Suitability of Different Probability Distributions for Performing Schedule Risk Simulations in Project Management

J. Krige Visser
Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa

***Abstract*--Project managers are often confronted with the question on what is the probability of finishing a project within budget or finishing a project on time. One method or tool that is useful in answering these questions at various stages of a project is to develop a Monte Carlo simulation for the cost or duration of the project and to update and repeat the simulations with actual data as the project progresses. The PERT method became popular in the 1950's to express the uncertainty in the duration of activities. Many other distributions are available for use in cost or schedule simulations.**

This paper discusses the results of a project to investigate the output of schedule simulations when different distributions, e.g. triangular, normal, lognormal or betaPert, are used to express the uncertainty in activity durations. Two examples were used to compare the output distributions, i.e. a network with 10 activities in sequence and a network where some of the activities are performed in parallel. The results indicate that there is no significant difference in the output distributions when different input distributions with the same mean and variance values are used.

I. INTRODUCTION

A. Background

Many factors influence the total project outcome or performance. The three dimensional goal for any project is well known, i.e. to finish the project on time, within budget and with satisfactory performance or quality [11]. Large and complex projects typically have high uncertainty in the final project outcome, e.g. space exploration (International space station, Mars Curiosity lander), passenger airplane development (Airbus A380, Boeing 787) and bridge construction (Akashi-Kaikyo, Millau). Such complex projects are often late and exceed the original budget. Good project and risk management therefore involves the reduction of uncertainty in the outcome of the project.

One method of reducing general uncertainty in the project duration is to model the uncertainty in the duration of individual activities by means of probability distributions and to run a simulation with multiple trials to produce a probability distribution for the total project duration. An initial simulation is usually performed before project implementation and for projects with a long duration the simulation can be repeated at various stages or milestones of the project. Actual duration values for completed activities are then used in place of the initial estimates and 80 or 90% certainty of completion are compared with the original values derived from the first simulation. This was the approach used for the schedule risk management of the Øresund bridge project that was completed 5 months ahead of schedule in 2001 [1].

The Project Evaluation and Review Technique (PERT), in conjunction with the Critical Path Method (CPM), were developed in the 1950's to address the uncertainty in project duration for complex projects [11], [13]. The expected value or mean value for each activity of the project network was calculated by applying the beta distribution and three estimates for the duration of the activity. The total project duration was determined by adding all the duration values of the activities on the critical path. The Monte Carlo simulation (MCS) provides a distribution for the total project duration and is therefore more useful as a method or tool for decision making.

B. Research objectives

The main objective of this study was to compare the output of a project schedule risk simulation when alternative input distributions are used to express the uncertainty in the duration of individual activities of the project. This objective is supported by the following sub-objectives.

- Perform a literature search on comparison of probability distributions used in general Monte Carlo simulation.
- Define two test 'projects', one with a number of activities to be performed in sequence and one with parallel activities, that can be used to compare different distributions.
- Run simulations using different input schedule distributions for a series and parallel network with a simulation add-in for MS Excel and standalone simulation software.
- Compare the total project duration outputs graphically and through statistical analysis.

The following hypothesis was defined.

H_0: The distribution for the total duration of a project comprising ten individual activities that each have a triangular distribution is not significantly different to the distribution for the total project duration obtained when the ten activities have a normal, lognormal, Gumbel or Weibull distribution, 95% confidence level.

H_1: The distribution for the total duration of a project comprising ten individual activities that each have a triangular distribution is significantly different to the distribution for the total project duration obtained when the ten activities have a normal, lognormal, Gumbel or Weibull distribution, at a 95% confidence level.

II. LITERATURE AND THEORY

A. Project scheduling

Planning and scheduling are tools that assist in achieving the three crucial goals of project management. The work breakdown structure (WBS) is one of the outputs of the planning function and enables the development of the project cost breakdown and the project schedule. Single point estimates are used for the duration of activities and the activity network enables a calculation of the critical path and total project duration. In modern projects, a schedule simulation is often used to complement the single point deterministic approach by incorporating uncertainty in the duration of the activities [2].

B. Project risk simulation

MCS methods have become more popular amongst project managers and planners in the last decade due to the availability of fast desktop computers and software that is freely available, especially as add-ins for spreadsheets like MS Excel. The method is well documented and discussed in various textbooks [11], [2], [14], and [6]. Wood [16] says MCS can be "applied in many diverse fields that require outcomes to be quantified statistically under conditions of uncertainty to aid in decision-making."

Various software programs for performing simulations are available to model uncertainty in the cost or duration of activities. Two standalone software programs that use discrete event simulation are Arena and GoldSim. Some add-ins for performing simulations with MS Excel are @Risk, Crystal Ball, SimVoi and RiskAmp. GoldSim [5] and SimVoi [15] were selected for this research project.

C. Probability distributions

Triangular distribution

The triangular distribution is bounded on the left and right and can be symmetric or skew depending on the values of the three parameters that determine the shape of the distribution. The three parameters of the triangular distribution are typically:

- a = lower bound (minimum)
- m = most likely value (mode)
- b = upper bound (maximum)

The density and cumulative distribution functions for the triangular distribution are not provided by MS Excel but can be calculated using the formulas from [3]. The random variates for the triangular distribution are also not provided in MS Excel but were calculated with (1) to (3) below.

$$T(\rho,a,m,b) = a + \sqrt{\rho(m-a)(b-a)} \quad for\ 0 < \rho < P(a,m,b) \tag{1}$$

$$T(\rho,a,m,b) = m - \sqrt{(1-\rho)(m-a)(m-b)} \quad for\ P(a,m,b) \le \rho \le 1 \tag{2}$$

where $P(a,m,b) = (b-a)/(m-a)$ (3)

and ρ is the random number obtained with the RAND() function of MS Excel or the RandUniform(0,1) function provided by SimVoi.

Normal distribution

The normal distribution, also know as the Gaussian distribution, is a continuous and symmetric distribution defined by two parameters, i.e. the mean value μ and the standard deviation, σ. The density function of the normal distribution is defined over -∞ to +∞ and care should be taken when using this distribution in simulation where the variable is only defined from 0 - ∞, e.g. duration of an activity which cannot be negative. The normal distribution is symmetric around the mean value whereas the other distributions investigated are non-symmetric. Random variates for the normal distribution can be calculated with the NORMINV(RAND(), μ, σ) function of MS Excel or the RandNormal(μ, σ) function provided by SimVoi.

Lognormal distribution

The lognormal distribution is a continuous, non-symmetric distribution that is often used to model the duration of activities or tasks. It applies mostly to novice artisans or workers that have to perform non-standard and complex tasks. These tasks often have an overflow especially when something goes wrong. It has two parameters, i.e. the mean value μ and the standard deviation σ. The random variates can be calculated using the MS Excel function LOGNORM.INV(RAND(), μ ,σ) or the RandLognormal(μ, σ) function provided by SimVoi.

Gumbel distribution

The Gumbel distribution is also known as the Smallest Extreme Value (SEV) distribution (Type I) and has a positive location parameter μ and a positive scale parameter β. A random variate, T, for the Gumbel distribution was calculated using (4) and the RandUniform(0,1) function to generate a random number. This distribution is mainly used in the analysis of extreme values and in survival analysis.

$$T(\rho,\mu,\beta) = \mu - \beta \cdot LN\left[LN\left(\frac{1}{\rho}\right)\right] \tag{4}$$

where ρ is a random number.

Frechet distribution

The Fréchet distribution, also known as the inverse Weibull distribution, belongs to the broader family of extreme value distributions and has a positive shape parameter α and a positive scale parameter, s. It is often applied for natural events or disasters like earthquakes, storms, floods and volcanic eruptions. The parameters of the Fréchet distribution cannot be calculated from the mean and standard deviation using analytical equations and a numerical equation solver was used to determine the parameters. A random variate, T, for the Fréchet distribution was calculated using (5).

$$T(\rho,s,\alpha) = -s \cdot [LN(\rho)]^{-1/\alpha} \tag{5}$$

where ρ is a random number.

Weibull distribution

The Weibull distribution is well known for modelling reliability of physical assets and humans due to its versatility with regard to failure rate. Two and three parameter distributions are typically used but the 2-parameter version is simpler and easier to apply. The 'shape' parameter, α, determines whether an item exhibits a decreasing, constant or decreasing failure rate. The second parameter, β, is known as the 'characteristic life' when used in reliability applications.

The Weibull distribution is also useful to model task or activity duration in projects since it is one of a few distributions that is skewed towards the left, i.e. a negative skewness factor. The general assumption is that the distribution of task duration is skewed to the right and therefore some historical data is needed on specific task durations to warrant the use of the Weibull distribution. A random variate, T, for the Weibull distribution was calculated using (6).

$$T(\rho,\beta,\eta) = \eta\left[LN\left(1-\frac{1}{1-\rho}\right)\right]^{\frac{1}{\beta}} \quad (6)$$

where ρ is a random number.

BetaPert distribution

The betaPert distribution was developed in the 1950's to describe the uncertainty in activity durations of complex projects and used the beta distribution as the basis. The beta distribution is rather complex but quite versatile [9]. The parameters of the beta distribution are not easy to estimate for the duration of activities but the betaPert simplifies this process by introducing 3 parameters, i.e. an optimistic, most likely (mode) and pessimistic estimate for the duration. The OPERT software add-in for MS Excel was used to generate random variates from the betaPert distribution [8].

D. Comparison of distributions

Ferson et al. [4] were of the opinion that 'the results of probabilistic risk analyses are known to be sensitive to the choice of distributions used as inputs, an effect which is undoubtedly even stronger for the tail probabilities'. In some simulation applications, this may be the case, especially when a large number of trials are used. The occurrence of very small or very large random variates is more likely for fat-tailed distributions like the Gumbel, Fréchet and Fisk distributions than for the normal or lognormal distributions.

Hajdu and Bokor [7] tested simulations with uniform, triangular and beta distributions and indicated that 'the use of different distributions with the same three-point estimation has a smaller effect on the project duration than a 10% difference in the values of the three-point estimation'.

Sherer et al. [12] investigated the application of a symmetric triangular distribution as an approximation for the normal distribution and a non-symmetric triangular distribution for the lognormal distribution. The authors concluded that the triangular distribution provides a good approximation for the normal distribution in the range of the mean ± 2.44σ.

Wood [16] performed simulations with the same input data and found that the output cumulative distributions for the uniform, normal, lognormal and triangular distributions could vary as much as 10%. This contradicts the findings of Hajdu and Bokor [7].

III. METHODOLOGY

A. Overview

Two theoretical experiments were performed using a simulation software add-in for MS Excel, i.e. SimVoi™ as well as the GoldSim™ standalone software program. The schedule simulation can be done with MS Excel without additional software or add-ins using standard functions provided, e.g. RAND() for generating random numbers between 0 and 1 and FREQUENCY(Array1, Array2,) to produce a histogram of random variates. The variates can be sorted from smallest to largest and a cumulative distribution values can be determined. For the total duration of the project. Random variates from the distributions used in this study can be determined from analytical expressions except for the normal distribution for which the built-in function of MS Excel can be used. However, the simulation add-ins or standalone software automate the simulation process and is more convenient, easier and faster to use.

Two project activity networks, each comprising 10 activities were selected for this study i.e. a 'series' network where all activities are performed in sequence and a 'parallel' network where some activities take place in parallel. As a starting point values were assumed for the parameters of the triangular distribution for each of the ten activities. The two networks are shown in Fig. 1 and Fig. 2 below.

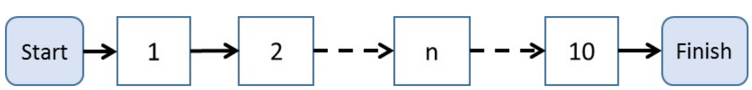

Figure1: Series activity network

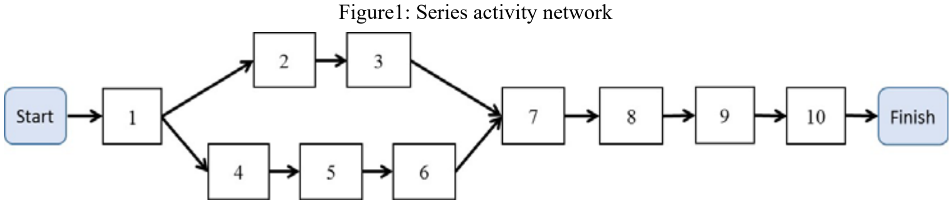

Figure2: Parallel activity network

The values assumed for the parameters of the triangular distribution as well as the mean and variance values are shown in Table 1. The values for the parameters a, m and b in Table 1 are arbitrary time units which could typically be days or weeks.

TABLE 1: VALUES ASSUMED FOR DURATION OF ACTIVITIES 1-10

Activity	a	m	b	Mean	Variance
1	5	7	12	8.000	2.167
2	7	10	15	10.667	2.722
3	10	12	17	13.000	2.167
4	6	7	15	9.333	4.056
5	7	10	16	11.000	3.500
6	10	11	16	12.333	1.722
7	5	8	13	8.667	2.722
8	6	10	20	12.000	8.667
9	17	20	29	22.000	6.500
10	3	4	8	5.000	1.167

The values for the triangular distributions were selected to give a right skewed distribution (positive skewness). Many of the standard distributions like the lognormal, gamma and Gumbel are right skewed. The normal distribution (symmetric) and Weibull distribution (left skewed if the shape parameter is greater than 1) are exceptions.

B. Data gathering

A spreadsheet simulation model for each of the two examples, series and parallel activities, was formulated in MS Excel. The mean and variance values calculated for each of the activities as given in Table 1 were used to determine the parameters of the lognormal, normal, Gumbel, Weibull and Fréchet distributions. A simulation with 20000 trials for SimVoi and 10000 trials for GoldSim was then performed for each activity network with the different input distributions for the activity durations. The output of the simulation is a distribution for the total project duration.

The basic output of some add-in software is a list of all the random variates determined for various random numbers. For this study it was an array comprising 20000 values (10000 for GoldSim) for the total project duration. Some add-in software offer further analysis of this basic data to provide the mean, standard deviation, skewness and other statistical parameters of the total project duration. Values for the cumulative distribution are also provided by most simulation software.

C. Data analysis

The output distributions for different input distributions for the activity durations were compared using a number of different measures. The simplest comparison is to compare the array of duration values for one distribution as input against the array of duration values for another set of input distributions. The triangular distribution was chosen as reference and all other distributions were therefore compared against this one. Two distributions that are both close to a normal distribution can be compared by performing a two sample independent t-test which produces a p-value. For a p-value greater than 0.05 (typical value used) the null hypothesis, i.e. that the two distributions that are compared are significantly different, can be rejected in favour of the alternative hypothesis that they are the same. Another method that was used is to compare two scenarios was the Kolmogorov-Smirnov (K-S) test which does not require normality of the two distributions that are compared.

The cumulative distribution values provided by the software add-ins are determined through interpolation of the cumulative data and the total duration values at the 80 and 90% probability values (P80 and P90) were used for comparison. These values are often the most important output of a schedule risk simulation and are used for critical schedule related decisions throughout the project execution. The percentage variation of the P80 and P90 values as compared with the values for the triangular distribution were used to determine how well the simulation produced useful outputs.

IV. RESULTS AND FINDINGS

A. Overview

The SimVoi and GoldSim software provide the raw data, i.e. the random variates calculated for the total project duration in this study. Most software also provides some statistical data of the output distribution. SimVoi and GoldSim also provide graphical output, typically the probability density and cumulative distribution function of the output distribution.

The results are provided separately for the two cases, i.e. the series activity network and the parallel activity network. The descriptive statistics are given for the SimVoi add-in only since GoldSim does not provide all this detail. The P80 and P90 duration values are compared against the triangular distribution to assess the effect of the tails of the input distributions.

Statistical tests were performed to determine how well the output distributions for total project duration agreed when the Gumbel, lognormal, normal, Weibull and Fréchet input distributions were used in comparison with the output when triangular distribution inputs were used. The difference in cumulative probability at about 20 data points were squared, added together and the square root was calculated. The Kolmogorov-Smirnov test for two samples was done with the cumulative probability values for the output distributions.

B. Series activity network

The SimVoi simulation software add-in provides some statistical data on the output distribution. The results for the different input distributions are given in Table 2 below.

TABLE 2: DESCRIPTIVE STATISTICS FOR SERIES ACTIVITY NETWORK, SIMVOI SIMULATION

	Triangular	Gumbel	Lognormal	Normal	Weibull	Fréchet
Mean	112.05	112.00	112.06	111.98	112.01	111.93
Std Dev	5.94	5.90	5.98	5.94	5.93	5.89
First Quartile	107.94	107.83	107.89	107.95	108.08	107.80
Median	111.90	111.57	111.91	111.96	112.12	111.29
Third Quartile	116.00	115.75	116.02	116.01	116.01	115.36
Skewness	0.132	0.412	0.193	-0.006	-0.112	0.830

The mean and standard deviation values for all 5 input distributions are very similar, varying only a fraction of a percent. The largest difference is seen in the skewness which varies from -0.112 for the Weibull distribution inputs to +0.830 for the Fréchet input distributions.

P80 and P90 values

The difference in the P80 and P90 values of the other input distributions and that of the triangular distribution is shown in Fig. 3 below.

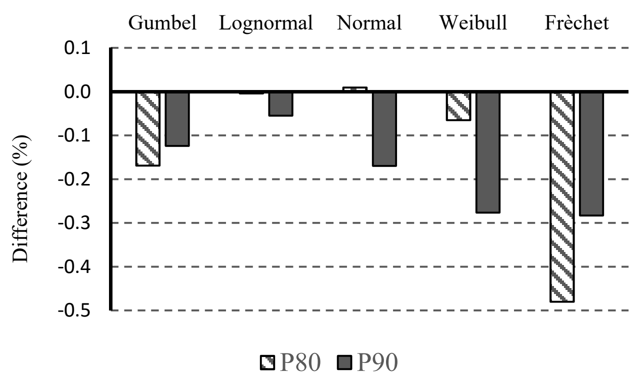

Figure 3: Difference in P80 and P90 values for series network

For the normal, lognormal and Weibull distributions the P80 values differ very slightly (< 0.1%) from the P80 values of the triangular distribution. The P90 values differ somewhat more (~ 0.3%), except for the Fréchet distribution that differs by ~0.5%. It is clear that the longer right hand tail of the Fréchet does make some difference in the P80 and P90 values.

Sum of the squares

The cumulative distributions produced by the simulation can be compared by calculating the square root of the sum of the squares of the difference for a number of duration values. Twenty data points, corresponding to a duration interval of 2.5 time units, were used for the comparison and the results are shown in Fig. 4 for both simulation programs.

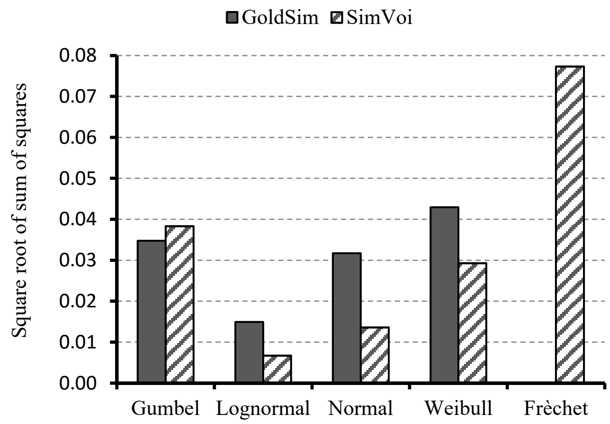

Figure 4: Square root of the sum of the squares of differences in CDF

The lognormal distribution provided the smallest values and therefore gives the closest result in comparison with the triangular distribution. The normal distribution, even though it's a symmetric distribution, provided the 2nd best result. The Fréchet distribution gave the worst result vs the triangular distribution. The Goldsim software does not provide the Fréchet distribution and could therefore not be used to perform the simulation.

Two sample t-test

The p-values as determined by the 2-sample independent t-test (T.TEST function of MS Excel) for different input distributions are shown in Table 3. The distributions were compared against the triangular distribution.

TABLE 3: P-VALUES FOR TWO SAMPLE T-TEST FOR SERIES NETWORK

	Gumbel	Lognormal	Normal	Weibull	Fréchet
SimVoi	0.423	0.756	0.295	0.589	0.044
GoldSim	0.889	0.811	0.130	0.186	

The p-values for the Gumbel, lognormal, normal and Weibull distributions are greater than 0.05 which means that the output with these input distributions are not significantly different than the output if the triangular distribution is used. However, the p-value for the Fréchet as input distribution is less than 0.05 and the null hypothesis cannot be accepted meaning the Fréchet could produce an output that is significantly different than the output for the triangular distribution as input.

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (K-S) test can be used to compare two distributions that are not necessarily normal distributions. The difference in values for the cumulative distribution for two cases is determined and compared against a critical factor, D_{crit} [10]. The K-S test uses the maximum vertical deviation between the two CDF curves as the statistic D_{stat}. As for the K-S two-sample test, the null hypothesis (at significance level α) is rejected if the difference, ΔD, is positive. This difference is determined with (7).

$$\Delta D = D_{stat} - D_{crit} \quad (7)$$

where $D_{stat} = Maximum(D_m - D_n)$ (8)

and D_{crit} is the critical value obtained from standard K-S tables for the number of data points.

The difference ΔD for the K-S-test is shown in Fig. 5 for the different input distributions. The triangular distribution was used as reference input distribution.

All values in Fig. 5 are negative and therefore the null hypothesis can be accepted at a 95% significance level, i.e. the output distributions from the simulation does not differ significantly from the outputs obtained with the triangular distribution as input.

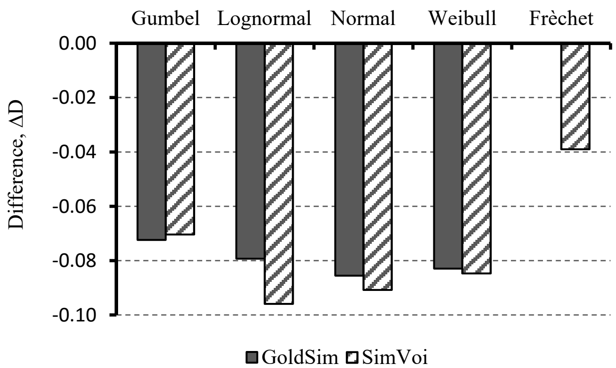

Figure 5: ΔD values for K-S test for series activity network

C. Parallel activity network

The same values for the parameters of the triangular distribution as used for the series activities were also used for the parallel situation as shown in Fig.2. The descriptive statistics as calculated by the SimVoi add-in for the total project duration are given in Table 4.

As seen in Table 4 the mean and standard deviation (Std. Dev.) values for the output distributions are very close. As noticed with the series network the largest difference is in the skewness factors which vary from -0.163 for the Weibull input distributions to +1.333 for the Fréchet input distributions. The skewness value for the lognormal distribution was the closest to that of the triangular distribution, i.e. 0.233 vs, 0.144.

P80 and P90 values

The P80 and P90 values as provided by the SimVoi software were analysed and the percentage difference from the triangular distribution values were determined. The results are shown in Fig. 6 below.

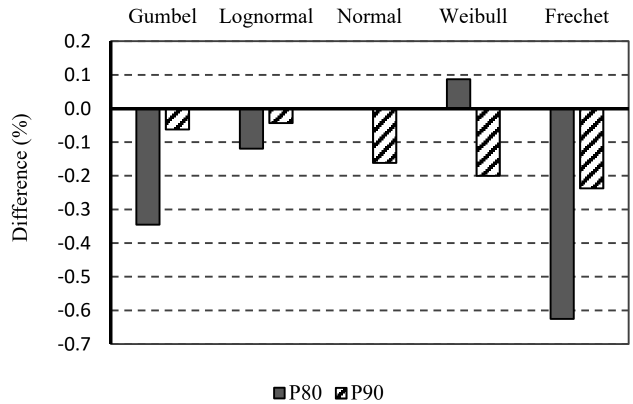

Figure 6: Difference in P80 and P90 values for parallel network

For the normal distribution, the P80 values are nearly identical to the triangular values while the P90 value is lower by less than 0.2%. For the Gumbel, lognormal and Fréchet distributions, the P90 values are closer to the triangular values than the P80 values. The largest difference is for the Fréchet distribution where the P80 value is about 0.62% less. The P80 value for the Weibull distribution is nearly 0.1% larger than the P80 value for the triangular distribution but the P90 value is about 0.2% less. The explanation for this is because the Weibull distribution is left skewed (negative skewness) and drops sharply to the right of the mode of the distribution.

TABLE 4: DESCRIPTIVE STATISTICS FOR PARALLEL ACTIVITY NETWORK, SIMVOI SIMULATION

	Triangular	Gumbel	Lognormal	Normal	Weibull	Fréchet
Mean	88.33	88.33	88.38	88.39	88.38	88.34
Std. Dev.	5.52	5.50	5.50	5.50	5.50	5.54
First Quartile	84.52	84.49	84.58	84.69	84.72	84.49
Median	88.20	88.00	88.23	88.36	88.58	87.65
Third Quartile	92.01	91.75	91.92	92.09	92.20	91.46
Skewness	0.144	0.441	0.233	0.004	-0.163	1.333

Sum of the squares

The cumulative distributions produced by the simulation can be compared by calculating the square root of the sum of the squares of the difference for a number of duration values. Twenty data points, corresponding to a duration interval of 2.5 time units, were used for the comparison and the results are shown in Fig. 7 for 2 simulation programs.

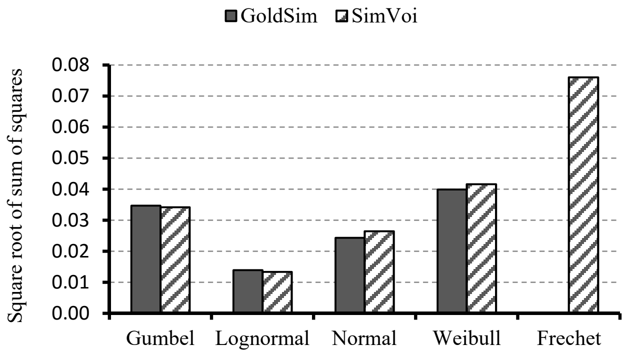

Figure 7: Square root of the sum of the squares of differences in CDF

The lognormal distribution provided the smallest values of about 0.013 and therefore gives the closest result in comparison with the triangular distribution. The normal distribution, even though it's a symmetric distribution, provided the 2^{nd} best result. The Fréchet distribution gave the worst result vs the triangular distribution. The Goldsim software does not provide the Fréchet distribution and could not be used to run the simulation.

Two sample T-test

The p-values for the parallel network as determined by the 2-sample independent t-test are shown in Table 5.

TABLE 5: P-VALUES FOR TWO SAMPLE T-TEST FOR PARALLEL NETWORK

	Gumbel	Lognormal	Normal	Weibull	Fréchet
SimVoi	0.875	0.423	0.348	0.453	0.981
GoldSim	0.930	0.384	0.234	0.218	

The p-values for all the input distributions are greater than 0.05 which supports the output with these input distributions are not significantly different than the output if the triangular distribution is used.

Kolmogorov-Smirnov test

The difference, ΔD, from the K-S test for the parallel network is shown in Fig. 8 for different input distributions.

The lognormal distribution has the largest negative ΔD and is therefore the closest fit to the triangular distribution. The Fréchet distribution has the smallest negative ΔD value and is therefore the worst fit vs. the triangular distribution. All 5 input distributions have negative ΔD values and therefore the hypothesis that the simulation output from these 5 distributions does not differ significantly from the output of the triangular distribution at a 95% confidence level can be accepted.

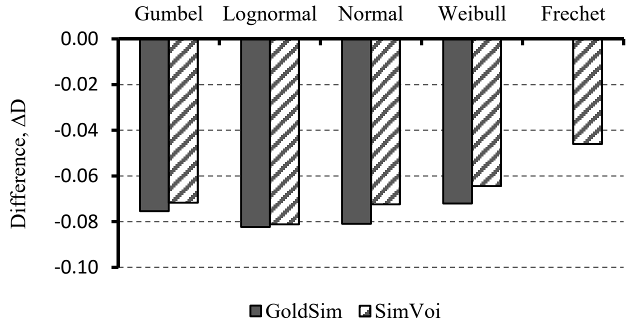

Figure 8: ΔD values from K-S test for parallel activity network

D. The betaPert vs triangular distribution

The betaPert distribution was developed in the 1950's to model the high uncertainty in the duration of activities of highly complex projects. It is derived from the beta distribution to facilitate the input of three duration values, i.e. optimistic (a), most likely (m) and pessimistic (b) estimates. The output of a simulation for betaPert input values will not provide the same output distribution as the triangular distribution for the same input values since the mean and standard deviation (Std Dev.) values are different as seen in (9) to (12) below.

BetaPert distribution:

$$mean = \frac{a+4m+b}{6} \quad (9)$$

$$Std\ Dev = \frac{b-a}{6} \quad (10)$$

Triangular distribution:

$$mean = \frac{a+m+b}{3} \quad (11)$$

$$StdDev = \frac{a^2+m^2+b^2-ab-am-bm}{18} \quad (12)$$

The betaPert distribution assigns a greater weight to the most likely value in comparison with the triangular distribution where the 3 parameters have equal weight.

The data given in Table 1 for the network with parallel activities was used to compare these two distributions, but without activities 4, 6 and 10. The cumulative distributions (CDF) for the total duration for both input distributions are shown in Fig. 9.

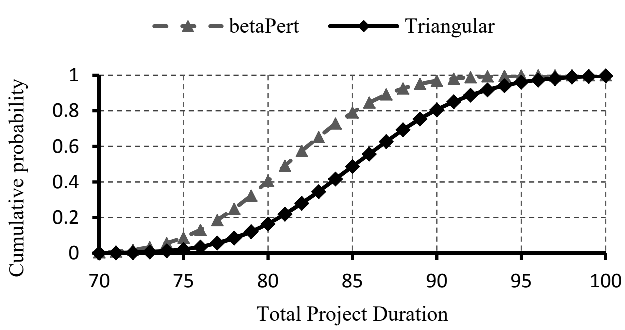

Figure 9: CDF for triangular and betaPert input distributions with the same values for 3 parameters, a; m and b

For the second example the mode, m, of the triangular distribution was also used as the mode for the betaPert distribution. However, the values of the scale parameters, a and b, for the betaPert distribution were calculated to give the same mean value and standard deviation as for the triangular distribution. The output distribution with the betaPert as input is compared with the output with the triangular distribution as input in Fig. 10.

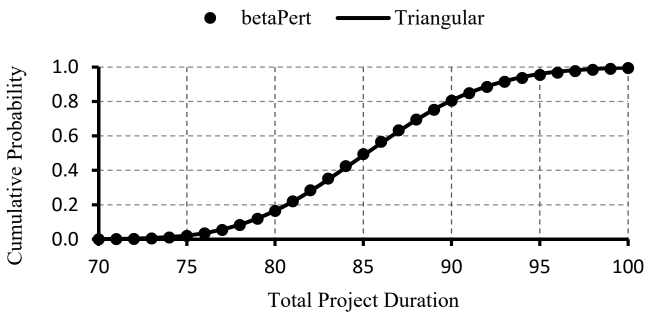

Figure 10: CDF for triangular and betaPert input distributions for the same mode, mean and standard deviation

The output distributions are very similar for these input distributions. The square root of the sum of the squares of the difference between the two cumulative distributions in Fig. 10 was 0.023 and the p-value for the t-test was 0.81. The P80 values differ by 0.03% and the P90 values differ by 0.11%. The ΔD value for the K-S test was -0.052 which supports the hypothesis that the betaPert distribution does not differ significantly from the triangular distribution when the same mode, mean and standard deviation values are used. The p-value and similarity in the P80 and P90 values indicate that the two output distributions are very similar and that both these input distributions can be used for a schedule simulation.

E. Summary of findings

The difference in the P80 and P90 values for the Gumbel, lognormal, normal, Weibull and Fréchet input distributions compared against the triangular input distribution was less than 1% for the series and parallel network. The lognormal provided the smallest difference for P80 and P90 for both activity networks and the Fréchet the largest difference for both.

The square root of the sum of the differences for about 20 points of the cumulative distribution function when compared with the triangular distribution was smaller than 0.08 for the series and parallel networks. The lognormal had the smallest difference and the Fréchet had the largest difference for both.

The ΔD values for the K-S test were negative for the series and parallel networks for all input distributions compared with the triangular distribution. The lognormal had the largest negative and the Fréchet the smallest negative values. This indicated that the lognormal compares the best to the triangular input distribution.

V. CONCLUSIONS

Two project activity networks were investigated in this study, i.e. a series network with ten activities and a parallel network with ten activities. Duration values were accepted for the triangular distribution for each activity and Monte Carlo simulations were performed for this distribution and 5 other distributions. The distributions for the total project duration were compared to determine whether there is a significant difference in the output.

The results of this study indicate that the choice of probability distribution for the duration of activities is not critical for performing a schedule simulation for the total project. The distributions for the total project duration differed only slightly for the 6 different input distributions (triangular, lognormal, normal, Gumbel, Fréchet and Weibull) as indicated by the results of the t-test and the Kolmogorov-Smirnov test for similarity. The mean and standard deviation values for all input distributions were exactly the same but the skewness values were not. The Weibull distribution is skewed towards the left (negative skewness) for the values used in this study, the normal distribution is symmetric and the other distributions are skewed towards the right (positive skewness). For the two network examples used it can be concluded that the skewness of the input distribution does not have a significant effect on the results of the simulation, except that the skewness of the distributions for the total project duration vary slightly.

The special case of the betaPert input distribution was also investigated to determine whether the output distribution differed significantly when compared with the triangular input distribution. If the same values for the three parameters of the triangular distributions are also used for the betaPert distributions, the output distribution will be different. However, if the same mean, standard deviation and mode are used for both the betaPert and the triangular, the output distributions will be very similar.

Considering the results obtained this study, it can be concluded that the choice of input distribution does not have a significant effect on the output distribution of a schedule simulation when these input distributions are symmetric or slightly skewed. The results cannot be generalized to highly skewed distributions where the different tails of the distributions might lead to different results. This paper confirms what might be expected for a small network with slightly skewed input distributions for the duration of the activities. A focus for further study and analysis could be a larger network and highly skewed distributions. The output distribution might be distorted in the 80-95% probability region for this scenario.

The results of this study provide confidence to risk and project managers that they can use any of the input distributions analysed in this study, together with duration data provided by experts, for schedule risk simulations.

VI. RECOMMENDATIONS

Two sets of input data were used for this study, one for series activities and the other for activities that can be performed in parallel. To make the findings more general, at least one more set of data with highly skewed distributions should be analysed. It would be interesting to see how well the Weibull (left skewed) and normal distribution (symmetric) compare with right skewed distributions (lognormal, Gumbel and Fréchet) for highly skewed input distributions.

The schedule simulations were performed with two simulation software packages, i.e. SimVoi and GoldSim. It is recommended that one more software program, @Risk™ or RiskAmp™) be used to repeat these simulations and compare the results with the results given in this paper.

ACKNOWLEDGMENTS

I thank the Department of Engineering and Technology Management at the University of Pretoria for the use of the facilities and administrative support. I also thank the Microsoft, Treeplan Software and GoldSim companies for the use of their software for this project.

REFERENCES

[1] Christensen, P.J. and J. Rydberg, "Overcoming Obstacles: Strategic risk management in the Øresund bridge project", *PM Network*, 15, (11) pp. 30-37, 2001.

[2] Cooper, D., P. Bosnich, S. Grey, G. Purdy, G. Raymond, P. Walker and M. Wood, *Project Risk Management Guidelines. Managing Risk with ISO 31000 and IEC 62198*. 2nd Edition, John Wiley & Sons Ltd. United Kingdom, 2014.

[3] Evans, M., N. Hastings and B. Peacock. *Statistical distributions*, 3rd Edition, John Wiley & Sons, New York. 2000.

[4] Ferson, S, Ginzberg, L. and R. Akçakaya, "Whereof one cannot speak: When input distributions are unknown," Retrieved 11/01/15 World Wide Web, http://www.ramas.com/whereof.pdf).

[5] GoldSim Technology Group, "GoldSim™ Monte Carlo Simulation software for decision and risk management", Retrieved 01/27/16 World Wide Web, http://www.goldsim.com/Home/.

[6] Grey, S. *Practical Risk Assessment for Project Management*. John Wiley & Sons, Chichester, England. 1995.

[7] Hajdu, M. and O. Bokor, "The Effects of Different Activity Distributions on Project Duration in PERT Networks," *Procedia - Social and Behavioral Sciences* 119, pp. 766-775, 2014.

[8] Hayes, C. and J. Jacobs, "Openpert Reference Guide", Retrieved 11/01/15 World Wide Web, https://openpert. googlecode.com/files/openpert_reference_guide.pdf.

[9] McDonald, J.B. and Y.J. Xu. "A generalization of the beta distribution with applications," *Journal of Econometrics* 66, pp. 133-152, 1995.

[10] Pham, H. *Handbook of Engineering Statistics*, Springer-Verlag, UK. 2006.

[11] Raydugin, Y., *Project Risk Management: Essential Methods for Project Teams and Decision Makers*, John Wiley & Sons Inc., New Jersey, 2013.

[12] Scherer, W.T., T.A. Pomroy and D.N. Fuller, "The triangular density to approximate the normal density," *Reliability Engineering and System Safety* Vol. 82, pp. 331–341, 2003.

[13] Shankar. N.R., S.S. Babu, Y.L.P. Thorani and D. Raghuram, "Right skewed distribution of activity times in PERT", *International Journal of Engineering Science and Technology* 3, pp. 2932-2938, 2011.

[14] Steyn, H.P.J. et al. *Project Management: A multi-disciplinary approach.* 4th Edition, FPM Publishing, Pretoria, South Africa, 2016.

[15] Treeplan Software, SimVoi™ Monte Carlo Simulation Add-in. Retrieved 01/30/16 World Wide Web, http://simvoi.com.

[16] Wood, D.A., "Risk Simulation Techniques to aid project cost-time planning and management," *Risk Management* 4 (1) pp. 41-60, 2002.