Appendix A

Selecting and Using Probability Distributions

In this appendix

- Understanding probability distributions
- Selecting a probability distribution
- Using basic distributions
- Using continuous distributions

- Using discrete distributions
- Using the custom distribution
- Truncating distributions
- Comparing the distributions

This appendix explains probability and probability distributions. Understanding these concepts will help you select the right probability distribution for your spreadsheet model. This section describes in detail the distribution types Crystal Ball uses and demonstrates their use with real-world examples.

Understanding probability distributions

For each uncertain variable in a simulation, you define the possible values with a probability distribution. The type of distribution you select depends on the conditions surrounding the variable. For example, some common distribution types are:
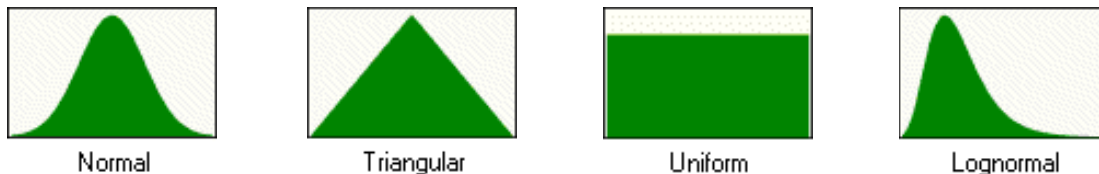

Figure A.1 Common distribution types

During a simulation, the value to use for each variable is selected randomly from the defined possibilities.

A simulation calculates numerous scenarios of a model by repeatedly picking values from the probability distribution for the uncertain variables and using those values for the cell. Commonly, a Crystal Ball simulation calculates hundreds or thousands of scenarios in just a few seconds.

A probability example

To begin to understand probability, consider this example: You want to look at the distribution of non-exempt wages within one department of a large company. First, you gather raw data, in this case the wages of each non-exempt employee in the department. Second, you organize the data into a meaningful format and plot the data as a frequency distribution on a chart. To create a frequency distribution, you divide the wages into groups (also called intervals or bins) and list these intervals on the chart's horizontal axis. Then you list the number or frequency of employees in each interval on the chart's vertical axis. Now you can easily see the distribution of non-exempt wages within the department.

A glance at the chart illustrated in Figure A.2 reveals that the most common wage range is $12.00 to $15.00.

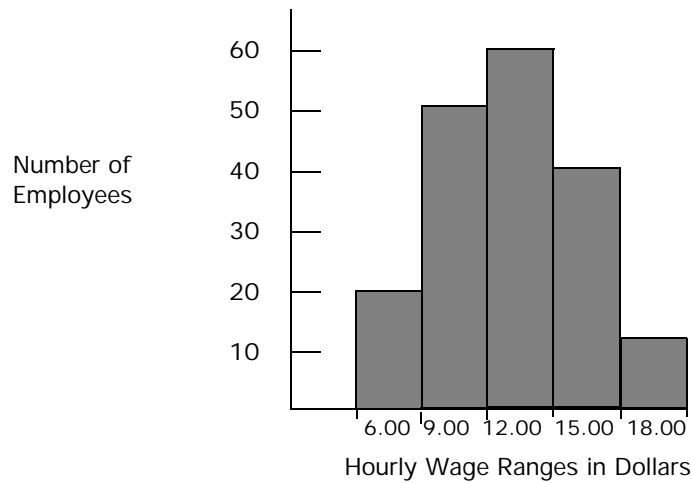Approximately 60 employees (out of a total of 180) earn from $12 to $15.00 per hour.

Figure A.2 Raw frequency data for a probability distribution

You can chart this data as a probability distribution. A probability distribution shows the number of employees in each interval as a fraction of the total number of employees. To create a probability distribution, you divide the number of employees in each interval by the total number of employees and list the results on the chart's vertical axis.

The chart illustrated in Figure A.3 shows you the number of employees in each wage group as a fraction of all employees; you can estimate the likelihood or probability that an employee drawn at random from the whole group earns a wage within a given interval. For example, assuming the same conditions exist at the time the sample was taken, the probability is 0.33 (a 1 in 3 chance) that an employee drawn at random from the whole group earns between $12 and $15 an hour.
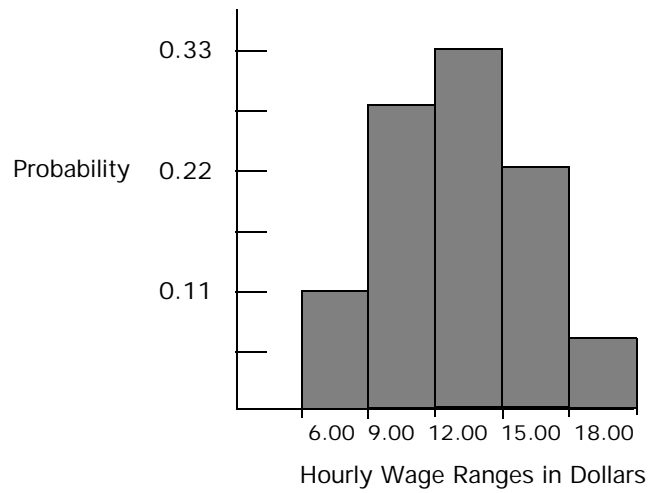
Figure A.3 **Probability distribution of wages**

Compare the probability distribution in the example above to the probability distributions in Crystal Ball (Figure A.4).
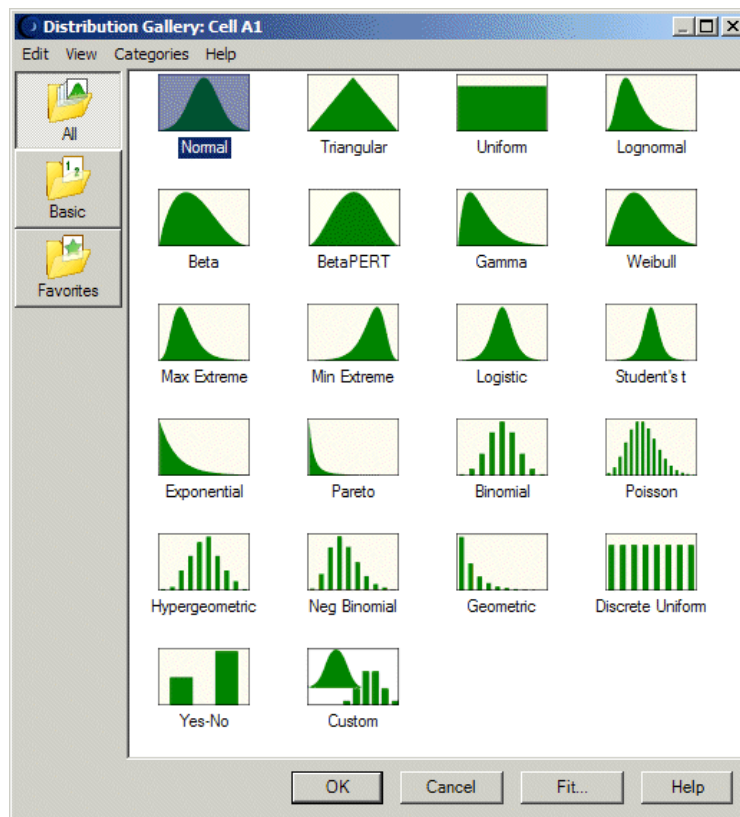

Figure A.4 **Distribution Gallery dialog**

The probability distribution in the example in Figure A.3 has a shape similar to many of the distributions in the Distribution Gallery. This process of plotting data as a frequency distribution and converting it to a probability distribution provides one starting point for selecting a Crystal Ball distribution. Select the distributions in the gallery that appear similar to your probability distribution, then read about those distributions in this chapter to find the correct distribution.

For information about the similarities between distributions, see "Comparing the distributions" on page 334. For a complete discussion of probability distributions, refer to the sources listed in the bibliography.

Discrete and continuous probability distributions

Notice that the Distribution Gallery shows whether the probability distributions are discrete or continuous. Discrete probability distributions describe distinct values, usually integers, with no intermediate values and are shown as a series of vertical columns, such as the binomial distribution at the bottom of Figure A.4 on page 266. A discrete distribution, for example, might describe the number of heads in four flips of a coin as 0, 1, 2, 3, or 4.

Continuous probability distributions, such as the normal distribution, describe values over a range or scale and are shown as solid figures in the Distribution Gallery. Continuous distributions are actually mathematical abstractions because they assume the existence of every possible intermediate value between two numbers. That is, a continuous distribution assumes there is an infinite number of values between any two points in the distribution.

However, in many situations, you can effectively use a continuous distribution to approximate a discrete distribution even though the continuous model does not necessarily describe the situation exactly.

In the dialogs for the discrete distributions, Crystal Ball displays the values of the variable on the horizontal axis and the associated probabilities on the vertical axis. For the continuous distributions, Crystal Ball does not display values on the vertical axis since, in this case, probability can only be associated with areas under the curve and not with single values.

For more information on the separate probability distributions and how to select them, see these sections:
- Continuous distribution descriptions beginning on page 271
- Discrete distribution descriptions beginning on page 301
- Custom distribution description beginning on page 316

Crystal Ball Note: Initially, the precision and format of the displayed numbers in the probability and frequency distributions come from the cell itself. To change the format, see "Customizing chart axes and axis labels" on page 143.

Selecting a probability distribution

Plotting data is one guide to selecting a probability distribution. The following steps provide another process for selecting probability distributions that best describe the uncertain variables in your spreadsheets.

To select the correct probability distribution:

- Look at the variable in question. List everything you know about the conditions surrounding this variable.

   You might be able to gather valuable information about the uncertain variable from historical data. If historical data are not available, use your own judgment, based on experience, to list everything you know about the uncertain variable.

   For example, look at the variable "patients cured" that was discussed in the Vision Research tutorial in Chapter 2 of the Crystal Ball Getting Started Guide. The company must test 100 patients. You know that the patients will either be cured or not cured. And, you know that the drug has shown a cure rate of around 0.25 (25%). These facts are the conditions surrounding the variable.

- Review the descriptions of the probability distributions.

   This chapter describes each distribution in detail, outlining the conditions underlying the distribution and providing real-world examples of each distribution type. As you review the descriptions, look for a distribution that features the conditions you have listed for this variable.

- Select the distribution that characterizes this variable.

   A distribution characterizes a variable when the conditions of the distribution match those of the variable.

   The conditions of the variable describe the values for the parameters of the distribution in Crystal Ball. Each distribution type has its own set of parameters, which are explained in the following descriptions.

   For example, look at the conditions of the binomial distribution, as described on page 302:

   - For each trial, only two outcomes are possible: success or failure.

- The trials are independent. What happens on the first trial does not affect the second trial, and so on.

- The probability of success remains the same from trial to trial.

Now check the "patients cured" variable in Tutorial 2 in the Crystal Ball Getting Started Guide against the conditions of the binomial distribution:

- There are two possible outcomes: the patient is either cured or not cured.

- The trials (100) are independent of each other. What happens to the first patient does not affect the second patient.

- The probability of curing a patient 0.25 (25%) remains the same each time a patient is tested.

Since the conditions of the variable match the conditions of the binomial distribution, the binomial distribution would be the correct distribution type for the variable in question.

• If historical data are available, use distribution fitting to select the distribution that best describes your data.

Crystal Ball can automatically select the probability distribution that most closely approximates your data's distribution. The feature is described in detail in "Fitting distributions to data" beginning on page 29. You can also populate a custom distribution with your historical data.

After you select a distribution type, determine the parameter values for the distribution. Each distribution type has its own set of parameters. For example, there are two parameters for the binomial distribution: trials and probability. The conditions of a variable contain the values for the parameters. In the example used, the conditions show 100 trials and 0.25 (25%) probability of success.

In addition to the standard parameter set, each continuous distribution (except uniform) also lets you choose from alternate parameter sets, which substitute percentiles for one or more of the standard parameters. For more information on alternate parameters, see "Alternate parameter sets" on page 27.

Using basic distributions

This section describes distributions in the Basic category of the Distribution Gallery.
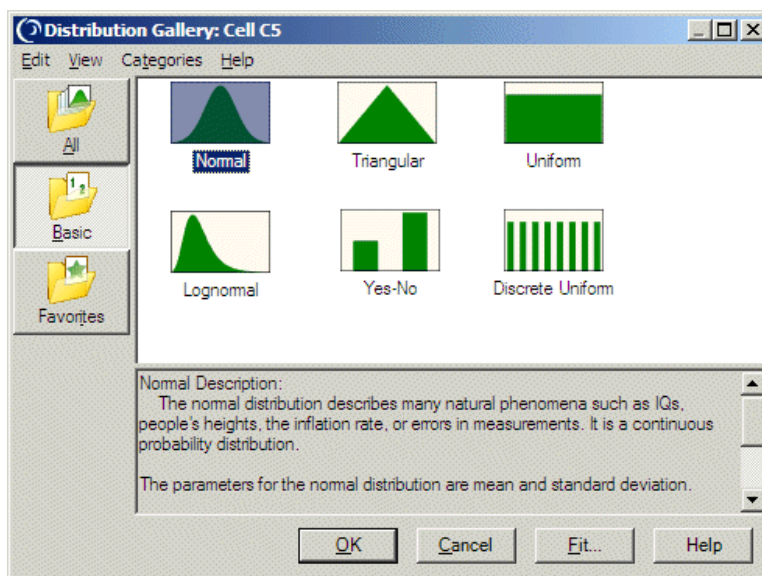

Figure A.5 Distribution Gallery, Basic category

Basic distributions are listed below in the same order they appear above. For details, see the page references below the names.
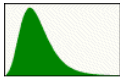
Table A.1 Summary of basic distributions

Shape	Name	Summary
	Normal (page 290)	The normal distribution is the most important distribution in probability theory because it describes many natural phenomena, such as people's IQs or heights. Decision-makers can use the normal distribution to describe uncertain variables such as the inflation rate or the future price of gasoline.
	Triangular (page 294)	The triangular distribution describes a situation where you know the minimum, maximum, and most likely values to occur. For example, you could describe the number of cars sold per week when past sales show the minimum, maximum, and usual number of cars sold.
	Uniform (page 297)	In the uniform distribution, all values between the minimum and maximum occur with equal likelihood.

Table A.1 Summary of basic distributions (Continued)

Shape	Name	Summary
Lognormal	Lognormal (page 285)	The lognormal distribution is widely used in situations where values are positively skewed, for example in financial analysis for security valuation or in real estate for property valuation.
Yes-No	Yes-no (page 314)	The yes-no distribution is a discrete distribution that describes a set of observations that can have only one of two values, such as yes or no, success or failure, true or false, or heads or tails.
Discrete Uniform	Discrete uniform (page 304)	In the discrete uniform distribution, all integer values between the minimum and maximum are equally likely to occur. It is the discrete equivalent of the continuous uniform distribution.

Using continuous distributions

Continuous probability distributions describe values over a range or scale and are shown as solid figures in the Distribution Gallery. Continuous distributions are actually mathematical abstractions because they assume the existence of every possible intermediate value between two numbers. That is, a continuous distribution assumes there is an infinite number of values between any two points in the distribution.

In many situations, you can effectively use a continuous distribution to approximate a discrete distribution even though the continuous model does not necessarily describe the situation exactly. For a comparison of continuous and discrete distributions, see page 267.

The continuous distributions listed in Table A.2 are described later in this section in alphabetical order. Page references appear below the names.
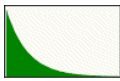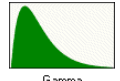
Table A.2 Summary of continuous distributions
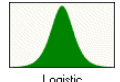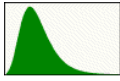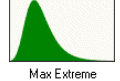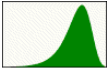
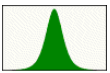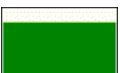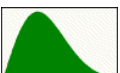| Shape	Name	Summary
Beta	Beta (page 274)	The beta distribution is a very flexible distribution commonly used to represent variability over a fixed range. It can represent uncertainty in the probability of occurrence of an event. It is also used to describe empirical data and predict the random behavior of percentages and fractions.
BetaPERT	BetaPERT (page 276)	The betaPERT distribution describes a situation where you know the minimum, maximum, and most likely values to occur. For example, you could describe the number of cars sold per week when past sales show the minimum, maximum, and usual number of cars sold. It is similar to the triangular distribution, described on page 294, except the curve is smoothed to reduce the peak. The betaPERT distribution is often used in project management models to estimate task and project durations.
Exponential	Exponential (page 278)	The exponential distribution is widely used to describe events recurring at random points in time or space, such as the time between failures of electronic equipment, the time between arrivals at a service booth, or repairs needed on a certain stretch of highway. It is related to the Poisson distribution, which describes the number of occurrences of an event in a given interval of time or space.
Gamma	Gamma (page 280)	The gamma distribution applies to a wide range of physical quantities and is related to other distributions: lognormal, exponential, Pascal, Erlang, Poisson, and chi-squared. It is used in meteorological processes to represent pollutant concentrations and precipitation quantities. The gamma distribution is also used to measure the time between the occurrence of events when the event process is not completely random. Other applications of the gamma distribution include inventory control, economics theory, and insurance risk theory.
Logistic	Logistic (page 283)	The logistic distribution is commonly used to describe growth (the size of a population expressed as a function of a time variable). It can also be used to describe chemical reactions and the course of growth for a population or individual.
Lognormal	Lognormal (Basic) (page 285)	The lognormal distribution is widely used in situations where values are positively skewed, for example in financial analysis for security valuation or in real estate for property valuation.
Max Extreme	Maximum extreme (page 287)	The maximum extreme distribution is commonly used to describe the largest value of a response over a period of time: for example, in flood flows, rainfall, and earthquakes. Other applications include the breaking strengths of materials, construction design, and aircraft loads and tolerances. This distribution is also known as the Gumbel distribution. and is closely related to the minimum extreme distribution, its "mirror image."

Table A.2 Summary of continuous distributions (Continued)

Shape	Name	Summary
Min Extreme	Minimum extreme (page 288)	The minimum extreme distribution is commonly used to describe the smallest value of a response over a period of time: for example, rainfall during a drought. This distribution is closely related to the maximum extreme distribution.
Normal	Normal (Basic) (page 290)	The normal distribution is the most important distribution in probability theory because it describes many natural phenomena, such as people's IQs or heights. Decision-makers can use the normal distribution to describe uncertain variables such as the inflation rate or the future price of gasoline.
Pareto	Pareto (page 291)	The Pareto distribution is widely used for the investigation of distributions associated with such empirical phenomena as city population sizes, the occurrence of natural resources, the size of companies, personal incomes, stock price fluctuations, and error clustering in communication circuits.
Student's t	Student's t (page 293)	The Student's t distribution is used to describe small sets of empirical data that resemble a normal curve, but with thicker tails (more outliers). For sets of data larger than 30, you can use the normal distribution instead.
Triangular	Triangular (Basic) (page 294)	The triangular distribution describes a situation where you know the minimum, maximum, and most likely values to occur. For example, you could describe the number of cars sold per week when past sales show the minimum, maximum, and usual number of cars sold.
Uniform	Uniform (Basic) (page 297)	In the uniform distribution, all values between the minimum and maximum occur with equal likelihood.
Weibull	Weibull (Rayleigh) (page 299)	The Weibull distribution describes data resulting from life and fatigue tests. It is commonly used to describe failure time in reliability studies, and the breaking strengths of materials in reliability and quality control tests. Weibull distributions are also used to represent various physical quantities, such as wind speed.

Crystal Ball Note: As you work with the Crystal Ball probability distributions, you can use the Parameters menu found in the distribution menubar to specify different combinations of parameters. For more information, see "Alternate parameter sets" on page 27.
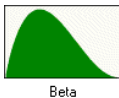
Beta distribution

Parameters

Minimum, Maximum, Alpha, Beta

Conditions

The uncertain variable is a random value between the minimum and maximum value.

The shape of the distribution can be specified using two positive values (Alpha and Beta parameters).

Description

The beta distribution is a very flexible distribution commonly used to represent variability over a fixed range. One of the more important applications of the beta distribution is its use as a conjugate distribution for the parameter of a Bernoulli distribution. In this application, the beta distribution is used to represent the uncertainty in the probability of occurrence of an event. It is also used to describe empirical data and predict the random behavior of percentages and fractions.

The value of the beta distribution lies in the wide variety of shapes it can assume when you vary the two parameters, alpha and beta. If the parameters are equal, the distribution is symmetrical. If either parameter is 1 and the other parameter is greater than 1, the distribution is J-shaped. If alpha is less than beta, the distribution is said to be positively skewed (most of the values are near the minimum value). If alpha is greater than beta, the distribution is negatively skewed (most of the values are near the maximum value). Because the beta distribution is very complex, the methods for determining the parameters of the distribution are beyond the scope of this manual. For more information about the beta distribution and Bayesian statistics, refer to the texts in the Bibliography.

Example

A company that manufactures electrical devices for custom orders wants to model the reliability of devices it produces. After analyzing the empirical data, the company knows that it can use the beta distribution to describe the reliability of the devices if the parameters are alpha = 10 and beta = 2.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the beta distribution:

- The reliability rate is a random value somewhere between 0 and 1.
- The shape of the distribution can be specified using two positive values: 10 and 2.

These conditions match those of the beta distribution.
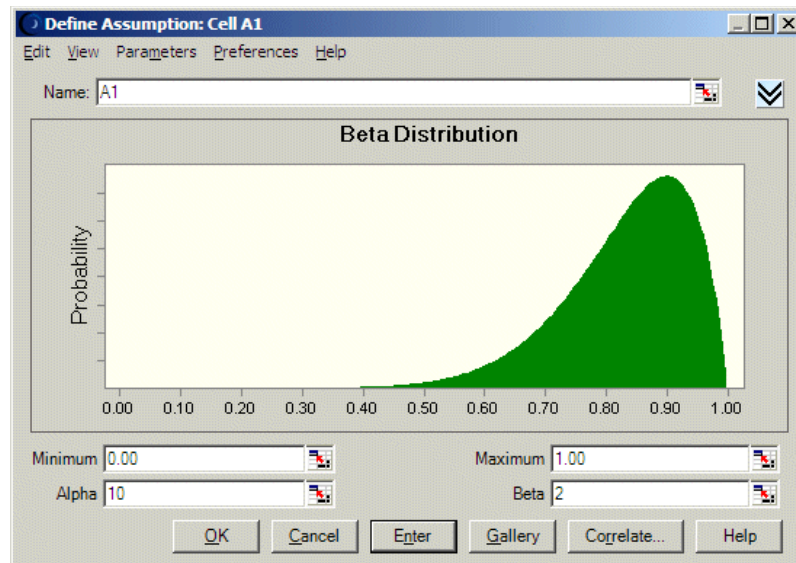
Figure A.6 Beta distribution

Figure A.6 shows the beta distribution with the alpha parameter set to 10, the beta parameter set to 2, and Minimum and Maximum set to 0 and 1. The reliability rate of the devices will be x.

Statistical Note: Models that use beta distributions will run more slowly because of the inverse CDF and alternate parameter calculations that take place when random numbers are handled as part of beta distributions.

BetaPERT distribution

Parameters

Minimum, Likeliest, Maximum

Conditions

The minimum number of items is fixed.

The maximum number of items is fixed.

The most likely number of items falls between the minimum and maximum values, forming a smoothed distribution on the underlying triangle. It shows that values near the minimum and maximum are less likely to occur than those near the most likely value.

Description
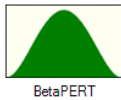
The betaPERT distribution describes a situation where you know the minimum, maximum, and most likely values to occur. This distribution is popular among project managers for estimating task durations and the overall length of a project. For example, you could estimate the duration of a project task which historically takes 24 days to complete, on average, but has taken as few as 18 days under favorable conditions and as long as 32 days in some extreme circumstances.

The betaPERT can also be used in the same situations where a triangular distribution would be used. However, the underlying distribution is smoothed to reduce the peakedness of a standard triangular distribution.

For a discussion of how this distribution relates to the beta distribution, see the description of the betaPERT distribution in Chapter 2 of the Crystal Ball Reference Manual (available through the Crystal Ball Help menu).

Example

A project manager wants to estimate the time (number of days) required for completion of a project. From the manager's past experience, similar projects typically take 7 days to finish, but can be finished in 5 days given favorable conditions, and can take as long as 12 days if things do not happen as expected. The project manager wants to estimate the probability of finishing within 9 days.

The first step in selecting a probability distribution is matching available data with a distribution's conditions. Checking the betaPERT distribution for this project:

- The minimum number of days for completion is 5.

- The maximum number of days for completion is 12.

- The most likely number of days for completion is 7, which is between 5 and 12.

These conditions match those of the betaPERT distribution shown in Figure A.7.
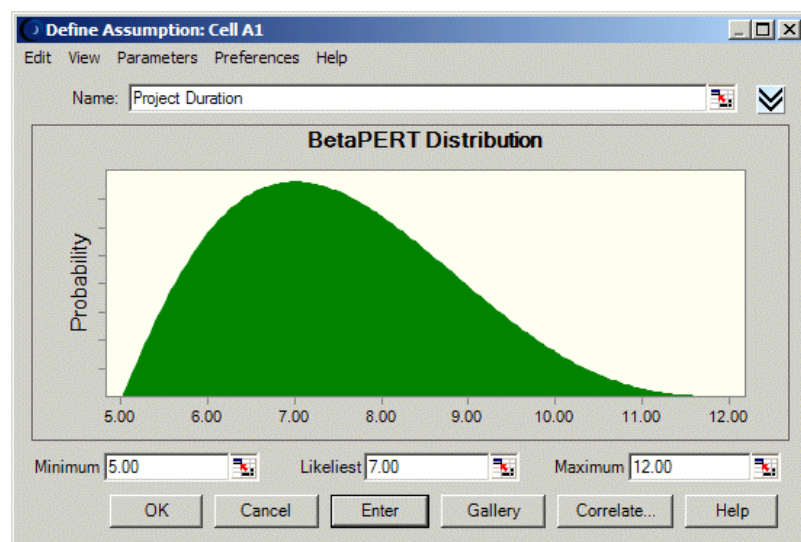

Figure A.7 BetaPERT distribution

When a forecast with formula $=$A1 is created, simulation results show there is about an 88% probability of the project completing within 9 days.

If the same forecast is calculated using a triangular distribution instead of a betaPERT, the probability of completing within 9 days is about 73%.
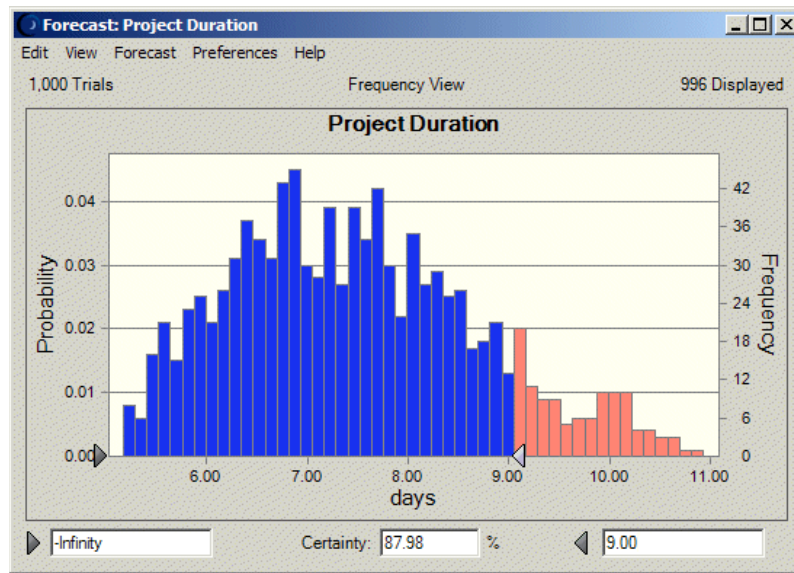
Figure A.8 Project duration based on betaPERT distribution

Exponential distribution

Parameter

Rate

Conditions

The exponential distribution describes the amount of time between occurrences.

The distribution is not affected by previous events.

Description
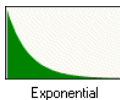
The exponential distribution is widely used to describe events recurring at random points in time or space, such as the time between failures of electronic equipment, the time between arrivals at a service booth, or repairs needed on a certain stretch of highway. It is related to the Poisson distribution, which describes the number of occurrences of an event in a given interval of time or space.

An important characteristic of the exponential distribution is the "memoryless" property, which means that the future lifetime of a given object has the same distribution, regardless of the time it existed. In other words, time has no effect on future outcomes.

Example one

A travel agency wants to describe the time between incoming telephone calls when the calls are averaging about 35 every 10 minutes. This same example was used for the Poisson distribution to describe the number of calls arriving every 10 minutes.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the exponential distribution:

- The travel agency wants to describe the time between successive telephone calls.

- The phone calls are not affected by previous history. The probability of receiving 35 calls every 10 minutes remains the same.

The conditions in this example match those of the exponential distribution.

The exponential distribution has only one parameter: rate. The conditions outlined in this example include the value for this parameter: 35 (calls) every minute or a rate of 35. Enter this value to set the parameter of the exponential distribution in Crystal Ball.
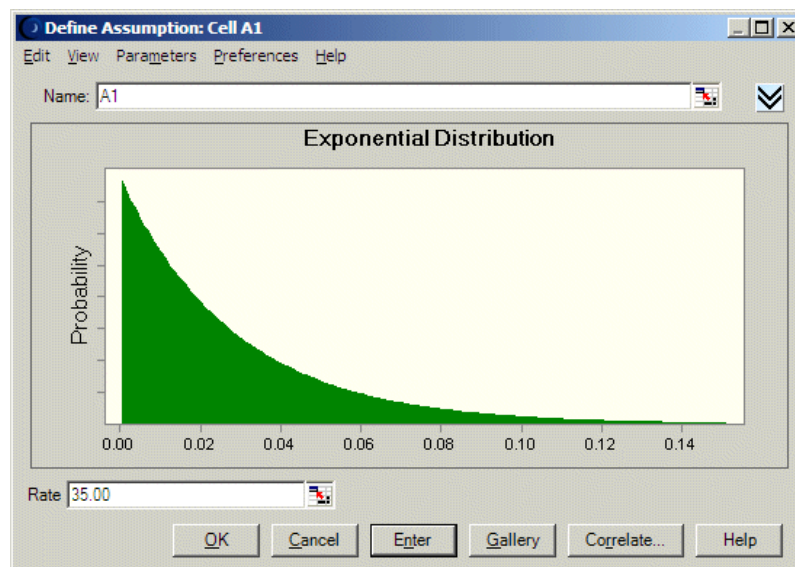

Figure A.9 Exponential distribution

The distribution in Figure A.9 shows the probability that x number of time units (10 minutes in this case) will pass between calls.

Example two

A car dealer needs to know the amount of time between customer drop-ins at his dealership so that he can staff the sales floor more efficiently. The car dealer knows an average of 6 customers visit the dealership every hour.

Checking the exponential distribution:

- The car dealer wants to describe the time between successive customer drop-ins.
- The probabilities of customer drop-ins remain the same from hour to hour.

These conditions fit the exponential distribution.

The resulting distribution would show the probability that x number of hours will pass between customer visits.

Gamma distribution (also Erlang and chi-square)

Parameters

Location, Scale, Shape

Conditions

The gamma distribution is most often used as the distribution of the amount of time until the r th occurrence of an event in a Poisson process. When used in this fashion, the conditions underlying the gamma distribution are:

- The number of possible occurrences in any unit of measurement is not limited to a fixed number.
- The occurrences are independent. The number of occurrences in one unit of measurement does not affect the number of occurrences in other units.
- The average number of occurrences must remain the same from unit to unit.

Description
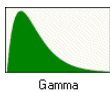
Gamma

The gamma distribution applies to a wide range of physical quantities and is related to other distributions: lognormal, exponential, Pascal, Erlang, Poisson, and chi-square. It is used in meteorological processes to represent pollutant concentrations and precipitation quantities. The gamma

distribution is also used to measure the time between the occurrence of events when the event process is not completely random. Other applications of the gamma distribution include inventory control, economics theory, and insurance risk theory.

Example one

A computer dealership knows that the lead time for re-ordering their most popular computer system is 4 weeks. Based upon an average demand of 1 unit per day, the dealership wants to model the number of business days it will take to sell 20 systems.

Checking the conditions of the gamma distribution:

- The number of possible customers demanding to buy the computer system is unlimited.

- The decisions of customers to buy the system are independent.

- The demand remains constant from week to week.

These conditions match those of the gamma distribution. (Note that in this example the dealership has made several simplifying assumptions about the conditions. In reality, the total number of computer purchasers is finite, and some might have influenced the purchasing decisions of others.)

The shape parameter is used to specify the r th occurrence of the Poisson event. In this example, you would enter 20 for the shape parameter (5 units per week times 4 weeks). The result is a distribution showing the probability that x number of business days will pass until the 20th system is sold.

Figure A.10 illustrates the gamma distribution.
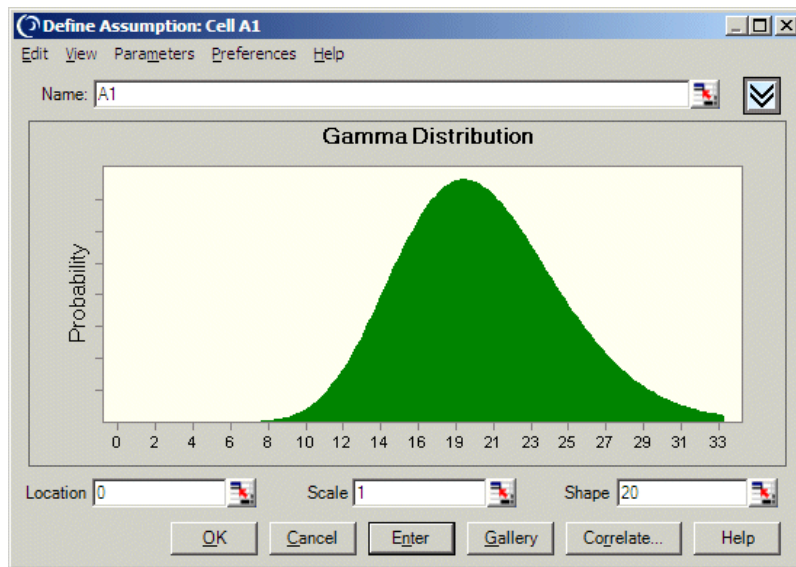
Figure A.10 Gamma distribution

Example two

Suppose a particular mechanical system fails after receiving exactly 5 shocks to it from an external source. The total time to system failure, defined as the random time occurrence of the 5th shock, follows a gamma distribution with a shape parameter of 5.

Some characteristics of the gamma distribution:

- When shape = 1, gamma becomes a scalable exponential distribution.

- The sum of any two gamma-distributed variables is a gamma variable.

- If you have historical data that you believe fits the conditions of a gamma distribution, computing the parameters of the distribution is easy. First, compute the mean $(\bar{x})$ and variance (s^2) of your historical data. Then compute the distribution's parameters:

  shape parameter = $\bar{x}^2/s^2$
  scale parameter = $s^2/\bar{x}$

Chi-square and Erlang distributions

You can model two additional probability distributions, the chi-square and Erlang distributions, by adjusting the parameters entered in the Gamma Distribution dialog. To model these distributions, enter the parameters as described below:

Chi-square distribution

With parameters N and S, where N = number of degrees of freedom and $S = \text{scale}$, set your parameters as follows:

$$\text{shape} = \frac{N}{2} \qquad \text{scale} = 2S^2$$

The chi-square distribution is the sum of the squares of N normal variates.

Erlang distribution

The Erlang distribution is identical to the gamma distribution, except the shape parameter is restricted to integer values. Mathematically, the Erlang distribution is a summation of N exponential distributions.

Logistic distribution

Parameters

Mean, Scale

Description

The logistic distribution is commonly used to describe growth (i.e., the size of a population expressed as a function of a time variable). It can also be used to describe chemical reactions and the course of growth for a population or individual.
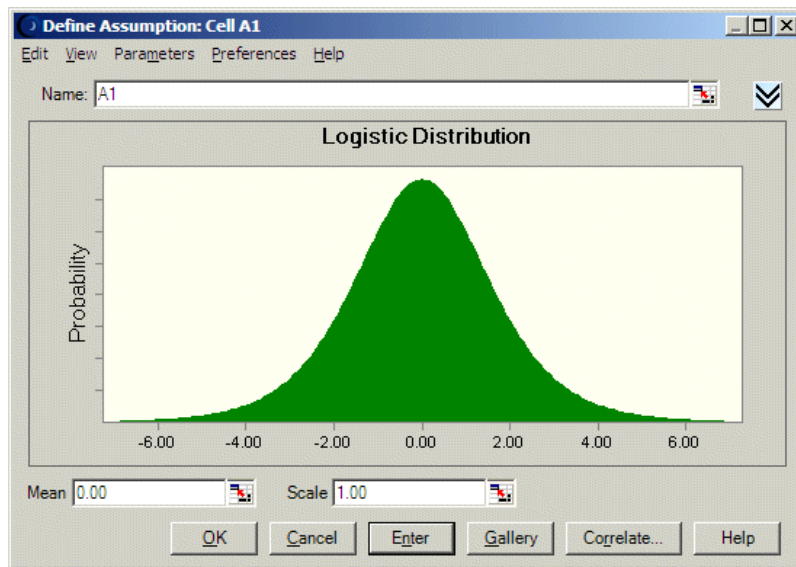
Figure A.11 Logistic distribution

Calculating parameters

There are two standard parameters for the logistic distribution: mean and scale. The mean parameter is the average value, which for this distribution is the same as the mode, since this is a symmetrical distribution.

After you select the mean parameter, you can estimate the scale parameter. The scale parameter is a number greater than 0. The larger the scale parameter, the greater the variance.

To calculate a more exact scale, you can estimate the variance and use the equation:

$$\alpha = \sqrt{\frac{3 \cdot \text{variance}}{\pi^2}}$$

where α is the scale parameter.

Lognormal distribution

Parameters

Mean, Standard Deviation

Conditions

The uncertain variable can increase without limits, but cannot fall below zero.

The uncertain variable is positively skewed with most of the values near the lower limit.

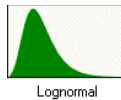**The natural logarithm of the uncertain variable yields a normal distribution.**

Description

The lognormal distribution is widely used in situations where values are positively skewed, **for example in financial analysis for security valuation or in real estate for property valuation.**

Stock prices are usually positively skewed, rather than normally (symmetrically) distributed. Stock prices exhibit this trend because they cannot fall below the lower limit of zero, but might increase to any price without limit.

Similarly, real estate prices illustrate positive skewness **since property values cannot become negative.**

Glossary Term:
skewed, positively—
A distribution in which most of the values occur at the lower end of the range.

Example

The lognormal distribution can be used to model the price of a particular stock. You purchase a stock today at $50. You expect that the stock will be worth $70 at the end of the year. If the stock price drops at the end of the year, rather than appreciating, you know that the lowest value it can drop to is $0.

On the other hand, the stock could end up with a price much higher than expected, thus implying no upper limit on the rate of return. In summary, your losses are limited to your original investment, but your gains are unlimited.

Using historical data, you can determine that the standard deviation of the stock's price is $12.

Statistical Note: If you have historical data available with which to define a lognormal distribution, it is important to calculate the mean and standard deviation of the logarithms of the data and then enter these log parameters using the Parameters menu (Log Mean and Log Standard Deviation). Calculating the mean and standard deviation directly on the raw data will not give you the correct lognormal distribution. Alternatively, use the distribution fitting feature described on page 29.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the lognormal distribution:

- The price of the stock is unlimited at the upper end but cannot drop below $0.

- The distribution of the stock price is positively skewed.

- The natural logarithm of the stock price yields a normal distribution.

These conditions match those of the lognormal distribution (Figure A.12).

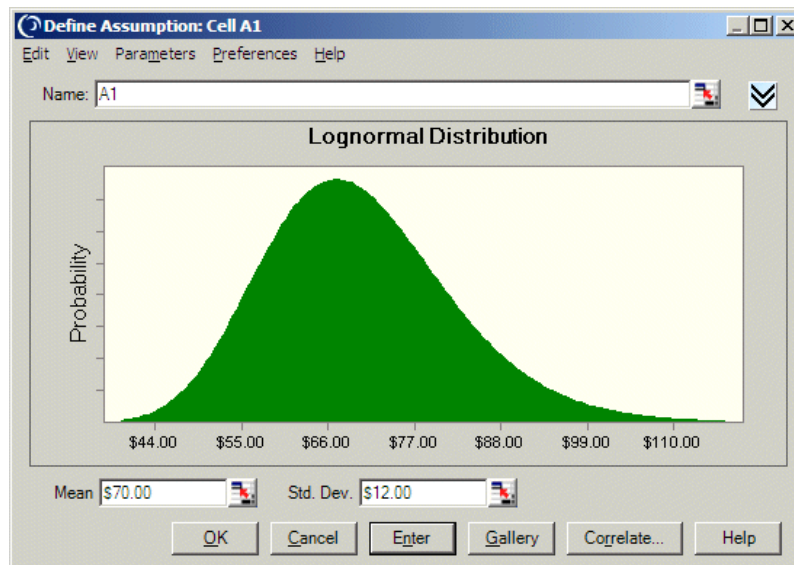Figure A.12 Lognormal distribution

In the lognormal distribution, the mean parameter is set at $70.00 and the standard deviation set at $12.00. This distribution shows the probability that the stock price will be $x.

Lognormal parameter sets

By default, the lognormal distribution uses the arithmetic mean and standard deviation. For applications where historical data are available, it is more

appropriate to use the logarithmic mean and logarithmic standard deviation or the geometric mean and geometric standard deviation. These options are available from the Parameters menu in the menubar.
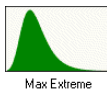
For more information on these alternate parameters, see "Lognormal distribution" in the "Equations and Methods" chapter of the online Crystal Ball Reference Manual. For more information about this menu, see "Alternate parameter sets" on page 27.

Maximum extreme distribution

Parameters

Likeliest, Scale

Description

The maximum extreme distribution is commonly used to describe the largest value of a response over a period of time: for example, in flood flows, rainfall, and earthquakes. Other applications include the breaking strengths of materials, construction design, and aircraft loads and tolerances. The maximum extreme distribution is also known as the Gumbel distribution. This distribution is closely related to the minimum extreme distribution, described beginning on page 288.
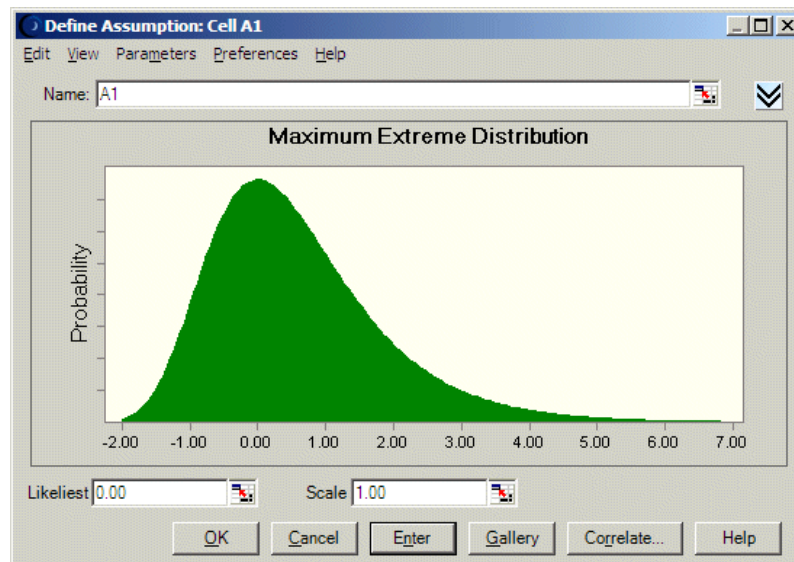

Figure A.13 Maximum extreme distribution

Calculating parameters

There are two standard parameters for the maximum extreme value distribution: Likeliest and Scale. The Likeliest parameter is the most likely value for the variable (the highest point on the probability distribution, or mode).

After you select the Likeliest parameter, you can estimate the Scale parameter. The Scale parameter is a number greater than 0. The larger the Scale parameter, the greater the variance.

To calculate a more exact scale, you can estimate the mean and use the equation:

$$\alpha = \left| \frac{\text{mean} - \text{mode}}{0.57721} \right|$$

where α is the Scale parameter.

Or estimate the variance and use the equation:

$$\alpha = \sqrt{\frac{6 \cdot \text{variance}}{\pi^2}}$$

where α is the Scale parameter.

Minimum extreme distribution
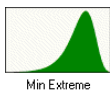
Parameters

Likeliest, Scale

Description

Min Extreme

The minimum extreme distribution is commonly used to describe the smallest value of a response over a period of time: for example, rainfall during a drought. This distribution is closely related to the maximum extreme distribution, described beginning on page 287.
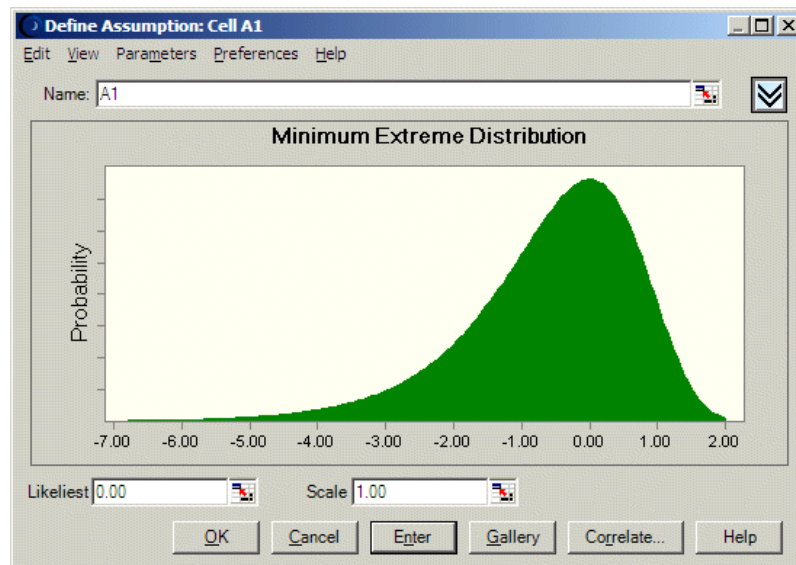
Figure A.14 Minimum extreme distribution

Calculating parameters

There are two standard parameters for the minimum extreme value distribution: Likeliest and Scale. The Likeliest parameter is the most likely value for the variable (the highest point on the probability distribution, or mode).

After you select the Likeliest parameter, you can estimate the Scale parameter. The Scale parameter is a number greater than 0. The larger the Scale parameter, the greater the variance.

To calculate a more exact scale, you can estimate the mean and use the equation:

$$\alpha = \left| \frac{\text{mean} - \text{mode}}{0.57721} \right|$$

where α is the Scale parameter.

Or estimate the variance and use the equation:

$$\alpha = \sqrt{\frac{6 \cdot \text{variance}}{\pi^2}}$$

where α is the Scale parameter.

Normal distribution

Parameters

Mean, Standard Deviation

Conditions

Some value of the uncertain variable is the most likely (the mean of the distribution).

The uncertain variable could as likely be above the mean as it could be below the mean (symmetrical about the mean).

The uncertain variable is more likely to be in the vicinity of the mean than far away.

Statistical Note: Of the values of a normal distribution, approximately 68% are within 1 standard deviation on either side of the mean. The standard deviation is the square root of the average squared distance of values from the mean.

Description
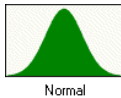
The normal distribution is the most important distribution in probability theory because it describes many natural phenomena, such as people's IQs or heights. Decision-makers can use the normal distribution to describe uncertain variables such as the inflation rate or the future price of gasoline.

The following example shows a real-world situation that matches (or closely approximates) the normal distribution conditions. A more detailed discussion of calculating standard deviation follows this example.

Example

The normal distribution can be used to describe future inflation. You believe that 4% is the most likely rate. You are willing to bet that the inflation rate could as likely be above 4% as it could be below. You are also willing to bet that the inflation rate has a 68% chance of falling somewhere within 2% of the 4% rate. That is, you estimate there is approximately a two-thirds chance that the rate of inflation will be between 2% and 6%.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the normal distribution:

- The mean inflation rate is 4%.
- The inflation rate could as likely be above or below 4%.

- The inflation rate is more likely to be close to 4% than far away. In fact, there is approximately a 68% chance that the rate will lie within 2% of the mean rate of 4%.

These conditions match those of the normal distribution.

The normal distribution uses two parameters: Mean and Standard Deviation. Figure A.15 shows the values from the example entered as parameters of the normal distribution in Crystal Ball: a mean of 0.04 (4%) and a standard deviation of 0.02 (2%).
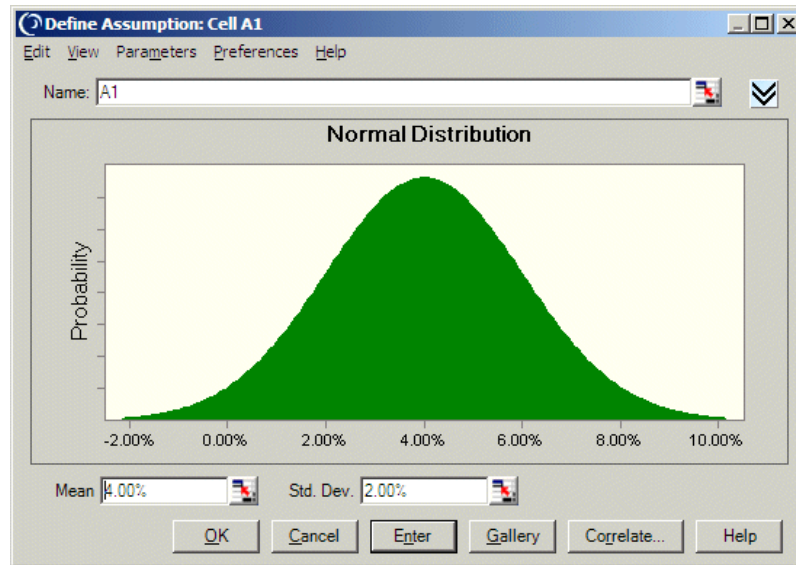
Figure A.15 Normal distribution

The distribution in Figure A.15 shows the probability of the inflation rate being a particular percentage.

Pareto distribution

Parameters

Location, Shape

Description
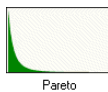
The Pareto distribution is widely used for the investigation of distributions associated with such empirical phenomena as city population sizes, the occurrence of natural resources, the size of companies, personal incomes, stock price fluctuations, and error clustering in communication circuits.
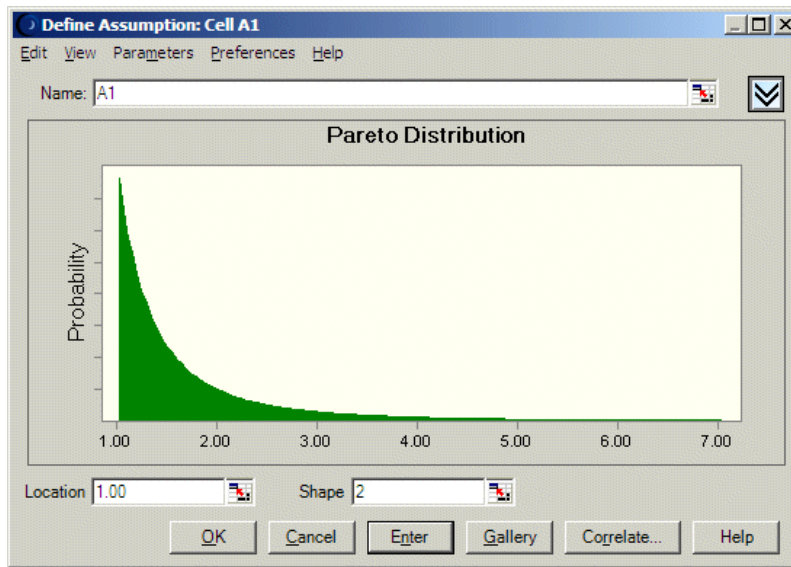
Figure A.16 Pareto distribution

Calculating parameters

There are two standard parameters for the Pareto distribution: Location and Shape. The Location parameter is the lower bound for the variable.

After you select the Location parameter, you can estimate the Shape parameter. The Shape parameter is a number greater than 0, usually greater than 1. The larger the Shape parameter, the smaller the variance and the thicker the right tail of the distribution appears.

To calculate a more exact shape, you can estimate the mean and use the equation (for shapes greater than 1):

$$\text{mean} = \frac{\beta \cdot L}{\beta - 1}$$

where β is the Shape parameter and L is the Location parameter. You can use Excel Solver to help you calculate this parameter, setting the constraint of $\beta > 1$.

Or estimate the variance and use the equation (for shapes greater than 2):

$$\text{variance} = \frac{\beta \cdot L^2}{(\beta - 2)(\beta - 1)^2}$$

where β is the Shape parameter and L is the Location parameter. You can use Excel Solver to help you calculate this parameter, setting the constraint of $\beta > 2$.
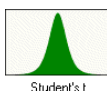
Student's t distribution

Parameters

Midpoint, Scale, Degrees of Freedom

Conditions

The values are distributed symmetrically about the midpoint.

The likelihood of values at the extreme ends is greater than those of the normal distribution.

Description

In classical statistics, the Student's t distribution is used to describe the mean statistic for small sets of empirical data when the population variance is unknown. Classically, degrees of freedom is typically defined as the sample size minus 1.

For purposes of simulation, the Student's t distribution resembles a normal curve, but with thicker tails (more outliers) and greater peakedness (high kurtosis) in the central region. As degrees of freedom increase (at around 30), the distribution approximates the normal distribution. For degrees of freedom larger than 30, you should use the normal distribution instead. The Student's t is a continuous probability distribution.

Since the Student's t distribution has an additional parameter than controls the shape of the distribution (Degrees of Freedom) over the normal distribution, the greater flexibility of the Student's t distribution is sometimes preferred for more precise modeling of nearly normal quantities found in many econometric and financial applications.

The default parameters for the Student's t distribution are Midpoint, Scale, and Degrees of Freedom.
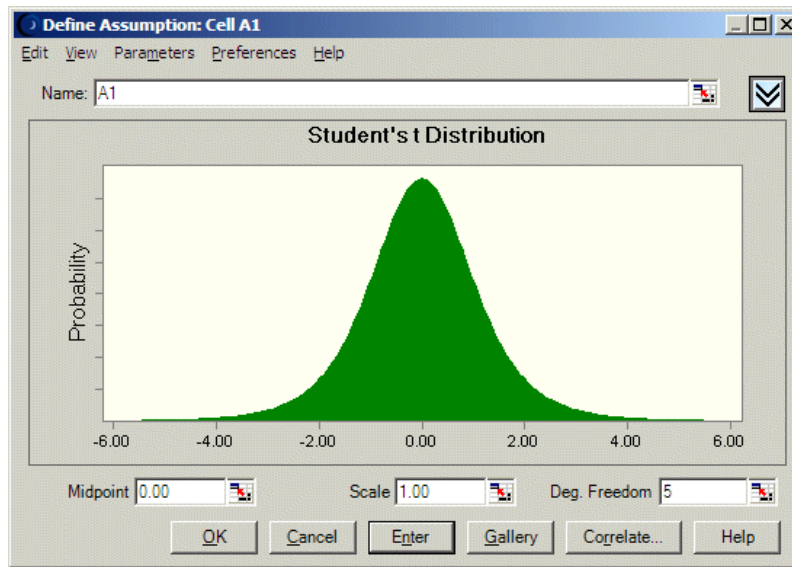
Figure A.17 Student's t distribution

The Midpoint parameter is the central location of the distribution (also mode), the x-axis value where you want to place the peak of the distribution. The Degrees of Freedom parameter controls the shape of the distribution. Smaller values result in thicker tails and less mass in the center. The Scale parameter affects the width of the distribution by increasing the variance without affecting the overall shape and proportions of the curve. Scale can be used to widen the curve for easier reading and interpretation. For example, if the midpoint were a large number, say 5000, the scale could be proportionately larger than if the midpoint were 500.

Example

For examples, see "Normal distribution" on page 290. The uses are the same except that the sample degrees of freedom will be < 30 for the Student's t distribution.

Triangular distribution

Parameters

Minimum, Likeliest, Maximum

Conditions

The minimum number of items is fixed.

The maximum number of items is fixed.

The most likely number of items falls between the minimum and maximum values, forming a triangular-shaped distribution, which shows that values near the minimum and maximum are less likely to occur than those near the most likely value.

Description

The triangular distribution describes a situation where you know the minimum, maximum, and most likely values to occur. For example, you could describe the number of cars sold per week when past sales show the minimum, maximum, and usual number of cars sold.

Triangular

Example one

An owner needs to describe the amount of gasoline sold per week by his filling station. Past sales records show that a minimum of 3,000 gallons to a maximum of 7,000 gallons are sold per week, with most weeks showing sales of 5,000 gallons.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the triangular distribution:

- The minimum number of gallons is 3,000.

- The maximum number of gallons is 7,000.

- The most likely number of gallons (5,000) falls between 3,000 and 7,000, forming a triangle.

These conditions match those of the triangular distribution.

The triangular distribution has three parameters: Minimum, Likeliest, and Maximum. The conditions outlined in this example contain the values for these parameters: 3,000 (Minimum), 5,000 (Likeliest), and 7,000 (Maximum). You would enter these values as the parameters of the triangular distribution in Crystal Ball.

The following triangular distribution shows the probability of x number of gallons being sold per week.
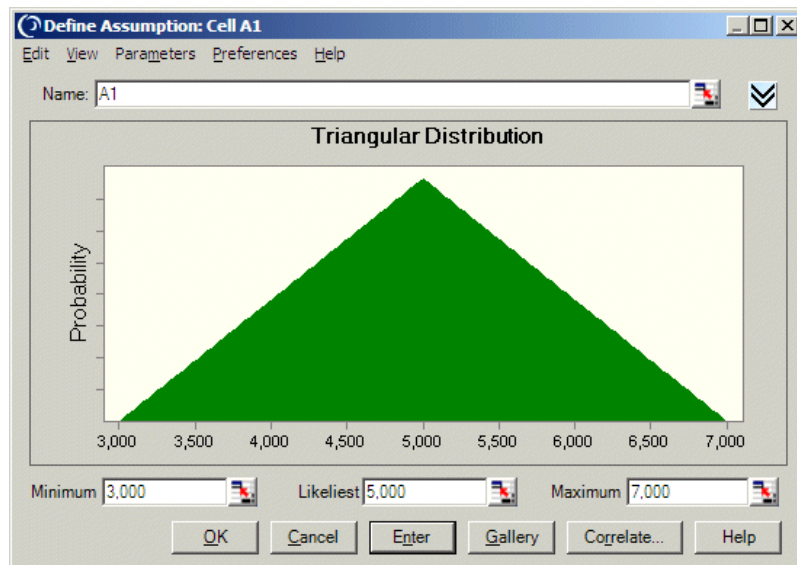
Figure A.18 Triangular distribution

Example two

The triangular distribution also could be used to approximate a computer-controlled inventory situation. The computer is programmed to keep an ideal supply of 25 items on the shelf, not to let inventory ever drop below 10 items, and not to let it ever rise above 30 items.

Check the triangular distribution conditions:

- The minimum inventory is 10 items.
- The maximum inventory is 30 items.
- The ideal level most frequently on the shelf is 25 items.

These conditions match those of the triangular distribution.

The result would be a distribution showing the probability of x number of items in inventory.

Uniform distribution
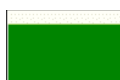
Parameters

Minimum, Maximum

Conditions

The minimum value is fixed.

The maximum value is fixed.

All values between the minimum and maximum occur with equal likelihood.

Description

Uniform

In the uniform distribution, all values between the minimum and maximum occur with equal likelihood.

Example one

An investment company interested in purchasing a parcel of prime commercial real estate wants to describe the appraised value of the property. The company expects an appraisal of at least $500,000 but not more than $900,000. They believe that all values between $500,000 and $900,000 have the same likelihood of being the actual appraised value.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. In this case:

- **The minimum value is $500,000.**

- **The maximum value is $900,000.**

- **All values between $500,000 and $900,000 are equally possible.**

These conditions match those of the uniform distribution. The uniform distribution has two parameters: the Minimum ($500,000) and the Maximum ($900,000). You would enter these values as the parameters of the uniform distribution in Crystal Ball.
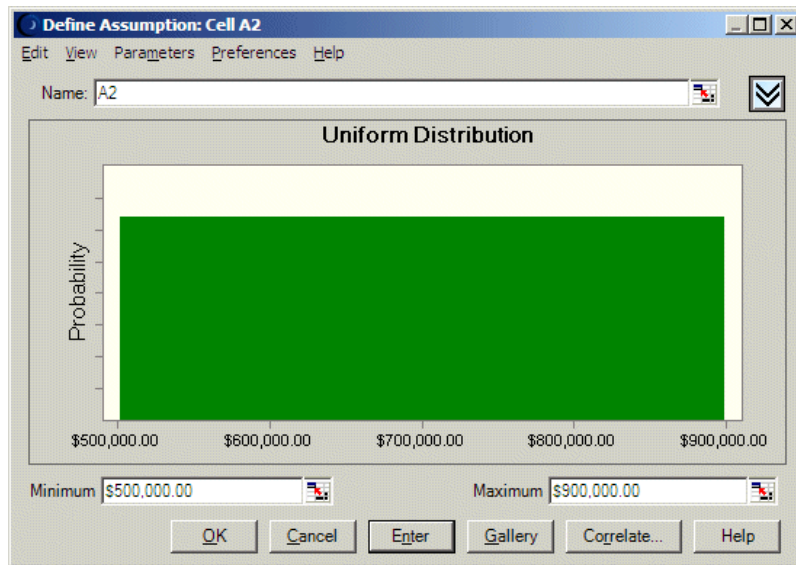
Figure A.19 Uniform distribution

The distribution in Figure A.19 shows that all values between $500,000 and $900,000 are equally possible.

Example two

A manufacturer determines that he must receive 10% over production costs—or a minimum of $3 per unit—to make the manufacturing effort worthwhile. He also wants to set the maximum price for the product at $6 per unit, so that he can gain a sales advantage by offering the product for less than his nearest competitor. All values between $3 and $6 per unit have the same likelihood of being the actual product price.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the uniform distribution:

- The minimum value is $3 per unit.

- The maximum value is $6 per unit.

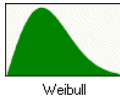- All values between $3 and $6 are equally possible.

You would enter these values in Crystal Ball to produce a uniform distribution showing that all values from $3 to $6 occur with equal likelihood.

Weibull distribution (also Rayleigh distribution)

Parameters

Location, Scale, Shape

Description

The Weibull distribution describes data resulting from life and fatigue tests. It is commonly used to describe failure time in reliability studies, and the breaking strengths of materials in reliability and quality control tests. Weibull distributions are also used to represent various physical quantities, such as wind speed.

The Weibull distribution is a family of distributions that can assume the properties of several other distributions. For example, depending on the shape parameter you define, the Weibull distribution can be used to model the exponential and Rayleigh distributions, among others.

The Weibull distribution is very flexible. When the Weibull Shape parameter is equal to 1.0, the Weibull distribution is identical to the exponential distribution. The Weibull Location parameter lets you set up an exponential distribution to start at a location other than 0.0. When the Shape parameter is less than 1.0, the Weibull distribution becomes a steeply declining curve. A manufacturer might find this effect useful in describing part failures during a burn-in period.
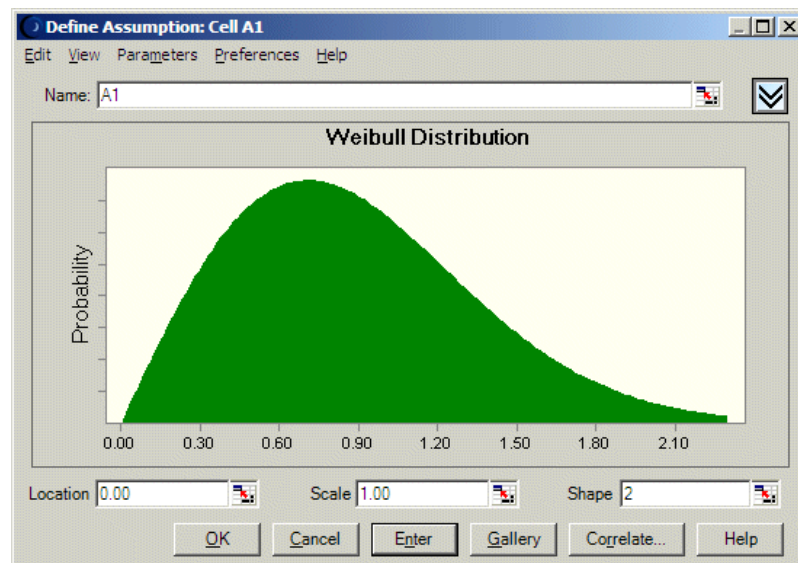

Figure A.20 Weibull distribution

When the Shape parameter is equal to 2.0, as in Figure A.20, a special form of the Weibull distribution, called the Rayleigh distribution, results. A researcher might find the Rayleigh distribution useful for analyzing noise problems in communication systems or for use in reliability studies.

Calculating parameters

There are three standard parameters for the Weibull distribution: Location, Scale, and Shape. The Location parameter is the lower bound for the variable.

The Shape parameter is a number greater than 0, usually a small number less than 10. When the Shape parameter is less than 3, the distribution becomes more and more positively skewed until it starts to resemble an exponential distribution (shape < 1). At a shape of 3.25, the distribution is symmetrical, and above that value, the distribution becomes more narrow and negatively skewed.

After you select the Location and Shape parameter, you can estimate the Scale parameter. The larger the scale, the larger the width of the distribution.

To calculate a more exact scale, estimate the mean and use the equation:

$$\alpha = \frac{\text{mean} - L}{\Gamma\left(1 + \frac{1}{\beta}\right)}$$

where α is the scale, β is the shape, L is the location, and Γ is the gamma function. You can use the Excel GAMMALN function and Excel Solver to help you calculate this parameter.

Statistical Note: For this distribution, there is a 63% probability that x falls between α and $\alpha + L$.

Or estimate the mode and use the equation:

$$\alpha = \frac{\text{mode} - L}{\left(1 - \frac{1}{\beta}\right)^{\frac{1}{\beta}}}$$

where α is the scale, β is the shape, and L is the location.

Example

A lawn mower company is testing its gas-powered, self-propelled lawn mowers. They run 20 mowers, and keep track of how many hours each mower runs until its first breakdown. They use a Weibull distribution to describe the number of hours until the first failure.

Using discrete distributions

Discrete probability distributions describe distinct values, usually integers, with no intermediate values and are shown as a series of vertical bars, such as the binomial distribution at the bottom of Figure A.4 on page 266. A discrete distribution, for example, might describe the number of heads in four flips of a coin as 0, 1, 2, 3, or 4.

The following discrete distributions are described later in this section in alphabetical order. Page references appear below the names.

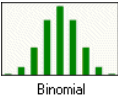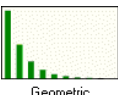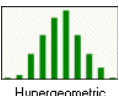**Table A.3 Summary of discrete distributions**

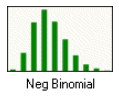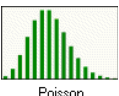| Shape	Name	Summary
Binomial	Binomial (page 302)	The binomial distribution describes the number of times a particular event occurs in a fixed number of trials, such as the number of heads in 10 flips of a coin or the number of defective items in 50 items.
Discrete Uniform	Discrete uniform (Basic) (page 304)	In the discrete uniform distribution, all integer values between the minimum and maximum are equally likely to occur. It is the discrete equivalent of the continuous uniform distribution.
Geometric	Geometric (page 306)	The geometric distribution describes the number of trials until the first successful occurrence, such as the number of times you need to spin a roulette wheel before you win.
Hypergeometric	Hypergeometric (page 307)	The hypergeometric distribution is similar to the binomial distribution; both describe the number of times a particular event occurs in a fixed number of trials. However, binomial distribution trials are independent, while hypergeometric distribution trials change the success rate for each subsequent trial and are called "trials without replacement."
Neg Binomial	Negative binomial (page 310)	The negative binomial distribution is useful for modeling the distribution of the number of trials until the r th successful occurrence, such as the number of sales calls you need to make to close a total of 10 orders. It is essentially a super-distribution of the geometric distribution.
Poisson	Poisson (page 312)	The Poisson distribution describes the number of times an event occurs in a given interval, such as the number of telephone calls per minute or the number of errors per page in a document.

Table A.3 Summary of discrete distributions (Continued)

Shape	Name	Summary
Yes-No	Yes-no (Basic) (page 314)	The yes-no distribution is a discrete distribution that describes a set of observations that can have only one of two values, such as yes or no, success or failure, true or false, or heads or tails.

Binomial distribution

Parameters

Probability, Trials

Statistical Note: The word "trials," as used to describe a parameter of the binomial distribution, is different from "trials" as it is used when running a simulation in Crystal Ball. Binomial distribution trials describe the number of times a given experiment is repeated (flipping a coin 50 times would be 50 binomial trials). A simulation trial describes a set of 50 coin flips (10 simulation trials would simulate flipping 50 coins 10 times).

Conditions

For each trial, only two outcomes are possible.

The trials are independent. What happens in the first trial does not affect the second trial, and so on.

The probability of an event occurring remains the same from trial to trial.
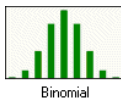
Description

The binomial distribution describes the number of times a particular event occurs in a fixed number of trials, such as the number of heads in 10 flips of a coin or the number of defective items in 50 items.

Example one

You want to describe the number of defective items in a total of 50 manufactured items, 7% of which (on the average) were found to be defective during preliminary testing.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the binomial distribution:

- There are only two possible outcomes: the manufactured item is either good or defective.

- The trials (50) are independent of one another. Any given manufactured item is either defective or not, independent of the condition of any of the others.

- The probability of a defective item (7%) is the same each time an item is tested.

These conditions match those of the binomial distribution.

The parameters for the binomial distribution are Probability and Trials. In example one, the values for these parameters are 50 (Trials) and 0.07 (7% Probability of producing defective items). You would enter these values to specify the parameters of the binomial distribution in Crystal Ball.
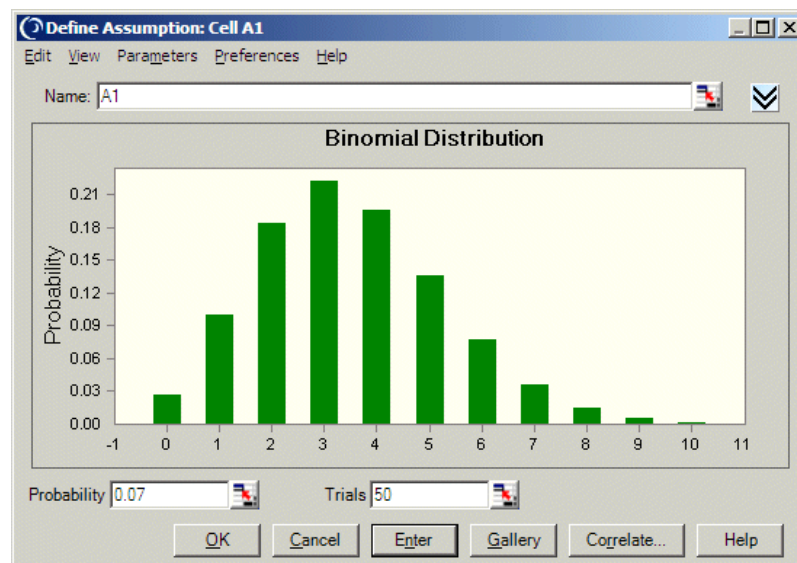

Figure A.21 Binomial distribution

The distribution illustrated in Figure A.21 shows the probability of producing x number of defective items.

Example two

A company's sales manager wants to describe the number of people who prefer the company's product. The manager conducted a survey of 100

consumers and determined that 60% prefer the company's product over the competitor's.

Again, the conditions fit the binomial distribution with two important values: 100 (trials) and 0.6 (60% probability of success). These values specify the parameters of the binomial distribution in Crystal Ball. The result would be a distribution of the probability that x number of people prefer the company's product.

Discrete uniform distribution

Parameters

Minimum, Maximum

Conditions

The minimum value is fixed.

The maximum value is fixed.

All integer values between the minimum and maximum are equally likely to occur.

Description

In the discrete uniform distribution, all integer values between the minimum and maximum are equally likely to occur. It is a discrete probability distribution. The discrete uniform distribution is very similar to the uniform distribution (page 297) except it is discrete instead of continuous; all its values must be integers.

The discrete uniform distribution can be used to model rolling a six-sided die. In that case, the minimum value would be 1 and the maximum 6.
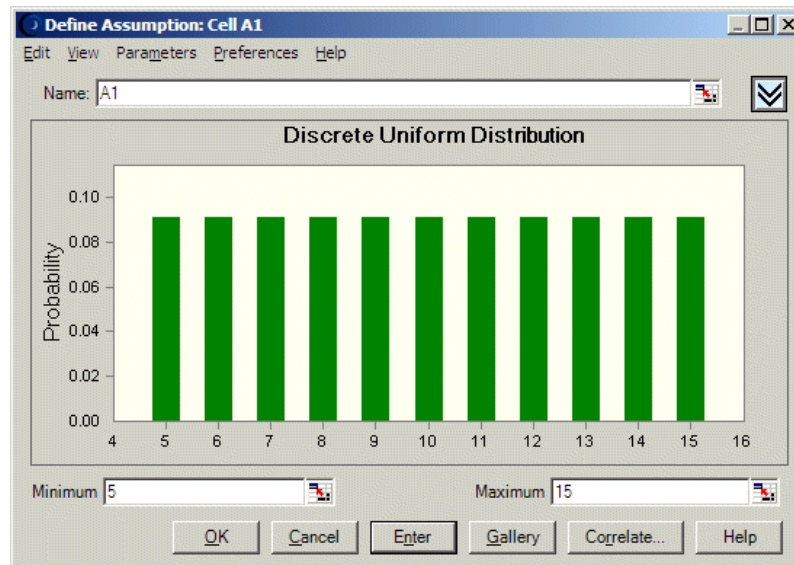
Figure A.22 Discrete uniform distribution

Example

A manufacturer determines that he must receive 10% over production costs—or a minimum of $5 per unit—to make the manufacturing effort worthwhile. He also wants to set the maximum price for the product at $15 per unit, so that he can gain a sales advantage by offering the product for less than his nearest competitor. All values between $5 and $15 per unit have the same likelihood of being the actual product price, however he wants to limit the price to whole dollars.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the uniform distribution:

- The minimum value is $5 per unit.

- The maximum value is $15 per unit.

- All integer values between $5 and $15 are equally possible.

You would enter these values in Crystal Ball to produce a discrete uniform distribution showing that all whole dollar values from $5 to $15 occur with equal likelihood. Figure A.22 illustrates this scenario.

Geometric distribution

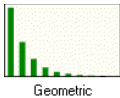## Parameter

Probability

Conditions

The number of trials is not fixed.

The trials continue until the first success.

The probability of success is the same from trial to trial.

Description

Geometric

The geometric distribution describes the number of trials until the first successful occurrence, such as the number of times you need to spin a roulette wheel before you win.

Example one

If you are drilling for oil and want to describe the number of dry wells you would drill before the next producing well, you would use the geometric distribution. Assume that in the past you have hit oil about 10% of the time.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the geometric distribution:

- The number of trials (dry wells) is unlimited.
- You continue to drill wells until you hit the next producing well.
- The probability of success (10%) is the same each time you drill a well.

These conditions match those of the geometric distribution.

The geometric distribution has only one parameter: Probability. In this example, the value for this parameter is 0.10, representing the 10% probability of discovering oil. You would enter this value as the parameter of the geometric distribution in Crystal Ball.

The distribution illustrated in Figure A.23 shows the probability of x number of wells drilled before the next producing well.
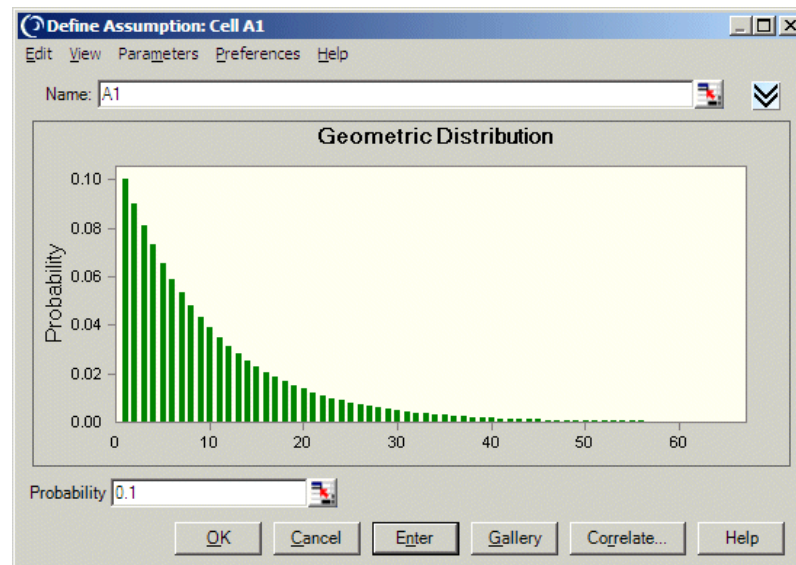
Figure A.23 Geometric distribution

Example two

An insurance company wants to describe the number of claims received until a "major" claim arrives. Records show that 6% of the submitted claims are equal in dollar amount to all the other claims combined.

Again, identify and enter the parameter values for the geometric distribution in Crystal Ball. In this example, the conditions show one important value: a 0.06 (6%) probability of receiving that "major" claim. The result would be a distribution showing the probability of x number of claims occurring between "major" claims.

Hypergeometric distribution

Parameters

Success, Trials, Population

Statistical Note: The word "trials," as used to describe a parameter of the hypergeometric distribution, is different from "trials" as it is used when running a simulation in Crystal Ball. Hypergeometric distribution trials describe the number of times a given experiment is repeated (removing 20 manufactured parts from a box would be 20 hypergeometric trials). A simulation trial describes the removing of 20 parts (10 simulation trials would simulate removing 20 manufactured parts 10 times).
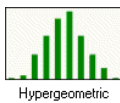
Conditions

The total number of items or elements (the population size) is a fixed number: a finite population. The population size must be less than or equal to 1000.

The sample size (the number of trials) represents a portion of the population.

The known initial success rate in the population changes slightly after each trial.

Description

The hypergeometric distribution is similar to the binomial distribution in that both describe the number of times a particular event occurs in a fixed number of trials. The difference is that binomial distribution trials are independent, while hypergeometric distribution trials change the success rate for each subsequent trial and are called "trials without replacement."

For example, suppose a box of manufactured parts is known to contain some defective parts. You choose a part from the box, find it is defective, and remove the part from the box. If you choose another part from the box, the probability that it is defective is somewhat lower than for the first part because you have removed a defective part. If you had replaced the defective part, the probabilities would have remained the same, and the process would have satisfied the conditions for a binomial distribution.

Example one

You want to describe the number of consumers in a fixed population who prefer Brand X. You are dealing with a total population of 40 consumers, of which 30 prefer Brand X and 10 prefer Brand Y. You survey 20 of those consumers.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the hypergeometric distribution:

- The population size (40) is fixed.

- The sample size (20 consumers) represents a portion of the population.

- Initially, 30 of 40 consumers preferred Brand X so the initial success rate is 30. This rate changes each time you question one of the 20 consumers, depending on the preference of the previous consumer.

The conditions in this example match those of the hypergeometric distribution.

Statistical Note: If you have a probability from a different-sized sample instead of a success rate, you can estimate initial success by multiplying the population size by the probability of success. In this example, the probability of success is 75% (.75 x 40 = 30 and 30/40 = .75).

The three parameters of this distribution are initial Success, number of Trials, and Population size. The conditions outlined in this example contain the values for these parameters: a population Size of 40, sample size (Trials) of 20, and initial Success of 30 (30 of 40 consumers will prefer Brand X). You would enter these values as the parameters of the hypergeometric distribution in Crystal Ball.
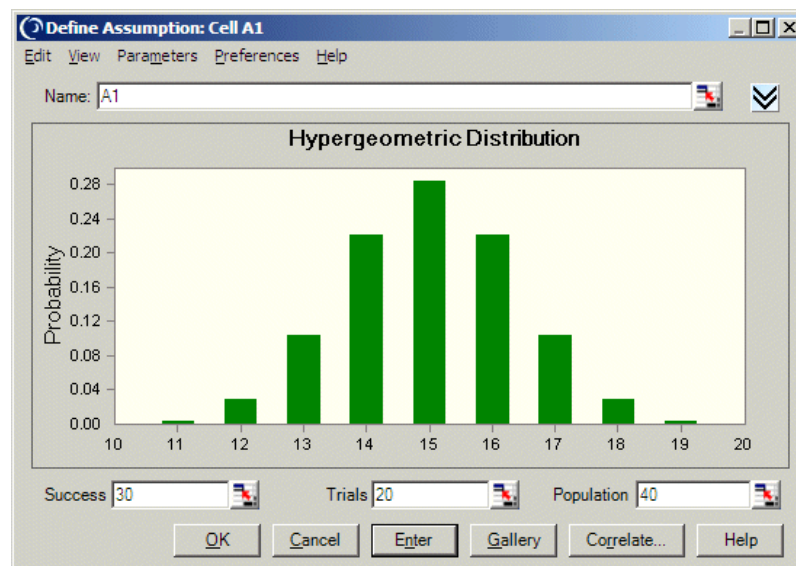
Figure A.24 Hypergeometric distribution

The distribution illustrated in Figure A.24 shows the probability that x number of consumers prefer Brand X.

Example two

The U.S. Department of the Interior wants to describe the movement of wild horses in Nevada. Researchers in the department travel to a particular area in Nevada to tag 100 horses in a total population of 1,000. Six months later the researchers return to the same area to find out how many horses remained in the area. The researchers look for tagged horses in a sample of 200.

Check the data against the conditions of the hypergeometric distribution. The parameter values for the hypergeometric distribution in Crystal Ball are the population size of 1,000, sample size of 200, and an initial success rate of 100

out of 1,000 (or a probability of 10% — 0.1 — of finding tagged horses. The result would be a distribution showing the probability of observing x number of tagged horses.

Crystal Ball Note: If you used this distribution in a model created in Crystal Ball 2000.x, you might notice slight data changes when running that model in the current version of Crystal Ball. This is because some rounding might occur when converting the probability parameter used in previous releases to the success parameter used in this version of Crystal Ball.

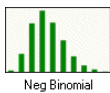# Negative binomial distribution

Parameters

Probability, Shape

Conditions

The number of trials is not fixed.

The trials continue until the r th success.

The probability of success is the same from trial to trial.

Statistical Note: The total number of trials needed will always be equal to or greater than r.

Description

Neg Binomial

The negative binomial distribution is useful for modeling the distribution of the number of trials until the r th successful occurrence, such as the number of sales calls you need to make to close a total of 10 orders. It is essentially a super-distribution of the geometric distribution.

Example

A manufacturer of jet engine turbines has an order to produce 50 turbines. Since about 20% of the turbines do not make it past the high-velocity spin test, the manufacturer will actually have to produce more than 50 turbines.

Matching these conditions with the negative binomial distribution:

- The number of turbines to produce (trials) is not fixed.

- The manufacturer will continue to produce turbines until the 50th one has passed the spin test.

- The probability of success (80%) is the same for each test.

These conditions match those of the negative binomial distribution.

The negative binomial distribution has two parameters: Probability and Shape. The Shape parameter specifies the r th successful occurrence. In this example you would enter 0.8 for the Probability parameter (80% success rate of the spin test) and 50 for the Shape parameter (Figure A.25).
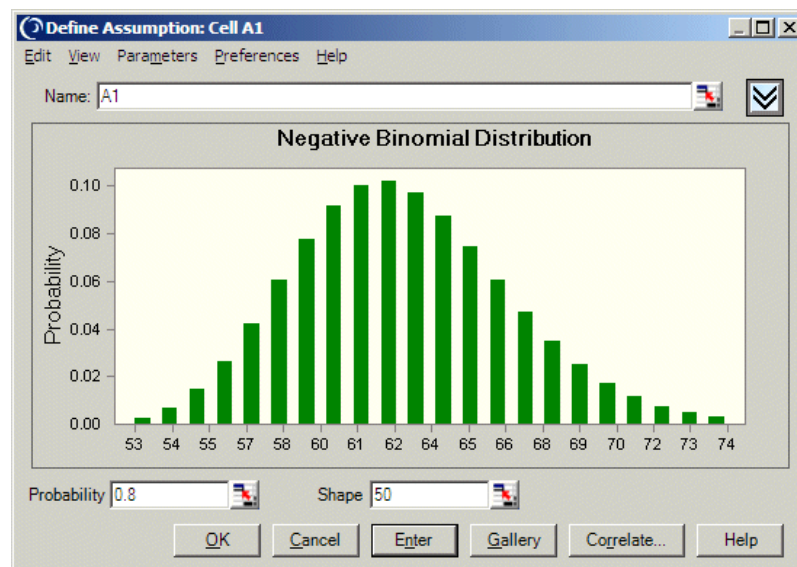
Figure A.25 Negative binomial distribution

Some characteristics of the negative binomial distribution:

- When Shape = 1, the negative binomial distribution becomes the geometric distribution.

- The sum of any two negative binomial distributed variables is a negative binomial variable.

- Another form of the negative binomial distribution, sometimes found in textbooks, considers only the total number of failures until the r th successful occurrence, not the total number of trials. To model this form of the distribution, subtract out r (the value of the shape parameter) from the assumption value using a formula in your worksheet.

Poisson distribution

Parameter
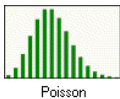
Rate

Conditions

The number of possible occurrences in any interval is unlimited.

The occurrences are independent. The number of occurrences in one interval does not affect the number of occurrences in other intervals.

The average number of occurrences must remain the same from interval to interval.

Description

Poisson

The Poisson distribution describes the number of times an event occurs in a given interval, such as the number of telephone calls per minute or the number of errors per page in a document.

Example one

An aerospace company wants to determine the number of defects per 100 square yards of carbon fiber material when the defects occur an average of 8 times per 100 square yards.

The first step in selecting a probability distribution is matching your data with a distribution's conditions. Checking the Poisson distribution:

- Any number of defects is possible within 100 square yards.

- The occurrences are independent of one another. The number of defects in the first 100 square yards does not affect the number of defects in the second 100 square yards.

- The average number of defects (8) remains the same for each 100 square yards.

These conditions match those of the Poisson distribution.

The Poisson distribution has only one parameter: Rate. In this example, the value for this parameter is 8 (defects). You would enter this value to specify the parameter of the Poisson distribution in Crystal Ball.
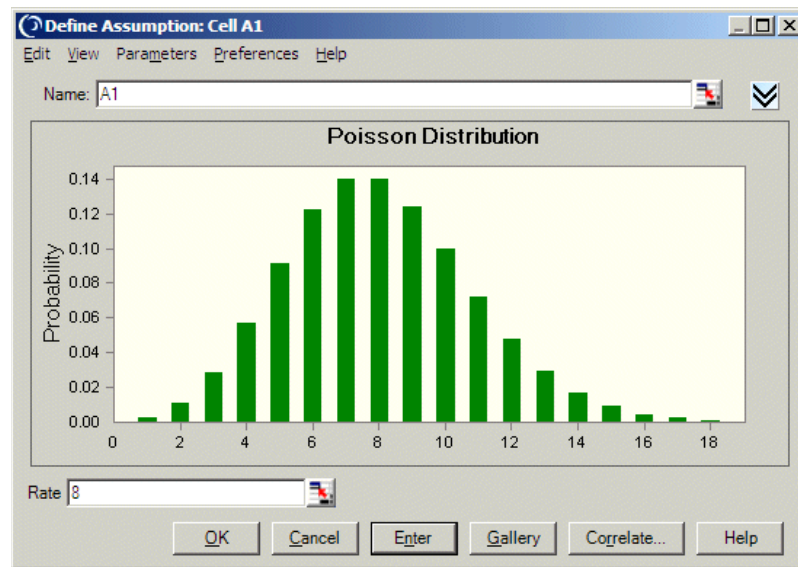
Figure A.26 Poisson distribution

The distribution illustrated in Figure A.26 shows the probability of observing x number of defects in 100 square yards of the carbon fiber material.

Statistical Note: The size of the interval to which the rate applies, 100 square yards in this example, has no bearing on the probability distribution; the rate is the only key factor. If needed for modeling a situation, information on the size of the interval must be encoded in your spreadsheet formulas.

Example two

A travel agency wants to describe the number of calls it receives in 10 minutes. The average number of calls in 10 minutes is about 35.

Again, you begin by identifying and entering the values to set the parameters of the Poisson distribution in Crystal Ball. In this example, the conditions show one important value: 35 calls or a rate of 35. The result would be a distribution showing the probability of receiving x number of calls in 10 minutes.

Yes-no distribution

Parameter

Probability of Yes (1)

Conditions

The random variable can have only one of two values, for example, 0 and 1.

The mean is p, or probability (0 < p < 1).

Description

The yes-no distribution is also called the Bernoulli distribution in statistical textbooks. This distribution describes a set of observations that can have only one of two values, such as yes or no, success or failure, true or false, or heads or tails. It is a discrete probability distribution.

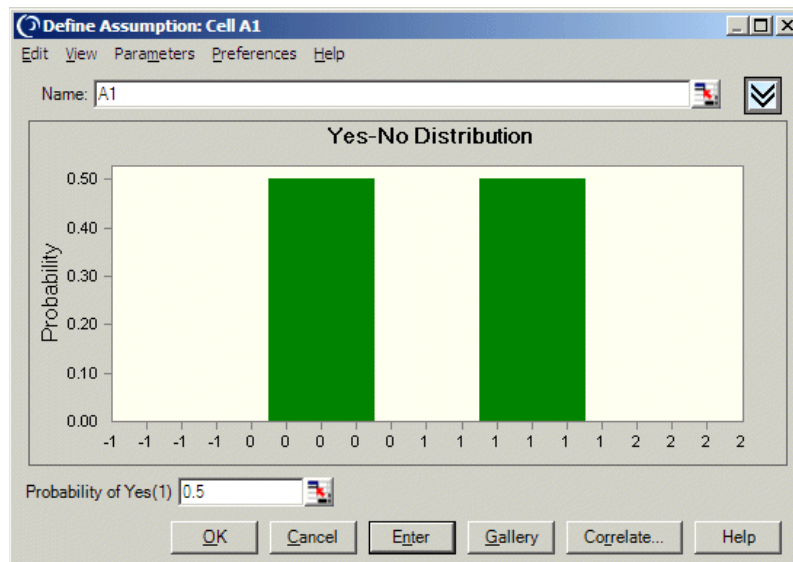**The yes-no distribution has one parameter, Probability of Yes (1).**

Figure A.27 Yes-no distribution

Example

A machine shop produces complex high-tolerance parts with a .02 probability of failure and a .98 probability of success. If a single part is pulled from the line, Figure A.28 shows the probability that the part is good.
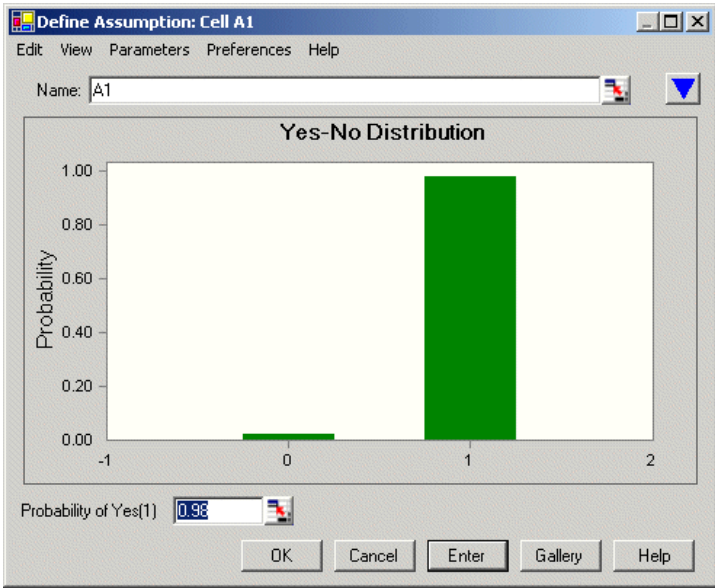
Figure A.28 Probability of pulling a good part

Using the custom distribution

If none of the provided distributions fits your data, you can use the custom distribution to define your own. For example, a custom distribution can be especially helpful if different ranges of values have specific probabilities. You can create a distribution of one shape for one range of values and a different distribution for another range.

The following sections explain how to use the custom distribution and provide examples of its use.

Custom distribution
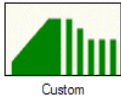
With Crystal Ball, you can use the custom distribution to represent a unique situation that cannot be described using other distribution types: you can describe a series of single values, discrete ranges, or continuous ranges. This section uses real-world examples to describe the custom distribution.

Crystal Ball Note: For summaries of the data entry rules used in the examples plus additional rules, see "Entering tables of data into custom distributions" beginning on page 327 and "Other important custom distribution notes" beginning on page 332.

Since it is easier to understand how the custom distribution works with a hands-on example, you might want to start Crystal Ball and use it to follow the examples. To follow the custom distribution examples, first create a new Excel workbook then select cells as specified.

Example one

Before beginning example one, open the Custom Distribution dialog as follows:

1. Click cell D11.

2. Select Define > Define Assumption.

    The Distribution Gallery dialog appears.

3. Click the All category to select it.

4. Scroll to find the custom distribution, then click it.

5. Click OK.

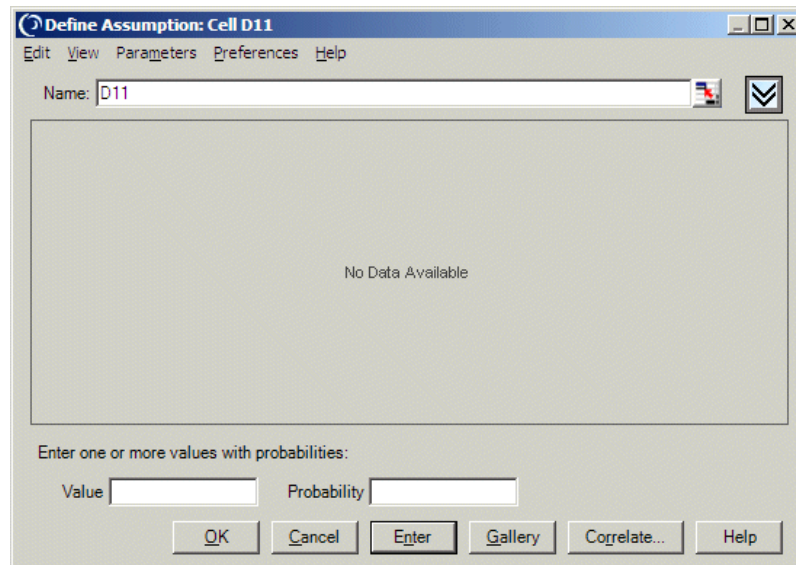   Crystal Ball displays the Define Assumption dialog.

Figure A.29 Define Assumption dialog for custom distributions

Using the custom distribution, a company can describe the probable retail cost of a new product. The company decides the cost could be $5, $8, or $10. In this example, you will use the custom distribution to describe a series of single values.

To enter the parameters of this custom distribution:

1. Type 5 in the Value field and click Enter.

    Since you do not specify a probability, Crystal Ball defaults to a relative probability of 1.00 for the value 5. A single value bar displays the value 5.00.

Statistical Note: Relative probability means that the sum of the probabilities does not have to add up to 1. So the probability for a given value is meaningless by itself; it makes sense only in relation to the total relative probability. For example, if the total relative probability is 3 and the relative probability for a given value is 1, the value has a probability of 0.33.

2. Type 8 in the Value field.

3. Click Enter.

    Since you did not specify a probability, Crystal Ball defaults to a relative probability of 1.00 (displayed on the scale to the left of the Custom Distribution dialog) for the value 8. A second value bar represents the value 8.

4. Type 10 in the Value field.

5. Click Enter.

   Crystal Ball displays a relative probability of 1.00 for the value 10. A third single value bar represents the value 10.

   Figure A.30 shows the value bars for the values 5, 8, and 10, each with a relative probability of 1.00.
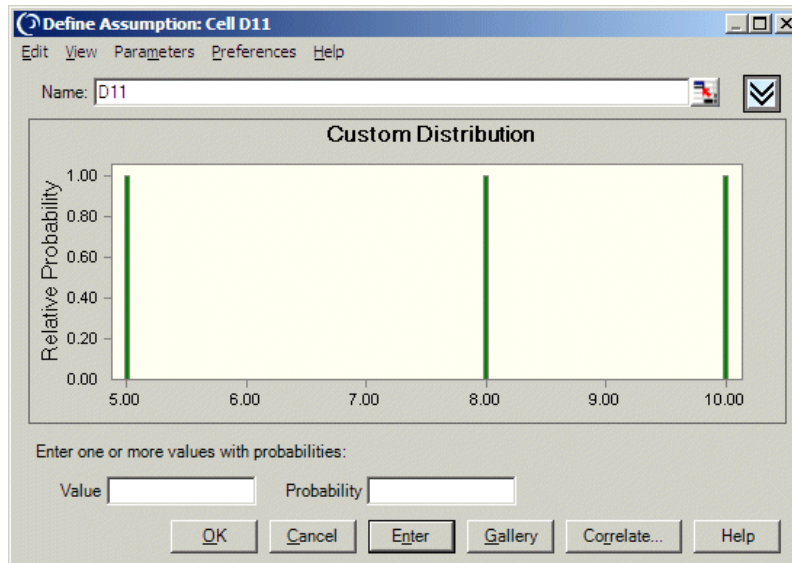
Figure A.30 Single values

Now, each value has a probability of 1. However, when you run the simulation, their total relative probability becomes 1.00 and the probability of each value is reset to .3333.

If you want to reset their probabilities before you run the simulation, follow these steps:

1. Click the bar with a value of 5.00.

   Its value appears in the Value field.

2. Type the probability as the formula = 1/3 in the Probability field and click Enter.

   You could also enter a decimal — for example, 0.333333 — but the formula is more exact.

3. Follow steps 6 and 7 for the other two bars.

Crystal Ball rescales each value to a relative probability of 0.33 on the left side of the screen.
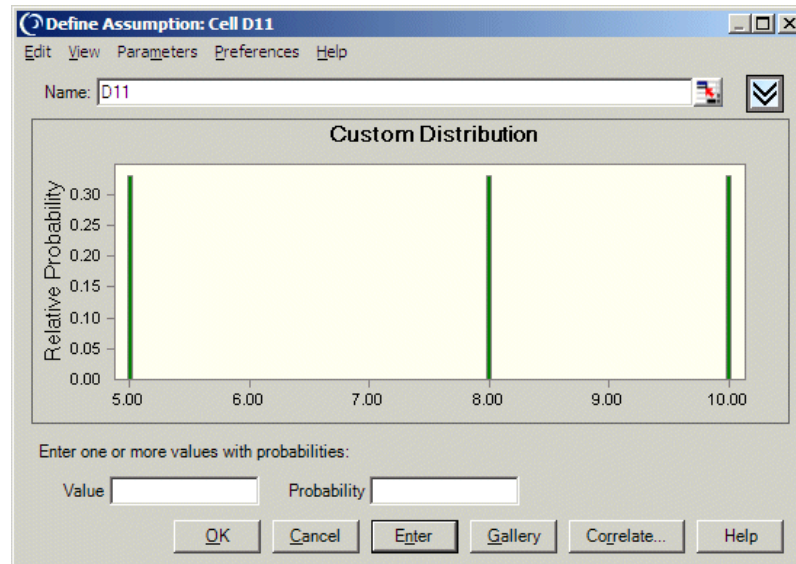
Figure A.31 Single values with adjusted probabilities

Example two

Before beginning example two, clear the values entered in example one as follows:

1. Right-click in the chart and choose Clear Distribution from the right-click menu.

In this example, you will use the custom distribution to describe a continuous range of values, since the unit cost can take on any value within the specified intervals.

1. Choose Parameters > Continuous Ranges to enter value ranges.

2. Enter the first range of values:

    *   Type 5 in the Minimum field.

    *   Type 15 in the Maximum field.

    *   Type .75 in the Probability field. This represents the total probability of all values within the range.

3. Click Enter.

Crystal Ball displays a continuous value bar for the range 5.00 to 15.00, as in Figure A.32, and returns the cursor to the Minimum field. Notice that the height of the range is 0.075. This represents the total probability divided by the width (number of units) in the range, 10.
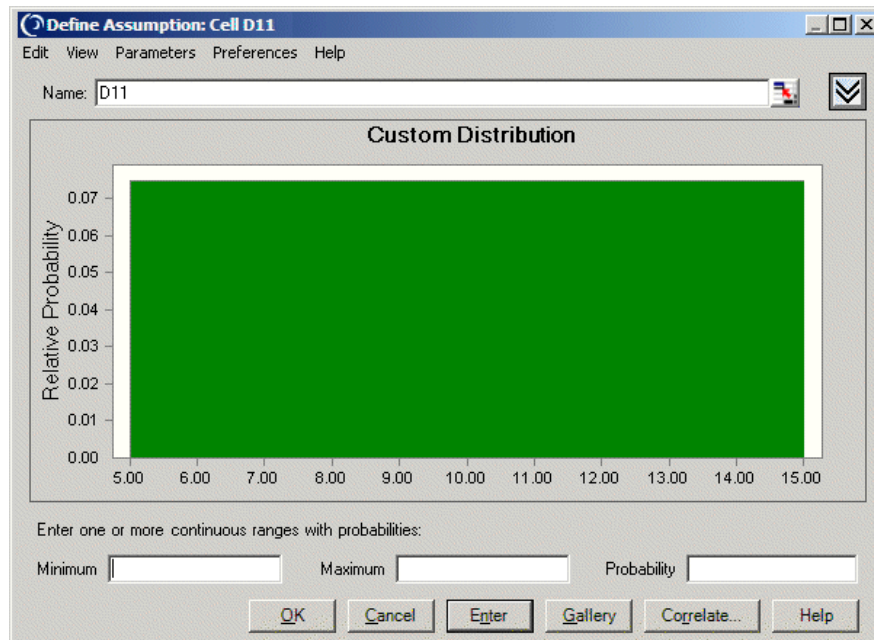
Figure A.32 A continuous custom distribution

4. Enter the second range of values:

   •   Type 16 in the Minimum field.

   •   Type 21 in the Maximum field.

   •   Type .25 in the Probability field.

   •   Click Enter.

Crystal Ball displays a continuous value bar for the range 16.00 to 21.00. Its height is .050, equal to .25 divided by 5, the number of units in the range. Both ranges now appear in the Custom Distribution dialog (Figure A.33).
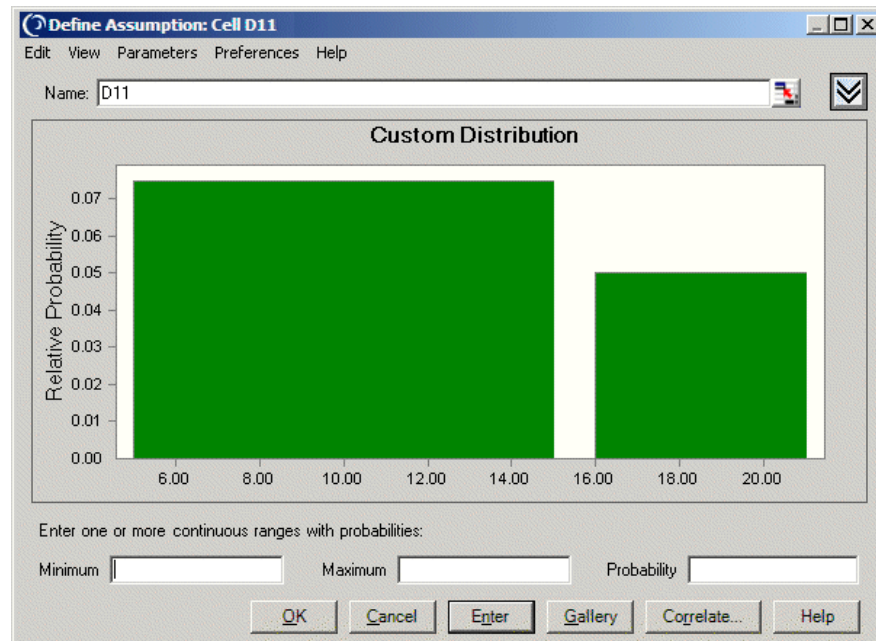
Figure A.33 Custom distribution with two continuous ranges

You can change the probability and slope of a continuous range, as described in the following steps:

1. Click anywhere on the value bar for the range 16 to 21.

    The value bar changes to a lighter shade.

2. Choose Parameters > Sloping Ranges.

    Additional parameters appear in the Custom Distribution dialog.

Figure A.34 Sloping Range parameters, Custom Distribution dialog

3. Set the Height of Min. and Height of Max. equal to what currently appears in the chart, 0.05.

    This can be an approximate value. The Height of Min. is the height of the range Minimum and the Height of Max. is the height of the range Maximum.

4. Click Enter.

    The range returns to its original color and its height appears unchanged.

5. Click in the range again to select it and set the Height of Max. to 0.025. Then, click Enter.

   The right side of the range drops to half the height of the left, as shown in Figure A.35. The range is selected to show its parameters after the change.
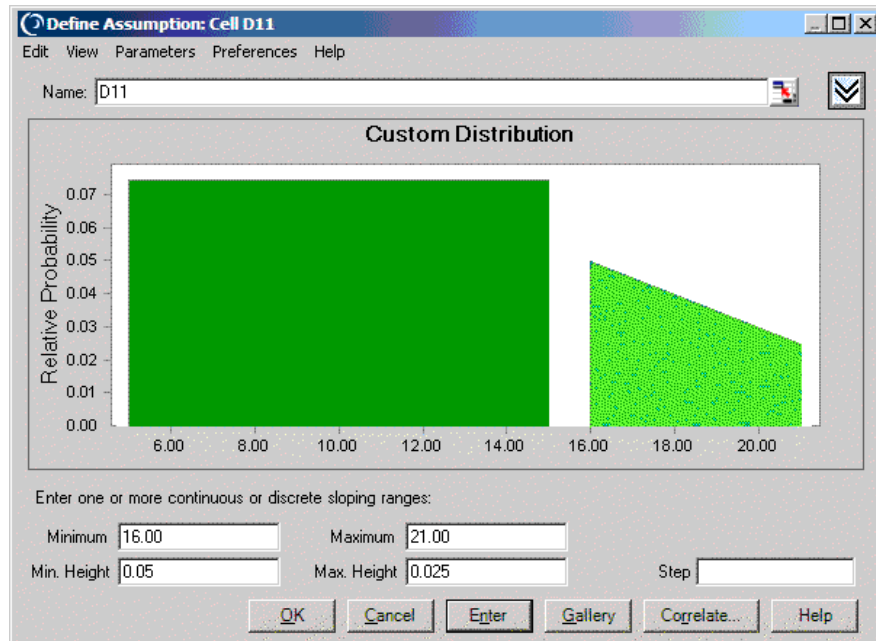
Figure A.35 Sloping continuous value range

6. You can change the range from continuous to discrete values by adding a step value. Type .5 in the Step field and click Enter.

   The sloped range is now discrete. Separate bars appear at the beginning and end of the range and every half unit in between (16, 16.5, 17, 17.5 and so on until 21), as shown in Figure A.36 on page 323. If the discrete range represented money, it could only include whole dollars and 50-cent increments.

Crystal Ball Note: You can enter any positive number in the Step field. If you entered 1 in this example, the steps would fall on consecutive integers, such as whole dollars.

Leave the Step parameter blank for continuous ranges.

Figure A.36 A sloped discrete range with steps of .5

Although the bars have spaces between them, their heights and the width of the range they cover are equal to the previous continuous sloped range and the total probability is the same.
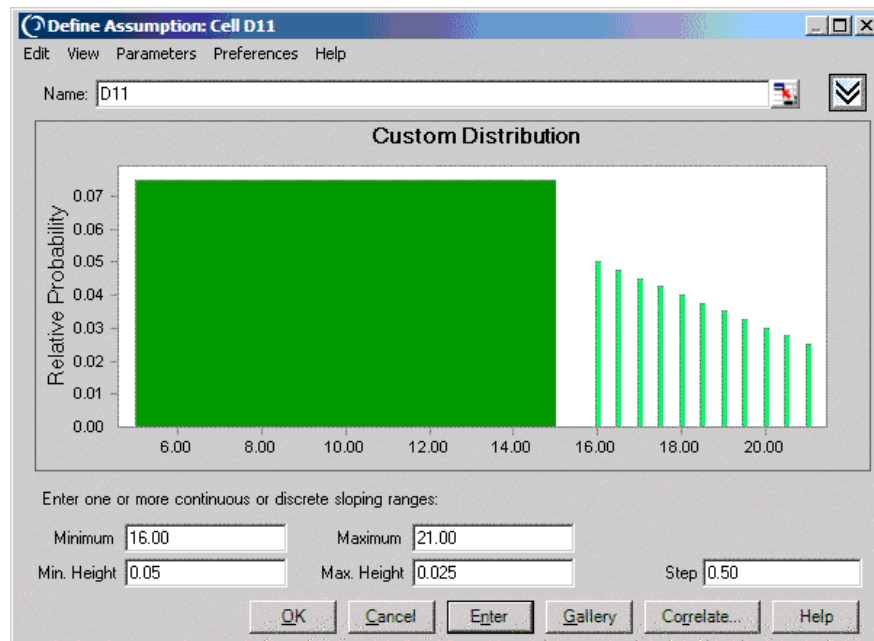
Crystal Ball Note: While a second continuous range could have extended from 15 to 20, the second range in this example starts at 16 rather than 15 to illustrate a discrete range because, unlike continuous ranges, discrete ranges cannot touch other ranges.

With Crystal Ball, you can enter single values, discrete ranges, or continuous ranges individually. You also can enter any combination of these three types in the same Custom Distribution dialog as long as you follow these guidelines: ranges and single values cannot overlap one another; however, the ending value of one continuous range can be the starting value of another continuous range.

Example three

This example describes a special feature on the Custom Distribution dialog: the Load Data button, which lets you pull in numbers from a specified cell range (grouped data) on the worksheet. This example is not a hands-on exercise, but the illustrations will guide you through the procedure. After you read this section, you can experiment with your own data by pulling in numbers from specified cell ranges on your worksheet.

In this example, the same company decides that the unit cost of the new product can vary widely. The company feels it has a 20% chance of being any number between $10 and $20, a 10% chance of being any number between $20 and $30, a 30% chance of being any number between $40 and $50, a 30% chance of being a whole dollar amount between $60 and $80, and there is a 5% chance the value will be either $90 or $100. All the values have been entered on the worksheet in this order: range minimum value, range maximum value (for all but Single Value ranges), total probability, and step (for the Discrete Range only).

	A	B	C	D	E	F	G
1	Minimum	Maximum	Prob.	Step			
2	$10	$20	0.2			Continuous Range	
3	$20	$30	0.1			Continuous Range	
4	$40	$50	0.3			Continuous Range	
5	$60	$80	0.3	1		Discrete Range	
6	$90		0.05			Single Value	
7	$100		0.05			Single Value	

Figure A.37 Four-column custom data range

In this case, discrete ranges have the most parameters. So, you can create an assumption, choose Custom Distribution, and then choose Parameters > Discrete Ranges before loading the data.

Crystal Ball Note: If your data also included discrete sloping ranges, you could choose Parameters > Sloping Ranges before loading the data. The data table would then have five columns and could accommodate all data types.

Once the Parameters setting has been made, you can follow these steps to complete the data load:

1. Click the More button to the right of the Name field.

    The Custom Distribution dialog expands to include a data table, as shown in Figure A.38.
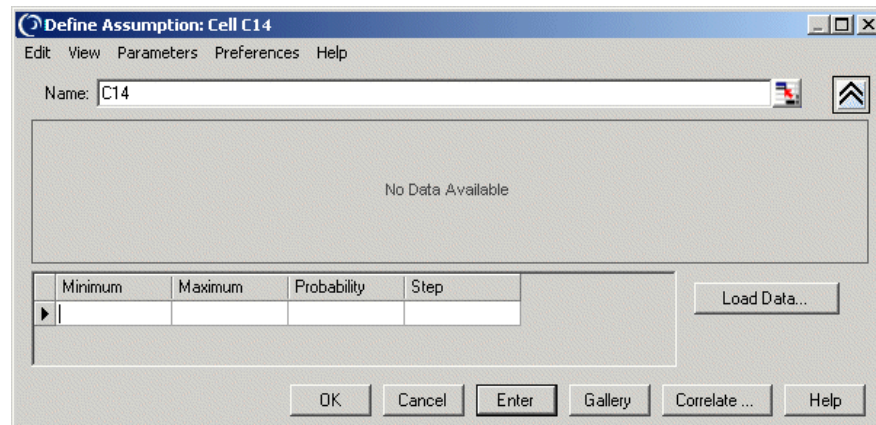
Figure A.38 Custom distribution with data table

A column appears for each parameter in the current set (selected using the Parameters menu). Parameters > Discrete Ranges was set before viewing the data table, so there is a column in the data table for each discrete range parameter. Because the single value and continuous ranges have subsets of the same group of parameters, their parameters will also fit into the table.

2. Since the values are already on the worksheet, you can click Load Data to enter them into the Custom Distribution dialog.

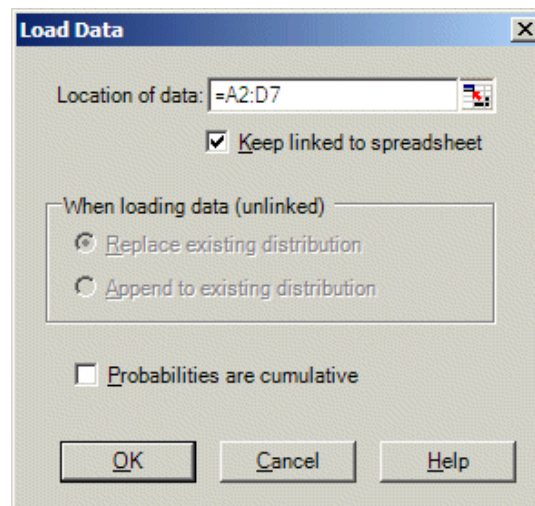The Load Data dialog appears, as shown in Figure A.39.

Figure A.39 Load Data dialog, Custom Distribution

The default settings are appropriate for most purposes, but the following other options are available:

- When loading unlinked data, you can choose to replace the current distribution with the new data or append new data to the existing distribution.

- If probabilities are entered cumulatively into the spreadsheet you are loading, you can check Probabilities Are Cumulative. Then, Crystal Ball determines the probabilities for each range by subtracting the previous probability from the one entered for the current range. You will need to choose View > Cumulative Probability to display the data cumulatively in the assumption chart.

3. Enter a location range for your data. When all settings are correct, click OK.

Crystal Ball enters the values from the specified range into the custom distribution and plots the specified ranges, as shown in Figure A.40.
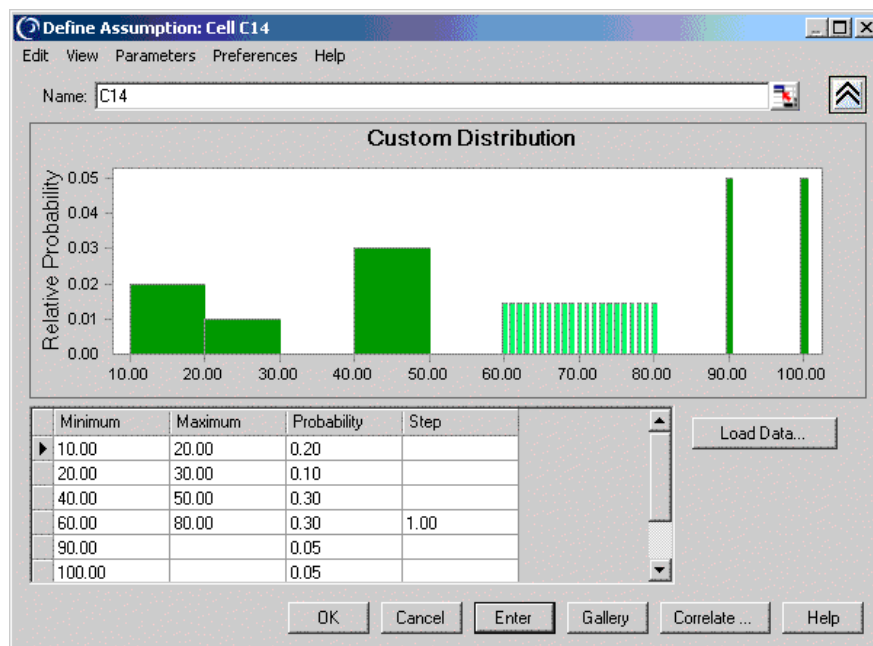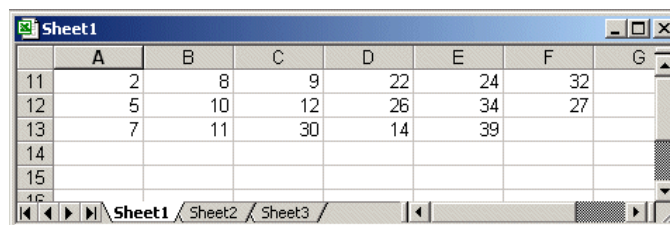
Figure A.40 Custom data from worksheet

Entering tables of data into custom distributions

Follow the rules in this section for loading data.

Unweighted values

Single values are values that don't define a range. Each value stands alone. For a series of single values with the same probabilities (unweighted values), use a one-column format or more than five columns. The values go in each cell and the relative probabilities are all assumed to be 1.0. Choose Parameters > Unweighted Values to enter these.

Figure A.41 Single values with the same probability

Figure A.42 Unweighted values loaded in a custom distribution

Weighted values

For a series of single values all with different probabilities, use a two-column format. The first column contains single values, the second column contains the probability of each value.

Figure A.43 Single values with different probabilities (weighted values)

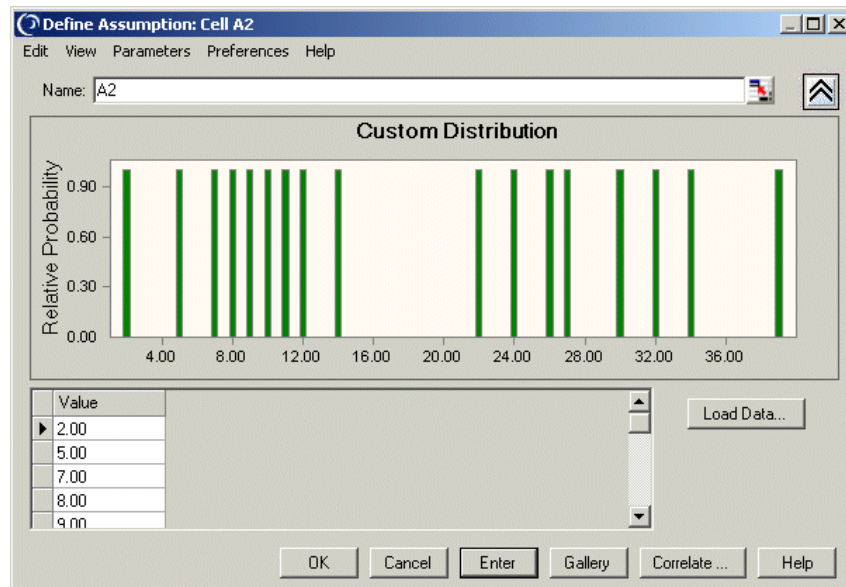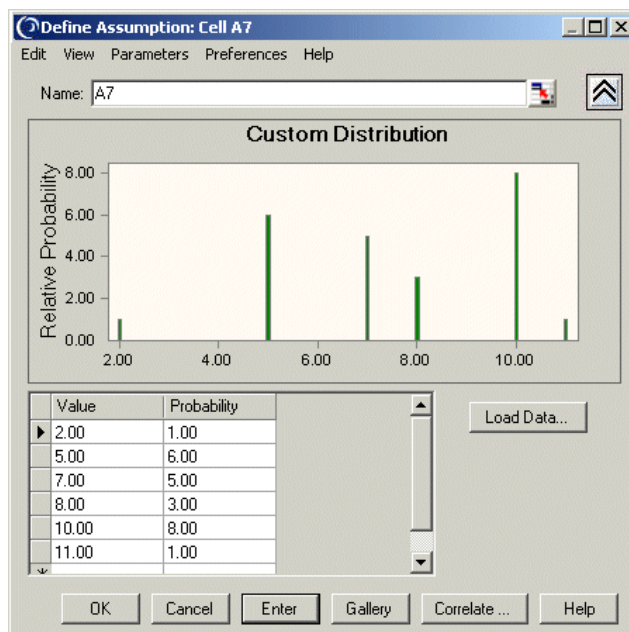**Figure A.44 Weighted values loaded in a custom distribution**

Mixed single values, continuous ranges, and discrete ranges

For any mixture of single values and continuous ranges, use a three-column format, obtained by choosing Parameters > Continuous ranges. The three-column format is the same as using the first three columns shown in Figure A.38, Figure A.39, and Figure A.40 beginning on page 325.

If the mix includes uniform (non-sloping) discrete ranges, use a four-column format, as in the first four columns of Figure A.45 and Figure A.46. To obtain four columns, choose Parameters > Discrete Ranges.

Mixed ranges, including sloping ranges

If sloping ranges are included in a mix of ranges, choose Parameters > Sloping Ranges to display a five-column data table. The first column contains the range Minimum value, the second column contains the range Maximum value, the third column contains Height of Min. (the relative probability — height — at the Minimum value), the fourth column contains Height of Max. (the relative probability at the Maximum value), and the fifth column contains the Step value for discrete sloping ranges. For continuous sloping ranges the fifth column (Step) is left blank.

Note that if there are uniform discrete ranges, their first three columns contain the Minimum, Maximum, and Probability as in a four-column format but the fourth column is left blank and Step is entered in the fifth column.

	A — Custom Value Description	B — Minimum	C — Maximum	D — Height of Min.	E — Height of Max.	F — Step
13						
15	Discrete Value = 0.5, Probability = 1.0	0.5				
16	Continuous Range = 1-2, Uniform Probability = 1.0	1	2			
17	Discrete Value = 2.5, Probability = 5	2.5		5		
18	Continuous Range = 3-5, Uniform Probability = 4	3	5	4		
19	Continuous Range = 5-7, Sloping Probability = 3-6	5	7	3	6	
20	Continuous Connected Range = 7-10, Sloping Probability = 6-8		10		8	
21	Discrete Range = 12-14, Uniform Probability = 2, Step = 0.5	12	14	2		0.5
22	Discrete Range = 15-17, Sloping Probability = 1-4, Step = 1	15	17	1	4	1
23						

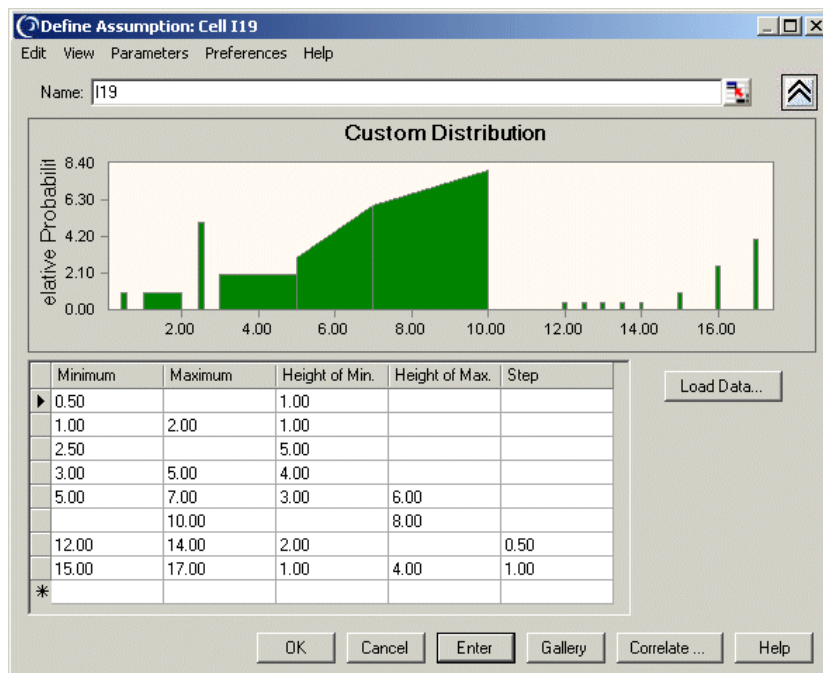**Figure A.45 Mixed ranges, including sloping ranges**

Figure A.46 Mixed ranges loaded in a custom distribution

Connected series of ranges (sloping)

For a connected series of sloping continuous ranges, choose Parameters >
Sloping Ranges to use a five-column format. The first column contains the
lowest Minimum value of the right-most range, the second column contains
the Maximum value of each connected range, the third column contains the
Height of Min. (relative probability of the Minimum value) if it differs from
the previous Height of Max. (otherwise it is left empty), and the fourth
column contains Height of Max. (relative probability of the Maximum value)
for that range. The fifth column is left blank for continuous ranges but a fifth
column is necessary to indicate that these are sloping ranges.

For example, row 20 in Figure A.45 shows a connected continuous sloping
range. The Minimum cell is blank because the Minimum value is equal to 7,
the previous Maximum. The Height of Min. is blank because it is equal to 6,
the previous Height of Max.

Connected series of continuous uniform ranges (cumulative)

For a connected series of continuous uniform ranges specified using
cumulative probabilities, use a three-column format with the common
endpoints of the ranges in the second column and the cumulative
probabilities in the third column. The first column is left blank except for the

minimum value of the first range, beside the maximum in the second column. Be sure to check Probabilities Are Cumulative in the Load Data dialog.

	A	B	C
	Minimum of First Range	Maximums of Continuous Ranges	Cumulative Probabilities
1			
2	5	10	0.25
3		15	0.50
4		25	0.75
5		30	0.90

Figure A.47 Connected continuous uniform ranges
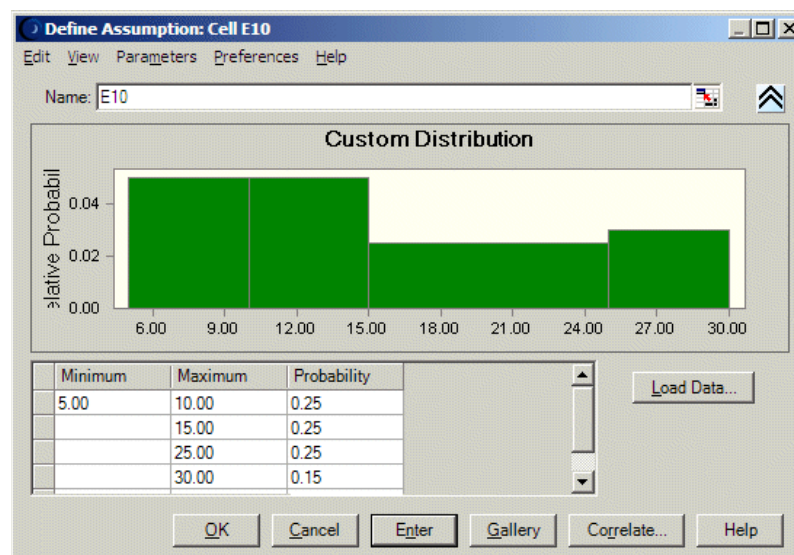
Figure A.48 Connected continuous uniform ranges after loading

Other data load notes

You can load each type of range separately or you can specify the range type with the greatest number of parameters and load all types together. Other rules are.

- Cumulative probabilities are supported for all but sloping ranges.

- Blank probabilities are interpreted as a relative probability of 1.0.

- Ranges or values with 0 probabilities are removed. Sloping ranges with Height of Min. and Height of Max. equal to 0 are also removed.

- For continuous connected ranges, for either endpoint values or probabilities, if the starting cell is blank, the previous end value is used as the start for this range.

- When you load a discrete value that exists in the table already, its probability is incremented by 1. For continuous ranges, this is not allowed; an error message about overlapping ranges appears.

Changes from Crystal Ball 2000.x (5.x)

In previous versions of Crystal Ball, discrete values with the same probability could be entered in ranges with five columns or more. Now, they cannot be entered in ranges with five columns but can only be entered in single columns or ranges with six or more columns (to distinguish them from sloping ranges).

In previous versions of Crystal Ball, continuous uniform ranges with cumulative probabilities could be entered in a two-column format. Now a three-column format is required, discussed in "Connected series of continuous uniform ranges (cumulative)" on page 330.

The three-column sloping range format used in previous versions of Crystal Ball has been replaced by a five-column format, described in "Mixed ranges, including sloping ranges" on page 329 and the section that follows it, "Connected series of ranges (sloping)."

Other important custom distribution notes

Even if you don't load data from the spreadsheet into the Custom Distribution dialog, you can still add and edit data using the data table. To do this, click the More button to display the data table. Then, you can:

- Enter a different value in the data table and click Enter to change the data.
- Type the minimum, maximum, probability, and step (if discrete data) into a blank row and click Enter to add new data.
- To delete a single range of data, select that row of data, right-click, and choose Delete Row.
- To clear all data rows, right-click within the data table and choose Clear Distribution.
- To delete a single range of data without using the data table, click the range to select it and either:
  - Set the Probability or Height of Min. and Height of Max. to 0, or
  - Choose Edit > Delete Row or right-click and choose Delete Row.
- Statistics for custom distributions are approximate.

Truncating distributions

You can change the bounds or limits of each distribution, except the custom distribution, by dragging the truncation grabbers or by entering different numeric endpoints for the truncation grabbers. This "truncates" the distribution. You can also exclude a middle area of a distribution by crossing over the truncation grabbers to "white out" the portion you want to exclude.

Crystal Ball Note: To display the truncation grabbers, open an assumption in the Define Assumption dialog and click the More button.

For example, suppose you want to describe the selling price of a house up for auction after foreclosure. The bank that holds the mortgage will not sell for less than $80,000. They expect the bids to be normally distributed around $100,000 with a standard deviation of $15,000. In Crystal Ball you would specify the mean as 100,000 and the standard deviation as 15,000 and then move the left grabber to set the limit of 80,000. The grabber "whites out" the portion you want to exclude, as shown in Figure A.49.

Be aware...

Each adjustment changes the characteristics of the probability distribution. For example, the truncated normal distribution in Figure A.49 will no longer have an actual mean of $100,000 and standard deviation of $15,000. Also, statistics values will be approximate for truncated distributions.

When using alternate percentile parameters, the actual percentiles calculated for a truncated distribution will differ from the specified parameter values. For example, a normal distribution specified with $10^{th}/90^{th}$ percentiles and truncated on either side of the distribution will have actual $10^{th}/90^{th}$ percentiles greater or less than the specified percentiles.

Crystal Ball Note: Showing the mean line of the distribution is useful when truncating distributions. However, the mean line value might differ from the Mean parameter field. The mean line shows the actual mean of the truncated distribution while the Mean parameter field shows the mean of the complete distribution.
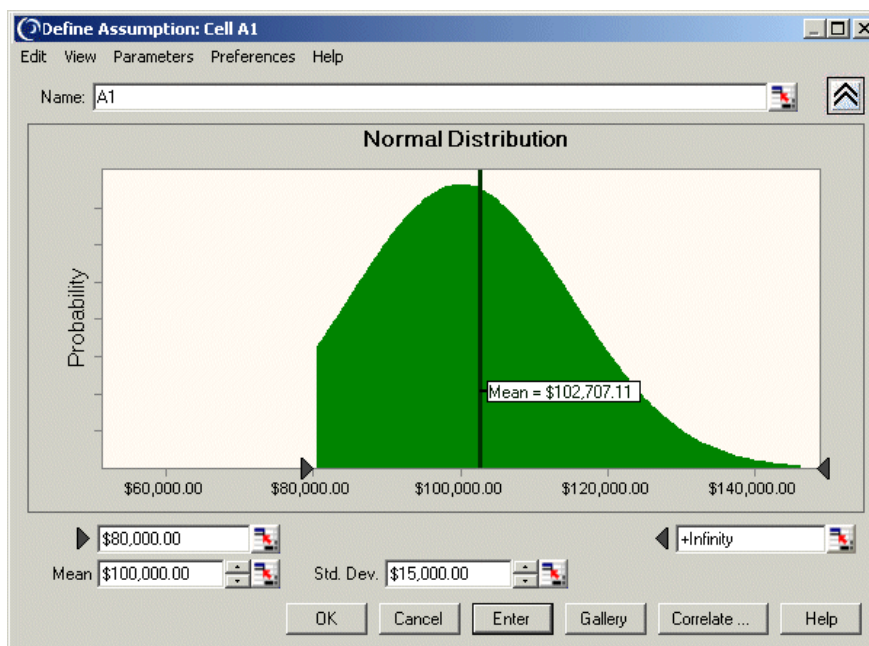
Figure A.49 Truncated distribution example

Comparing the distributions

Many of the distributions discussed in this chapter are related to one another in various ways. For example, the geometric distribution is related to the binomial distribution. The geometric distribution represents the number of trials until the next success while the binomial represents the number of successes in a fixed number of trials. Similarly, the Poisson distribution is related to the exponential distribution. The exponential distribution represents the amount of time until the next occurrence of an event while the Poisson distribution represents the number of times an event occurs within a given period of time.

In some situations, as when the number of trials for the binomial distribution becomes very large, the normal and binomial distributions become very similar. For these two distributions, as the number of binomial trials approaches infinity, the probabilities become identical for any given interval. For this reason, you can use the normal distribution to approximate the binomial distribution when the number of trials becomes too large for Crystal Ball to handle (more than 1000 trials). You also can use the Poisson distribution to approximate the binomial distribution when the number of trials is large, but there is little advantage to this since Crystal Ball takes a comparable amount of time to compute both distributions.

Likewise, the normal and Student's t distributions are related. With Degrees of Freedom > 30, Student's t closely approximates the normal distribution.

The binomial and hypergeometric distributions are also closely related. As the number of trials and the population size increase, the hypergeometric trials tend to become independent like the binomial trials: the outcome of a single trial has a negligible effect on the probabilities of successive observations. The differences between these two types of distributions become important only when you are analyzing samples from relatively small populations. As with the Poisson and binomial distributions, Crystal Ball requires a similar amount of time to compute both the binomial and hypergeometric distributions.

The yes-no distribution is simply the binomial distribution with Trials = 1.

The Weibull distribution is very flexible. Actually, it consists of a family of distributions that can assume the properties of several distributions. When the Weibull shape parameter is 1.0, the Weibull distribution is identical to the exponential distribution. The Weibull location parameter lets you set up an exponential distribution to start at a location other than 0.0. When the shape parameter is less than 1.0, the Weibull distribution becomes a steeply declining curve. A manufacturer might find this effect useful in describing part failures during a burn-in period. When the shape parameter is equal to 2.0, a special form of the Weibull distribution, called the Rayleigh distribution, results. A researcher might find the Rayleigh distribution useful for analyzing noise problems in communication systems or for use in reliability studies. When the shape parameter is set to 3.25, the Weibull distribution approximates the shape of the normal distribution; however, for applications when the normal distribution is appropriate, us it instead of the Weibull distribution.

The gamma distribution is also a very flexible family of distributions. When the shape parameter is 1.0, the gamma distribution is identical to the exponential distribution. When the shape parameter is an integer greater than one, a special form of the gamma distribution, called the Erlang distribution, results. The Erlang distribution is especially useful in the areas of inventory control and queueing theory, where events tend to follow Poisson processes. Finally, when the shape parameter is an integer plus one half (e.g., 1.5, 2.5, etc.), the result is a chi-squared distribution, useful for modeling the effects between the observed and expected outcomes of a random sampling.

When no other distribution seems to fit your historical data or accurately describes an uncertain variable, you can use the custom distribution to simulate almost any distribution. The Load Data button on the Custom Distribution dialog lets you read a series of data points or ranges from value cells in your worksheet. If you like, you can use the mouse to individually alter

the probabilities and shapes of the data points and ranges so that they more accurately reflect the uncertain variable.

Using probability functions

For each of the Crystal Ball distributions, there is an equivalent Excel function. You can enter these functions in your spreadsheet directly instead of defining distributions using the Define Assumption command. Be aware, though, that there are a number of limitations in using these functions. These are listed below.

To view these functions and their parameters, choose Insert > Function in Excel, and then be sure the category is set to Crystal Ball 7.

Excel 2007 Note: In Excel 2007, choose Formulas > Insert Function.
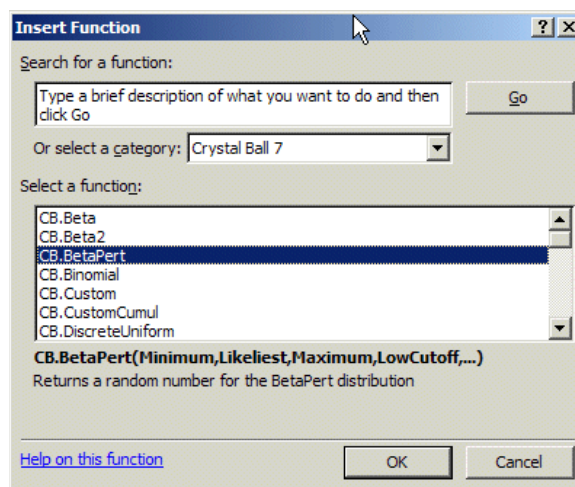

Figure A.50 Crystal Ball functions in Excel

Parameters and a brief description appear below the list of functions. The Cutoff parameters let you enter truncation values, while NameOf is the assumption name. For parameter descriptions and details on each distribution, see the entry for that distribution earlier in this appendix.

Crystal Ball Note: The beta distribution changed from previous versions to Crystal Ball 7. Both the original and revised functions appear for compatibility. CB.Beta has three parameters but CB.Beta2 is the Crystal Ball 7 version with Minimum and Maximum instead of Scale.

Limitations of probability functions

Distributions defined with probability functions differ from those entered with the Define Assumption command in these ways:

- You can't correlate them.

- You can't view charts or statistics on them.

- You can't extract data from them or include them in reports.

- They are not included in sensitivity analyses or charts.

Probability functions and random seeds

"Sampling preferences" on page 85 describes how you can use the Sampling tab of the Run Preferences dialog to use the same sequence of random numbers for each simulation.

If you use Define > Define Assumption or the Define Assumption toolbar button to define assumptions, the same sequence of random numbers is used for each simulation, even if you switch from Extreme speed to Normal speed or back to Extreme speed. If you use the probability functions to define assumptions, one sequence of random numbers is used for Extreme speed and a different sequence is used for Normal speed.

Appendix B

Maximizing Your Use of Crystal Ball

In this appendix

- Simulation accuracy
- Simulation speed
- Sample size
- Correlated assumptions

This chapter contains information that you can use to improve the overall performance of Crystal Ball. These improvements occur in terms of the accuracy of your model or speed of the results.