION