Jeremy Tejada
ISE 441 - Introduction to Simulation

Lesson Plan for Simulation with Spreadsheets (8/31/11 & 9/7/11)

Learning Outcomes:

1. Students will be able to list and define the different components of a spreadsheet simulation
2. Students will be able to determine when to build a spreadsheet simulation and when to use Arena
3. Students will be able to construct spreadsheet simulations to model a variety of problems
4. Students will be able to analyze and interpret the results of spreadsheet simulation models

8:05-8:20 AM: In Class Quiz #1 and Introduce Spreadsheet Simulation

In Class Quiz # 1

Spreadsheet Simulation can be used to model SIMPLE static and dynamic simulation problems
        Static – Time does not play a role
        Dynamic – Time plays a critical role in the problem and in computing performance measures

In the next two class periods we will construct two spreadsheet simulation models
        1. A static spreadsheet simulation model of a single period inventory problem
        2. A dynamic spreadsheet simulation of a M/U/1 Single Server Queue

Components of a Spreadsheet Simulation Model:
        1. Inputs – Random variables that describe the distribution of parameters such as time between arrivals, service times, or costs
            a. Either given in the problem statement or data must be collected and a distribution must be determined base on that data
        2. Spreadsheet Logic – Formulas that are required in order to manipulate input variables in order to describe different aspects of the problem
            a. The most difficult component to simulation modeling with spreadsheets
            b. Different for every problem that you encounter
            c. Combine your knowledge of the problem with your knowledge of spreadsheets to manipulate information to get the proper results
        3. Results – Generally confidence intervals on performance measures and graphs that describe performance measures of interest.

8:20 -9:30 AM: The News Vendor Problem – Solution by Spreadsheet Simulation

Problem Statement:

Rupert sells daily newspapers on the street corner
Each day he must buy the same fixed number of copies of the news paper (q)

Cost (c = $0.55 each)
Sales Price (r = $1.00 each)

He has written down his daily demand for the last year, and determined that demand for one day is independent of the next, and is a random variable ($D \sim$ NORMAL(135.7, 27.1) – Can demand be negative? [No]. How do we account for this using spreadsheet logic? [MAX function]. Is the normal distribution discrete or continuous? [Continuous]. How can we map this into demand? [ROUND function]

If $D < q$, Rupert can satisfy all customers and will have $q - D \geq 0$ papers left over
    Salvage Value ($s = \$0.03$)
If $D > q$, Rupert cannot satisfy all customers and will miss out on $D - q \geq 0$ sales
    We assume there is no "loss of goodwill" cost ($\pi = 0$)

Each day starts fresh and is independent of any other day, so this is a single period problem

Is this static or dynamic problem? [Static]. Does it matter what time customers come or just that they come? Is this model discrete or continuous? [Discrete]. Deterministic or stochastic? [Stochastic]

The question of interest: How many papers (q) should Rupert buy each morning in order to maximize his expected profit? If he buys too few he will miss out on 45 cents profit for lost sales, if he buys too many he will have scrap left over at a loss of 52 cents each. He must decide on a FIXED value of q papers to purchase every day.

Based on what I just discussed, do you think Rupert should in general order more papers than the mean demand, or fewer papers than the mean demand? [Fewer, having too many papers causes greater loss than having too few papers]

ISE 408 Aside:

Let us recall something we learned is ISE 408 about the single period inventory problem before we begin constructing the simulation model.

$$P\{StockOut\} = \frac{c-s}{r-s+\pi} = \frac{0.55-0.03}{1.00-0.03+0} = \frac{0.52}{0.97} = 0.5361$$

$$P\{D < q\} = 1 - 0.5361 = 0.4639$$

Since Demand is Normal:

$$\Phi\left(\frac{q-135.7}{27.1}\right) = 0.4639$$

Using the Normal Tables:

$$\Phi(x) = 0.4639 \Rightarrow x = -0.09$$

Thus:

$$\frac{q-135.7}{27.1} = -0.09 \Rightarrow q = 135.7 - 0.09(27.1) = 133.26 \approx 133$$

However, this is the optimal number of papers to order in a given day. However, we might want to know the expected value and variance of daily profit given we use this order quantity q. We also may wish to know if we could order 120 or 140 papers and get similar profits. Simulation will allow us to determine the answers to these types of questions.

Construction of the Spreadsheet Simulation Model:

Open the News Vendor spreadsheet in Excel so we can begin to construct a simulation model of this system

We want to determine the best value of the order quantity q, with the goal being to maximize the expected daily profit. Thus, for a number of alternative values of q, we want to find a method for determining the expected daily profit.

1. Create an area in the spreadsheet to store input parameters for the problem (done for you) and fill in this area with the appropriate values.

2. We need a method for generating demand on each day. We know that this is a normally distributed random variable with a mean of 135.7 and standard deviation of 27.1. Generally, we would need to explicitly define the inverse cumulative distribution function (Inverse CDF), and then use the method of inversion to generate normally distributed random variables (more on this for the M/U/1 spreadsheet model). However, a closed form expression for the Inverse Normal CDF does not exist, but Excel has a function that approximates the inverse normal CDF:

   NORMINV(U, μ, σ)
   > U = number between 0 and 1 that defines the probability value on the y-axis
   > μ = the mean of the normal distribution
   > σ = the standard deviation of the normal distribution
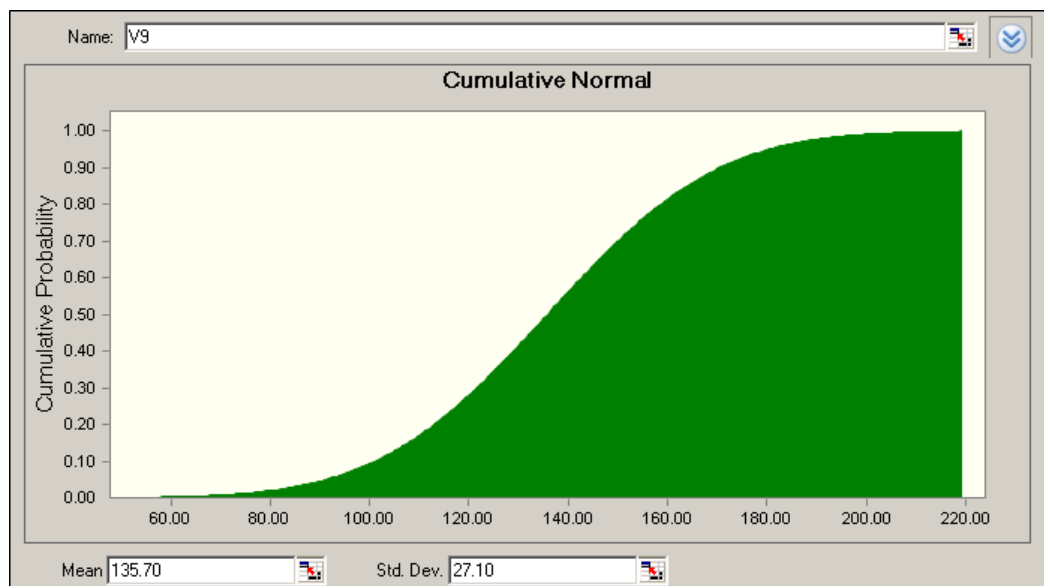
In order to generate random variates that represent demand, we need to generate uniformly random numbers between 0 and 1 and substitute them for the value of U in the above formula. Excel's function for generating uniformly distributed random numbers between 0 and 1 is:

RAND ()

Thus, out formula becomes:

NORMINV(RAND(), 135.7, 27.1)

Is this the complete formula for demand? [No]. The demand must be an integer, because a fraction a person will not come to purchase a paper. Thus, our formula becomes:

ROUND(NORMINV(RAND(), 135.7, 27.1), 0)

In addition, the normal distribution can take on all values from negative infinity to positive infinity no matter what the mean and variance of the distribution are. Thus, the final version of the formula is:

MAX(ROUND(NORMINV(RAND(), 135.7, 27.1), 0), 0)

Aside: Calculation Refresh

In order to keep excel from recalculating the demand every time you alter a cell, go to Excel Options → Formulas → Calculations → Manual Refresh (Uncheck re-calculate upon saving)

Excel will now only refresh demand random variables when you press the F9 key

3. Now we need to compute the number sold on each day. This is relatively easy. Consider both cases ($D<q$ and $D>q$), the number sold is the minimum of these two values:

Minimum(D, q)

4. We also need to compute the number of papers scrapped on each day. This is also relatively easy. Consider both cases ($D<q$ and $D>q$).

If $D < q$, Rupert will have to scrap $q - D \geq 0$ papers

If $D > q$, Rupert will not have to scrap any papers

Thus, the formula for number scrapped can be written as:

Maximum($q - D$, 0)

5. Lastly, we need to derive a formula for daily profit as a function of the order quantity q. We will call this daily profit $W(q)$. Profit = Revenue – Cost. Rupert's daily cost is easy to compute:

Cost = cq

Revenue:

If $D < q$,
    Rupert will sell D papers for r each, for a Revenue of Dr
    Rupert will have to scrap $q - D$ papers, for a Revenue of $(q – D)s$

If $q < D$,
    Rupert will sell q papers for r each, for a Revenue of qr
    Rupert will not scrap any papers

Thus, the formula for Profit = Revenue - Cost is:

$$r*(\min\{D,q\})+s*(\max\{q-D,0\})-cq$$

Now we need to insert these formula into Excel using $ in the proper place so that we can copy and paste formulas easily. Now that we have simulated 30 days of this single period inventory problem, we want to analyze the results and make an intelligent statement about how many papers Rupert should buy.

1. We start by computing the average profit over the 30 day time horizon for each value of q we chose using the AVERAGE() function.

2. We would also like to know how much variation there is around that expected value, so we compute a two-sided 95% confidence interval around the expected profit using the following formula:

$$CI = \overline{X} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

Thus, we will used the STDEV() function to compute the standard deviation of daily profit, the TINV() function to compute the appropriate t-value, and the COUNT() function to compute the sample size (n).

3. We may also be interested in the probability of losing money on any given day, which can be done using the COUNT() and COUNTIF() functions.

4. Finally, we create a graph of the mean, upper limit, and lower limit of expected profit as a function of the order quantity q. This given us a visual means for comparing the confidence intervals on E[$W(q)$].

Discussion Points:

1. Common random numbers for daily demand are used as a variance reduction technique. We want to detect the differences due to different order quantities, not due to the differences in daily demand. Thus, we compare across the same set of daily demands. You can think of this like a "blocking" factor in a statistical experiment.

2. What value would you choose for order quantity? [120-160]. Does simulation provide you with a single numerical answer, or something more valuable? We can also plug in our optimal value of 133 from analytical techniques to check the variability involved in using this "optimal" value.

3. Notice higher values of q are associated with higher levels of risk.

4. Number of Runs Calculations to Extend this spreadsheet:

$$n^* = n\left(\frac{H}{H*}\right)^2$$

9:30 -9:55 AM: In-Class Assignment #1: Problems 2-11 and 2-12 from the Textbook:

1. Increase the sample size on the news vendor problem in Section 2.7.1 for 30 days to 150 days. As in the original problem, use the same realized demands in column E for all values of q. You do not need to change the histograms at the bottom of the columns. Based on your results, what sample size (in days) would be needed to bring the maximum half-width of all five confidence intervals down to ± $1.50? Do not actually carry this out, just perform the necessary calculations and make your final answer obvious in the spreadsheet.

Submit this file to Wolfware before leaving class with In Class Assignment 1 Part 1 in the filename. The latest time to submit the file is 10:15 AM. If the file has not been submitted by then, you will receive a zero for this in-class assignment.

2. Based on the spreadsheet you created in problem 1, try to hone in more precisely on the optimal value of q. When you are finished, you should leave the 5 best values of q at the top of your spreadsheet. Keep the sample size at 150 days.

   Submit this file to Wolfware before leaving class with In Class Assignment 1 Part 2 in the filename. The latest time to submit the file is 10:15 AM. If the file has not been submitted by then, you will receive a zero for this in-class assignment.

Discussion of Homework 1 and time for answering questions

1.   Discuss the definition of time averages and explain why they are an effective average measure

A Single Server Queue – Solution by Spreadsheet Simulation

Aside: Kendall's Notation: A/B/C/K/N/D which is abbreviated A/B/C assuming (K=∞, N=∞, D=FIFO)

A = Arrival Time Distribution {M, D, U, E_k, Ph, G}
B = Service Time Distribution {M, D, U, E_k, Ph, G}
C = Number of Servers {1,2,...}
K = System Capacity (including those in service)
N = Size of the calling population
D = Queue Discipline {FIFO, LIFO, SIRO, Priority}

Problem Statement:

This problem is very similar to the drilling center in the hand simulation problem
This is a very simple dynamic simulation model made possible only because of Lindly's recursion

Customers arrive one at a time (Customer 1 at time zero with empty and idle conditions
Customers may or may not have to wait in a First in First Out (FIFO) Queue
Customers are then served one at a time by a single server, and then depart

A_i = Inter-Arrival Time of ith customer ~ EXPO(1.6) minutes (independent of each other)
        λ = Arrival Rate = (1/1.6) = 0.625 arrivals per minute
S_i = Service Time of ith customer ~ UNIFORM(0.27, 2.29) minutes
        Independent of each other and the inter-arrival times

How would we write the queue in terms of Kendall's Notation? [M/U/1].

We are interested in the Waiting Times WQ_i for each of the first 50 customers (i = 2, 3, ..., 50)
        $WQ_1 = 0$ → The first customer does not have to wait (arrives to empty/idle system)

Since the only performance measure we are interested are the waiting times in the queue, we can use Lindley's recursion to compute waiting time in the queue for each customer:

$$WQ_i = \max\{WQ_{i-1} + S_i - A_i, 0\} \text{ for } i = 2,3,...,50$$

The question of interest: What is the average time a customer can expect to spend waiting in the queue for this system?

Construction of the Spreadsheet Simulation Model:

In order to simulate this, we need to be able to sample for the exponential and uniform distributions.

Arrival Times – Exponential Distribution:

$$f(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

$$CDF = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \lambda e^{-\lambda x}dx = \lambda \int_{0}^{x} e^{-\lambda x}dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_{0}^{x} = \lambda \left[\frac{e^{-\lambda x}}{-\lambda} - \frac{1}{-\lambda}\right] = 1 - e^{-\lambda x}$$

To find the Inverse CDF, we set the CDF = U and solve for x:

$$1 - e^{-\lambda x} = U \implies e^{-\lambda x} = 1 - U \implies -\lambda x = \ln(1 - U) \implies x = \left(-\frac{1}{\lambda}\right)\ln(1 - U)$$

In order to generate exponential random variables who mean is λ, we generate a uniform random variable between 0 and 1 and substitute that value for U in the above equation.

Service Times – Uniform Distribution:

$$f(x) = \frac{1}{b - a} \text{ for } a \geq x \geq b$$

$$CDF = \int_a^x \frac{1}{b-a} dx = \frac{1}{b-a}[x]_a^x = \frac{1}{b-a}(x-a) = \frac{x-a}{b-a}$$

To find the Inverse CDF, we set the CDF = U and solve for x:

$$\frac{x-a}{b-a} = U \implies x - a = (b-a)U \implies x = a + (b-a)U$$

In order to generate exponential random variables which has a minimum of a and a maximum of b, we generate a uniform random variable between 0 and 1 and substitute that value for U in the above equation.

In addition, for a uniformly distributed random variable:

$$E[U] = \frac{a+b}{2}, \quad \text{Var}[X] = \frac{(b-a)^2}{2}$$

Now we will open the spreadsheet and begin constructing out simulation model

1. Create an area in the spreadsheet to store input parameters for the problem (done for you) and fill in this area with the appropriate values. We can also compute the mean and variance of the exponential distribution.

    a. We will perform 5 independent replications of this simulation of 50 customers

2. For the first customer, there is not inter-arrival time. However, we can put in our formula for sampling service times and our formula for determining the waiting time in the queue.

3. For the second customer, we can copy the two formulas from above for the service time and waiting time in the queue, but we also must add the formula for inter-arrival time for customers 2 to 50.

4. We can now compute the average for each customer across replications in column T

5. We can also compute the average inter-arrival time, service time, and waiting time in the queue for across all 50 customers for each of the 5 replications

    a.   The inter-arrival time average and service time average should be used to ensure you are using the correct formula for sampling random variables. The mean for these columns should be CLOSE to, but not exactly, the mean for the random variable you are sampling

6. We can compute a 95% confidence interval on the mean number time spent waiting in the queue.

    a.   Which number should be used to create this confidence interval? Why?

    b.   Why do we not compute the variance of the waiting times across all 50 customers for each replication? [Downward Bias].

7. Lastly, using results from queuing theory, we can compute the long-run steady state expected waiting time in the queue:

$$E[WQ_\infty] = \lambda\left(\frac{\text{Var}[S] + E[S]^2}{2(1 - \lambda E[S])}\right) = 0.625\left(\frac{0.34 + 1.28^2}{2(1 - 0.625 * 1.28)}\right) = \textbf{3.0913 Mins}$$

8. We can also check to see if our 95% CI covers the steady state expected waiting time in the queue… Do it? Why/Why not?

9. Discuss the graph on the left and the results

10. Discuss how this is very inadequate for dynamic simulation models. Also discuss add-ins such as crystal ball that make simulation with spreadsheets easier.

9:15 -9:55 AM: In-Class Assignment #2: Problem 2-15 from the Textbook:

1. Supermarket customers load their carts with goods totaling between $5 and $200, uniformly (and continuously) distributed; call this the *raw order amount*. Assume that customers purchase independently of each other. At checkout, 63% of customers have a loyalty card that gives them a 4% discount of their *raw order amount*. Also at checkout, 18% of customers have coupons that give them 7% off their *raw order amount*. These two discounts occur independently of one another, and a given customer could have one of them, both of them, or neither of them, to get their *net order amount*, the amount they actually end up paying. Construct a spreadsheet simulation to simulate 100 customers and collect statistics on the net order amount; these statistics should include the average, standard deviation, minimum, and maximum. You should also form a 95% CI for the mean *raw order amount*, the mean *net order amount*, the mean customer loyalty discount, and the mean coupon discount. (Hint: Use the Excel IF function in combination with the RAND() function to determine whether or not customers get discounts. Also, the number of people who have each type of discount will change when you hit F9 so you should find a command that allows you to account for this in your 95% CI.)

Submit this file to Wolfware before leaving class with In Class Assignment 2 in the filename. The latest time to submit the file is 10:15 AM. If the file has not been submitted by then, you will receive a zero for this in-class assignment.