# Monte Carlo Simulation Practicum

**S. David Alley, P.E.**

**ANNA, Inc (annainc.com)**

## Abstract

Monte Carlo analysis is commonly used to predict the cost of future activity by acknowledging the uncertainty of the assumptions that contribute to the cost estimate and enabling the expression of those uncertainties in clever ways that bring forth and capture the subtle information that does exist.

This paper will empower those not familiar with Monte Carlo simulation with the knowledge of how the mechanics of the Monte Carlo simulation works and how to use this valuable tool effectively and with confidence. The approach in the presentation is to lead the audience through some simple yet powerful examples that clearly convey the principals and applications of Monte Carlo simulation.

The paper avoids the use of fancy words and transfers the gist of the knowledge to the audience. This presentation is designed for those just now introduced to the concept of Monte Carlo analysis applied to cost estimation and is quite useful to those who wish to have the means to explain the analysis to others such as co-workers, supervisors or clients. The materials distributed in the presentation are suitable for use as a reference and for instruction to others; the materials may be copied freely as long as the presenter's information remains in the footer of the document. The objective of the presenter is to empower all with the knowledge and skill to use this analysis technique effectively and with confidence.

# A Cost Estimate Example

We will start with an example and work our way through creating a cost estimate using the best guess technique and then create a cost estimate with the same example using Monte Carlo simulation techniques.

Let's estimate the cost of an automobile repair: the cost of the next service shop work on your 125,000 mile luxury car. You have heard that familiar squeaking sound in your brakes and you have a miss in the engine that you have been told is due to spark plug wire problems. So for the purpose of this example, assume you will have to have a brake job and a spark plug wire replacement and that you will have to have a car from the dealer to drive while yours is in the shop. What is the cost of your repair?

Of course, you cannot answer the question yet because the cost of your repair is the amount you will pay the repair shop in the future, and you will know that only when you pay—so the best we can do now is to estimate the cost of this future expenditure. As a Nils Bohr once stated: "Predictions are difficult, especially when they involve the future."

Create a cost estimate that lists three variables and a total, and fill in your best guess. It may look like this:

**Automobile repair cost estimate—**
**first guess ($)**

| | | |
|---|---|---|
| 1 | Brake job | 600 |
| 2 | Spark plug wire replacement | 200 |

| | | | |
|---|---|---|---|
| 3 | Rental car | | 100 |
| | | Total | 900 |

Take a look at this simple spreadsheet model; it is simply a sum of three numbers, each of which is a best guess without extensive research. We could investigate the costs, get quotes from various vendors, discuss parts costs with the auto parts supply stores, or solicit the opinion of an acquaintance. For the next steps of this practicum and in an effort to improve the quality of our estimate, we will use an ROM (Rough Order of Magnitude) technique to bracket our estimate.

Here is our estimate with the ROM brackets:

**Automobile repair cost estimate ($)**
**Rough Order of Magnitude**
**($)**

| | | Best guess | 50% | 200% |
|---|---|---|---|---|
| 1 | Brake job | 600 | 300 | 1200 |
| 2 | Spark plug wire replacement | 200 | 100 | 400 |
| 3 | Rental car | 100 | 50 | 200 |
| | Total | 900 | 450 | 1800 |

Now we can begin to get a feel for the use of a bracketing technique—it allows us to view the total cost with the realization that the cost could be significantly higher (our usual fear). By the way, this is an ANSI/PMI: 99-001-2004 standard cost estimating technique, which is used to create ROMs from single-point estimates. The intent is to capture the possibility that we could be off by a factor of two in either direction. Unfortunately, our estimates can be off by more than a factor of two and this technique does not capture the likelihood that a cost will be at a certain point within the bracket except for the implicit expectation that the best guess is more likely than the maximum or minimum. Therefore this technique, while useful, is not as sophisticated as Monte Carlo Simulation.

## Experience and Reference Points

If you were to review your past experience in more detail you will probably find references from you own personal experience in an auto repair shop or those of others which were either extremely high or extremely low. You may have had an uncle that took a car in for a tune up and ended up spending $2000 in repairs. Last week, I was told about a trade-in car that experienced a transmission failure a few days before the swap; the seller paid for a $1500 repair bill—that was $1500 more (infinitely more) than the expected repair cost for that week. A few months ago, I took my Subaru in for exhaust system diagnosis and repair . . . that cost me only $40 rather than the expected $600 because the exhaust noise was caused by a loose muffler clamp rather than a failed exhaust system—that was only 7% of the expected cost.

Perhaps the range of estimates can be modified to encompass these new reference points so the table would now look like the following, but that does not really help because it does not capture the likelihood that the cost will be a certain amount:

**Automobile repair cost estimate ($)**
**Rough Order of Magnitude**

| | | Best guess | Extreme low side: 7% | 50% | 200% | Extreme high side: +$1500 |
|---|---|---|---|---|---|---|
| 1 | Brake job | 600 | 42 | 300 | 1200 | 2100 |
| 2 | Spark plug wire replacement | 200 | 14 | 100 | 400 | 1700 |
| 3 | Rental car | 100 | 7 | 50 | 200 | 1600 |
| | Total | 900 | 63 | 450 | 1800 | 5400 |

Now, at least we are probably capturing the very high and very low cost that we are "sure" captures the extreme ends of possibility. However, what can we do with this data? We can wring our hands and hope for the best but we really do not believe the cost could be close to either end of the extreme case values. So, we are likely to dismiss these as unlikely and simply focus on the original best guess.

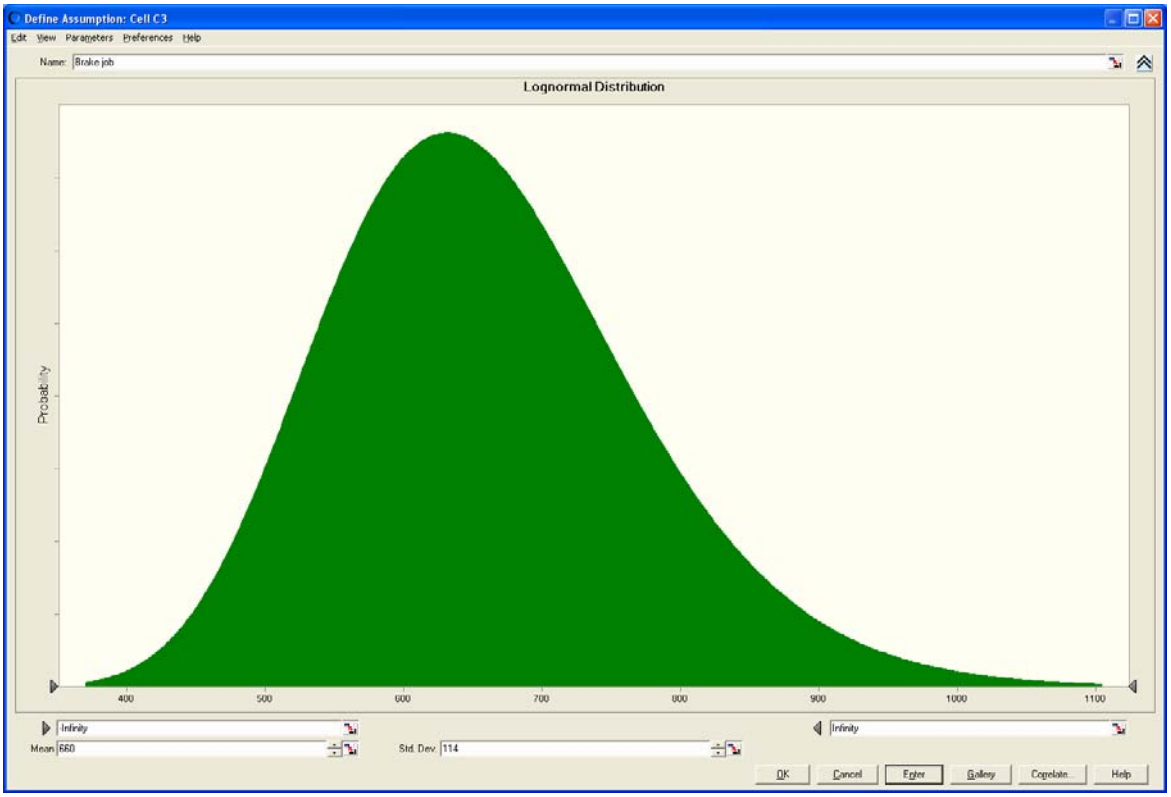
## Capturing the Likelihood of the Costs

Monte Carlo simulation allows us to capture more of our experience than the single-point numbers in the last table—it allows us to capture our feelings about the likelihood that costs within the expected range turn out to be the actual cost.

What words would you use to describe the likelihood that the cost of the brake job is $300? Would you use some of the following words:

- Likely
- Unlikely
- Very likely
- Very unlikely
- No way would it be that high (low)
- Impossible
- Could happen
- Less likely than __________

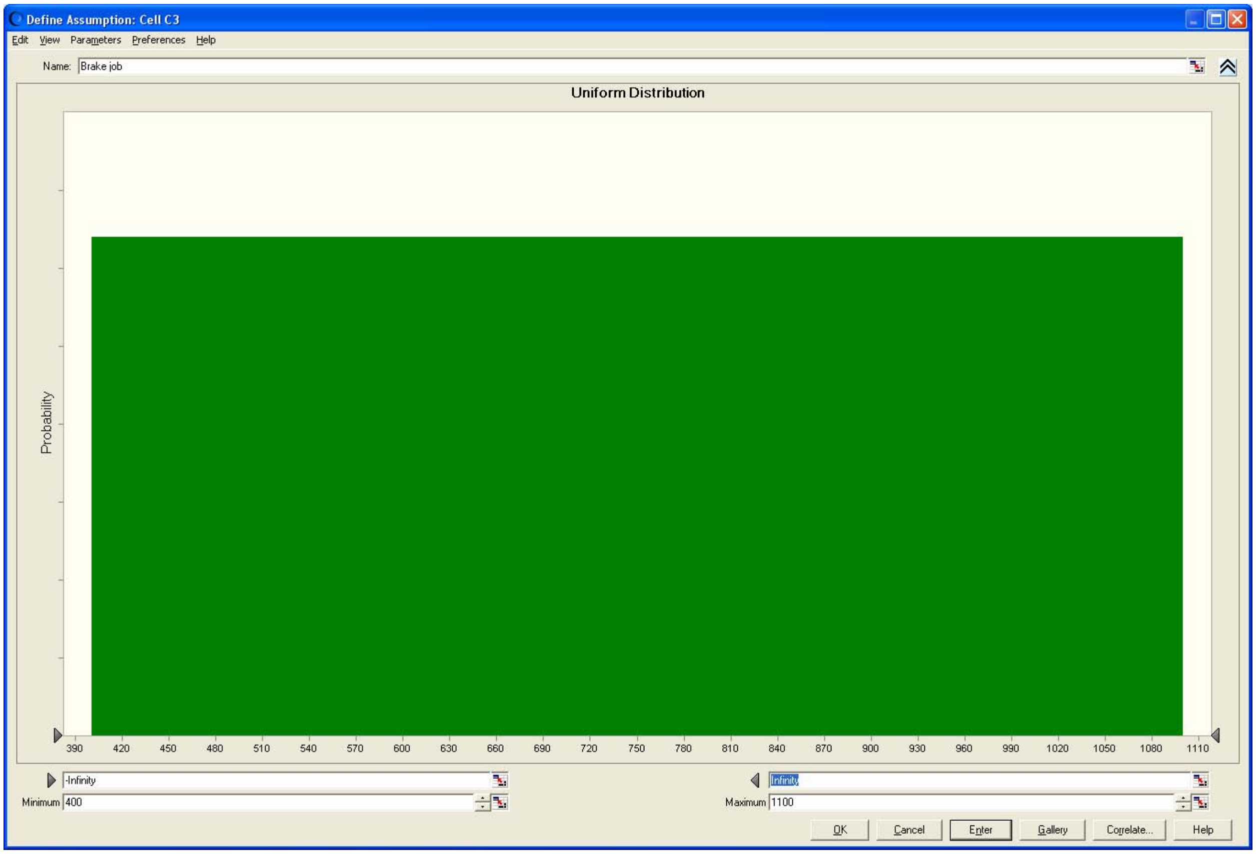
In fact, these words can be used to develop a mathematical representation of you assessment of the likelihood of a range of possible values—this is called a probability distribution. An example follows of my probability distribution of the cost of a brake job; it reflects my belief that I will seldom get out of the repair shop for less than $300 and there is no way that the cost would top $1000, and the most likely cost is centered around my initial thought of $600, maybe a bit higher.



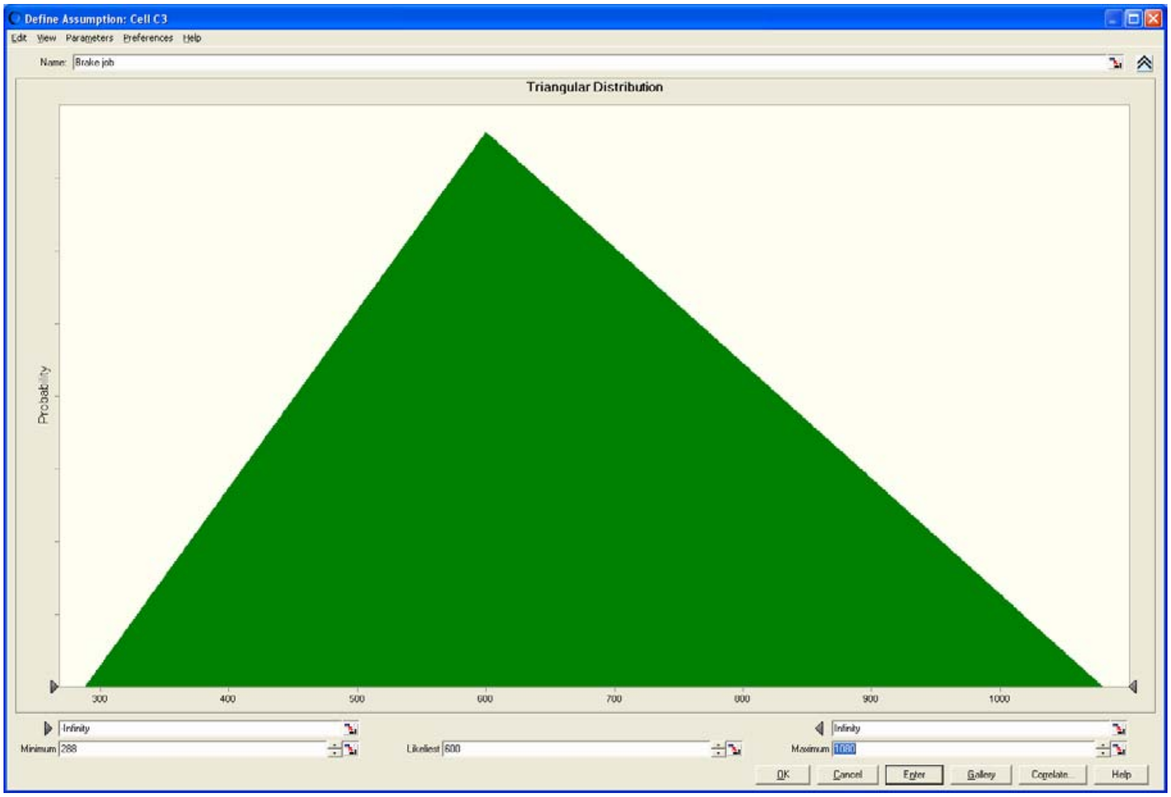

The diagram above, called a probability distribution, is named: Lognormal Distribution. This name has meaning and significance to a statistician but we are simply using the curve that fits our understanding of the data. In this practicum, we are not going to get into the math of these various curve types—instead, we will simply use the curves that visually seem to fit our understanding of the data.

This is not a practicum in analyzing historical data and developing models that describe the past in the hopes that it will be an accurate prediction of the future—we are simply trying to find a way of expressing what we do know about the likelihood of the future cost. In fact, all you have to know about probability and statistics, probability distributions and the various names that go along with them is that if the curve is high, the point under that part of the curve (along the bottom scale) is more likely to occur than a value under the lower part of the curve. In the diagram above, I liked the fact that it seems to lean a bit toward the low side yet still allows for the very unlikely higher numbers over $1000.

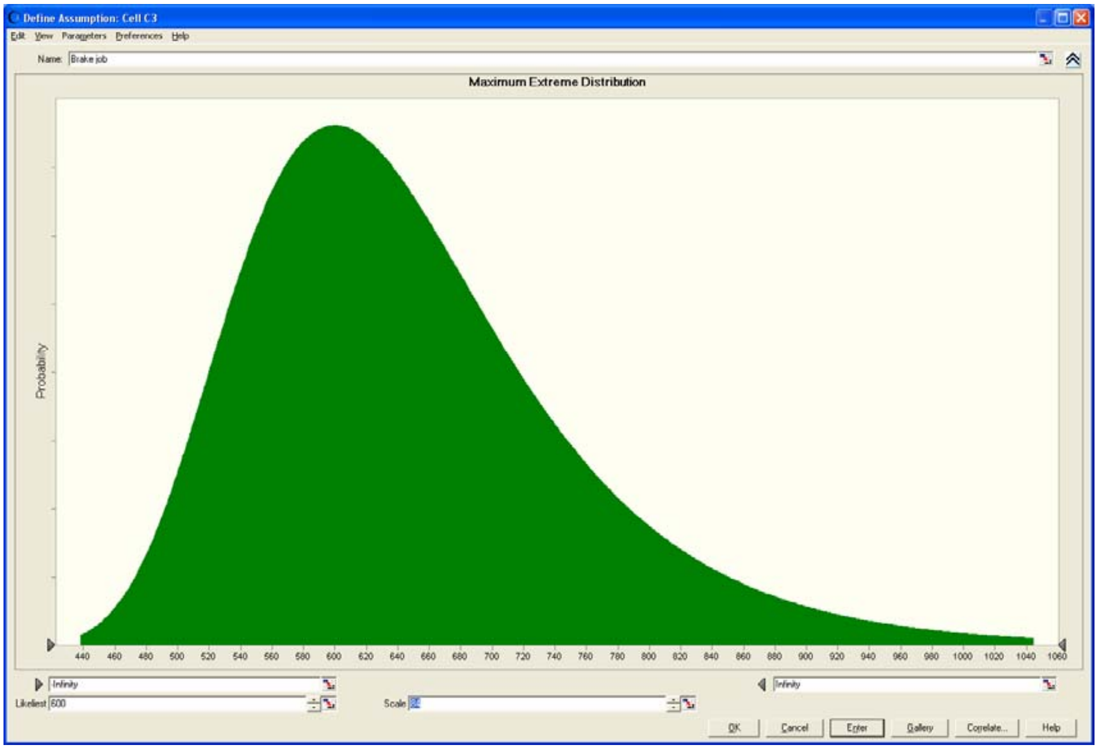
The software (Crystal Ball in this case) has all these curves in an array from which to choose. I looked at several curves and rejected several before I landed on this one. Here are a few I rejected:




That one was immediately rejected because it simply provided a range, with each value as likely as the next from the minimum to the maximum. .



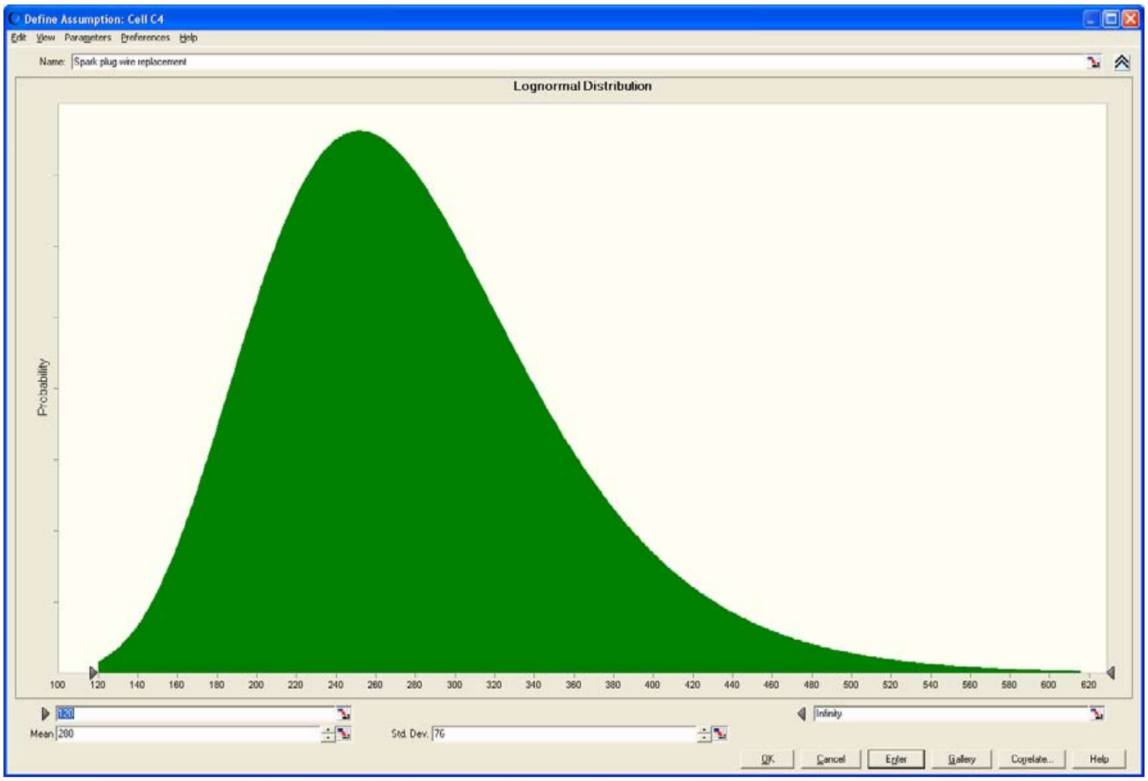

Not bad, but I felt it gave too much likelihood to the values at the high and low end—after all, I felt the cost would be close to $600 . . . it has been a number of times in the past.



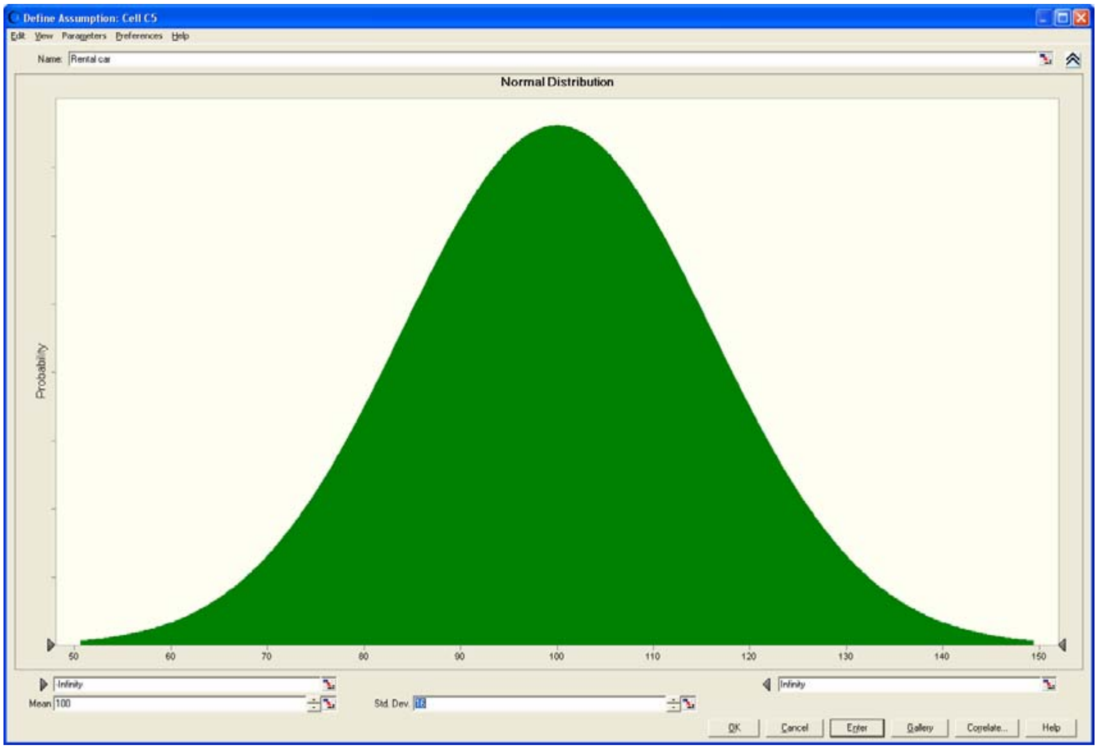
Not bad but I felt it would allow too much likelihood in the range below $600; after all, I very rarely get out of the shop for less than $400 when I have squeaky brake noise.

Notice that there is no scale along the left side. The scale is not needed since the software is smart enough to use the information in the right way.

In a similar fashion, I selected the following probability distributions that I felt were representative of the likelihood and range of the variables. For the cost of replacing sparkplug wires, I used the following:

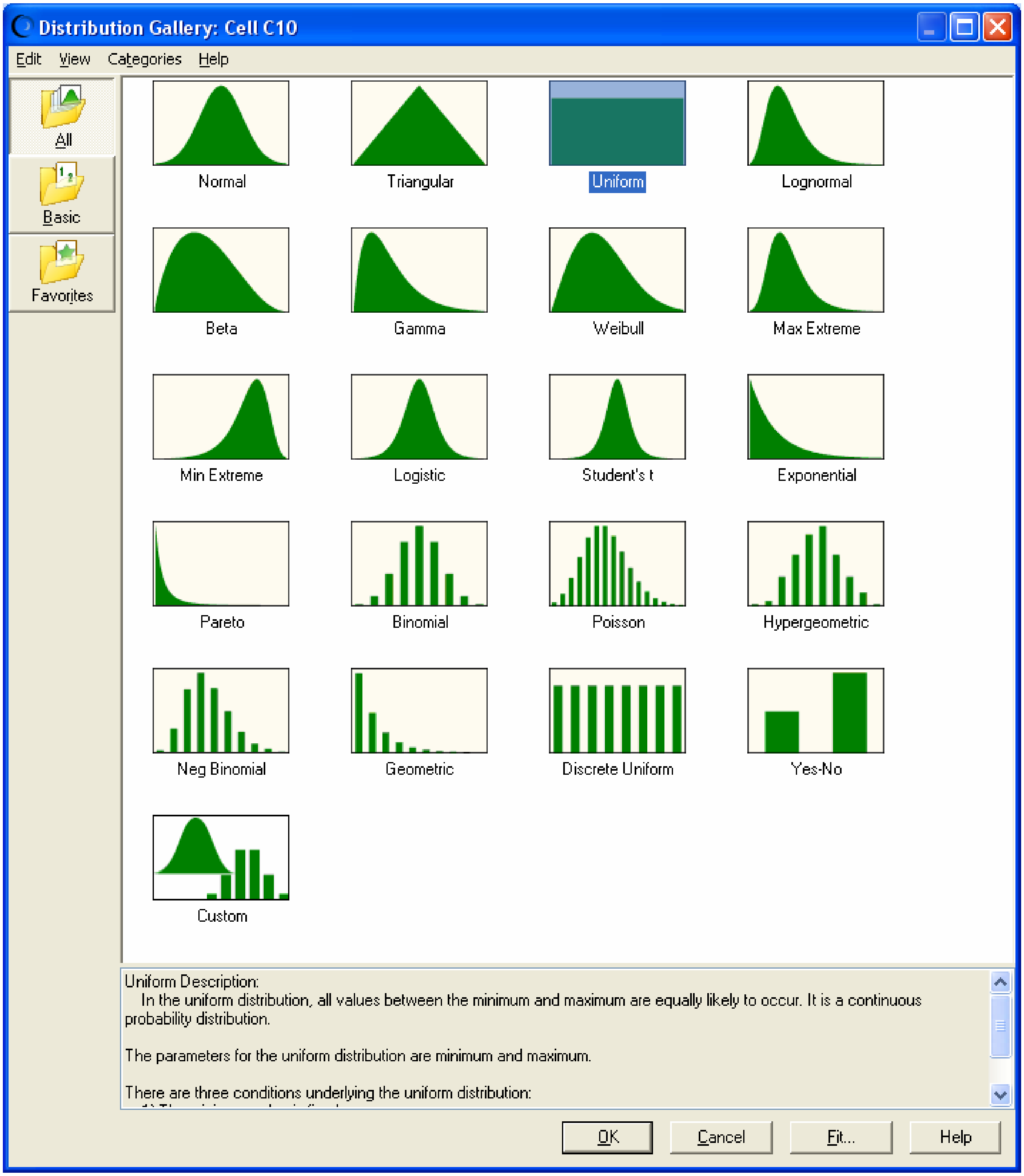



My thought process was that it would be impossible to get the job done for less than $120 and that the job, in my wildest imagination, would not top the $650 range—more than three times my best guess. Surely, this model captured the possible range and the likelihood of the various costs.




For the cost of the rental car, I chose the probability distribution above since it seemed to be centered about $100 and showed possibilities, equally unlikely that the cost could be $50 or $150. This seemed to satisfy my understanding of the cost of rental cars and the likelihood that I could get a reasonable car for a hundred bucks.

You may ask how it is that I find these probability distributions to select—well, that is easy, simply click the Define Assumption buttons and visually select a sample from the gallery, which looks like this:




So you see that there a lot of options, including a custom selection that allows a lot of flexibility. Now we can look at how the software puts these probability distributions together to create a useful tool.

## The Model

The model now looks like this.  Notice that the assumption cells, which contain the probability distributions are light green while the forecast or result cell (the total estimate) is blue.

**Automobile repair cost estimate ($)**
**Monte Carlo Model**

| | | Best guess |
|---|---|---|
| 1 | Brake job | 600 |
| 2 | Spark plug wire replacement | 200 |
| 3 | Rental car | 100 |
| | Total | 900 |

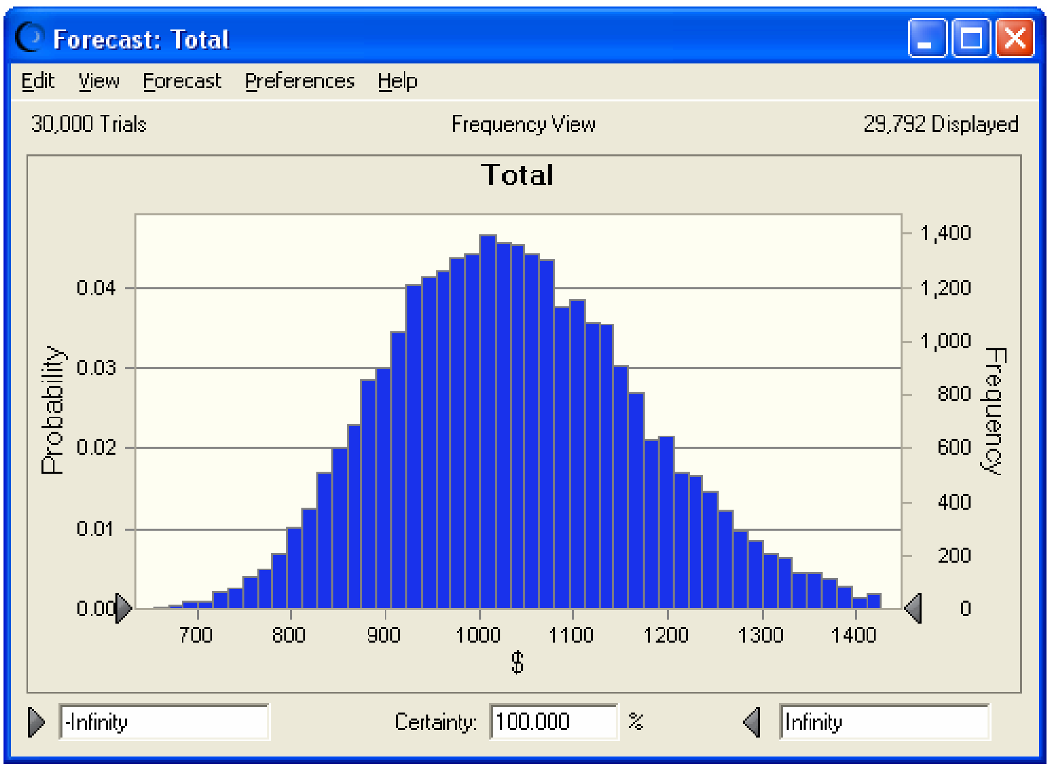
## Simulation

A simulation is the next step.  For the fist of many simulations, the software starts down the list of costs and selects at random a value for each of the three assumption variables.  For instance, in the first pass, the software selected $422 for the brake job, $197 for the Spark plug replacement and $141 for the rental car for a total of $760.

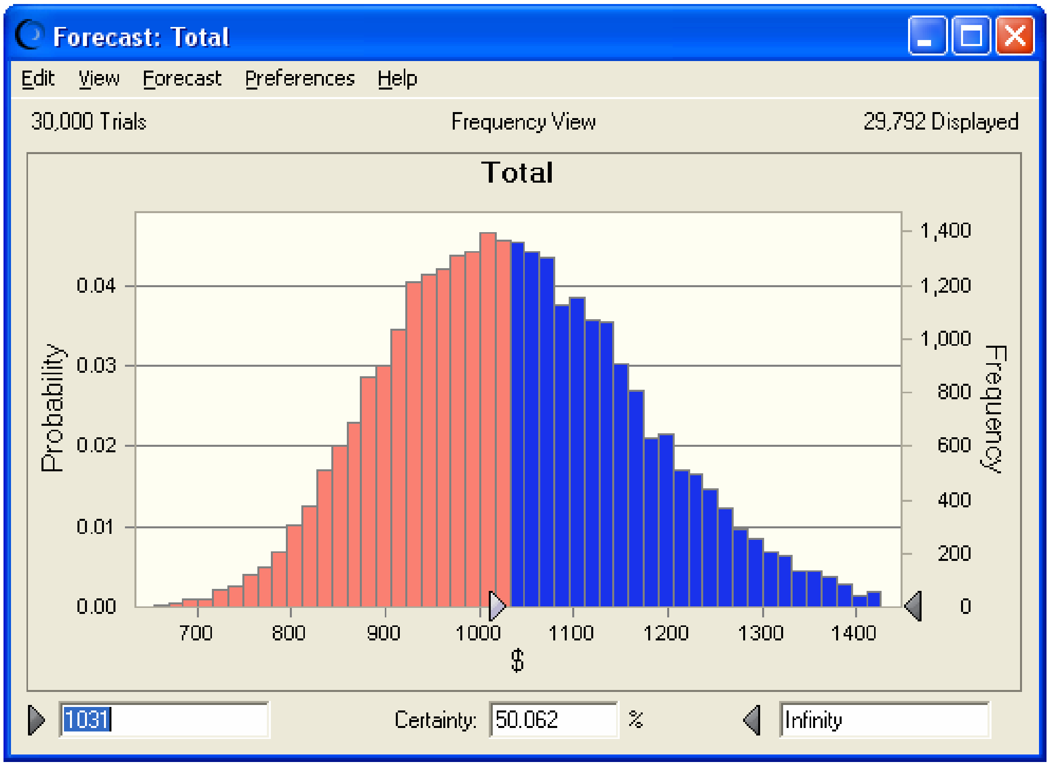
The second simulation will be performed like the first EXCEPT that the software does not make purely random selections—it makes random selections biased to approximate the probability distribution curve if all the selections were stacked together. In other words, as selections are made, the software selects values so that after 1000 selections or so, the values near the high points in the probability distributions are selected much more often than the values at either end of the probability distributions.

## Results

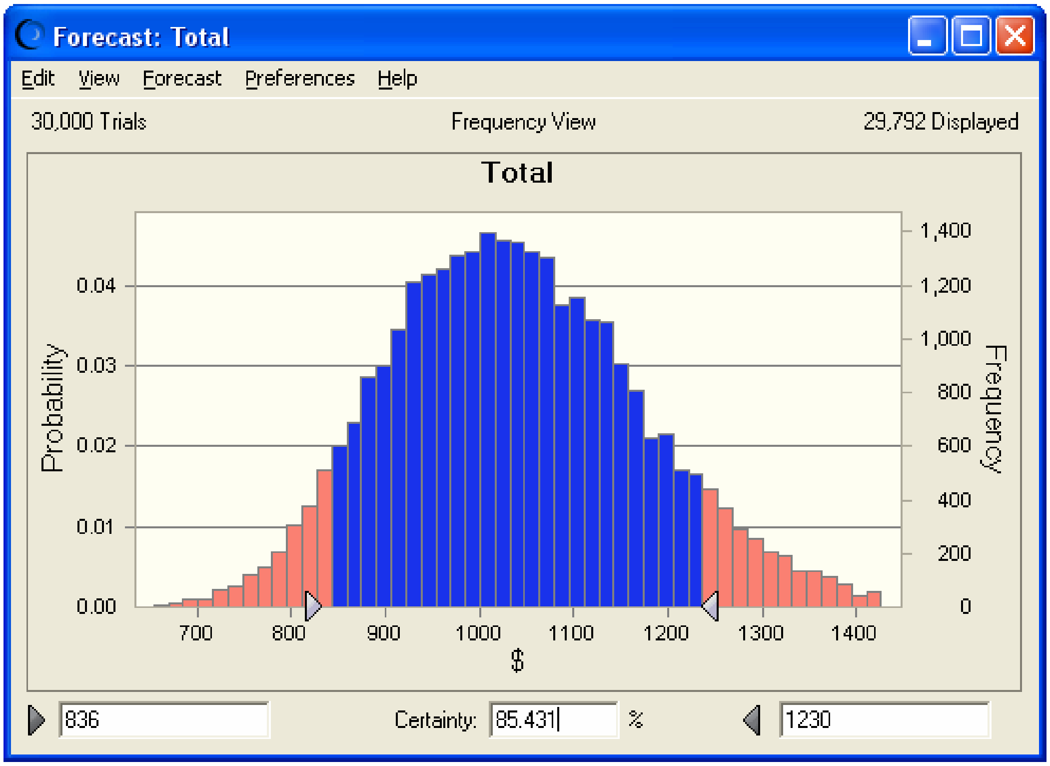
After 30,000 trials (simulations), the following was the result. Notice that this is the combination of the three curves in proportion to the relative value of each and that the lowest expected values are around $650 and the highest expected values are around $1450. Recall that these values are similar to the sums of the three low end values and the three high end values of the three distributions used to make up this model.



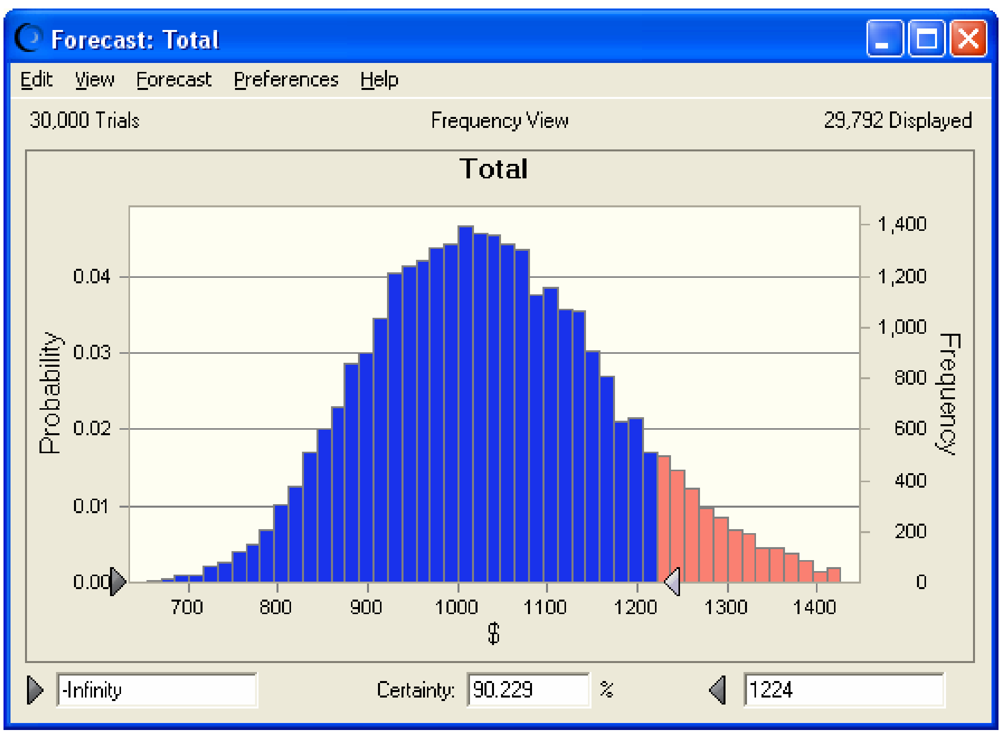

By dragging cursers left and right, we find that it is a 50% likelihood that the cost will be more than $1031 and just as likely that the cost will be less than $1031. Of course, we may as well call this $1000 because our estimates are surely not accurate to the nearest dollar or tens of dollars. Here is the frequency view again, with the left curser pulled over to identify the 50% point:




Another way to slice the data is to find the middle 50% value. The following frequency view tells us that it is 85% likely to be between $836 and $1230. I recommend rounding those to $800 and $1200.




A similar view of the data is provided below. Note that the display indicates that it is a 90% probability that the value is greater than approximately $1200.



Now, for one more view of the data. Here is a listing of each percentiles or certainty levels. A certainty level of 80% means that it is 80% likely that the value is less than approximately 1150.

| Forecast values | |
|---|---|
| 0% | 615 |
| 10% | 870 |
| 20% | 922 |
| 30% | 961 |
| 40% | 997 |
| 50% | 1,031 |
| 60% | 1,067 |
| 70% | 1106 |
| 80% | 1,151 |
| 90% | 1,223 |
| 100% | 1,929 |

Remember that in all Monte Carlo models, the accuracy of the results depends entirely upon the accuracy of the input probability distributions. And, since these are loosely created, the accuracy of the results are not so good. What is remarkably good is that you now have some handle on nebulous data in a much more quantifiable manner than a single point estimate.

A technique that greatly improves accuracy of the input probability distributions is to use a group to develop a consensus for each assumption cell. And, remember that humans like us tend to leave out the extremes, so encourage those working with you to think broadly about the possibilities that are real, even if unlikely.

## The final word

OK, the example above is an actual example. Would you believe that the cost of replacing the spark plug wires alone was $1,350? It was. The distributor is located below the water pump and the front of the car has to be removed to get to that area. The brakes were approximately $850. The rental car was not needed since I was able to use my mountain bike.

The bottom line is that Monte Carlo simulation is a useful tool for getting a handle on nebulous data but take the results as a dressed up version of your qualitative thoughts.