

In This Issue

1. Learn how to use **Risk Simulator** to perform a distributional fitting model.

“How do you know which distribution to use for a particular input variable in a model?”

Contact Us

Real Options Valuation, Inc.
4101F Dublin Blvd., Ste. 425,
Dublin, California 94568 U.S.A.
admin@realoptionsvaluation.com
www.realoptionsvaluation.com
www.rovusa.com

Theory

Another powerful simulation tool is distributional fitting; that is, which distribution does an analyst or engineer use for a particular input variable in a model? What are the relevant distributional parameters? If no historical data exist, then the analyst must make assumptions about the variables in question. One approach is to use the Delphi method, where a group of experts are tasked with estimating the behavior of each variable. For instance, a group of mechanical engineers can be tasked with evaluating the extreme possibilities of a spring coil's diameter through rigorous experimentation or guesstimates. These values can be used as the variable's input parameters (e.g., uniform distribution with extreme values between 0.5 and 1.2). When testing is not possible (e.g., market share and revenue growth rate), management can still make estimates of potential outcomes and provide the best-case, most-likely case, and worst-case scenarios, whereupon a triangular or custom distribution can be created.

However, if reliable historical data are available, distributional fitting can be accomplished. Assuming that historical patterns hold and that history tends to repeat itself, then historical data can be used to find the best-fitting distribution with their relevant parameters to better define the variables to be simulated. Figures 1, 2, and 3 illustrate a distributional-fitting example. The following illustration uses the **Data Fitting** file in the examples folder.

Procedure

Use the following steps to perform a distributional fitting model:

- Open a spreadsheet with existing data for fitting (e.g., use the *Data Fitting* example file).
- Select the data you wish to fit, not including the variable name (data should be in a single column with multiple rows).
- Select *Risk Simulator* | *Tools* | *Distributional Fitting (Single-Variable)*.
- Select the specific distributions you wish to fit to or keep the default where all distributions are selected and click **OK** (Figure 1).
- Review the results of the fit, choose the relevant distribution you want, and click **OK** (Figure 2).

Results Interpretation

The null hypothesis (H_0) being tested is such that the fitted distribution is the same distribution as the population from which the sample data to be fitted come. Thus, if the computed p-value is lower than a critical alpha level (typically 0.10 or 0.05), then the distribution is the wrong distribution. Conversely, the *higher the p-value, the better the distribution fits the data*. Roughly, you can think of p-value as a *percentage explained*; that is, if the p-value is 0.9727 (Figure 2), then setting a normal distribution with a mean of 99.28 and a standard deviation of 10.17 explains about 97.27% of the variation in the data, indicating an especially good fit. The data was from a 1,000-trial simulation in **Risk Simulator** based on a normal distribution with a mean of 100 and a standard deviation of 10. Because only 1,000 trials were simulated, the resulting distribution is fairly close to the specified distributional parameters, and in this case, about a 97.27% precision.

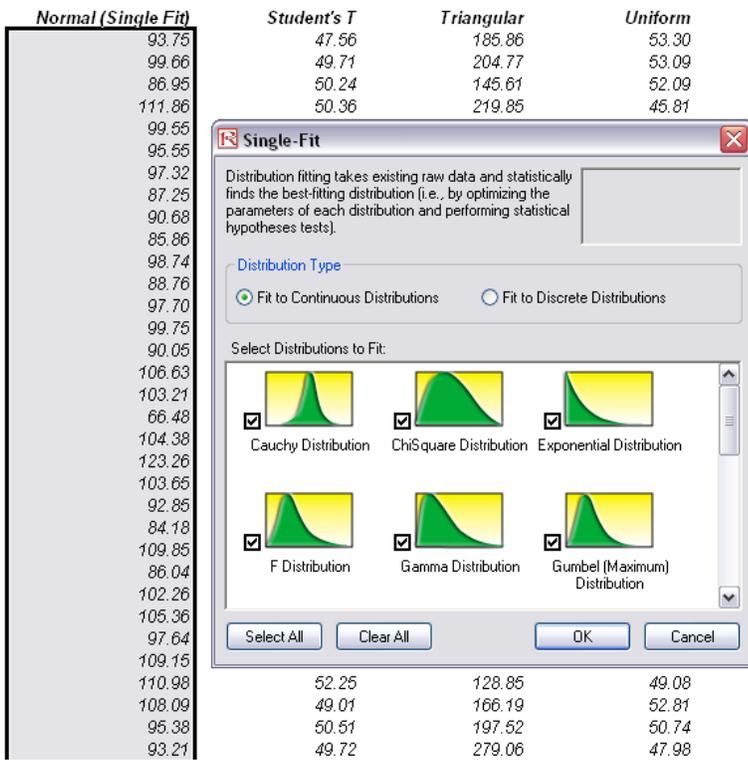


Figure 1. Single-Variable Distributional Fitting

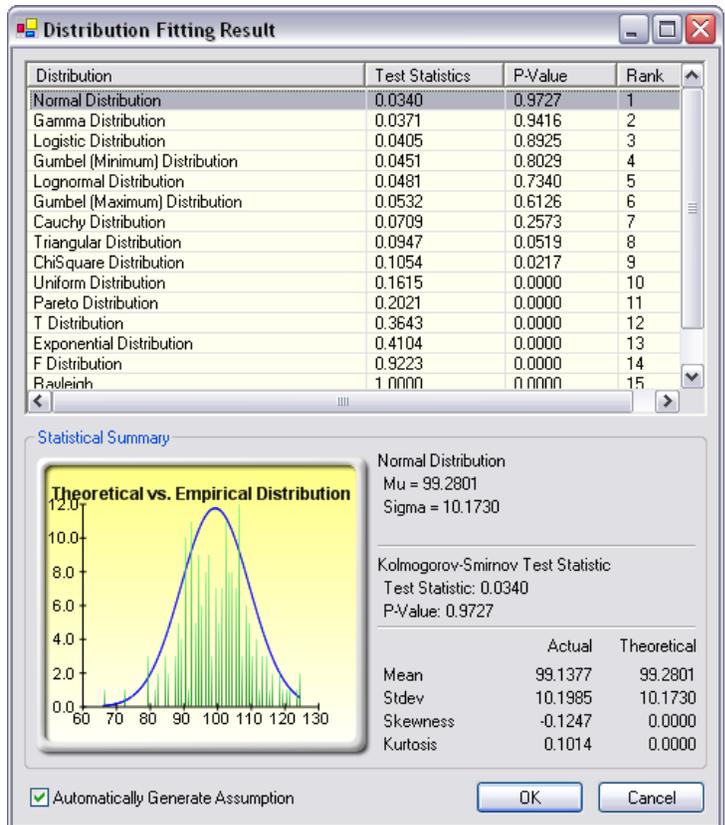
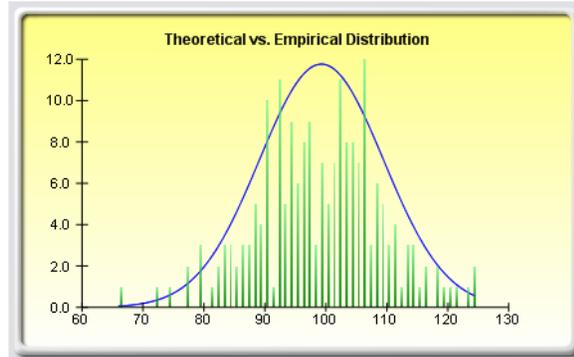


Figure 2. Distributional Fitting Result

Single Variable Distributional Fitting

Statistical Summary

Fitted Assumption	99.14	
Fitted Distribution	Normal Distribution	
Mu	99.28	
Sigma	10.17	
Kolmogorov-Smirnov Statistic	0.03	
P-Value for Test Statistic	0.9727	
	Actual	Theoretical
Mean	99.14	99.28
Standard Deviation	10.20	10.17
Skewness	-0.12	0.00
Excess Kurtosis	0.10	0.00



Original Fitted Data

93.75	99.66	86.95	111.86	99.55	95.55	97.32	87.25	90.68	85.86	98.74	88.76	97.70
99.75	90.05	106.63	103.21	66.48	104.38	123.26	103.65	92.85	84.18	109.85	86.04	102.26
105.36	97.64	109.15	110.98	108.09	95.38	93.21	83.86	100.17	110.17	103.72	120.52	95.09
115.18	83.64	90.23	92.44	92.37	92.70	110.81	72.67	104.23	96.47	121.15	94.92	77.26
103.45	96.75	93.91	101.91	124.14	90.95	107.13	92.02	96.43	96.35	88.30	108.48	113.50
101.40	104.72	102.43	113.59	124.15	109.24	105.34	104.57	97.83	94.39	116.19	84.66	101.17
106.13	107.17	95.83	106.67	92.42	79.64	94.15	106.00	113.45	92.63	94.51	93.05	96.19
100.85	83.34	111.82	118.12	87.17	103.66	106.93	82.45	102.74	86.82	106.68	89.61	94.56
101.34	91.32	102.02	82.51	104.46	84.72	105.05	108.40	106.59	109.43	92.49	94.52	94.00
105.92	88.13	96.41	101.45	79.93	89.68	102.91	114.95	92.58	94.05	107.90	111.05	90.58
97.09	105.44	94.95	102.55	77.41	108.53	90.54	100.41	106.83	99.63	79.72	89.32	116.30
98.27	101.73	90.84	74.45	102.24	103.34	96.51	114.55	93.94	106.29	102.95	112.73	98.09
108.20	105.80	106.48	102.88	104.93	103.00	99.10	108.52	101.31	88.17	90.62	96.53	106.03
109.12	104.23	90.34	95.12	102.03	100.00	118.17	99.06	81.89	104.29	92.68	114.89	102.49
119.21	106.20	88.26	92.45	105.15	103.79	100.84	95.19	85.10	97.25	87.65	97.58	111.44
99.52	89.83	97.86	90.96	97.14								

Figure 3. Distributional Fitting Report

Both the results (Figure 2) and the report (Figure 3) show the test statistic, p-value, theoretical statistics (based on the selected distribution), empirical statistics (based on the raw data), the original data (to maintain a record of the data used), and the assumption complete with the relevant distributional parameters (i.e., if you selected the option to automatically generate assumption and if a simulation profile already exists). The results also rank all the selected distributions and how well they fit the data.

Fitting Multiple Variables

For fitting multiple variables, the process is fairly similar to fitting individual variables. However, the data should be arranged in columns (i.e., each variable is arranged as a column) and all the variables are fitted. The same analysis is performed when fitting multiple variables as when single variables are fitted. The difference here is that only the final report will be generated and you do not get to review each variable's distributional rankings. If the rankings are important, run the single-variable fitting procedure instead, on one variable at a time.

Procedure

- Open a spreadsheet with existing data for fitting.
- Select the data you wish to fit (data should be in multiple columns with multiple rows).
- Select *Risk Simulator* | *Tools* | *Distributional Fitting (Multi-Variable)*.
- Review the data, choose the types of distributions you want to fit to, and click *OK*.

(See "Notes" on next page.)

Notes

Notice that the statistical ranking methods used in the distributional fitting routines are the Chi-Square test and Kolmogorov-Smirnov test. The former is used to test discrete distributions and the latter, continuous distributions. Briefly, a hypothesis test coupled with the maximum likelihood procedure with an internal optimization routine is used to find the best-fitting parameters on each distribution tested, and the results are ranked from the best fit to the worst fit. There are other distributional fitting tests such as the Anderson-Darling, Shapiro-Wilks, etc. However, these tests are very sensitive parametric tests and are highly inappropriate in Monte Carlo simulation distribution-fitting routines when different distributions are being tested. Due to their parametric requirements, these tests are most suited for testing normal distributions and distributions with normal-like behaviors (e.g., binomial distribution with a high number of trials and symmetrical probabilities) and will provide less accurate results when performed on non-normal distributions. Take great care when using such parametric tests. The Kolmogorov-Smirnov and Chi-Square tests employed in *Risk Simulator* are nonparametric and semiparametric in nature and are better suited for fitting normal and non-normal distributions.