Chapter 10
Simulation

Sumit Kunnumkal

1 Introduction

The goal of this chapter is to provide an understanding of how simulation can be an effective business analytics technique for informed decision making. Our focus will be on applications and to understand the steps in building a simulation model and interpreting the results of the model; the theoretical background can be found in the reference textbooks described at the end of the chapter. Simulation is a practical approach to decision making under uncertainty in different situations. For example: (1) We have an analytical model and we would like to compare its output against a simulation of the system. (2) We do not have an analytical model for the entire system but understand the various parts of the system and their dynamics well enough to model them. In this case, simulation is useful in putting together the various well-understood parts to examine the results. In all these cases, the underlying uncertainty is described, the model developed in a systematic way to model the decision variables, when necessary describe the dynamics of the system, and use simulation to capture values of the relevant outcomes. This chapter sets out the steps necessary to do all the above in a systematic manner.

S. Kunnumkal (✉)
Indian School of Business, Hyderabad, Telangana, India
e-mail: sumit.kunnumkal@queensu.ca

2 Motivating Examples

We will use the following example throughout the chapter: Consider a fashion retailer who has to place an order for a fashion product well in advance of the selling season, when there is considerable uncertainty in the demand for the product. The fashion item is manufactured in its factories overseas and so the lead time to obtain the product is fairly long. If the retailer orders too large a quantity, then it is possible that the retailer is left with unsold items at the end of the selling season, and this being a fashion product loses a significant portion of its value at the end of the season. On the other hand, if the retailer orders too little, then it is possible that the product may be stocked out during the selling season. Since the lead time of the product is long compared to the length of the selling season (typically 12 weeks), the retailer is unable to replenish inventory of the product during the selling season and stock outs represent a missed sales opportunity. The retailer would like to understand how much to order factoring in these trade-offs.

The example described above is an example of a business problem where decisions have to be made in the presence of uncertainty, considering a number of different trade-offs. In this chapter, we will study Monte Carlo simulation, which is an effective technique to make decisions in the presence of uncertainty.

What Is Simulation?

At a high level, a simulation is a "virtual mirror" where we build a model of our business on a computer. A simulation model has three broad purposes. The first is to model an existing business situation. Then, to understand the impact of making a change to the business and finally to understand what may be the optimal intervention. For example, a manufacturing company may be interested in building a simulation model to describe its current production system and understand where the bottlenecks are and where material may be piling up. Next, it may build on this model to understand the impact of making a process change or augmenting the capacity of a particular machine in its production line. Finally, it may use this simulation model to understand the optimal process flow or capacity expansion plan in its production facility. As another example, a manager of a customer service center may want to build a simulation model of its existing system to understand the average on-hold times for its customers. Next, she may want to use the simulation model to understand the impact of an intervention to bring down the on-hold times, say, by augmenting the number of agents during peak hours. Finally, she may be considering a range of interventions from augmenting capacity to better agent training, and may like to use the model to understand the effectiveness of the different intervention strategies in bringing down the on-hold times. She can then use a cost–benefit analysis to decide how to best reduce the on-hold times of the customers.

Computer Simulation vs Physical Simulation

Throughout this chapter, we use simulation to refer to a computer simulation where we model a business problem on a computer. We distinguish this from a physical simulation, which typically involves building a small-scale replica or running a small pilot study. As an example of a physical simulation, consider testing a scale model of a new aircraft wing design in a wind tunnel to understand its performance. As another example, a fast-food chain may want to carry out a small pilot study at one of its restaurants to understand the impact of process redesign on the customer wait times. An advantage of a physical simulation compared to a computer simulation is that it requires fewer assumptions as to how the system functions. For example, in the case of the fast-food chain, the customer wait times can be obtained by simply measuring how long customers wait in the pilot study. It requires minimal assumptions on the nature of the arrival and service processes and how the system functions.

In a computer simulation, on the other hand, we model the business problem on a computer and this naturally requires a greater understanding of the process, the constraints, and the relations between the different variables. Therefore, in a computer simulation the modeler has to capture the most relevant aspects of the business problem and think critically about the set of assumptions underpinning the model. Some advantages of a computer simulation over a physical simulation are the shorter implementation times and the reduced costs.

Applications

While simulation has its origins in World War II, it continues to find new and interesting applications: HARVEY is a biomedical engineering software that simulates the flow of blood throughout the human body based on medical images of a patient. This can be a useful tool to inform surgical planning or to design new drug delivery systems (Technology Review 2017). GE uses a simulation model of a wind farm to inform the configuration of each wind turbine before the actual construction (GE Look ahead 2015). UPS has developed a simulation software called ORION ride that it uses to simulate the effectiveness of new package delivery routes before actually rolling them out (Holland et al. 2017). Jian et al. (2016) describe an application to bike-sharing systems, while Davison (2014) describes applications in finance.

The continued interest in building simulation models to answer business questions stems from a number of reasons. For one, the greater availability of data allows for models that are able to describe the underlying uncertainty more accurately. A second reason is the growing complexity of business problems in terms of the volume and frequency of the transactions as well as the nonlinear nature of the relationships between the variables. Nonlinear models tend to quickly become challenging to analyze mathematically and a simulation model is a particularly effective technique in such situations. Furthermore, the performance metrics obtained by simulation models, such as expected values, are also more appropriate for business situations involving a large volume of transactions. Expected values

can be interpreted as long run averages and are more meaningful when applied to a large number of repeated transactions. Finally, advances in computing hardware and software also make it possible to run large and complex simulation models in practice.

Advantages and Disadvantages

One of the main advantages of a simulation model is that it is easy to build and follow. A simulation model is a virtual mirror of the real-world business problem. It is therefore easy to communicate what the simulation model is doing, since there is a one-to-one correspondence between elements in the real-world problem and those in the simulation model. A simulation model is also flexible and can be used to model complex business processes, where the mathematical analysis becomes difficult. It is also possible to obtain a range of performance metrics from the simulation model and this can be particularly helpful when there are multiple factors to be considered when making decisions.

One disadvantage of a simulation model is that it often tends to be a "black-box" model. It describes what is happening, but does not provide insight into the underlying reasons. Since simulation models are easy to build, there is a tendency to add a number of extraneous features to the model. This drives up the complexity, which makes it difficult to derive insight from the model. On the other hand, it usually does not improve the quality of the solution.

3   Simulation Modeling Method

Simulation modeling comprises of a number of logically separate steps. The first step in building a simulation model is to identify the input and the output random variables and the relationship between the two. We think of the input random variables as the underlying drivers of uncertainty. We think of the output random variables as the performance metrics of interest. The output depends on the input and since the input is random, so is the output. Mathematically, if X is the input random variable and Y is the output random variable, then we let $Y = f(X)$ to denote the dependence of Y on X. The function f(.) relates the output to the input and we say that the output is a function of the input. We note that simulation is most helpful in situations where the function f(.) is nonlinear. In the fashion retail example described previously, the demand for the product is the input random variable. The performance metric of interest is the profit, which is a function of the input since profits depend on sales. On the other hand, in the customer service center example, the random inputs are the arrival times and service times of the customers. The performance metric of interest is the on-hold time, which depends on the number of customers requesting service and the amount of time it takes to serve them.

The next step is to generate a random sample of the input and use the relation between the input and the output to generate a random sample of the output.

Let $X_1, \ldots, X_n$ be a random sample of size n of the input. The random sample $X_1, \ldots, X_n$ is drawn from the distribution of X. In the appendix, we describe how this can be done on a computer. Given the random sample of the input, we obtain a random sample of size n of the output: $Y_1, \ldots, Y_n$ by using the relation $Y_i = f(X_i)$. The function $f(.)$ *will depend on the application.*

The third step is to interpret and make sense of the output of the simulation. Here we use the sample of the output to estimate the performance metrics of interest. We note that since we use a random sample of the input, the sample of the output is random as well and the estimates of the output that we obtain from the simulation change with the input sample. Moreover, there is sampling error since we work with a finite sample. Therefore, the results of the simulation are both random and approximate and we use concepts from statistical sampling theory to interpret the results.

We illustrate the steps using the fashion retail example.

Example 10.1 (Fashion Retailer): A fashion retailer purchases a fashion product for Rs (Indian rupees) 250 and sells it for Rs 500. The retailer is not sure what the demand for this product will be. Based on experience (past sales for similar products, market judgment etc.), she thinks that the demand for the product over the selling season will be normally distributed with a mean of 980 and a standard deviation of 300. (In the appendix, we describe how input distributions can be obtained from historical data.) The retailer also incurs a fixed cost of Rs 150,000 (administrative salaries, fixed overhead charges, etc.), which is independent of the sales volume. Assume for the sake of simplicity that this is the only product that the retailer sells and that the retailer can meet all of the demand for the product. Further, assume that any unsold product can be returned to the manufacturer and the cost recouped. Based on these assumptions, we would like to answer the following questions:

(a) What is the retailer's expected profit over the selling season?
(b) What is the standard deviation of the retailer's profit?
(c) What is the probability that the retailer will not break even at the end of the selling season?
(d) What is the probability that the retailer's profits will exceed Rs. 100,000?

Solution: We first determine the input and the output random variables and the relation between the two.

- Step 1: Specify input—demand. We denote demand by the random variable X. We are given that X is normally distributed with a mean of 980 and a standard deviation of 300.
- Step 2: Specify output—profit. We denote profit by the random variable Y.
- Step 3: Relate the input to the output: The unit profit margin is Rs. 250. We multiply this by the demand and subtract off the fixed costs to obtain the profit. Therefore, $Y = 250X - 150,000$.

Before we go and build a simulation model, we note that we can answer the questions exactly in this case since the relation between the input and the output is linear. We therefore do the exact analysis first so that we have a benchmark to compare the simulation results later.

Exact analysis: Since Y is a linear function of X, we can use linearity of expectations and conclude that $E[Y] = 250E[X] – 150,000 = Rs\ 95,000$. Therefore, the expected profit is Rs. 95,000.

We have $V[Y] = 250^2\ V[X] = 250^2 * 300^2$. The standard deviation of Y is therefore $250 * 300 = 75,000$. Therefore, the standard deviation of the retailer's profit is Rs. 75,000.

Since X is normally distributed and Y is a linear function of X, Y is also normally distributed with a mean of 95,000 and a standard deviation of 75,000. We can use z-tables or Excel functions to determine the answer. For example, using the Excel function NORM.DIST(.), we get $P(Y <= 0) = 0.10$ and $P(Y >= 100,000) = 0.47$. Therefore, there is about a 10% chance that the retailer would not break even and there is a 47% chance that the retailer's profits would exceed Rs. 100,000.

Simulation model: We next build a simulation model to answer the same questions and compare the results from the simulation model to those obtained from the exact analysis. We have already described the input and the output random variables and the relationship between the two. In a simulation model, we first generate a random sample of size n of the input. So let $X_1, \ldots, X_n$ be a random sample of size n of the demand drawn from a normal distribution with a mean of 980 and a standard deviation of 300. Given the sample of the input, we then obtain a sample of the output by using the relation between the input and the output. Let $Y_1, \ldots, Y_n$ be the sample of size n of the output (profit) where $Y_i = 250X_i – 150,000$. Finally, we use the sample of the output to estimate the performance measures of interest. In particular, we use the sample mean $\overline{Y} = \frac{Y_1 + \cdots + Y_n}{n}$ to estimate $E[Y]$. We use the sample variance $S_Y^2 = \frac{(Y_1 - \overline{Y})^2 + \cdots + (Y_n - \overline{Y})^2}{n-1}$ to estimate $V[Y]$ (refer to the Chap. 6 on basic inferences). We use the fraction of the output sample that is smaller than zero to estimate the probability $P(Y <= 0)$ and the fraction of the output sample that is larger than 100,000 to estimate the probability $P(Y >= 100,000)$. That is, letting $I_i = 1$ if $Y_i <= 0$ and $I_i = 0$ otherwise, we estimate $P(Y <= 0)$ using $\frac{I_1 + \cdots + I_n}{n}$. We estimate the probability $P(Y >= 100,000)$ in a similar way.

Implementing the simulation model: We describe how to implement the simulation model in Excel using *@Risk*. *@Risk* is an Excel add-in that is part of the Palisade DecisionTools Suite.[1] We note that there are a number of other software packages and programming languages that can be used to build simulation models. While some of the implementation details vary, the modeling concepts remain the same. We also note that Excel has some basic simulation capabilities and we describe this in the appendix.

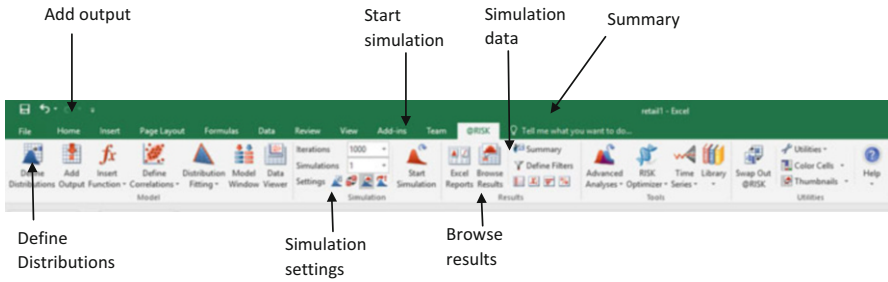[1]http://www.palisade.com/decisiontools_suite (accessed on Jan 25, 2018).
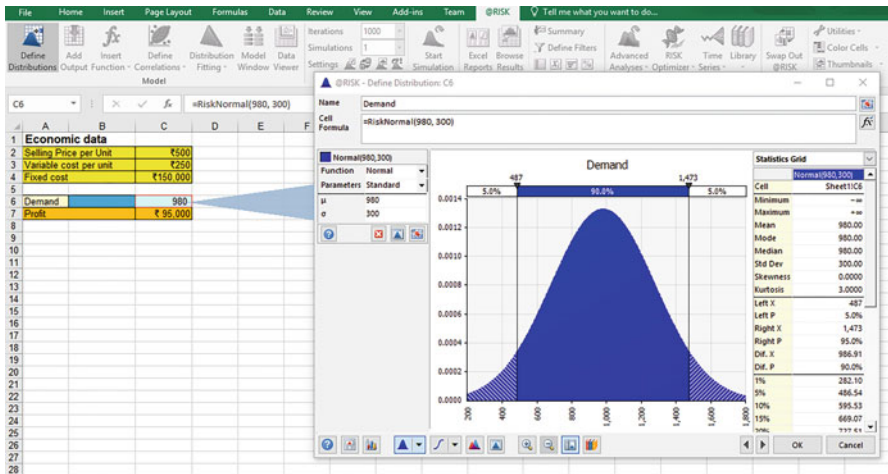
Fig. 10.1 @Risk toolbar

Fig. 10.2 Defining the input distribution

Implementation in @Risk

Once the Palisade DecisionTools Suite is installed, @Risk can be launched from the Start Menu in Windows. The @*Risk* toolbar is shown in Fig. 10.1. @*Risk* has many features and we will only cover the basic ones in this chapter. The more advanced features can be understood from the @*Risk* documentation.[2]

- Step 1: We specify the input cell corresponding to demand using the "Define Distributions" command from the @*Risk* toolbar (see Fig. 10.2). @Risk has a number of in-built distributions including the binomial, exponential, normal, Poisson, and uniform. It also allows for custom distributions through the "Discrete Distribution" option. In our example, demand is normally distributed. So we select the normal distribution from the list of options and specify the mean
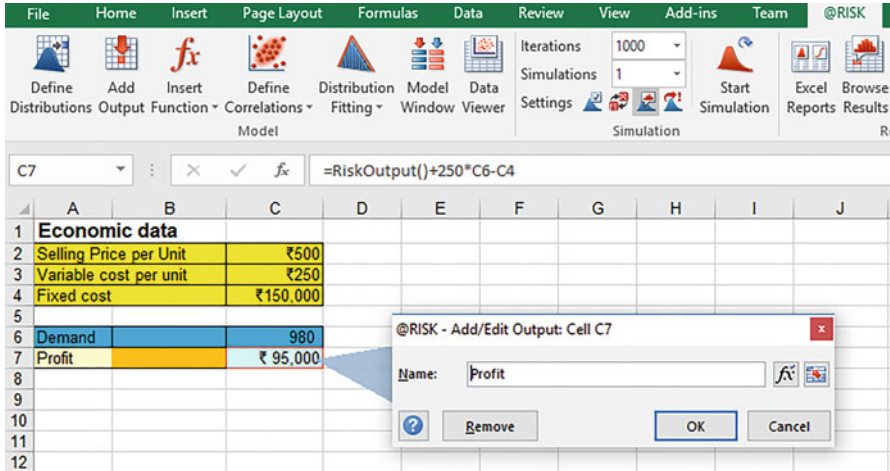
Fig. 10.3 Defining the output cell

(980) and the standard deviation (300) of the demand random variable. Note that alternatively, the demand distribution can be directly specified in the input cell by using the *@Risk* function *RISKNORMAL(.)*.

- Step 2: Next, we specify the relation between the input and the output in the output cell ($Y = 250X – 150,000$).
- Step 3: We use the "Add Output" command from the *@Risk* toolbar to indicate that the cell corresponding to profit is the output cell (see Fig. 10.3).
- Step 4: Before we run the simulation, we click the "Simulation Settings" button. Under the "General" tab, we set the number of iterations to be 1000 (see Fig. 10.4). The number of iterations corresponds to the sample size. Here, the sample size $n = 1000$. We later describe how we can determine the appropriate sample size for our simulation model. Under the "Sampling" tab, we set the sampling type to "Monte Carlo," we use the default option for the generator (Mersenne Twister) and fix the initial seed to 1 (see Fig. 10.5). We briefly comment on the generator as well as the seed. There are different algorithms to generate random numbers on a computer. *@Risk* has a number of such built-in algorithms and Mersenne twister is one such algorithm. We provide more details in the appendix. The seed provides the starting key for the random number generator. Briefly, the seed controls for the random sample of the input that is generated and fixing the seed fixes the random sample of size n (here 1000) that is drawn from the input distribution. It is useful to fix the seed initially so that it becomes easier to test and debug the simulation model. We provide more details in the appendix.
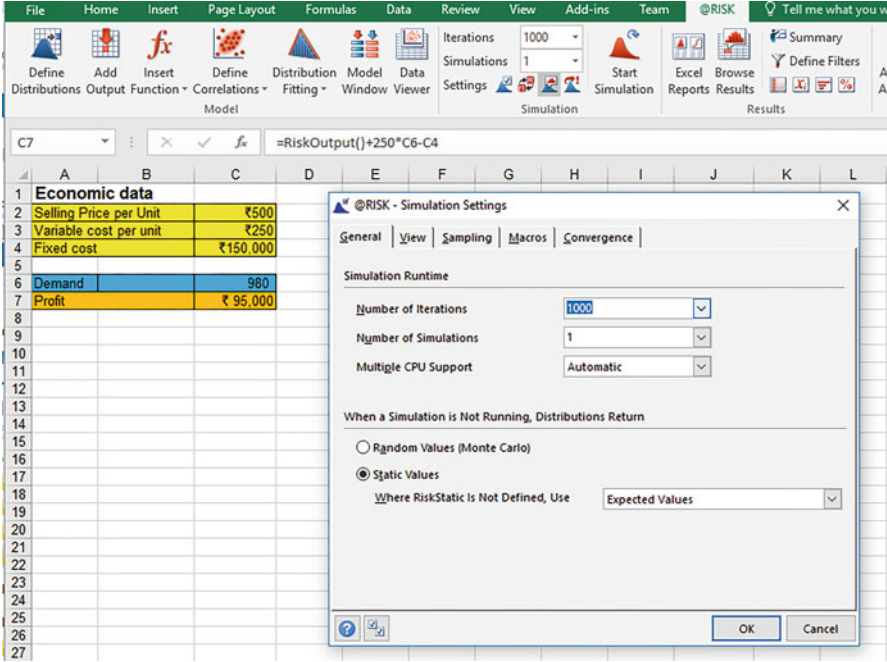- Step 5: We are now ready to run the simulation. We do so by clicking "Start Simulation."

Fig. 10.4 Simulation settings: setting the number of iterations

- Step 6: After @*Risk* runs the simulation, the raw data generated by @*Risk* can viewed by clicking on "Simulation data" (see Fig. 10.6). The column corresponding to the input (demand) shows the 1000 values drawn from the normal distribution with a mean of 980 and a standard deviation of 300. The column corresponding to output (profit) shows the output sample, which is obtained by applying the relation between the input and the output ($Y = 250X - 150,000$) to the input sample. The simulation data can be exported to an Excel file for further analysis. @*Risk* also provides summary statistics and this can be accessed, for example, through "Browse Results" (see Fig. 10.7).

Simulation Results

The summary statistics indicate that the sample mean of profit is Rs 94,448 and the sample standard deviation of profit is Rs 75,457. Therefore, our estimate of the retailer's expected profit from the simulation model is Rs 94,448 and that of the standard deviation is Rs 75,457. Interpreting the area under the histogram between an interval as an estimate of the probability that the random variable is contained in that range, we estimate $P(Y <= 0)$ as 9.9% and $P(Y => 100,000) = 47.8\%$. Therefore, based on the simulation model we think there is roughly a 10% chance of not breaking even and a 48% chance of the profits exceeding Rs 100,000.
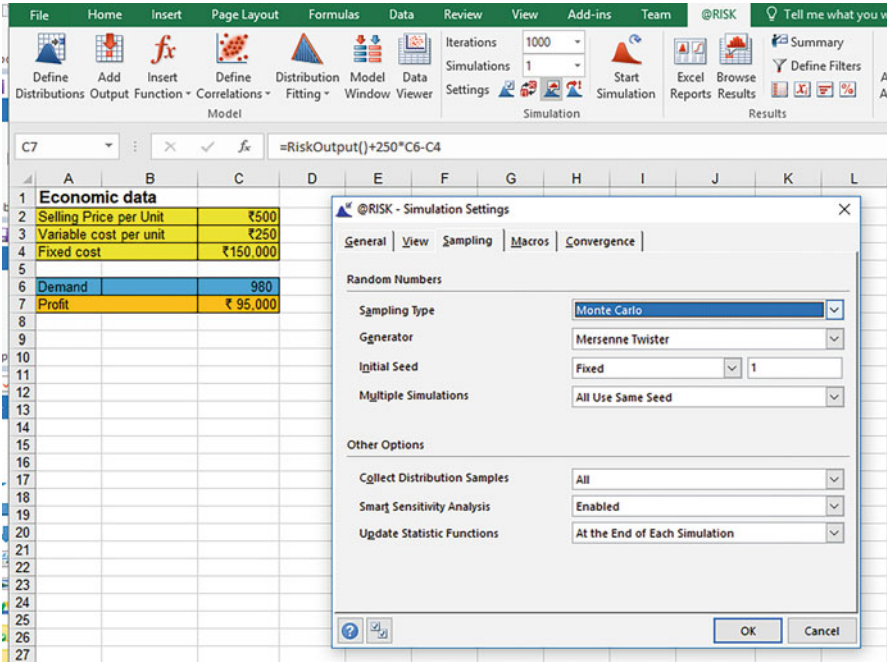
Fig. 10.5 Simulation settings: sampling type, generator, and initial seed
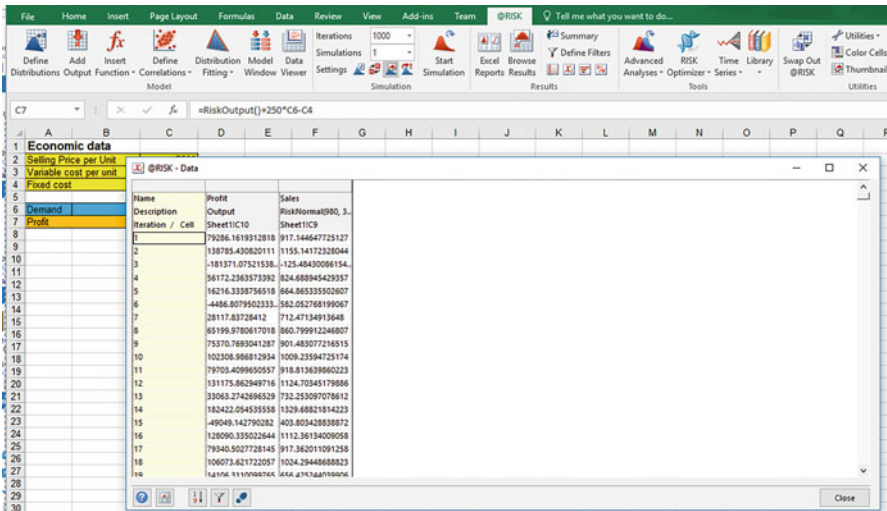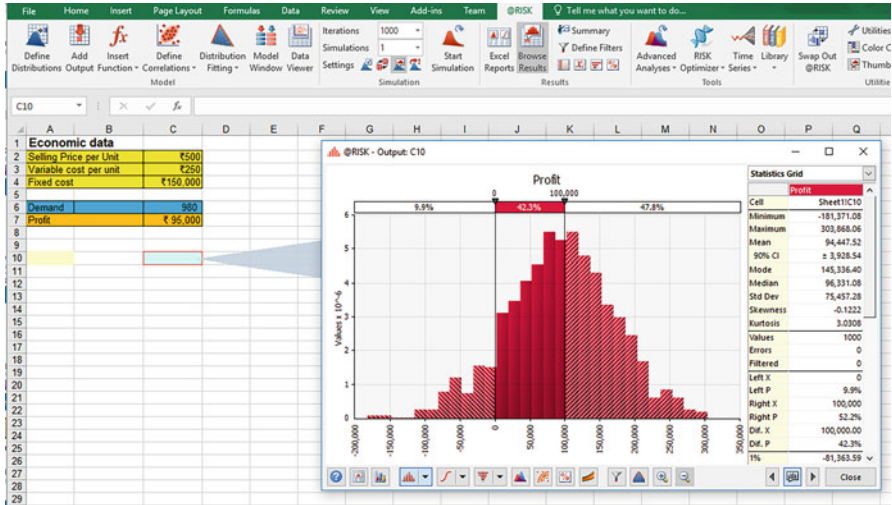

Fig. 10.6 Simulation data

Fig. 10.7 Simulation statistics

Table 10.1 Comparison of simulation results with the results from the exact analysis

	Simulation model	Exact analysis
Expected profit (in Rs)	94,448	95,000
Standard deviation of profit (in Rs)	75,457	75,000
Sample size (no. of iterations)	1000	
95% confidence interval	[89771, 99124]	
P(Y <= 0)	9.9%	10.3%
P(Y >= 100,000)	47.8%	47.3%

Comparison of the Simulation Results to the Exact Analysis

Table 10.1 summarizes the results obtained from the simulation model and compares it to the exact analysis. We observe that the simulation results are close but not exact. The error comes from finite sampling. Moreover, the results are random in that if we had taken a different random sample of size 1000 of the input, the results would have been slightly different (see Table 10.2). Therefore, we assess the accuracy of the simulation results by building a confidence interval. We use the formula $\bar{x} \pm 1.96\, s/\sqrt{n}$ to construct a 95% confidence interval for the expected profit (see, e.g., Stine and Foster (2014)).

Table 10.2 compares the simulation results obtained using different seeds. The second column shows the simulation results when the initial seed is set to 1, and the last column shows the simulation results when the initial seed is set to 2. Note that the estimates change a little when we change the seed since we change the sample of the input random variable.

Table 10.2 Comparison of simulation results for different values of the initial seed

	Simulation model Initial seed = 1	Simulation model Initial seed = 2
Expected profit (in Rs)	94,448	92,575
Standard deviation of profit (in Rs)	75,457	73,367
Sample size (no. of iterations)	1000	1000
95% confidence interval	[89771, 99124]	[88027, 97122]
P(Y <= 0)	9.9%	10.0%
P(Y >= 100,000)	47.8%	44.0%

Table 10.3 Comparison of simulation results for different sample sizes

Sample size (no. of iterations)	Sample mean (in Rs)	Sample standard deviation (in Rs)	Confidence interval
10	37,560	87,536	[−16696, 91816]
100	89,151	79,561	[73557, 104745]
1000	94,448	75,457	[89771, 99124]
10,000	94,392	74,536	[92931, 95853]

The initial seed is fixed to 1

We note that the question as to which is the "right" seed is not meaningful since the estimates always have sampling error associated with them. The more relevant question to ask is regarding the sample size (or the number of iterations in @Risk) since this determines the accuracy of the simulation results. Table 10.3 shows how the accuracy of the simulation estimates changes with the sample size. We observe that as the sample size increases, we obtain progressively more accurate estimates; the sample mean of profit we obtain from the simulation model is closer to the population mean (95,000) and the confidence intervals for the expected profit are also narrower. This is a natural consequence of the central limit theorem, which states that as the sample size increases, the sample mean gets more and more concentrated around the population mean (see, e.g., Ross (2013)). As a result, we get more accurate estimates as the sample size increases.

The natural question that then arises is: what should be the appropriate sample size? This in general depends on the nature of the business question being answered by the simulation model and so tends to be quite subjective. The main ideas come from statistical sampling theory. We first fix a margin of error, that is, a target width of the confidence interval that we are comfortable with. We then determine the sample size so that the actual width of the confidence interval matches with the target width. Therefore, if we let e be the target margin of error (for a 95% confidence interval), then we determine the sample size n as

$$n = \left[\frac{1.96 \, s_n}{e} \right]^2.$$

In the above equation, s_n is the sample standard deviation, which in turn depends on the sample size n. We break this dependence by running the simulation model for

a small number of iterations (say, 100 or 200) to obtain an estimate of the sample standard deviation s_n and use this estimate in the above formula to determine the appropriate value of n. For example, we might use the sample standard deviation corresponding to 100 iterations ($s_{100} = 79{,}561$) from Table 10.3 as an approximation to s_n in the above equation. So if we have a margin of error $e = 1000$, the required sample size n is

$$n = \left[\frac{1.96 * 79561}{1000} \right]^2 \approx 24317.$$

3.1 Decision Making Under Uncertainty

We have used simulation so far simply as an evaluation tool to obtain the summary statistics associated with a performance metric. We now build on this and add a decision component to our model. In particular, we consider business settings where we have to make decisions under uncertainty and see how we can use simulation to inform decision making.

Suppose that we have to pick an action "a" from a range of possible alternatives $\{a_1, \ldots, a_K\}$. In making this decision, we have a performance metric (output) in mind and we would like to choose the action that optimizes this performance metric. The output variable is affected not only by the action that we take (decisions) but also by uncertain events (inputs). That is, if we let Y denote the output variable, we have $Y = f(X, a)$, where X represents the random input and $f(.,.)$ is a function that relates the output to our decisions as well as the random inputs. If the function $f(.,.)$ is nonlinear, then the mathematical analysis quickly becomes challenging. In such cases, a simulation model can be a useful alternative.

The simulation approach remains quite similar to what we have discussed in the previous section. The first step is to identify the decision variable, as well as the input and the output random variables. We specify the range of the decision variable, the distribution of the input random variable and the relation between the output and the decision and the input variables. Once we do this, we evaluate the outcomes associated with each possible action using simulation and then pick the action that optimizes the output metric. That is, we pick an action a_k from the list of possible actions. We then generate a random sample of the input $X_1, \ldots, X_n$. Given the action a_k and the random sample of the input, we obtain a random sample of the output using $Y_i = f(X_i, a_k)$. We use the random sample of the output to estimate the performance metric of interest when we choose action a_k. We repeat this process for each action in the list and pick the action that optimizes the performance metric. Note that since the results of the simulation are random and approximate, the previous statements regarding the interpretation of the simulation results continue to hold.

We next build on the fashion retailer example to illustrate these steps.

Example 10.2 (Fashion Retailer, Continued)

A fashion retailer purchases a fashion product for Rs 250 and sells it for Rs 500. The retailer has to place an order for the product before the start of the selling season when the demand is uncertain. The demand for the product is normally distributed with a mean of 980 and a standard deviation of 300. If the retailer is left with unsold items at the end of the selling season, it disposes them off at a salvage value of Rs 100. The retailer also incurs a fixed cost of Rs 150,000, which is independent of the sales volume. How many units of the product should the retailer order? Assume that the retailer's objective is to maximize its expected profit.

Solution

We first specify the decision variable, as well as the input and the output random variables. Then we relate the output to the input and the decision variables.

- Step 1: Specify decision—stocking quantity. We let q denote the stocking decision. The theoretical range for the stocking quantity is $[0, \infty)$. However, for practical reasons (minimum batch size, budget constraints, etc.) we may want to impose lower and upper limits on the stocking quantity. Here, given that the demand is normally distributed, we consider order quantities that are within one standard deviation above and below the mean. So we will consider $q \in \{680, 730, \ldots, 1280\}$ (q increases in step size $= 50$). We note that we could have considered an extended range and even a more refined range for the decision variable. The trade-off is the increased solution time stemming from evaluating the profit (output) at each possible value of the decision.
- Step 2: Specify input—demand. We denote demand by the random variable X. X is normally distributed with a mean of 980 and a standard deviation of 300.
- Step 3: Specify output—profit. We denote profit by the random variable Y.
- Step 4: Relate the input and the decision to the output. The unit profit margin is Rs. 250. We have sales $= min(X, q)$ since we cannot sell more than what is demanded (X) and what we have in stock (q). It follows that the number of unsold items is the difference between the stocking quantity and the sales. That is, *Unsold = q − Sales*. We have *Revenue = 250 Sales + 100 Unsold*, where the first term captures the revenues from sales at the full price while the second term captures the salvage value of the unsold items. On the other hand, *Cost = 250 q + 150,000*. Therefore, we have that the profit *Y = Revenue − Cost = 250 Sales + 100 Unsold − 250q − 150,000*.

Now, we proceed to build a simulation model to determine the optimal stocking quantity. We note that the problem we are considering is an example of a newsvendor problem, which is a very well-studied model in the operations management literature and it is possible to obtain an analytical expression for the optimal ordering quantity (see, e.g., Porteus (2002)). We do not dwell on the exact mathematical analysis here and instead focus on the simulation approach. An advantage of the

simulation model is that it can be easily adapted to the case where the retailer may have to manage a portfolio of products as well as accommodate a number of other business constraints.

Simulation Model

We have already described how the decision variable together with the random input affects the output. We evaluate the expected profit corresponding to each stocking level in the list {680, 730, ..., 1280} and pick the stocking level that achieves the highest expected profit. So we pick a stocking level q_k from the above list. Then we generate a random sample of size n of the input. So, let $X_1, \ldots, X_n$ be a random sample of size n of the demand drawn from a normal distribution with a mean of 980 and a standard deviation of 300. Given the sample of the input, we then obtain a sample of the output by using the relation between the decision, the input, and the output. Let $Y_1, \ldots, Y_n$ be the sample of size n the output (profit) where

$$Y_i = 250 \, Sales_i + 100 \, Unsold_i - 250q_k - 150000$$
$$= 250 \min(X_i, q_k) + 100 \, (q_k - \min(X_i, q_k)) - 250q_k - 150000.$$

Finally, we use the sample of the output to estimate the expected profit corresponding to ordering q_k units. We repeat this process for all the possible stocking levels in the list to determine the optimal decision.

Implementing the simulation model in @Risk: We specify the decision cell corresponding to the stocking quantity using the RISKSIMTABLE(.) function. The argument to the RISKSIMTABLE(.) function is the list of stocking quantities that we are considering (see Fig. 10.8). The remaining steps are similar to the previous simulation model: we specify the input cell corresponding to demand using the RISKNORMAL(.) function and specify the output cell corresponding to profit using the RISKOUTPUT(.) function. We link the decision variable and the input to the output using the relation described above.

Before we run the simulation model, we specify the "Number of Iterations" in the "Simulation Settings" tab. We set the "Number of Iterations" to be 1000 as before. Now we would like to generate an input sample of size 1000 for each possible value of the decision variable. There are 13 possible stocking decisions that we are considering ({680, ..., 1280}) and so we would like an input sample of size 1000 to be generated 13 times, one for each stocking decision in the list. We specify this by setting the "Number of Simulations" to be 13 (see Fig. 10.9). In the "Sampling" tab, we set the "Sampling Type," "Generator," and "Initial Seed" as before. We also set "Multiple Simulations" to use the same seed (see Fig. 10.10). This ensures that the same input sample of size 1000 is used to evaluate the profits corresponding to all of the stocking decisions. That is, the same underlying uncertainty drives the outcomes associated with the different actions. As a result, it is more likely that any differences in the output (expected profit) associated with the different decisions (stocking quantity) are statistically significant. The benefit of using common random numbers can be formalized mathematically (see, e.g., Ross (2013)).
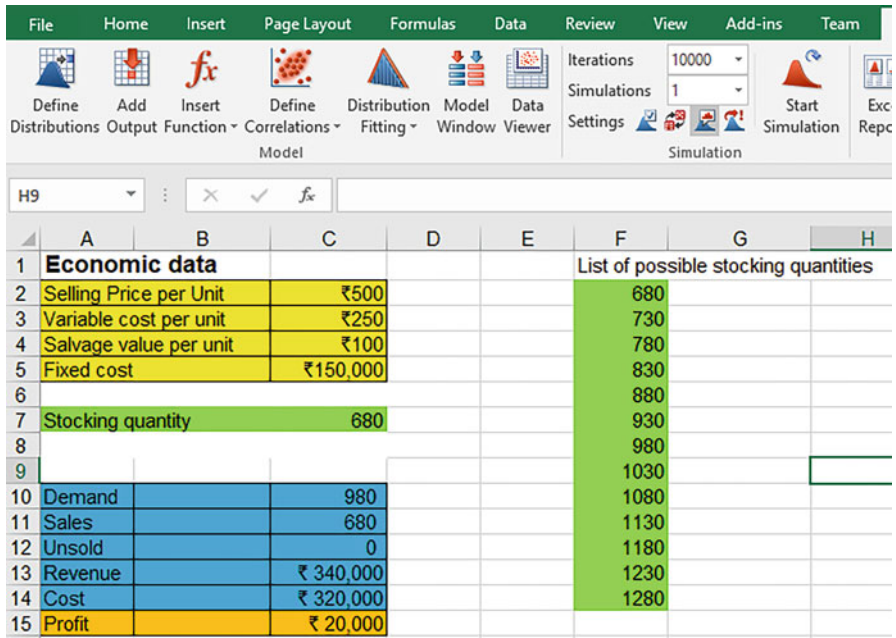
Fig. 10.8 Determining the optimal stocking quantity using simulation. The decision cell is specified using the RISKSIMTABLE(.) function and the argument to the function is the list of possible stocking quantities described in the list F2:F14 in the Excel spreadsheet

After running the simulation, we view the simulation results by clicking on the "Summary" button. The table summarizes the simulation statistics for the different values of the ordering quantity (see Fig. 10.11). It indicates that the expected profit is maximized by ordering 1080 units.

We make two observations regarding the simulation results. First, the expected profit for each stocking level is estimated as the sample average of a sample of size 1000. It remains to be verified that the results are statistically significant. That is, are the differences we see in the samples representative of a corresponding difference in the populations? This question can be answered using standard statistical tests (see, e.g., Stine and Foster (2014)). Second, we note that ordering 1080 units is the best choice from the list {680, . . . , 1280}. It is possible that we could further increase profits if we were not restricted to these ordering levels. We can evaluate the profits on a more refined grid of ordering levels to check if this is indeed the case. The trade-off is the increased computation time stemming from evaluating the profit for a larger number of stocking levels.

Our analysis so far was based on the assumption that the retailer was interested in maximizing its expected profit for a single product. We now extend the model in a couple of directions.
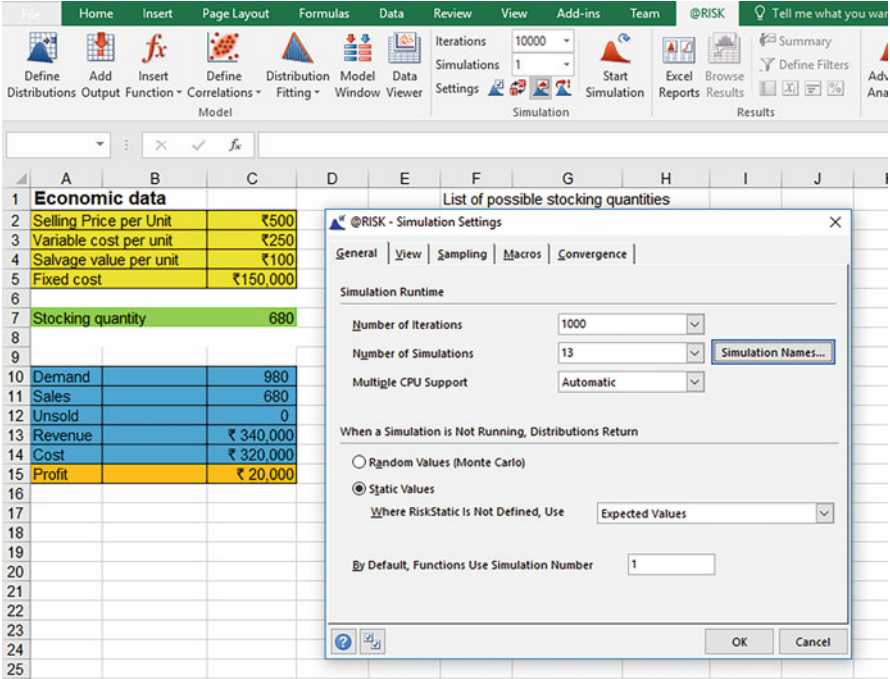
Fig. 10.9 Specifying "Number of Simulations"

Example 10.3 (Fashion Retailer, Stockouts)

Suppose that the retailer also cares for opportunities lost. That is, if the retailer runs out of inventory of the product then some customers who visit the store would be unable to purchase the product. If stockouts occur frequently and a large number of customers do not find the product available, the customers may take their business elsewhere and this would impact the profits of the retailer in the long run. Consequently, the retailer would also like to factor in lost demand when making the stocking decisions. Suppose that the retailer would like to ensure that the expected percentage unmet demand is no larger than 5%. How many units of the product should the retailer order in order to maximize expected profits while ensuring that the stockout constraint is satisfied?

Solution

We can easily modify our simulation model to answer this question. The decision variable and the input random variable remain the same. In addition to the profit metric, we add the expected percentage unmet demand as a second metric in our simulation model. We define the percentage unmet demand as $max(X - q, 0)/X$, where X is the demand and q is the stocking level. Note that there is unmet demand
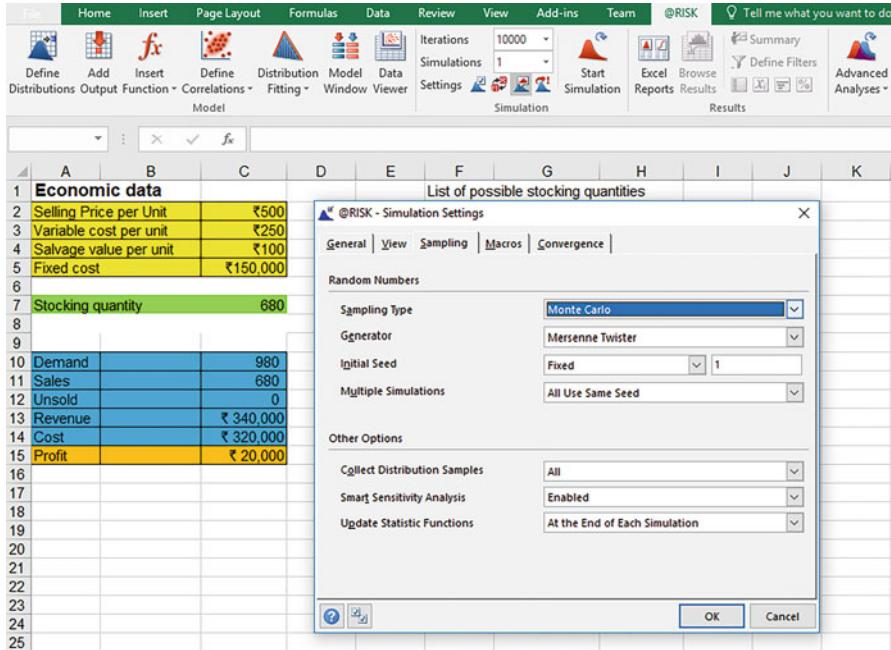
Fig. 10.10 Specifying the seed used for the different simulations. We set multiple simulations to use the same seed
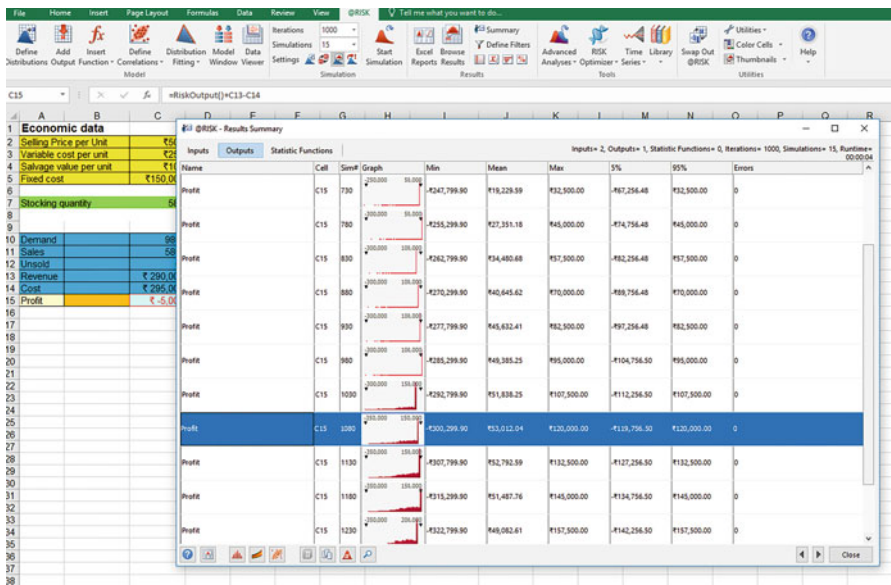

Fig. 10.11 Results summary

Table 10.4 Expected profit and expected % unmet demand as a function of the ordering quantity

Stocking quantity	Expected profit	Exp. % unmet demand
680	₹10,102	28.94
730	₹19,229	24.95
780	₹27,351	21.31
830	₹34,480	17.98
880	₹40,645	14.94
930	₹45,632	12.24
980	₹49,385	9.87
1030	₹51,838	7.83
1080	₹53,012	6.11
1130	₹52,792	4.70
1180	₹51,487	3.54
1230	₹49,082	2.61
1280	₹45,836	1.85

only when $X > q$ and so $X - q > 0$. We add the percentage unmet demand as a second output in our spreadsheet model (using the RISKOUTPUT(.) function).

Table 10.4 describes the summary results we obtain after running the simulation model again. The first column gives the stocking level, the second column gives the corresponding expected profit while the last column gives the expected percent unmet demand. We notice that the previous stocking level of 1080 which maximized the expected profit results in the expected percent unmet demand being around 6%, violating the 5% threshold. The feasible stocking levels, that is, those which satisfy the 5% stockout constraint, are {1130, 1180, 1230, 1280}. Among these stocking levels, the one which maximizes the expected profit is 1130. Therefore, based on our simulation model we should order 1130 units of the product, which results in an expected profit of around Rs 52,792 and the expected percent unmet demand being around 4.7%.

Example 10.4 (Fashion Retailer, Multiple Products)

We now consider the case where the retailer sells multiple products that are potential substitutes. That is, if a customer does not find the product that she is looking for on the shelf, she may switch and buy a product that is a close substitute. To keep things simple, let us assume that the retailer sells two products during the selling season. The base (primary) demands for the two products are independent, normally distributed random variables. The cost and demand characteristics of the two products are described in Table 10.5. In addition, the retailer incurs a fixed cost of Rs 150,000 that is independent of the sales volumes. The two products are substitutes in that if one product is stocked out then some of the customers interested in purchasing that product might switch over and purchase the other product (provided it is in stock). If product 1 is stocked out, then 70% of the unmet demand for that product shifts to product 2. On the other hand, if product 2 is stocked out, then 30% of the demand for that product shifts to product 1. Therefore, there is

Table 10.5 Cost and demand characteristics of the two products

	Product 1	Product 2
Unit selling price	Rs 500	Rs 300
Unit cost price	Rs 250	Rs 200
Unit salvage value	Rs 100	Rs 50
Demand	Normal with a mean of 980 and a standard deviation of 300	Normal with a mean of 2000 and a standard deviation of 500

a secondary demand stream for each product that is created when the other product is stocked out. What should be the stocking levels of the two products for the retailer to maximize its expected profits?

Solution

We determine the decision variable, as well as the input and the output random variables. We then describe the relation between the output and the decision and the input variables.

- Step 1: Specify decisions—stocking quantities of the two products. We let q_1 denote the stocking level of product 1 and q_2 denote the stocking level of product 2. For each product, we consider stocking levels that are within one standard deviation of the mean demand. Therefore, we consider $q_1 \in \{680, 730, \ldots, 1280\}$ and $q_2 \in \{1500, 1550, \ldots, 2500\}$. We again note that it is possible to work with an expanded range of values and also consider a more refined set of grid points, at the expense of greater computational effort.
- Step 2: Specify input—demands for the two products. We let X_1 denote the primary demand random variable for product 1 and X_2 denote the primary demand random variable for product 2. X_1 is normally distributed with a mean of 980 and a standard deviation of 300, while X_2 is normally distributed with a mean of 2000 and a standard deviation of 500. Furthermore, X_1 and X_2 are independent random variables.
- Step 3: Specify output—profit. We denote profit by the random variable Y.
- Step 4: Relate the input and the decision to the output. We have primary $Sales_i = min\ (X_i, q_i)$ for $i = 1, 2$, where primary sales refer to the sales generated from the primary demand for that product. The remaining inventory of product i is therefore $Inventory_i = q_i - Primary\ sales_i$. On the other hand, the portion of the demand that cannot be satisfied from the on-hand inventory, unmet $Demand_i = X_i - Primary\ sales_i$. Now, if there is unmet demand for product 2, then 30% of that is channeled to product 1. Therefore, the secondary sales of product 1 is the smaller of the remaining inventory of product 1 (which remains after satisfying the primary demand for product 1) and the secondary demand for product 2. That is, $Secondary\ Sales_1 = min(Inventory_1, 0.3 * Unmet\ Demand_2)$. By following a similar line of reasoning, the secondary sales of product 2, $Secondary\ Sales_2 = min\ (Inventory_2, 0.7 * Unmet\ Demand_1)$. The number of unsold items of product i is, therefore, $Unsold_i = q_i - Primary\ Sales_i - Secondary\ Sales_i$.

Tallying up the revenues and costs, we have that the *total revenue = 250 * Primary Sales$_1$ + 250 * Secondary Sales$_1$ + 100 * Unsold$_1$ + 200 * Primary Sales$_2$ + 200 * Secondary Sales$_2$ + 50 * Unsold$_2$*. The total cost = $250q_1 + 200q_2 + 150{,}000$. We obtain the profit, Y, as the difference between the total revenue and the total cost.

Implementing the Simulation Model in @Risk

Now we have two decision variables q_1 and q_2, where q_1 takes values in the range {680, 730, ..., 1280} and q_2 takes values in the range {1500, 1550, .., 2500}. So there are 273 (13 * 21) possible combinations of q_1 and q_2 that we have to consider. While it is possible to implement the model using the RISKSIMTABLE(.) function, this would involve creating a list of all the possible combinations of the two decision variables and can be cumbersome. An alternative way to implement the simulation model is using the RISK Optimizer function (see Fig. 10.12). Under "Model Definition" we specify the optimization goal (maximize) and the metric that is optimized (mean value of profit). Under "Adjustable Cell Ranges," we specify the decision variables and the range of values that they can take (between the minimum and the maximum stocking levels, in increments of 50, see Fig. 10.13). We start the optimization routine by clicking on the "Start" button under the RISK Optimizer tab (see Fig. 10.14) and after @Risk has finished, we obtain the results from the "Optimization Summary" report (see Figs. 10.14 and 10.15). From the optimization summary report, we see that the optimal stocking level for product 1 is 1230 units and the optimal stocking level for product 2 is 1600 units. The corresponding expected profit is Rs 228,724. It is interesting to note that the optimal stocking levels of the two products when considered together are different from the
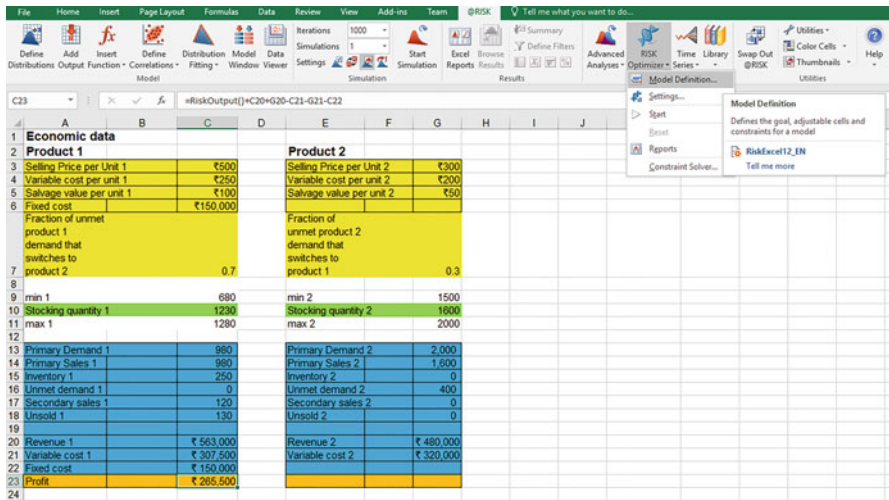

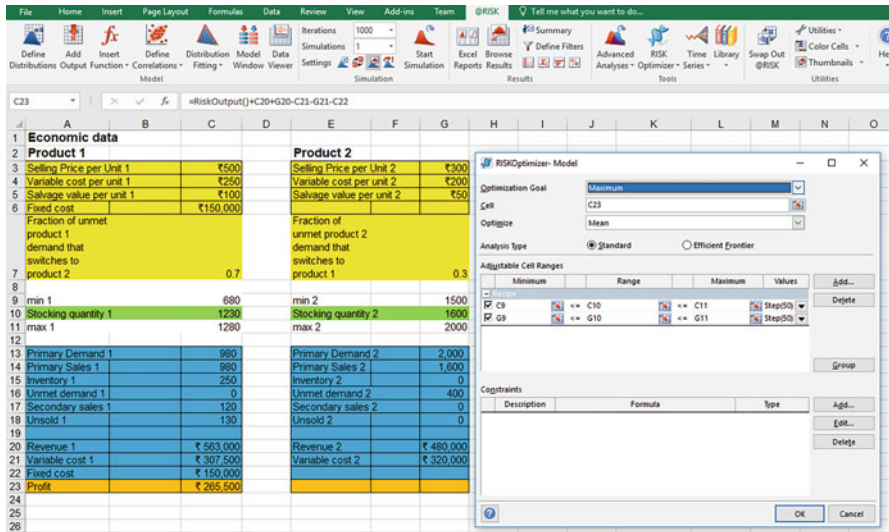
Fig. 10.12 Using RISK Optimizer

Fig. 10.13 Specifying the optimization objective and the decision variables in RISK Optimizer
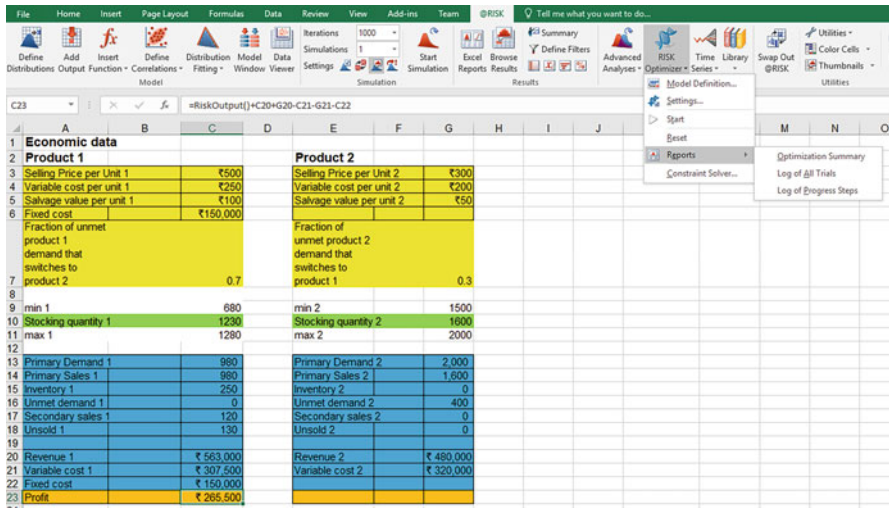

Fig. 10.14 Obtaining the results of the optimization

optimal stockings levels when considered independently. If product 1 is analyzed independently, we have from the earlier analysis (Example 10.2) that its optimal stocking level is 1080 units. On the other hand, if product 2 is considered in isolation, then its optimal stocking level turns out to be 1900 units. When both products are considered together, factoring in the substitution effects, the optimal stocking level of product 1 increases to 1230 units, while that of product 2 decreases to 1600 units.

Goal	
Cell to Optimize	Sheet1!C23
Statistic to Optimize	Mean
Type of Goal	Maximum

Results	
Valid Trials	143
Total Trials	143
Original Value	₹ 183,316
+ soft constraint penalties	0.00
= result	₹ 183,316
Best Value Found	₹ 228,724
+ soft constraint penalties	0.00
= result	₹ 228,724
Best Trial Number	37
Time to Find Best Value	0:00:36
Reason Optimization Stopped	Best solution found
Time Optimization Started	11/15/2017 15:26
Time Optimization Finished	11/15/2017 15:27
Total Optimization Time	0:00:41
Adjustable Cell Values	Sheet1!C10
Original	680
Best	1230
Adjustable Cell Values	Sheet1!G10
Original	2000
Best	1600

Fig. 10.15 Optimization summary report

4 Conclusion

In this chapter, we have given an overview of how simulation can be an effective technique to make decisions in the presence of uncertainty. Simulation is a particularly effective tool when there is a nonlinear relation between the input and the output random variables. In order to build a simulation model, we have to (1) specify the decision variables and their ranges, (2) the input random variables and their distributions, and (3) describe how the output depends on the decisions as well as the inputs. The output of a simulation is (1) random since we work with a random sample of the input and (2) approximate since we work with a finite sample of input. The results of the simulation model therefore have to be carefully interpreted using concepts from statistical sampling theory.

5 Solved Case Study "*Coloring the World*"

Daniel operates a successful civil work firm, *Coloring the World*. He has been running this decades-old family business for the last 10 years. *Coloring the World* provides painting and civil work services to commercial buildings and large apartments. Daniel has a dedicated sales team which is very active in identifying construction projects that may require the firm's services. He has another sales team which keeps following up with the businesses from established customers who may require refurbishment or painting work. Both the teams are quite active in generating leads. He also has a robust customer relationship system in place to engage with existing clients that helps in building long term association with the customers.

An existing customer, *Feel Good Fabrics*, a maker of cotton and linen cloth, has sent *Coloring the World* an RFP (Request for Proposal) to paint its entire manufacturing plant. Though Daniel has provided various small refurbishment work support to this client, he knows that this is a big requirement that can fetch good recognition and lead to a long-term association with the customer. With his earlier experience, Daniel knows that Brian Painters has also been approached for this offer and suspects that Whitney-White colors (W&W) is also trying hard to get empaneled on *Feel Good Fabrics'* vendors list. Daniel does not want to lose this opportunity to create an impactful relationship with *Feel Good's* commercial and operation team.

Daniel has competed with Brian and W&W for many other projects and believes he can more or less estimate the bidding strategies of these competitors. Assuming that these competing firms are bidding for this contract, Daniel would like to develop a bid that offers him a good shot at winning, but also does not result in a loss on the contract since the firm has many expenses including labor, paints, and materials.

Daniel estimates that *Brian's painters* bid could be anywhere between $450,000 and $575,000. As for *Whitney-White colors*, Daniel predicts the bid to be as low as $425,000 or as high as $625,000, but he thinks $550,000 is most likely. If Daniel bids too high, one of the competitors is likely to win the contract and Daniel's company will get nothing. If, on the other hand, he bids too low, he will probably win the contract but may have to settle for little or no profit, even a possible loss.

Due to the complexity in the plant structure of *Feel Good Fabrics*, Daniel in consultation with his service department estimates the direct cost to service the client at $300,000. Realizing that the costs are actually uncertain, Daniel takes this number to be expected value of the costs and thinks that the actual cost will be normally distributed around this mean value with a standard deviation of $25,000.

The preparation and budget estimation cost $10,000 to Daniel since that includes in-person visits and technical test clearance by the *Feel Good Fabrics* team. How much should Daniel bid in order to maximize his expected profit?

Solution

The underlying source of uncertainty is in the bids of the two competitors and in the direct costs. The decision is how much to bid for the project. Note that *Coloring the World* wins the project if they bid the lowest. Given the bids of the

competitors we also observe that *Coloring the World* will win the project for sure if they bid lower than \$425,000. So it does not make sense to bid anything lower than \$425,000 since it does not further improve their chances of winning the contract. On the other hand, bidding lower shrinks their profit margins. *Coloring the World* will also never win the project if they bid more than \$575,000. Therefore, the optimal bidding strategy must lie between \$425,000 and \$575,000.

Simulation Model:

Input random variables
Direct cost (D) = Normal (300000, 25000)
Brian painters' bid (B) = Uniform (450000, 575000)
W&W bid (W) = Triangular(425000, 550000, 625000)
Decision variable
Coloring the World's bid C, Range is between \$425,000 and \$575,000.
Output
Profit = IF (C < min(B, W), C – D, 0) – 10,000

From the simulation model, we find that a bid of \$450,000 is optimal. The expected profits will be \$137,000 on average. The 95% confidence interval is [\$135,000, \$140,000].

The number of simulations is number of times we would like to repeat the simulation. Here we would like to evaluate profits corresponding to the 13 different ordering quantities and so we set "Number of Simulations" to be 13.

Electronic Supplementary Material

All the datasets, code, and other material referred in this section are available in www.allaboutanalytics.net.

- Data 10.1: Priceline_Hotelbids.xlsx
- Data 10.2: Watch_bids.xlsx

Exercise Caselets

Caselet 10.1: Watch the Time

Christopher Luis is a consultant for an international architectural firm that specializes in building skyscrapers. He has consulted for various projects in China, India, USA, and other countries. Because of his profession, Chris travels to different countries to better understand the project location and local culture. All his travels help him customize and personalize designs based on the geography of the site and the local availability of material and labor.

Through his travel, Chris has developed a passion toward collecting antique watches. So much so, he is more famous as a watch collector among his family

and friends than as a famous architect. He has a collection of more than a hundred antique watches representing different cultures and manufacturing styles. He keeps looking for a new variety of mechanical watch to add to his collection during his travel across the globe. He has even subscribed to bidding websites that regularly add a variety of new watches to their product pages. He has been successful many a time in winning auctions but unfortunately at very high prices. Chris believes in paying the right price for the product and an added extra based on his emotional attachment to the product.

Over time, Chris has developed a bidding strategy that combines bidding at selective time points and at a specific price ratio considering the existing bids. He has started winning bids at lower prices in recent times, but he is a little disappointed as his new strategy did not work out with a few watches that he wanted to win desperately. Though, thankfully due to the new strategy he has not paid high prices on the winning deals.

Chris also notices that he is investing a lot of time and effort following up on multiple websites and placing the bid. Some websites even limit the number of bids for each customer and he is running out of number of bids very quickly. The other websites even charge per bid to the customers in order to restrain the customers from placing multiple bids of small differences. His experience tells that the winning bid ranges between $1.5\times$ and $3\times$ of the first few bid prices and follows a "regular" price trajectory that can help in estimating the final price. The price trajectories of the 100 products are provided on the book's website "Watch_bids.xlsx". The table contains the number of bids and the highest bid so far at 36 h before closing, 24 h, 12 h, and so on.

He wishes to develop a new strategy to place a single bid or a maximum of two bids at specific time points rather than following up multiple times and wasting time and effort monitoring the outcome.

- When should Chris bid in order to secure the deal?
- At what price ratio should Chris bid to secure the deal?
- How can he add extra value of emotional quotient to the bid in terms of timing and price?

Caselet 10.2: Winter Jackets

Monika Galore is excited as she graduates as a textile engineer from a premier engineering school in the USA. She received the best student award because of her outstanding innovation and contribution during her internship with a world-famous researcher, *Akihiko Karamorita.* Akihiko is a distinguished professor in the textile and chemical engineering department at her university. She mentored Monika in developing a fabric that not only can protect the skin from a harsh winter but also is light to wear. Alongside completing her dissertation, Monika and Akihiko patented the idea.

Many manufacturers offered them a huge price and life-time employment in research and development teams to sell their idea. *Monika* and *Akihiko* discussed commercializing the product rather than selling the idea to established manufac-

turers. They decided that they will contract the cloth and garment manufacturing to a quality and trusted supplier and sell the product online through e-commerce websites.

Under the women entrepreneurship scheme, Monika decided to set up her own plant to manufacture the fabric. Professor Akihiko approached her ex-students who had expertise in winter-wear manufacturing and designing. A few of them showed interest and also bid for the contract. Monika and Akihiko decided to contract the manufacturing of jackets to a third party. They outsourced the product manufacturing to *Vimal Jain*, a close friend of Monika and trusted student of Professor Akihiko, considering his family experience in textile manufacturing. *Vimal* proposed two designs—a Sports jacket and a Trendy jacket using the special fabric. Though he also proposed to mix this specific thread with another material in order to cater to different geographic needs, Monika and Akihiko rejected the idea and decided to target the niche segment of woolen cloths. Vimal agreed to design, manufacture, and supply the two types of jackets and in different sizes small, medium, large, and extra-large to the retail locations.

The product was initially sold only through an online channel. However, looking at the increasing demand, Monika and Akihiko decided to go offline and partnered with a large retail chain that had store presence across the world. The retailer negotiated a 20% profit margin on the items sold with a condition to return unsold products at 80% of the purchase price. Since the manufacturing of fabric and garment are done at different locations and by different manufacturers, it is essential to estimate the demand in advance to optimize inventory at various stages. Also, as this is a seasonal and/or fashion product, the excess inventory of unsold products may lead to deep discounts. The product demand also depends on the severity of weather.

They requested *Chris Alfo*, head of operations with the partner retailer to estimate the demand based on his experience with comparable products. Note: The numbers in demand estimate table are in thousands (Table 10.6).

Monika estimated the manufacturing cost of one jacket at $125 and fixed 40% markup when selling to the retailer. She thought that retailers would add 20% profit margin. Monika also found that unsold products can be sold at 50% of the cost in deep-discount outlets. She knew that all the stock has to be cleared within the same year considering the contemporary nature of fashion. Looking at historical weather reports, she estimated the probability of mild winter at 0.7 and cold at 0.3. The customers may switch between the product designs based on availability of the product. In case of extreme demand and unavailability, if winter is mild, the probability to switch from one design to another is 0.5 while in case of cold winter it is 0.9. Monika is planning how much to manufacture every year in order to procure

Table 10.6 Demand estimates for sizes and types of jackets

Type of jacket	Small	Medium	Large	Extra-large
Sports	150	300	180	90
Trendy	210	450	365	120

raw material and finalize the manufacturing contract with Vimal. She has to estimate demand in such a way that she does not end up with too much unsold product, as well as, does not lose the opportunity to sell more.

Caselet 10.3: Priceline—A Reverse Auction

Priceline popularized the name your own price (NYOP). For example, in its website,[3] it advertises "For Deeper Discounts Name Your Own Price®." In this model, called a reverse auction, the buyer specifies the product or service and names a price at which the buyer is willing to purchase the product. On the other side, sellers offer products or services at two or more prices and also the number of products that are available. For example, a room in a three-star hotel on September 21st in downtown Manhattan for a one-night stay could be a product. Several hotels would offer, say, anywhere from 1 to 5 rooms at rates such as $240 and $325. When a bid is made, the market-maker (in this case Priceline) picks a seller (a hotel that has made rooms available) at random and sees if there is a room that is available at a rate that is just lower than the bidder's price. If no such price is available, the market-maker chooses another seller, etc. More details about the model are found in Anderson and Wilson's article.[4]

The NYOP mechanism is fairly complex because it involves buyers, sellers, and the intermediary who are acting with limited information about the actual market conditions. For example, even the intermediary is not aware of the exact supply situation. How does this model benefit everyone? The buyer benefits because it creates a haggle-free environment: In case the bid fails the buyer cannot bid for a day, thus forcing the bidder to either reveal the true reservation price or be willing to forgo an opportunity to purchase in order to learn more about the model and act with delay. The seller benefits because the model avoids direct price competition. The intermediary benefits because it gets to keep the difference between the bidder's and the seller's price.

Barsing is a boutique hotel situated in downtown Manhattan. It has 100 rooms, out of which 65 are more or less indistinguishable with regard to size and amenities. It often finds some of its rooms remain unsold even during peak seasons due to the relative newness of the hotel and its small size. The variable cost of a room-night is around $75. Sometimes this cost may increase or decrease by 10% depending on the amount of cleaning and preparation necessary. Richard Foster who manages Barsing started offering rooms on an NYOP program called HotelsManhattan. He classifies Barsing's as a mid-range hotel with a family atmosphere. He feels that the program was intriguing and requires constant tinkering to get the price right.

Foster's assistant, Sarah, was tasked with reviewing the data and recommending an automatic approach to making rooms available on HotelsManhattan. Historical data on the number of bids made on the past 40 weekdays are available to her, see "Priceline_Hotelbids.xlsx" on book's website. Typically, 4–5 rooms are available to sell using the NYOP program.

---

[3]www.priceline.com (accessed on Aug 17, 2017).

[4]Anderson and Wilson (2011).

1. Assume that the bidders are fixed in number. If Sarah uses one price, what price maximizes the expected profit if four rooms are available?
2. Assume that the number of bidders is random. What is the optimal single price?
3. Assume that the number of bidders is random and Sarah can specify two prices. How should she set those prices to maximize her expected profit from four rooms?

Appendix 1: Generating Random Numbers on a Computer

Virtually all computer simulations use mathematical algorithms to generate random variables. These are potentially very fast, and can be replicated at will. However, it should be clear that no such sequence can be truly random. Consequently, we refer to such a sequence as a pseudo-random sequence. This is because one can in fact predict the sequence of numbers generated provided one had a sophisticated knowledge of the way the algorithm is designed to operate. This is not a huge concern from a practical standpoint as most of the popular algorithms generate sequences that are virtually indistinguishable from a truly random sequence.

A commonly used generator is the linear congruential generator (LCG), which obtains a sequence $x_0, x_1, \ldots$ of integers via the recursion

$$x_{n+1} = (ax_n + c) \mod m,$$

where a, c, and m are integers also known, respectively, as the multiplier, increment, and modulus. The number x_0 is called the initial seed. The mod operator applied to two numbers returns the remainder when the first number is divided by the second. So 5 mod 3 = 2 and 6 mod 3 = 0. Since each number in the sequence generated lies between 0 and $m - 1$, x_n/m is a number that lies between 0 and 1. Therefore, the sequence $x_1/m, x_2/m, \ldots$ has the appearance of a sequence of random numbers generated from the uniform distribution between 0 and 1. However, note that since each x_n lies between 0 and $m - 1$, the sequence must repeat itself after a finite number of values. Therefore, we would like to choose the values of a, c, and m so that a large number of values can be generated before the sequence repeats itself. Moreover, once the numbers a, c, m, and x_0 are known, the sequence of numbers generated by the LCG is completely deterministic. However, if these numbers were unknown to us and we were only observing the sequence generated by the LCG, it would be very hard for us to distinguish this sequence from a truly random sequence. For example, Fig. 10.16 shows the frequency histogram of the first 100 numbers of the sequence $x_1/m, x_2/m, \ldots$ obtained by setting $a = 16,807$, $m = 2,147,483,647$, $c = 0$, and $x_0 = 33,554,432$. It has the appearance of being uniformly distributed between 0 and 1, and it can be verified that the uniform distribution indeed is the distribution that best fits the data. By setting the seed of a random number generator, we fix the sequence of numbers that is generated by the algorithm. Thus, we are also able to easily replicate the simulation.
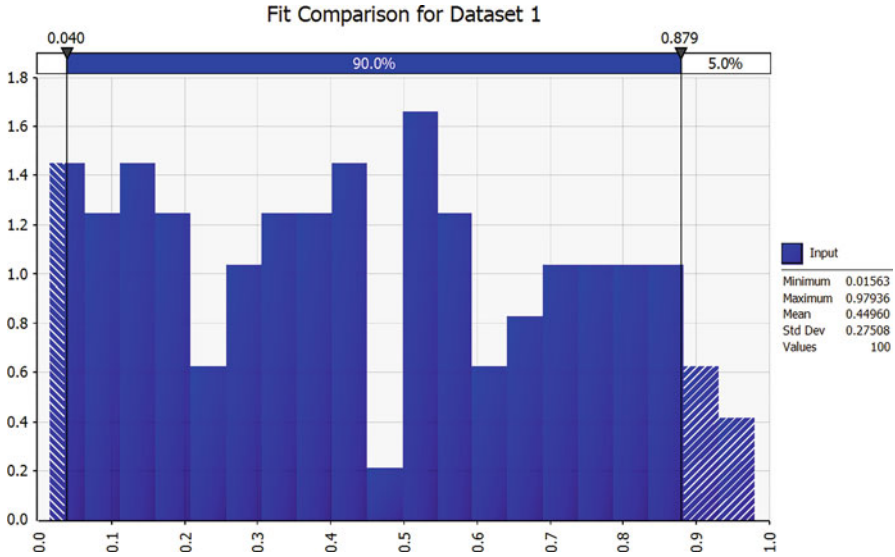
Fig. 10.16 Histogram of the sequence generated by the linear congruential generator with $a = 16,807$, $m = 2,147,483,647$, $c = 0$, and $x_0 = 33,554,432$

The LCG algorithm generates a sequence of numbers that have the appearance of coming from a uniform distribution. It is possible to build on this to generate pseudo-random numbers from other probability distributions (both discrete as well as continuous distributions). We refer the reader to Ross (2013) and Law (2014) for more details on the algorithms and their properties.

Appendix 2: Fitting Distributions to Data

When we build a simulation model, a key assumption is that we know the distributions of the input random variables. In practice, the distributions of the input random variables come from statistically testing the observed data to find out the distribution that best fits the data. The statistical tests are referred to as goodness-of-fit tests. The underlying idea is to compare the observed distribution with the hypothesized distribution and measure the discrepancy between the two. If the discrepancy is small, then it suggests that the hypothesized distribution is a good fit to the observed data. Otherwise, the hypothesized distribution is a poor fit. The error metrics and the hypothesis tests can be formalized; see Ross (2013) and Law (2014) for details. Here we only focus on how we can use @Risk to run the goodness-of-fit tests.

Given historical data, we can use the "Distribution Fitting" tool in @Risk to find out the distribution that best matches the data (see Fig. 10.17). After describing the
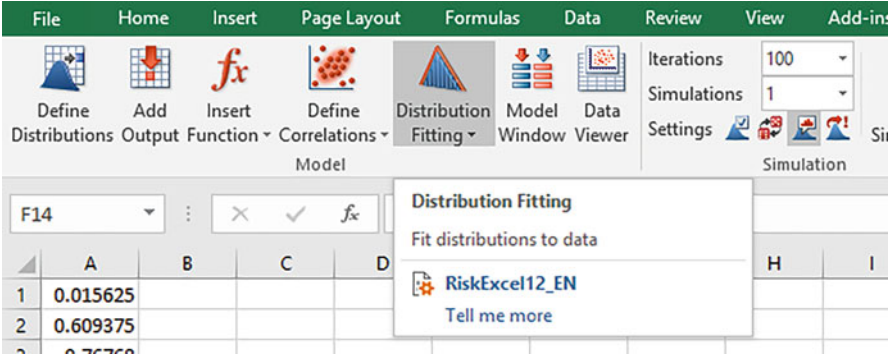
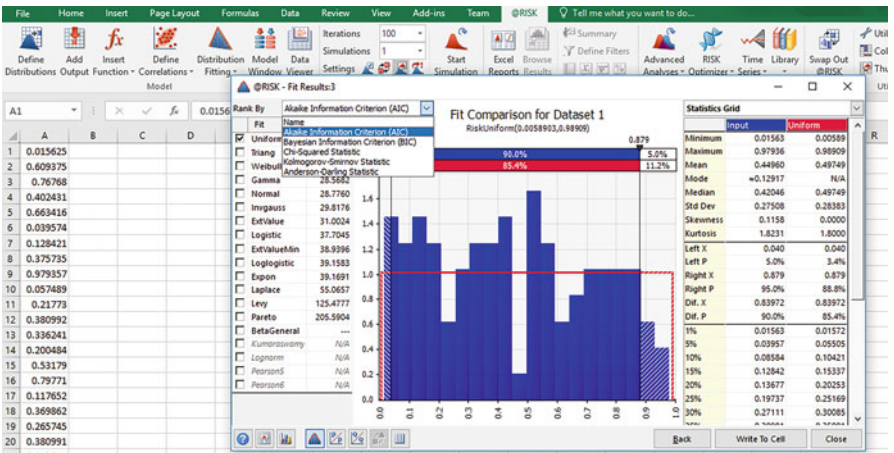**Fig. 10.17** Distribution Fitting tool in @Risk

Fig. 10.18 Goodness-of-fit test results

nature of the data set (discrete or continuous) as well as additional details regarding the range of values it can take (minimum, maximum values), we run the Distribution Fitting tool. This gives us a range of distributions and their corresponding fit values. We can broadly think of the fit values as measuring the error between the observed data and the hypothesized distribution and a smaller fit value indicates a better fit in general. There are different goodness-of-fit tests available. The fit values as well as the relative ranking of the distributions in terms of their fit can vary depending on the test that is used. We refer the reader to Ross (2013) and Law (2014) for more details regarding the different goodness-of-fit tests and when a given test is more applicable (Fig. 10.18).

Appendix 3: Simulation in Excel

Excel has some basic simulation capabilities. The RAND(.) function generates a (pseudo) random number from the uniform distribution between 0 and 1. The RANDBETWEEN(.,.) function takes two arguments a and b, and generates an integer that is uniformly distributed between a and b. There are methods that can use this sequence of uniform random numbers as an input to generate sequences from other probability distributions (see Ross (2013) and Law (2014) for more details). A limitation of using the RAND(.) and RANDBETWEEN(.,.) functions is that it is not possible to fix the initial seed and so it is not possible to easily replicate the results of the simulation.

As mentioned, there are a number of other packages and programming languages that can be used to build simulation models. For example, the code snippet below implements the fashion retailer simulation described in Example 10.1 in R:

Sales = rnorm(1000, 980, 300)

Profit = 250 * Sales − 150,000.

The first line generates a random input sample of size 1000 from the normal distribution with a mean of 980 and a standard deviation of 300. The second line generates a random sample of the output (profit) by using the random sample of the input and the relation between the profit and the sales. Note that the seed as well as the random number generator can be specified in R; refer the R documentation.

References

Anderson, C. K., & Wilson, J. G. (2011). Name-your-own price auction mechanisms – Modeling and future implications. *Journal of Revenue and Pricing Management, 10*(1), 32–39. https://doi.org/10.1057/rpm.2010.46.

Davison, M. (2014). *Quantitative finance: A simulation-based approach using excel*. London: Chapman and Hall.

GE Look ahead. (2015). *The digital twin: Could this be the 21st-century approach to productivity enhancements*. Retrieved May 21, 2017, from http://gelookahead.economist.com/digital-twin/.

Holland, C., Levis, J., Nuggehalli, R., Santilli, B., & Winters, J. (2017). UPS optimizes delivery routes. *Interfaces, 47*(1), 8–23.

Jian, N., Freund, D., Wiberg, H., & Henderson, S. (2016). Simulation optimization for a large-scale bike-sharing system. In T. Roeder, P. Frazier, R. Szechtman, E. Zhou, T. Hushchka, & S. Chick (Eds.), *Proceedings of the 2016 winter simulation conference*.

Law, A. (2014). *Simulation modeling and analysis*. McGraw-Hill Series in Industrial Engineering and Management.

Porteus, E. (2002). *Foundations of stochastic inventory theory*. Stanford business books, Stanford California.

Ross, S. (2013). *Simulation*. Amsterdam: Elsevier.

Stine, R., & Foster, D. (2014). *Statistics for business decision making and analysis*. London: Pearson.

Technology Review. (2017). September 2017 edition.