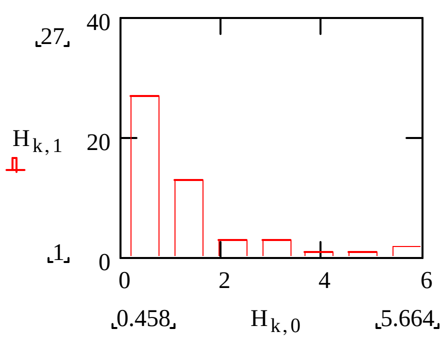
Final for CpE/EE/PEP 345 – Modeling and Simulation
Stevens Institute of Technology
Fall 2003

The final is open book/open notes. Total value is 100 points (30% of course grade). All questions are equally weighted. Do any 10 of the 12 questions. 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 additional problem sections are identified as extra credit problems.

1. Plot a histogram of the following data and identify the distribution that generated it. Find the mean and variance.

Numbers	Random values									
1-10	0.503	1.018	0.024	0.110	0.048	1.601	1.034	1.234	2.306	0.343
11-20	0.398	5.153	0.503	1.236	0.292	0.092	0.501	0.243	0.611	0.930
21-30	2.283	3.595	1.351	1.476	0.148	3.044	2.879	2.465	0.038	0.129
31-40	0.908	0.418	0.056	6.098	3.007	0.863	0.504	0.881	0.473	5.328
41-50	0.930	1.149	0.158	1.621	0.757	0.836	0.870	0.544	1.086	0.050

[2.5 pts] There are 50 points in the data set. There should be ~7 bins in the histogram. It will look like this:



[2.5 pts] The distribution is exponential.
[2.5 pts] The mean is: 1.243
[2.5 pts] The variance is: 1.932
[5 pts extra credit] for computing chi-squared, comparing this to an exponential, or other method of verifying distribution

2. A random number generator produces the following set of pseudo random numbers.
   a. How many runs are there in this set of data?
   b. How does this compare with the expected number of runs for a uniform distribution?

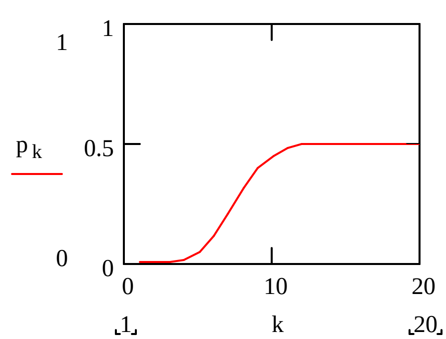
Numbers	Random values									
1-10	3.265	5.579	9.891	2.096	7.363	3.653	5.726	5.655	8.463	1.933
11-20	9.520	5.265	9.270	4.288	6.937	5.566	3.518	8.121	8.723	5.175
21-30	4.979	4.411	6.018	8.530	1.087	1.366	3.304	7.368	2.198	8.711
31-40	2.965	3.820	6.462	9.600	4.482	3.134	6.332	1.149	9.250	3.020
41-50	9.692	4.596	4.018	6.629	3.610	9.088	3.954	5.143	4.715	2.089

[5 pts] There are 36 runs in this data
[5 pts] The expected number of runs is: 33 with a variance of: 2.927
[5 pts extra credit] At a 5% significance level, the test statistic would be: 1.96, Z_0 for this set of data is 1.025, so the distribution appears to be uniform

3. Able and Baker have been hired by ARA to operate the Java City concession in the Burchard lobby. They have decided to split the job of servicing customers. Able prepares orders and Baker operates the cash register. Able can service 2 customers per minute while Baker can service 4 customers per minute, both with exponentially distributed service times. If the average customer inter-arrival time is exponentially distributed with a mean of 25 seconds, what is the probability that more than 15 customers will be waiting in line?

> [10 pts] 1.0. The fact that the two servers are operating in tandem means that each must be capable of serving the total load. In this case, Able is the bottleneck in the system. He cannot service the customer load, since his utilization needs to be more than one (25 second interarrival time with 30 second average service time). This means that the average queue length grows without bound, so the probability of more than N customers waiting for any finite N is 1.0.

4. NSA (No Such Agency) needs to generate truly random binary numbers to be used for the encryption keys used in their secure communications systems. They have decided to base their random numbers on counting the number of clicks, C, emitted by a Geiger counter during a time interval T, monitoring background cosmic radiation – if C is odd, a "1" will be generated, if C is even, a "0" will be generated. If the number of radioactive particles detected by the Geiger counter is a Poisson process with mean arrival rate λ, find an expression for the probability that C is odd. If it makes things simpler, you may assume λ = 3/second and T=5 seconds. (Extra credit: How could you design a truly random generator for your PC using a related technique, not based on radioactive decay but some other naturally occurring process that the PC could readily gain access to)

> [10 pts] For a poisson process, the probability that the number of arrivals in an interval t, N(t), is equal to n is:
>
> $$P[N(t)=n] = \frac{e^{-lt}(lt)^n}{n!} \quad \text{for } t \geq 0 \quad \text{and } n = 0,1,2,\ldots$$
>
> So if we are interested in the probability, p, that n is odd, we can sum the probabilities over the range of all odd numbers:
>
> $$p = \sum_{i=1}^{\infty} \frac{e^{-lT}(lT)^{2i-1}}{(2i-1)!}$$
>
> You could evaluate the infinite summation to see if it converges – obviously, the summation has to converge: for any finite λT, the numerator is increasing by a fixed λT^2 for each term, while the denominator is increasing by factors of i at each term. If we compute a few terms in the series, using the suggested values for λ and T, we find that the summation converges to .5 with less than λT=15 terms. Here is the estimate of p with various number of terms in the series:
>
> 
>
>
> [5 pts extra credit] Count the number of disk accesses or key strokes during an interval of time and use the evenness/oddness to determine a bit.

5. A linear congruential random number generator with m=53, a=17, and c=2 generates the following sequence of random numbers: {15,45,25,3}. Compute the next value in the sequence.

   > [10 pts] The next value is 0.

6. Two adjacent components in an electrical device have failure rates that seem somehow associated with each other. After testing 10 lots of circuit boards that have failed in service, the following data are observed:

	Lifetime in thousands of hours									
Lot #:	1	2	3	4	5	6	7	8	9	10
Comp#1	0.68	0.67	0.52	0.35	0.88	0.75	0.59	0.82	0.34	0.02
Comp#2	1.11	1.14	1.05	0.71	1.85	1.68	1.10	1.80	0.79	0.87

   Compute the correlation between lifetimes of Component 1 and Component 2 to determine if there is an association.

   > [10 pts] The correlation is 0.823, which suggests that there is an association between the failures of these two components.

7. Customer service calls arrive at Macrofirm's customer help line according to a Poisson process with an average of 20 calls per hour. What is the probability that 30 calls arrived between 11:30 am and 1:00 pm when the customer service agent took their lunch hour? If we know that 10 calls arrived between 11:30 am and 12:00pm, what is the probability that 20 calls arrived between 12:00 pm and 1:00 pm?

   > [5 pts] The number of calls arriving in a given interval is Poisson. We can calculate the probability directly from $P\left[N(t)=n\right]=\frac{e^{-lt}(l\,t)^n}{n!}$ where n=30, l=20 and t=1.5. In this case, the probability is .073.
   > [5 pts]We know that a given number of calls arrived in a subinterval, we have to compute the probability for the remainder of the interval. In this case, the problem becomes, "what is the probability that 20 busses arrive with t=1, and l=20?" This probability is .089

8. You are presented with a set of data that is hypothesized to be normally distributed. You want to use the χ^2 test to determine whether or not the hypothesis can be refuted. Describe each step you would go through to perform the test. For example, how would you find the expected number of values in each class interval?

   > a. [1 pt] The mean and variance of the data values are calculated.
   > b. [1 pt] The data is sorted, so that the number of values in each class can be found.
   > c. [2 pts] The number of class intervals is determined, on the order of the square root of the number of data values.
   > d. [4 pts] Since the distribution is not uniform, the class intervals are chosen so that the expected number of data values in each class is the same. Using the variance and mean calculated in step a, the inverse cumulative distribution function for a gaussian variable is used to find the class interval boundaries. The number of values in each class is the total number of data values divided by the number of classes
   > e. [1 pt] Calculate the terms of the chi-squared parameter as (E-O)^2/E and sum them
   > f. [1 pt] compare to critical value, e.g. from table.

9. How does a Quartile-Quartile plot differ from a X-Y scatter plot and what advantage (if any) does the Q-Q plot offer over a scatter plot? What do you expect to observe in a Q-Q plot to use it to make decisions?

[5 pts] A Q-Q plot displays sorted values from observed data versus the corresponding points in a distribution one is comparing to. An X-Y scatter plot merely plots one set of data against another, with no implied ordering of the data. While an X-Y scatter plot of observed data versus a known distribution may resemble a Q-Q plot, the unsorted order of the data means that corresponding sections of the two distributions cannot be directly compared to each other.

[5 pts] If the experimental data distribution matches the type and parameters of the known distribution it is plotted against, one should observe a generally straight line with a slope of 1 that intersect the axes at (0,0)

10. At Snevets University, the registrar's office has seating for 4 students and one person who fixes problems with student records. If 4 students are waiting, any additional arriving students must leave and come back later. The average service time for a student to have their registration corrected is 15 minutes. Students arrive at the rate of 3 per hour. What is the average number of students processed per hour?

[10 pts] This is a limited capacity queue, since only 4 students can be waiting. Because there are students who are turned away, the effective arrival rate is reduced and therefore, the average number of students processed per hour must be reduced from the arrival rate.

The probability that n students are waiting is:

$$P(n) = (1 - r)\, r^{n}$$

Since the server utilization is

$$r = \frac{l}{m}$$

or .75, this means that the probability of 4 students waiting is .079. When customers are turned away, the effective arrival rate is reduced according to the formula:

$$l_{eff} = l\,(1 - P(n))$$

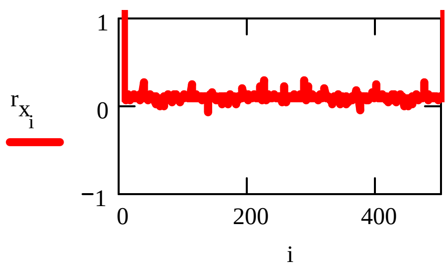
So, the effective arrival rate is 2.763. Since the server utilization is less than 1, there is no blocking in the queue, so the average departure rate is the same as the average effective arrival rate or 2.763 students per hour.

11. Two linear congruential pseudo random number generators have been designed. The first uses m=503, a=17, c=2, while the second uses m=503, a=19, c=13. You are trying to determine which is the best one to use. If X_j is the sequence of numbers generated by the generators, the following are plots of

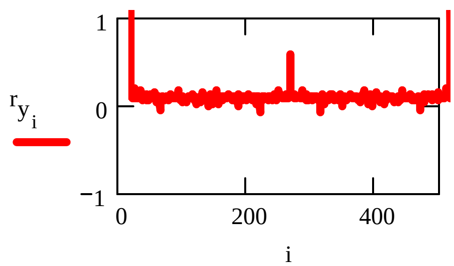
$$r_i = \frac{E(X_j X_{i+j}) - m^2}{s^2}$$

for i = 0…m. What is the physical meaning of the formula expressed above? What observations can you make about these two random number generators? What effects might you expect to see in simulations using one generator or the other?

$m_x = 503$ $a_x = 17$ $c_x = 2$



$m_y = 503$ $a_y = 19$ $c_y = 13$



[3 pts] The formula is calculating the correlation between the sequence of random number with itself with different amounts of time offset. This is the autocorrelation of the sequence.
[3 pts] The generator on the right shows an autocorrelation peak at an offset of about 250 samples. The value of correlation for this peak value is about ½ (actually .479) which is quite correlated. The generator on the left shows some autocorrelation peaks, but they are not as pronounced, the largest is value is 0.2
[4 pts] If the random numbers used in the simulation were generated by the generator at the left, the correlation in the sequence might have unexpected consequences – if the data being modeled were uncorrelated and the random generator created correlated random numbers, the validity of the simulation could not be assured.

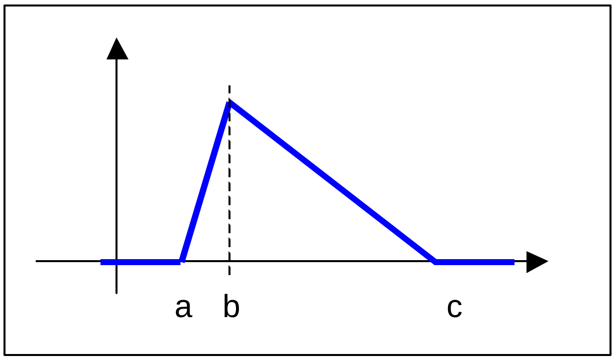
12. For a triangular distribution as shown below, we know that the mean of the distribution is:

$$m = \frac{(a+b+c)}{3}$$

In searching the Internet, we might find a formula for the variance of the triangular distribution:

$$s^{2} = \frac{(a^2 - b^2 + c^2 - ab + ac - bc + a - b + c)}{18}$$

Is this expression valid on its face?



[2 pts] The is not valid on its face.

[4 pts]First, the units of a, b, c, and σ should be the same. Here, the first 6 terms match, but the last three terms don't. There need to be a factor on these last three terms in the units of a/b/c to be able to add them to the squared and product terms.

[4 pts] Second, consider what happens when a=0, and b=c. The first three terms add to zero and the 4th and 5th terms are zero, since a is zero. The last three terms also add to zero. This means that σ^2 = -bc, or σ^2 = $-c^2$. For any value of c, this means that σ is imaginary, which is not possible.