Final exam for CpE/EE 345 – Modeling and Simulation
Stevens Institute of Technology
Fall 2002

Remember to sign pledge in blue book

December 9, 2002

The final is open book/open notes. Total value is 30 points. All questions are equally weighted. Do any 10 of 12 questions. Do more than 10 for extra credit. Some questions could be answered in more than one way. Only one answer is required, but extra credit will be given for identifying and explaining alternative answers. For questions that can be answered with qualitative written response only, extra credit will be given for a precise mathematical answer.

1. For an FM communications system, the modulation index b_f is defined as:

$$b_f = \frac{k_f A_m}{f_m}$$

Where k_f is frequency deviation constant in kiloHertz/Volt, A_m is the modulating signal amplitude in Volts, and the modulating signal frequency is f_m in kilohertz. While the actual bandwidth of the emitted signal is a complex function, the following equation has been suggested as a simple model for the upper bound for B_T, the bandwidth that contains 98% of the transmit energy:

$$B_T = 2\left(b_f - 1\right) f_m^{\ 2}$$

Is this a reasonable model for B_T? Why or why not?

Use face validity (dimensions, extreme conditions) to check if the model makes sense.

This is not a reasonable model. First, the units for B_T don't match. b_f is dimensionless, which means that, since f_m has the units of frequency, f_m^2 will have the units of frequency squared, which is not correct for bandwidth, B_T. In addition, consider the limiting conditions as b_f goes to zero. With decreasing b_f, B_T decreases, as expected but for $b_f = 1$, B_T goes to zero. Then with $b_f < 1$, B_T becomes negative, which doesn't make any sense.

2. You are using a simulation to study the steady-state average waiting time of customers in a complex system. However, you find that, because of the simulation complexity, the computer time needed to arrive at a steady-state value is excessive, leading to poor estimates of the steady-state values. How might you improve the speed at which the simulation converges?

Use an observation of real system to predict the average queue length. Use this as a starting position for the simulation. Alternatively, you could use prior steady-state values as initial conditions for future simulations. As a third alternative, you could create a simpler, faster simulation to estimate the steady-state values and use these as starting conditions.

3. You have been asked to design an experiment to determine steady-state parameters for a system being simulated. In the time available, you could either (a) run the simulation for 25 minutes of simulated time or (b) run 5 simulations, each for 5 minutes of simulated time. Which would you generally expect to be preferable to accurately determine the steady state parameters and why?

   (b) – the successive time blocks in (a) are correlated, biasing the estimates. With (b), the simulation can be started multiple times with independent values, leading to an unbiased estimate of the parameters.

   Note: I made no statement about the settling time of the simulation. While 5 minutes might be a short time for a system with average interarrival times of 3 minutes, for a system with events happening thousands of times per second, 5 minutes is an eternity.

4. Your director of sales is claiming that fast talking sales representatives make the most money for the company. As evidence, he offers data from ten sales representatives, in terms of annual sales and talking speed. Analyze the relationship between these pieces of information to support or refute the claim.

Sales rep #	Annual sales ($K)	Talking speed (words per minute)
1	103	54
2	122	48
3	171	57
4	191	53
5	194	42
6	178	31
7	152	41
8	108	54
9	191	31
10	204	35

To **analyze** the data, we need some measure of the relationship between the two sets of data. For a very large number of points, a graphical technique might do, but there aren't enough points to draw a meaningful plot. Compute the correlation between the series.

Here is how the series was created:

Generate two sets of 10 random numbers between 0 and 1 $N := 10$

$X_1 := \mathrm{runif}(N, 0, 1)$

$X_2 := \mathrm{runif}(N, 0, 1)$

Generate 10 random numbers between 30 and 60

$R_1 := X_1 \cdot 30 + 30$

Generate a second, correlated set

$\beta := -1$

$R_2 := \beta \cdot (R_1 - 45) + X_2 \cdot 100 + 100$

$R_1 := \mathrm{floor}(R_1 + .5)$ $R_2 := \mathrm{floor}(R_2 + .5)$

$R_1 =$

	0
0	54
1	48
2	57
3	53
4	42
5	31
6	41
7	54
8	31
9	35

$R_2 =$

	0
0	103
1	122
2	171
3	191
4	194
5	178
6	152
7	108
8	191
9	204

Here we calculate the correlation:

$$E(X) := \frac{\sum_{i=0}^{\text{length}(X)-1} X_i}{\text{length}(X)}$$

$\mu_1 := E(R_1)$ $\quad \mu_1 = 44.6$

$\mu_2 := E(R_2)$ $\quad \mu_2 = 161.4$

$$\text{cov}(X, Y) := \sum_{i=0}^{\text{length}(X)-1} \frac{(X_i - \mu_1)\cdot(Y_i - \mu_1)}{N}$$

$\sigma_1 := \sqrt{\text{var}(R_1)}$ $\quad \sigma_2 := \sqrt{\text{var}(R_2)}$ $\quad \sigma_1 = 9.457$ $\quad \sigma_2 = 35.917$

$$\text{correlation}(X, Y) := \frac{\text{cov}(X, Y)}{\sigma_1 \cdot \sigma_2}$$

$\text{correlation}(R_1, R_2) = -0.561$

The result is a somewhat anti-correlated set of values (-.561). The claim is refuted.

5. The following data is observed for the input process for a system. Calculate c^2 to compare this input data to a uniform distribution on [0,2]. For this problem, you are ***not*** being asked to compare c^2 to tabulated values to determine a confidence level, only to compute the value of the statistic.

1.247	1.523	0.889	0.988	1.310	0.819	0.849	0.404	1.923	0.891
0.911	0.734	0.671	1.154	1.508	0.907	1.733	0.827	1.109	1.540
0.686	1.032	1.070	1.011	1.759	1.228	0.455	0.966	0.313	0.483
1.279	0.094	1.705	1.294	1.198	1.179	0.377	1.473	1.529	0.552
1.077	0.592	1.424	1.468	1.815	1.237	1.393	1.773	0.793	0.641
1.677	0.707	1.483	0.641	1.298	0.866	1.141	1.274	0.882	1.801
0.810	1.137	0.990	0.733	0.361	0.953	1.159	0.730	1.134	0.452
0.997	1.231	0.863	1.721	0.953	1.021	1.353	1.276	0.322	1.865
1.156	1.554	1.588	0.980	1.304	1.285	0.852	1.404	0.932	0.406
1.135	0.808	0.985	1.064	1.221	0.896	1.529	1.113	1.370	1.392

Here's how the data was generated – with a triangular distribution on [0,2]

$N := 100$

$R := \mathrm{runif}(N, 0, 1)$

$j := 0 .. N - 1$

$X_j := \mathrm{if}\left[R_j > .5, 2 - \sqrt{2\cdot(1 - R_j)}, \sqrt{2\cdot R_j}\right]$

Here's the data:

$X1^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.247	1.523	0.889	0.988	1.31	0.819	0.849	0.404	1.923	0.891

$X2^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.911	0.734	0.671	1.154	1.508	0.907	1.733	0.827	1.109	1.54

$X3^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.686	1.032	1.07	1.011	1.759	1.228	0.455	0.966	0.313	0.483

$X4^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.279	0.094	1.705	1.294	1.198	1.179	0.377	1.473	1.529	0.552

$X5^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.077	0.592	1.424	1.468	1.815	1.237	1.393	1.773	0.793	0.641

$X6^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.677	0.707	1.483	0.641	1.298	0.866	1.141	1.274	0.882	1.801

$X7^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.81	1.137	0.99	0.733	0.361	0.953	1.159	0.73	1.134	0.452

$X8^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.997	1.231	0.863	1.721	0.953	1.021	1.353	1.276	0.322	1.865

$X9^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.156	1.554	1.588	0.98	1.304	1.285	0.852	1.404	0.932	0.406

$X10^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.135	0.808	0.985	1.064	1.221	0.896	1.529	1.113	1.37	1.392

Here is the c^2 computation.

$X_s := \mathrm{sort}(X)$

$k := 10 \quad p := \frac{1}{k} \qquad p = 0.1$

$i := 0 .. k - 1 \qquad a_i := \frac{2}{k}\cdot i$

$a_k := 2 \qquad E_i := p\cdot N$

$O := \mathrm{hist}(a, X_s) \qquad \mathrm{term}_i := \frac{(O_i - E_i)^2}{E_i}$

a =

	0
0	0
1	0.2
2	0.4
3	0.6
4	0.8
5	1
6	1.2
7	1.4
8	1.6
9	1.8
10	2

O =

	0
0	1
1	4
2	7
3	9
4	23
5	17
6	17
7	12
8	6
9	4

E =

	0
0	10
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10

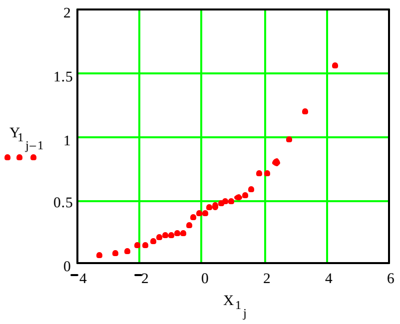
term =

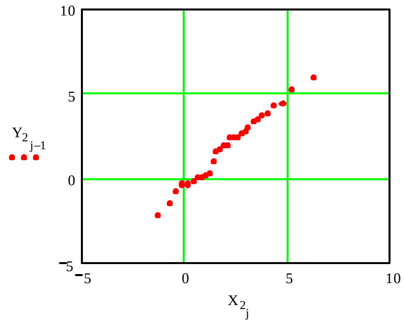
	0
0	8.1
1	3.6
2	0.9
3	0.1
4	16.9
5	4.9
6	4.9
7	0.4
8	1.6
9	3.6

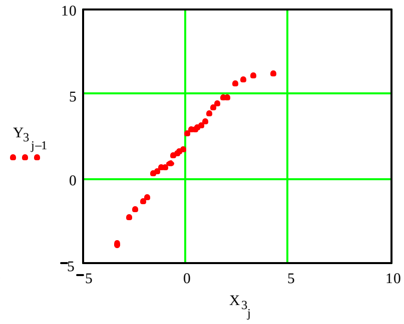
$$\chi_{sq} := \sum_{i=0}^{k-1} \frac{(O_i - E_i)^2}{E_i} \qquad \chi_{sq} = 45$$

With this large a value of c^2, this is not likely to be a uniform distribution.

6. Examine the three plots below. Y_1, Y_2, and Y_3 are derived from input processes to a system under study. X_1, X_2, and X_3 are generated from the inverse c.d.f. of a normal distribution with $s = 2$ and $m = 0$, 2, and 0, respectively. What is the significance of these plots and what conclusions can you draw from them?







These are quantile-quantile plots. The first and third plots demonstrate that the distributions being compared are unrelated. The first, because the slope is not constant, the third because the intercept is not 0. The middle plot shows a nearly constant slope with a zero intercept, suggesting that the second distribution is normal with $s = 2$ and $m = 2$. Here is how the plots were generated:

$N := 30 \quad \mu := 2 \quad \sigma := 3$

$R_1 := \mathrm{rexp}(N, \mu)$ $\quad$ $R_2 := \mathrm{rnorm}(N, \mu, \sigma)$ $\quad$ $R_3 := \mathrm{rnorm}(N, \mu, \sigma)$

$Y_1 := \mathrm{sort}(R_1)$ $\quad$ $Y_2 := \mathrm{sort}(R_2)$ $\quad$ $Y_3 := \mathrm{sort}(R_3)$

$\mathrm{mean}(R_1) = 0.501$ $\quad$ $\mathrm{mean}(R_2) = 2.066$ $\quad$ $\mathrm{mean}(R_3) = 2.308$

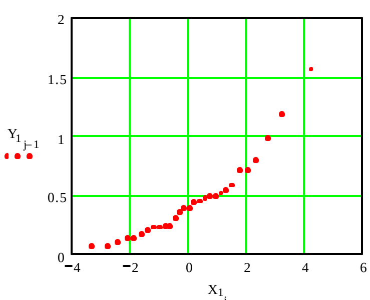
$\mathrm{var}(R_1) = 0.179$ $\quad$ $\sqrt{\mathrm{var}(R_2)} = 2.287$ $\quad$ $\sqrt{\mathrm{var}(R_3)} = 2.7$

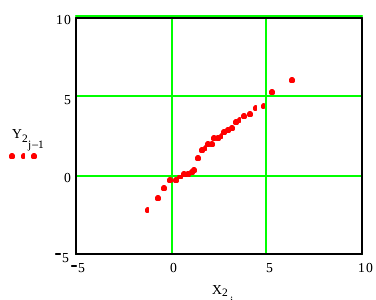
$j := 0 .. N - 1$

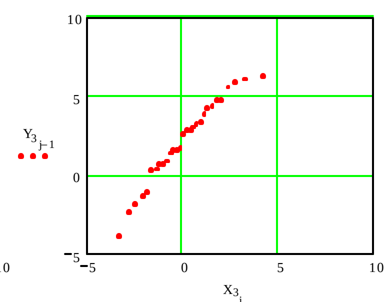
$$\gamma_j := \frac{j + 1 - \frac{1}{2}}{N}$$

$X_{1_j} := \mathrm{qnorm}(\gamma_j, 0, 2)$ $\quad$ $X_{2_j} := \mathrm{qnorm}(\gamma_j, 2, 2)$ $\quad$ $X_{3_j} := \mathrm{qnorm}(\gamma_j, 0, 2)$

$\mu - 2 = 0$







7. Plot a histogram of the following random data and identify the type of distribution that generated it:

1.221	0.841	0.904	0.792	0.952	0.817	0.726	1.723	1.097	0.707
0.977	1.254	0.668	0.928	0.580	0.525	1.030	0.523	0.436	1.316
0.335	0.504	1.298	0.328	1.129	1.454	0.454	1.442	1.442	0.727
1.569	1.003	0.797	0.735	1.243	1.195	0.872	1.347	0.945	1.181
1.036	0.936	1.144	0.927	0.981	1.488	0.704	0.670	1.330	1.211
0.612	1.441	0.483	0.273	0.856	1.006	0.714	1.328	0.943	0.843
1.030	0.771	0.577	1.096	0.227	1.827	0.829	0.669	0.852	0.720
0.941	0.947	0.457	0.748	1.395	1.169	1.350	0.721	1.048	0.326
1.100	0.902	1.593	1.620	1.133	0.991	1.012	1.018	1.390	0.252
1.652	1.018	0.966	0.492	0.596	0.652	0.913	1.470	1.050	0.669

There should be sqrt(N) bins, where N is the number of samples. This data is generated by a triangular distribution, so the histogram should look reasonably triangular. If the distribution is identified as normal or gaussian, this would also be accepted, since there isn't much in the way of tails of the distribution to tell the difference. If you look closely at the data, however, the telling point is that the data is strictly bounded by [0,2]. A normal distribution would likely have negative values or a few outliers.

Here is how the data was generated:

Generate a triangular distribution from 0 to 2 with a mode at 1, generate 100 poin

$N := 100$

$R := \mathrm{runif}(N, 0, 1)$

$j := 0 .. N - 1$

$X_j := \mathrm{if}\left[R_j > .5, 2 - \sqrt{2 \cdot (1 - R_j)}, \sqrt{2 \cdot R_j}\right]$

$i := 0 .. 9$

Here is the data:

$X1^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.221	0.841	0.904	0.792	0.952	0.817	0.726	1.723	1.097	0.707

$X2^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.977	1.254	0.668	0.928	0.58	0.525	1.03	0.523	0.436	1.316

$X3^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.335	0.504	1.298	0.328	1.129	1.454	0.454	1.442	1.442	0.727

$X4^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.569	1.003	0.797	0.735	1.243	1.195	0.872	1.347	0.945	1.181

$X5^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.036	0.936	1.144	0.927	0.981	1.488	0.704	0.67	1.33	1.211

$X6^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.612	1.441	0.483	0.273	0.856	1.006	0.714	1.328	0.943	0.843

$X7^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.03	0.771	0.577	1.096	0.227	1.827	0.829	0.669	0.852	0.72

$X8^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.941	0.947	0.457	0.748	1.395	1.169	1.35	0.721	1.048	0.326

$X9^T =$

	0	1	2	3	4	5	6	7	8	9
0	1.1	0.902	1.593	1.62	1.133	0.991	1.012	1.018	1.39	0.252

$X10^T =$

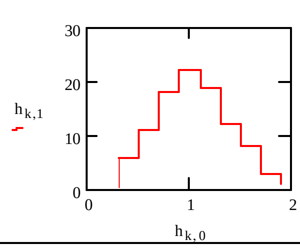
	0	1	2	3	4	5	6	7	8	9
0	1.652	1.018	0.966	0.492	0.596	0.652	0.913	1.47	1.05	0.669

And here is the histogram

$M := \mathrm{floor}(\sqrt{N} + .5)$ $\quad M = 10$

$h := \mathrm{histogram}(M, X)$

$k := 0 .. M - 1$



8. A linear congruential random number generator with parameters X_0=0, a=9, c=1, m=16 is used to generate random numbers for a simulation. Find the cycle length of this generator. Assuming that the value for m is restricted to the value given, are the other parameters for this generator chosen appropriately?

The cycle length is 16. This is the maximum length possible with m=16, so the parameters have been chosen appropriately.

This choice of parameters has been shown to be optimal for m a power of two, namely c relatively prime to m and a=4k+1, where k is an integer. Here are the states of the generator:

Linear Congruential Pseudo Random Number Generator

X0= 0
a= 9
c= 1
m= 16

B10=+MOD(B9*_a+_c,_m)

index	x(index)
0	0
1	1
2	10
3	11
4	4
5	5
6	14
7	15
8	8
9	9
10	2
11	3
12	12
13	13
14	6
15	7
16	0

9. The Third National Bank of Podunk has 6 tellers with an exponentially distributed service rate with a mean of 5 minutes. Customers arrive at the bank with exponentially distributed interarrival times with an average interarrival time of 1 minute. There is a single queue for customers to wait in, but the queue is restricted to 20 customers. What observation can you make about the departure rate of customers from the bank?

The system has a limited queue length. Utilization of the tellers is high, but less than 1 – the average utilization is $\lambda/(c*\mu)=5/6$. The effective arrival rate λ_e is reduced by customers who are turned away, so the departure rate must be reduced as well.

Extra credit for calculating the actual departure rate. This would be $l*(1-P_N)$ where N=20. Using the value for P0 from problem 16:

$$P_N := \frac{a^N}{c!\cdot c^{N-c}}\cdot P_0 \qquad P_N = 7.929\times 10^{-3}$$

$$\lambda_e := \lambda\cdot(1 - P_N) \qquad \lambda_e = 0.992$$

10. For the Third National Bank example in problem 9, what is the probability that a customer shows up at the bank and is immediately served?

This is an M/M/c/N queue. We can calculate P_0 to find the chances of no wait time.

$$c := 6 \qquad \lambda := 1 \qquad \mu := \frac{1}{5} \qquad a := \frac{\lambda}{\mu} \qquad N := 20$$

$$\rho := \frac{\lambda}{c\cdot\mu} \qquad \rho = 0.833$$

$$P_0 := \left(1 + \sum_{n=1}^{c} \frac{a^n}{n!} + \frac{a^c}{c!}\cdot \sum_{n=c+1}^{N} \rho^{n-c}\right)^{-1} \qquad P_0 = 4.691\times 10^{-3}$$

11. How many up/down runs are there in the following sequence of 30 uniformly distributed random numbers?

Numbers	Random values									
1-10	0.145	0.327	0.598	0.926	0.710	0.072	0.015	0.997	0.305	0.135
11-20	0.187	0.377	0.677	0.708	0.123	0.731	0.228	0.134	0.457	0.612
21-30	0.613	0.289	0.858	0.157	0.963	0.584	0.068	0.398	0.734	0.597

There are 8 up runs and 8 down runs.
Here is how the numbers were generated:

$N := 30$

$RUN := \mathrm{runif}(N, 0, 1)$

$i := 0..9$

$RUN1_i := RUN_i \qquad RUN2_i := RUN_{i+10} \qquad RUN3_i := RUN_{i+20}$

$RUN1^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.145	0.327	0.598	0.926	0.71	0.072	0.015	0.997	0.305	0.135

$RUN2^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.187	0.377	0.677	0.708	0.123	0.731	0.228	0.134	0.457	0.612

$RUN3^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.613	0.289	0.858	0.157	0.963	0.584	0.068	0.398	0.734	0.597

And here is how they were counted:

```
upruns(r) := | for i ∈ 0.. length(r) − 2
             |    x_i ← if(r_{i+1} > r_i, 1, 0)
             | k ← 1
             | ups ← if(x_0 = 1, 1, 0)
             | while (k < length(x))
             |    | ups ← if[(x_{k−1} = 0) ∧ (x_k = 1), ups + 1, ups]
             |    | k ← k + 1
             | return ups

downruns(x) := upruns(−x)

upruns(RUN) = 8
downruns(RUN) = 8
```

12. Customers calls arrive at a software support center with exponentially distributed interarrival times with an average time between arrivals of 15 minutes. What is the probability that 5 customers will arrive between 1 and 2:30 p.m.?

Customer arrival process is exponential with arrival rate λ=4/hour. This means that the number of customers arriving in a given time period is Poisson distributed.

$$P\left[N(t)-N(s)=n\right]=\frac{e^{-l\,(t-s)}\left[\,l\,(t-s)\right]^{n}}{n!}$$

The probability of 5 arrivals in 1½ hour is:

$$\frac{e^{-6}\left[6\right]^{5}}{5!}=\frac{.0025*7776}{120}=.1606$$