Final for CpE/EE 345 – Modeling and Simulation Stevens Institute of Technology Spring 2004	Exam number:

This final is open book/open notes. PCs or calculators may be used to solve problems or to lookup information in your notes for the course, but electronic communications with others in the class or outside is prohibited.

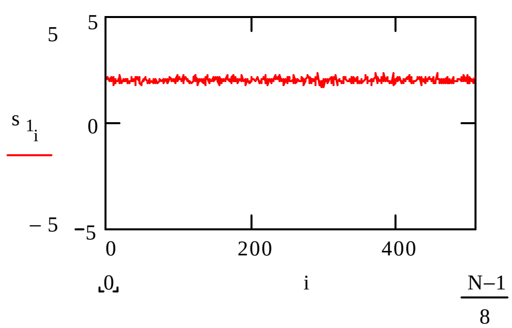
Total value is 100 points (30% of course grade). All questions are equally weighted. Do any 10 of the 12 question. Do more than 10 for extra credit. Some question can be answered in more than one way. Only one answer is required, but extra credit will be given for identifying and explaining alternate answers. Some questions ask for N answers. Extra credit will be given for more than N answers.

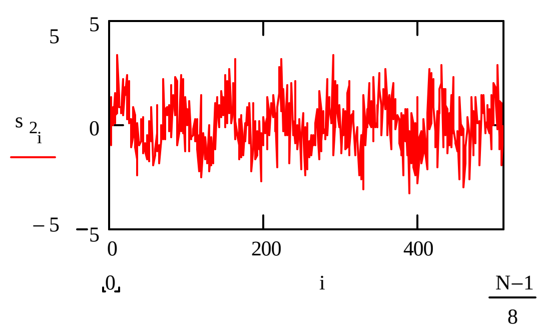
For ***most*** students in the class, this is a ***NO-RISK*** final. If your grade on this exam is lower than the projected grade assigned, that grade will be used in determining your grade for the course. This is ***not*** a no-risk final for students who did not take the midterm, who did not turn in the course project in time, or had a projected course score less than 60 points.

Be sure to pledge your exam booklet and put your name on it.

Also, you must write the exam number on the booklet – use the *Seat Number* space to write this in.

(1) You observe random behavior of the signals shown below at one particular point in an electronic system under two different operational conditions. Characterize the difference in the random behavior between these signals with respect to at least two of various statistical measures we have discussed in this class.





Here is how these two signals were generated with Mathcad:

$\mu_1 := 2 \quad \sigma_1 := .1 \qquad N := 4096$

$s_1 := \text{rnorm}(N, \mu_1, \sigma_1)$

$\mu_2 := 0 \quad \sigma_2 := 1$

$f_2 := 60 \qquad A_2 := 1$

$i := 0..\, N - 1$

$$p2_i := A_2 \cdot \sin\left(2 \cdot \pi \cdot f_2 \cdot \frac{i}{N}\right)$$

$s_2 := \text{rnorm}(N, \mu_2, \sigma_2) + p_2$

In the case of Condition A, the signal has a gaussian distribution with a non-zero mean and a small variance. In the case of Condition B, the signal also has a gaussian distribution, but in this case the mean is zero. The gaussian signal has been imposed on a sinusoidal 60 Hz signal. If we were to calculate the correlation of the resulting signal, we would expect time correlation in the aggegrate signal.
[5 pts for recognizing the difference in variance]
[5 pts for identifying the presence of correlation]
[5 pts for identifying the difference in mean value]
[5 pts for any other valid comparison I haven't listed here]

(2) Referring to the signal conditions shown in Question (1), describe how you would define a model to represent these signals in a simulation. Assume that you were given a table of values of the signals at various points in time. You do not have to define the parameters of the model, but rather how you would go about defining the overall model and finding the parameters.

The signal with the 60 Hz component will be a little tricky, you don't want the 60 Hz component to swamp the other parameters of the model, so you will need a way to eliminate that from the signal. Extra credit if you [5 pts] identify this issue and describe a workable way to deal with it. The best thing to do is to find the correlation of the signal and build a signal model to represent the correlation component. You will need to estimate the phase of the periodic component and adjust it to match the original signal. Next, you subtract this component from the original signal and model what remains, hopefully a signal with no correlation.
For signal A, and for signal B with the correlation removed, the steps are as follows:
[3 pts] identify the distribution, you can use the histogram or Q-Q plots to do this
[3 pts] estimate the parameters of the distribution but calculating them from the data
[4 pts] use the $\chi 2$ test, or other suitable means to test the goodness of fit
[2 pts EC] for identifying the proper number of degrees of freedom in doing the $\chi 2$ test where the parameters are estimated from the data.

(3) The Snevets University movie series is presenting a 12 hour marathon showing of the “Bored of the Thing” trilogy. Students will be arriving at the showing at an arrival rate of 20 students per minute with exponentially distributed interarrival times. Before they enter the theater, on average, half the students will be stopping at the Jabba Village snack bar and gift shop to stock up on provisions for the non-stop showing. The servers at the snack bar can serve an average of 4 students per minute with an exponentially distributed service rate. If the ticket taker at the entrace to the theater can process 15 students per minute at a constant service rate, what is the average number of students waiting to have their tickets taken?

By the Poisson splitting property, the arrival rate of the students directly going to the ticket taker will be 10 students per minute (half of 20). The Jabba Village servers will be operating at 100% utilization with a service rate of 4/minute and an arrival rate of 10/minute, so their output rate will be the 4/minute they can handle. Thus, the total arrival rate at the ticket taker will be 14 students/minute (10+4). For a M/G/1 queue (exponential arrival rate because of the Poisson pooling property, arbitrary service rate, one server), the server utilization, ρ, is λ/μ, in this case 14/15=.9333. The average queue length for an M/G/1 queue is:

$$L_Q = \frac{\rho^2(1+\sigma^2\mu^2)}{2(1-\rho)}$$

In this case, since the ticket taker’s service time is constant, σ=0, so:

$$L_Q = \frac{\rho^2}{2(1-\rho)}$$

The average queue length is 6.5.

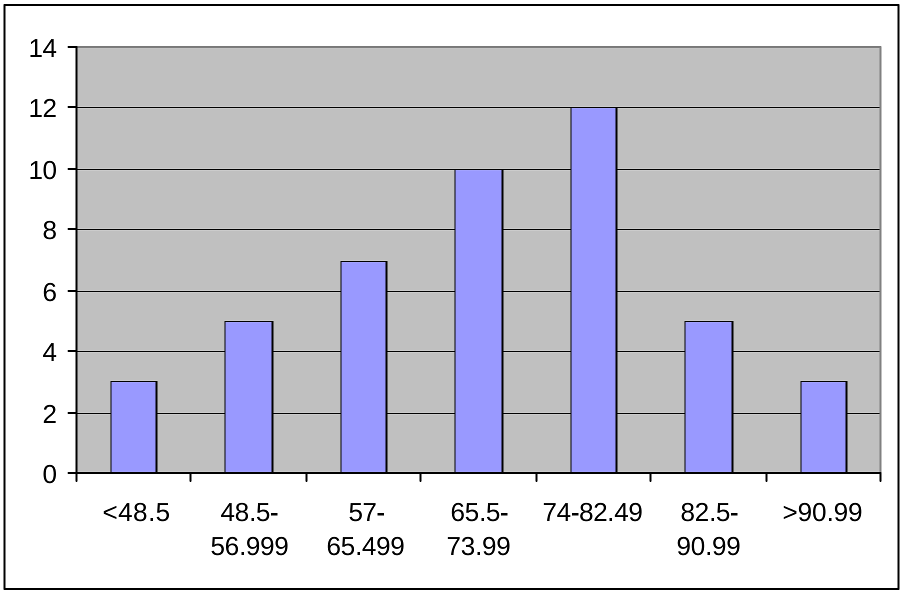
(4) One of the students waiting to get into the theater for Question (3) claims that he and all of the other students entering the theater had to wait in the ticket taker line for an average of 60 seconds. When he complained, he says the ticket taker told him that there were an average of 30 students in line. Is this reasonable?

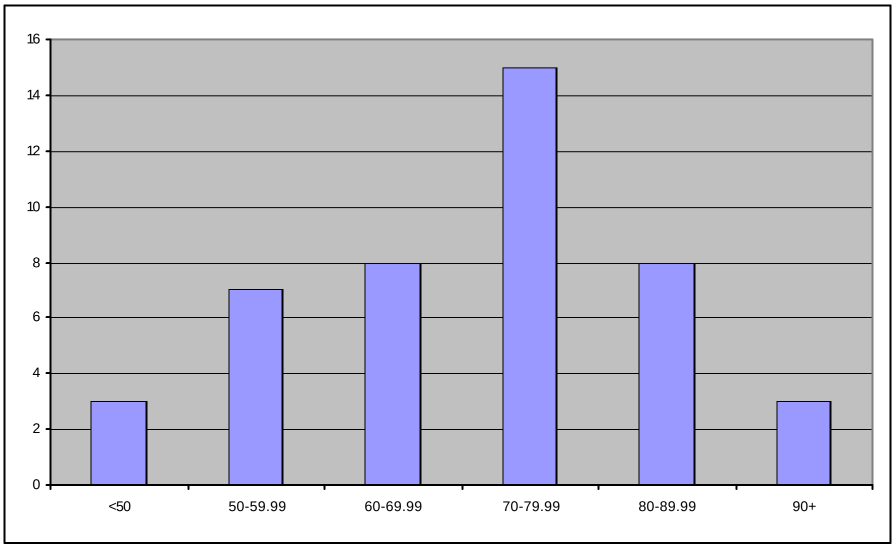
By Little’s equation, $L_Q = \lambda w_Q$, this would mean that the arrival rate at the ticket taker was 30 students per minute. This is more than the total arrival rate, ignoring the fact that the Jabba Village servers were limiting the arrival rate at the ticket taker. This is not reasonable.

(5) Prior to the Final exam in the Spring 2004 EE/PEP345 class, the projected scores for the class were as listed below (sorted according to midterm grade, from highest to lowest. A few data points have been deleted from the results). Plot a histogram of the data and suggest how you might model it in a simulation written in a language like C or C++ (e.g., what type of distribution would you use, how would you generate random values in your simulation?)

95.7	95.0	101.7	83.1	80.1	87.8	81.2
74.5	89.1	82.5	67.7	71.1	76.6	67.4
81.2	69.2	78.5	76.4	73.9	77.3	83.3
74.6	53.7	50.3	56.1	71.5	78.1	60.7
75.8	72.8	72.5	71.8	59.7	57.4	77.5
66.8	58.9	60.8	65.0	45.5	68.1	40.6
41.8	55.6					

First, notice that there are 44 values in the table. Sqrt(44) = 6.63, so we should use 6 or 7 bins in the histogram. Conveniently, the values range from about 40 to about 100, so we can use 6 bins, each 10 points wide, alternatively, we could use 7 bins, each about 8.5 points wide. The resulting histograms are shown here:





In both cases, the histogram looks rather Gaussian, which is the typical assumption made for grade distributions – "the Bell-shaped curve." To generate appropriate random values in a simulation, we would have to estimate the mean and standard deviation from the data. In this case, the mean value is 71.0 and the standard deviation is 13.91.

To generate Gaussian random values in a simulation program written in C or C++, it is very likely that there would be no built-in function to generate the random values directly. Instead, using the formula based on the Direct Transform technique, pairs of uniformly distributed random values could be generated with a built-in function like rand( ) or another linear congruential technique, and the uniform random numbers transformed to Gaussian values with the formula:

$$Z_1 = \sqrt{-2\ln(R_1)}\cos(2p R_2)$$

$$Z_2 = \sqrt{-2\ln(R_1)}\sin(2p R_2)$$

(6) I used the data in question (5) (x_A), along with the midterm scores (x_B) and calculated a function *R( )* using the following formula:

$$M(x) = \frac{1}{K}\sum_{i=1}^{K} x_k$$

$$R(x_A, x_B) = \frac{M\left((x_A - M(x_A)) \bullet (x_B - M(x_B))\right) \bullet (K-1)}{\left(\left(M(x_A \bullet x_A) - K \bullet M(x_A)^2\right) \bullet \left(M(x_B \bullet x_B) - K \bullet M(x_B)^2\right)\right)^{.5}}$$

The value calculated was .63. What might this suggest? Extra credit: Can you suggest a better way of analyzing the results?

M(x) is the mean value of the random variable x. R(x,y) is the correlation between x and y. In this case, the correlation between the midterm grades and the total score for the class, which was made up of the midterm grade, the homework grade and the project grade, was .63, which suggests a high correlation.

[EC] – A more reasonable way to analyze the results would be to remove the midterm from the total score and calculate the correlation between the midterm and the other course work. In this case, the correlation is .11, which might suggest that the values are not correlated with each other.

(7) The synchronization technique for a communications system transmits alternating 1's and 0's during the sync bit. The receiver must receive at least 8 of the last 10 bits correctly to decide that the system is in sync. If the channel bit error rate (probability of a bit being in error) is 10^{-1}, what is the probability that the receiver will declare a sync loss when the receiver is actually in sync? Extra credit: If the system is actually not in sync, what is the probability that the receiver will declare it to be in sync? State any assumptions you make in either case.

First, assume that the channel errors occur independently. Actually, this assumption is not always valid – for wireless systems, fading may occur, which tends to generate correlated errors. Ignoring this case, we can model this as a Bernouli process, where the probability of failure is 10^{-1} and the probability of success is $1-10^{-3}$.

The number of successes in a Bernouli trial is:

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

In this case, we must have 8, 9, or 10 successes (successful decoding of the sync bit) for the system to be in sync. This is:

$$P(valid\_sync) = \sum_{x=8}^{10} \binom{n}{x} p^x q^{n-x}$$

In this case the probability of the receiver declaring the link to be in sync is 0.93.

Extra credit: In this case, the receiver is looking at random data, so the probability of success for any 1 decision is 50%. Applying the same formula, the probability that the receiver will declare an in-sync condition where one does not exist is 0.055

(8) One of the examples cited for Homework #1 dealt with the number of drops of rain hitting a 12" (diameter) horizontal metal plate. Based on the material we have covered in this class, describe how you would calculate or simulate the instantaneous vertical force exerted on the plate by the rain. Assume: a 1 inch per hour rainfall, the average volume of a raindrop is .008 cubic inch (a 1/8th inch sphere), and there is an average layer of water on the plate 1/8th inch thick. State any assumptions you would make in approaching this problem.

First, it is reasonable to assume an exponential arrival rate for raindrops. Using the volume of an average raindrop and the volume of water deposited on the plate per hour, we can calculate an average drop arrival rate. In this case, l=3.927/sec. Next, we have to determine the force exerted by each drop of rain. We would need to estimate the velocity of the water drop. The best way I can think of to do this is to guestimate the vertical velocity based on the angle and horizontal velocity of a rainfall. It seems that rain is falling diagonally in 30-50 mph winds, so we could assume the vertical velocity of rain is about 40 mph, or about 18 meters/sec. The next step is pure guess work: To calculate the force exerted on the plate by a single raindrop, I will assume that the kinetic energy of the falling drop is dissipated on the plate, but this gives an energy, not a force. If we assume that the force is exerted through a distance equal to the diameter of the drop, we could convert the kinetic energy into a force through a distance.

Finally, don't forget the weight of the standing water on the plate. This is a simple F=ma, knowing the volume of the water, the density of water,

and the force of gravity.
The static forces and the force per droplet can be calculated analytically, but the dynamic force of the rain hitting the plate as a function of time is a random quantity. For this, we would need to simulate the rainfall and determine the resulting forces.

(9) According to research that I may have done on the Internet, there is a very accurate analytical model that gives the approximate force exerted on the metal plate in Question (8). The formula for the force is:

$$F(A, v, t, V_d) = \frac{d \cdot v}{t} V_d + d \cdot t \cdot g \cdot A$$

where d is the density of water, v is the rainfall vertical velocity, t is the thickness of the standing water, V_d is the volume of a drop of water, and g is 9.8 m/sec^2. Is this formula valid on its face?

No, I lied – I made up this formula. And, it's not valid on its face. The end units need to be that of force, e.g., newtons. The first term has the units of momentum over length (gm/sec) not that of force. If the first term were something like $\frac{1}{2}\frac{d \cdot v^2}{t} V_d$ it would look like the kinetic energy of a water drop interacting with the plate through a distance, which is the units of force. The second term does appear to be valid on its face, it has the units of force and looks like the steady state force exerted by the 1/8th standing water on the plate. There is one other thing that looks wrong with this formula – what happened to the effect of the rainfall rate? This has to enter the force formula somewhere.
Finally, what if the thickness of the standing water went to zero? This would cause the force to become infinite.

(10) For your class simulation project, you decided to simulate the average number of pounds of grass seed consumed per hour by a flock of sparrows on the Schaeffer lawn. After gathering data about the number of seeds consumed by the average bird and the average number of birds in a flock, and typical wind and weather conditions you are running your simulation to compare the results to the actual amount of grass seed put down at 7 am each morning by the Buildings and Grounds department. Your simulation captures 9 one-hour intervals on an average day, and you have run the simulation 5 times to correspond to a typical work week. You get the following results:

	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm
Monday	57	103	88	67	12	15	54	37	41
Tuesday	86	80	62	71	10	8	57	32	28
Wednesday	110	94	81	53	18	21	19	26	12
Thursday	74	67	59	40	23	26	37	22	10
Friday	91	74	53	48	34	28	21	19	6

When asked to indicate the average grass seed consumption per hour, how would you compute the value from this data? Describe why you think this an appropriate measure. Explain any anomalies in the simulation and why they might exist.

There appears to be some time-related change in the consumption. This may, perhaps, be because the birds that have consumed seed in one hour leave less for the birds that come in the next hour. Thus, we might expect correlation between hours of one run. On the other hand, each day starts off with a different volume consumed.
It would be appropriate to pick one sample value from each day, with no adjacent sample values from the same day, thus, we might average the following hourly values: Monday-8am, Tuesday-9am, Wednesday-10am, Thursday-11am, Friday-12pm, Monday-1pm, Tuesday-2pm, Wednesday-3pm, Thursday-4pm. Or, we could randomly pick a day and hour, so long as we captured all the days and all the different hours.
While there seems to be a drop off in the consumption across the day, there is a pronounced reduction in average consumption during the 12pm and 1pm measurement. You would have to go back to the data collection phase and see if there were any unusual conditions that might have consistently caused this – perhaps students feeding the birds, so they didn't eat grass seed or students playing Frisbee on the lawn, scaring them away. The data on Wednesday afternoon suggests it could be the latter.

(11) An electronic device has what appears to be an exponentially distributed failure rate with the observed values below. Compute the mean of the distribution and compute the $\chi 2$ statistic, comparing this distribution to an exponential distribution with the same mean.

605.231	86.945	8.651	145.23	12.239
112.888	632.339	254.743	171.55	13.814
68.644	384.545	140.161	73.305	231.003
189.265	40.231	751.117	149.158	22.071
115.511	9.973	65.041	495.31	24.908

$N := 25 \qquad \lambda := .005$

$R := \mathrm{rexp}(N, \lambda)$

$i := 0..4$

$D1_i := R_i \quad D2_i := R_{i+5} \quad D3_i := R_{i+10} \quad D4_i := R_{i+15} \quad D5_i := R_{i+20}$

$D1^T = (605.231 \quad 86.945 \quad 8.651 \quad 145.23 \quad 12.239)$

$D2^T = (112.888 \quad 632.339 \quad 254.743 \quad 171.55 \quad 13.814)$

$D3^T = (68.644 \quad 384.545 \quad 140.161 \quad 73.305 \quad 231.003)$

$D4^T = (189.265 \quad 40.231 \quad 751.117 \quad 149.158 \quad 22.071)$

$D5^T = (115.511 \quad 9.973 \quad 65.041 \quad 495.31 \quad 24.908)$

$$R_S := \text{sort}(R)$$

$$\lambda_{hat} := \frac{1}{\text{mean}(R)} \qquad \lambda_{hat} = 5.204 \times 10^{-3}$$

$$k := 6 \qquad p := \frac{1}{6}$$

$$i := 0..\,k-1 \qquad a_i := \frac{-1}{\lambda_{hat}} \cdot \ln(1 - i \cdot p)$$

$$a_k := \infty \qquad E_i := p \cdot N$$

$$O := \text{hist}(a, R_S) \qquad term_i := \frac{(O_i - E_i)^2}{E_i}$$

$$a = \begin{pmatrix} 0 \\ 35.034 \\ 77.912 \\ 133.192 \\ 211.104 \\ 344.295 \\ 1 \times 10^{307} \end{pmatrix} \quad O = \begin{pmatrix} 6 \\ 4 \\ 3 \\ 5 \\ 2 \\ 5 \end{pmatrix} \quad E = \begin{pmatrix} 4.167 \\ 4.167 \\ 4.167 \\ 4.167 \\ 4.167 \\ 4.167 \end{pmatrix} \quad term = \begin{pmatrix} 0.807 \\ 6.667 \times 10^{-3} \\ 0.327 \\ 0.167 \\ 1.127 \\ 0.167 \end{pmatrix}$$

$$\chi_{sq} := \sum_{i=0}^{k-1} \frac{(O_i - E_i)^2}{E_i} \qquad \chi_{sq} = 2.6$$

mean=192.16

(12) A linear congruential random number generator has the following parameters: X0=17, m=53, c=7, a=31. What are the next three outputs the generator will produce? How many outputs will be generated before the value 63 is output?

[3 pts] The generator outputs values between 0 and m-1=52. The value 63 will never be output.
[7 pts for the three correct values]

$X_0 := 17$ $m := 53$ $c := 7$ $a := 31$

$i := 1..3$

$X_i := \mathrm{mod}\left(X_{i-1} \cdot a + c, m\right)$

$$X = \begin{pmatrix} 17 \\ 4 \\ 25 \\ 40 \end{pmatrix}$$