**MSIT 431** **Final Exam**

**due at 8:00 am on Wednesday, December 7, 2011**

**The exam is an individual effort with no collaboration.**

Each question is worth 10 points. Errors from one problem to the next will not propagate, if you use in each problem the values of the parameters provided to you as possible answers to previous problems. These values are **not** equal to the actual answers.

Solve and turn in ONLY 10 problems. If you turn in 11 you will lose the highest scoring one.

It should take an average of 5:45 hours (with probably a large variance!)

1. The following table provides the joint probabilities P(X=k, Y=j) of the values that two random variables X and Y take (X takes the five values k=0, 1, 2, 3, 4 and Y takes the seven values (j=0.5, 1.5, 2.5, 3, 3.5, 4.5, 5.5). In this problem the Y represents an observation of the unknown variable X. Note that the values of Y are on the horizontal cells.

| Y=j | 0.5 | 1.5 | 2.5 | 3.0 | 3.5 | 4.5 | 5.5 | P(X=k) |
|---|---|---|---|---|---|---|---|---|
| X=k | | | | | | | | |
| 0 | 0.05 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 1 | 0.03 | 0.05 | 0.1 | 0.02 | 0.0 | 0.0 | 0.0 | |
| 2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | |
| 3 | 0.0 | 0.0 | 0.0 | 0.02 | 0.1 | 0.05 | 0.03 | |
| 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.05 | |
| P(Y=j) | | | | | | | | 1.0 |

Complete the table (the extra row and column shown) by finding the marginal probabilities of X and Y (P{X=k} and P{Y=j} when not considered together) and show that the means of X is exactly 2 and of Y is exactly 3. Find the variances, and standard deviations of X and Y.

2. In problem 1 find the correlation coefficient $r_{XY}$ between X and Y. (First find $R_{XY}$, then $C_{XY}$ and then use the relation $C_{XY} = r_{XY}\sigma_X\sigma_Y$ to find $r_{XY}$)

3. In problem 1, assume that the standard deviations of X and Y are found to be $\sigma_X = 1.2$ and $\sigma_Y = 1.3$ (these are close but not exact) and assume that the correlation coefficient you found in problem 2 is equal to $r_{XY} = 0.9$ (again close but not exact): find the regression for estimating **X from Y** and determine the variance of the residual error.

4. In order to find the drop probability of packets in a channel, we collected data by testing n=1,600 packets and observed that 400 were dropped. Let p denote the true probability we are looking for. Estimate the value of p and provide an error margin, m, at a confidence level C=98%. You may use the conservative value for the standard deviation, $\sigma$, of a single observation.

5. What would be the confidence level of our estimate in problem 4, if we wanted a **one-sided** error margin on the estimate equal to the margin of error found in that problem. Please indicate if this is a lower or upper one-sided margin of error.

6. In problem 4, what should be the number, n, of packets to be tested in the sample (instead of 1,600) so that we can have an error margin of 0.01 with 98% confidence level?

7. For the same data collected in problem 4, we wish to test which of the following hypotheses is true: $H_0$ (which assumes that $p = p_0 = 0.22$) or $H_1$ (which assumes that $p = p_1 > 0.22$). Determine a test that verifies that hypothesis $H_1$ is true at significance level $\alpha = 0.02$. Does the data (we counted 400 drops) that was provided in problem 4 support this hypothesis?

   If we counted 360 failures, would you decide in favor of $H_0$ or $H_1$?

   **NOTE:** In problem 7 you have exactly the same set-up as the coin problem that will be solved tomorrow (12/03/2011) in class for case (a): $p_1 > p_0$

8. Messages arrive to a processing center satisfying the Poisson assumption with arrival rate of $\lambda = 8$ messages/minute. They need to receive 60 messages before they are processed. What is the probability that the processing begins within 10 minutes?

   **HINT**: Use the Central Limit Theorem.

9. A channel with capacity 250 MGbps is serving n users. The maximum download speed is R=2 MGbps for each user. Each user uses the channel in an On-Off mode with 20% of the time being ON and using the maximum allowed speed. Find the maximum number of users that can be served if the probability of exceeding the channel capacity is to be less than 2%.

   **HINT**: This is an example of statistical multiplexing which you considered in homework 8, but without the different classes of users!

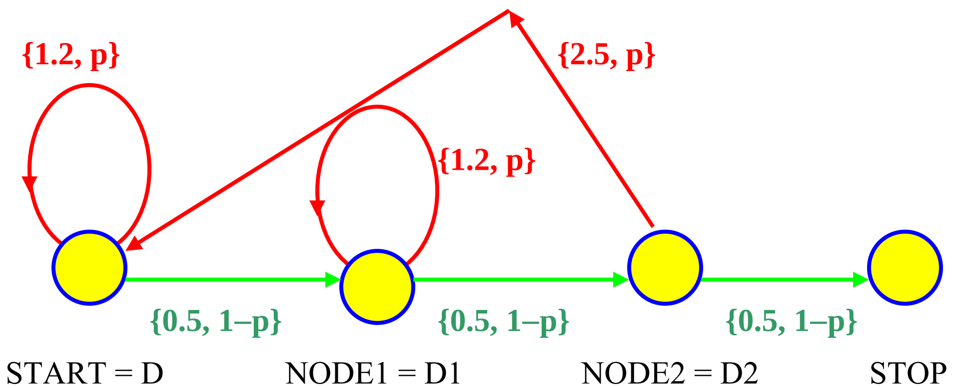
**10.** A network node with a large buffer to hold the queue of waiting messages may be connected to **ONLY one** of two sources. The node has a departure rate of $\mu$ messages per second. The sources have different arrival rates, where $\lambda_i$ is the rate for source #i such that $\lambda_1 = 0.75\mu$ and $\lambda_2 = 0.2\mu$. Let the events $A_i$ be defined by:

   $A_i$ = {node is connected to source #i}, with $P(A_1) = 0.1$ and $P(A_2) = 0.9$

   Define X as the number of messages in the queuing buffer at the node.

   Find the following probabilities: $P(X > 4)$ and $P(A_1|X > 4)$

11. A packet is being transmitted via three successive links, where at each link there is probability p of the transmission being **unsuccessful**. At each link, it takes 0.5 ms for successful transmission to the next node, while if unsuccessful the time required before retransmission is 1.2 ms for the first and second nodes, while it takes 2.5 ms for the third node. Note that when transmission is not completed at the third node the feedback goes all the way back to start, which needs to start over. The diagram is as shown:




Assume the **average** delay from START (to STOP) is D, from NODE1 (to STOP) is D1 and from NODE2 (to STOP) is D2. Write three equations involving the three average delays so that we **can** solve for D. Solve for p=0, p=0.1, and p=0.2.