



PR-155

Top-level Schedule Distribution Models

Peter Frederic

19 March 2013

2013 Professional Development & Training Workshop

New Orleans, LA

June 18-21, 2013

Tecolote Research, Inc.
420 Fairview Ave. Suite 201
Goleta, CA 93117-3626
(805) 571-6366

Tecolote Research, Inc.
415 E. Ocean Ave., Suite H
Lompoc, CA 93436
(805) 588-1330

Top-level Schedule Distribution Models

**Peter Frederic
Tecolote Research, Inc.**

Abstract

In the initial stages of a development project, it is sometimes necessary to build a summary-level schedule for planning and budgeting purposes before the day-by-day details of the project are fully defined or understood. However, when uncertainty assessments are performed on schedule networks containing few activities, the distribution forms chosen for individual activity durations can have a significant impact on the overall results. It is therefore important to choose uncertainty distribution forms that accurately represent the behavior of the sub-network of activities represented by each summary activity.

In this paper, we investigated probability theory to see if there were statistical distributions that were well suited to modeling the completion of typical schedule sub-networks consisting of multiple parallel activities. In order to test the applicability of the distributions investigated, we developed an Excel/@Risk tool to compare how various distributions behave versus simulated data from a simplified schedule network. We evaluated numerous distribution forms including: general Beta, PERT Beta, Log-normal, Weibull, Erlang, and Poisson distributions. We concluded that only the general BETA distribution could accurately model a sub-network consisting of multiple parallel paths. We propose additional research to develop Beta parameters to represent a variety of network topologies (e.g. mostly serial/mostly parallel, generous reserves/no reserves, many discrete risks/few discrete risks, etc.).

Table of Contents

1	INTRODUCTION	1
2	THE PROCESS	2
2.1	LOG-NORMAL DISTRIBUTION	3
2.2	WEIBULL DISTRIBUTION	4
2.3	ERLANG DISTRIBUTION	5
2.4	POISSON DISTRIBUTION	6
2.1	PERT-BETA DISTRIBUTION	7
2.1	BETA DISTRIBUTION	8
3	ADDITIONAL RESEARCH REQUIRED	10
4	CONCLUSIONS	11

1 INTRODUCTION

In the initial stages of a development project, it is sometimes necessary to build a summary-level schedule for planning and budgeting purposes before the day-by-day details of the project are fully defined or understood. However, when uncertainty assessments are performed on schedule networks containing few activities, the distribution forms chosen for individual activity durations can have a significant impact on the overall results. It is therefore important to choose uncertainty distribution forms that accurately represent the behavior of the sub-network of activities represented by each summary activity.

In the sections below, we evaluate numerous distribution forms including: Log-normal, Weibull, Erlang, Poisson, PERT Beta, and general Beta distributions.

2 THE PROCESS

We investigated probability theory to see if there were statistical distributions that were well suited to modeling the completion of multiple parallel activities. In order to test the applicability of the distributions investigated, we developed an Excel™/@Risk™ tool to compare how various distributions behave versus simulated data from a simplified schedule network. The model simulates 50 parallel activities feeding into one. Each of the 50 activities was modeled as a lognormal distribution with a mean duration of ten days and a standard deviation of four days. The test process is: 1) use @Risk to run 10,000 iterations of the sample network, 2) capture the final completion date in all 10,000 iterations, 3) bin the simulation results into a histogram, and 4) use the Excel Solver to fit PDFs for various distributions to the simulated data histogram. For each distribution type tested, the sections below provide a very brief description of the distribution as well as a probability density graph that shows how well the distribution fit the test data.

2.1 LOG-NORMAL DISTRIBUTION

From Wikipedia: "A log-normal distribution is a probability distribution of a random variable whose logarithm is normally distributed... A variable might be modeled as log-normal if it can be thought of as the multiplicative product of many independent random variables each of which is positive."

PDF:

$$\frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

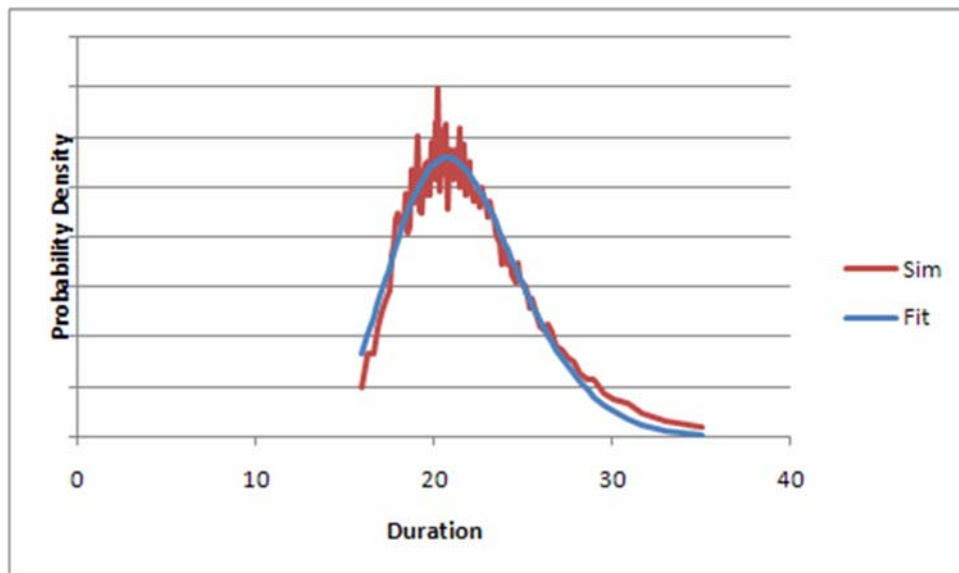
Where:

x = Duration

μ = Mean $\ln(\text{duration})$

σ = Standard deviation $\ln(\text{duration})$

The log-normal distribution matches the "50 parallel activities feeding into one" simulation data surprisingly well, but falls significantly short at the high tail of the distribution:



2.2 WEIBULL DISTRIBUTION

From Wikipedia: "The Weibull distribution is used:

- In survival analysis
- In reliability engineering and failure analysis
- In industrial engineering to represent manufacturing and delivery times"

PDF:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

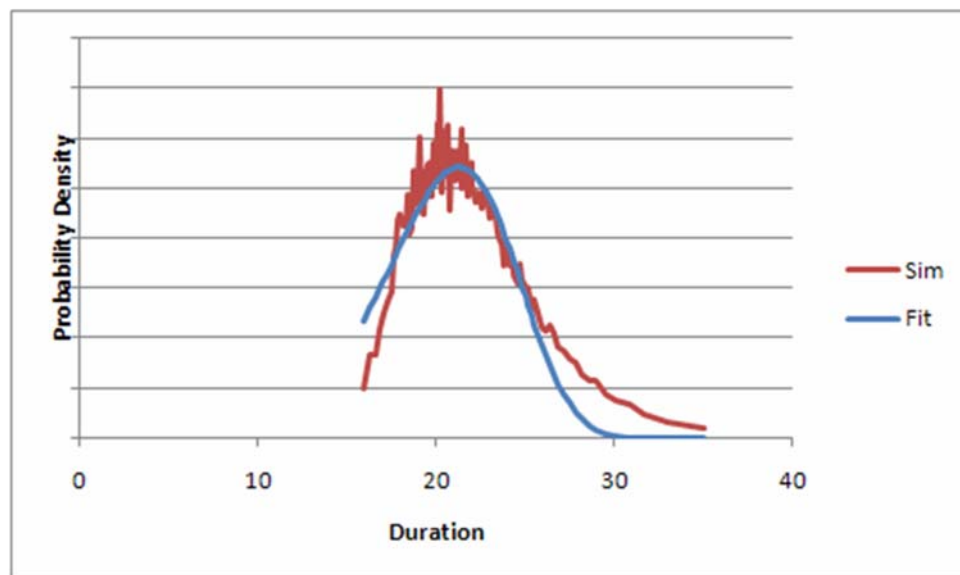
Where:

x = Duration

k = Shape parameter

λ = Scale parameter

The Weibull distribution does not match the "50 parallel activities feeding into one" simulation data well:



2.3 ERLANG DISTRIBUTION

From Wikipedia: "The Erlang distribution was developed by A. K. Erlang to examine the number of telephone calls which might be made at the same time to the operators of the switching stations. This work on telephone traffic engineering has been expanded to consider waiting times in queuing systems in general."

PDF:

$$\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

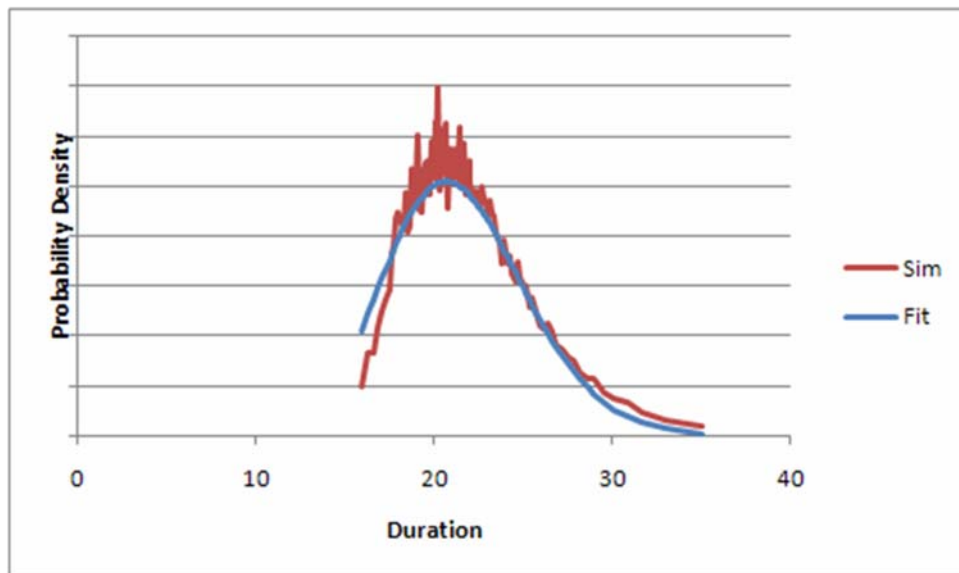
Where:

x = Duration

k = Shape parameter

λ = Rate parameter

The Erlang distribution matches the "50 parallel activities feeding into one" simulation data fairly well, but falls significantly short at the high tail of the distribution:



2.4 POISSON DISTRIBUTION

From Wikipedia: "...the Poisson distribution (pronounced pwason(or Poisson law of small numbers[1]) is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event." **Because it is a discrete distribution, it is not applicable to the finish time of multiple parallel activities.**

2.1 PERT-BETA DISTRIBUTION

From the @Risk™ help files: "The PERT distribution (meaning Program Evaluation and Review Technique) is rather like a Triangular distribution, in that it has the same set of three parameters. Technically it is a special case of a scaled Beta (or BetaGeneral) distribution. In this sense it can be used as a pragmatic and readily understandable distribution."

PDF:

$$f(y, a, m, b) = f(y; \alpha, \beta, a, b) = \frac{1}{B(\alpha, \beta)} \frac{(y - a)^{\alpha-1} (b - y)^{\beta-1}}{(b - a)^{\alpha+\beta-1}}.$$

Where:

y = Duration

μ = Mean = $(a + 4 * m + b) / 6$

α = Shape parameter = $6 * (\mu - a) / (b - a)$

β = Shape parameter = $6 * (b - \mu) / (b - a)$

a = Absolute minimum y

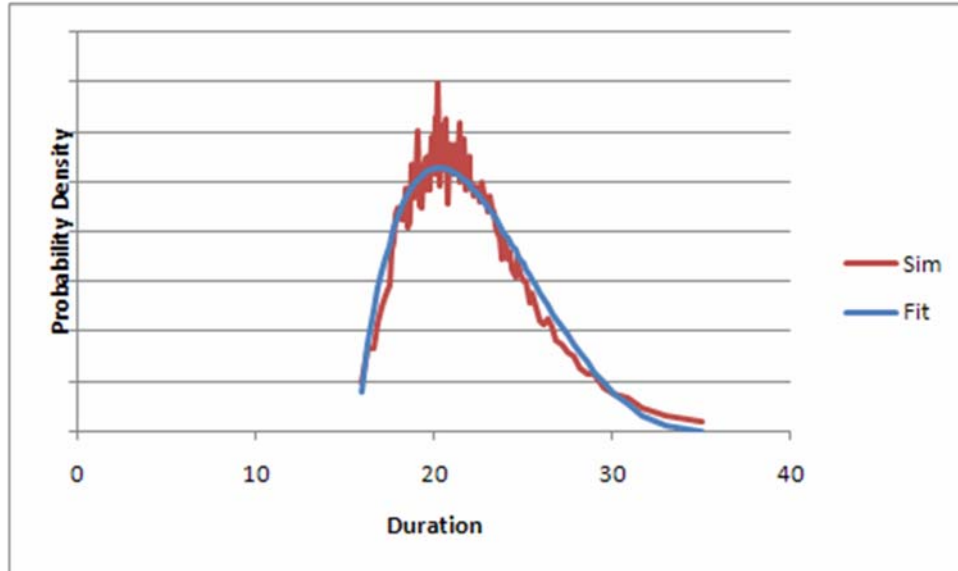
b = Absolute maximum y

m = Most likely value of y

and:

$$B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt$$

Because the shape parameters of the PERT form of the Beta curve are constrained, it is not able to match the "50 parallel activities feeding into one" simulation data well, especially at the high tail of the distribution:



2.1 BETA DISTRIBUTION

From Wikipedia: "The beta distribution can be used to model events which are constrained to take place within an interval defined by a minimum and maximum value. For this reason, the beta distribution — along with the triangular distribution — is used extensively in PERT, critical path method (CPM) and other project management/control systems to describe the time to completion of a task." The four parameter version of the Beta distribution is extremely flexible.

PDF:

$$f(y; \alpha, \beta, a, b) = \frac{1}{B(\alpha, \beta)} \frac{(y - a)^{\alpha-1} (b - y)^{\beta-1}}{(b - a)^{\alpha+\beta-1}}$$

Where:

y = Duration

α = Shape parameter

β = Shape parameter

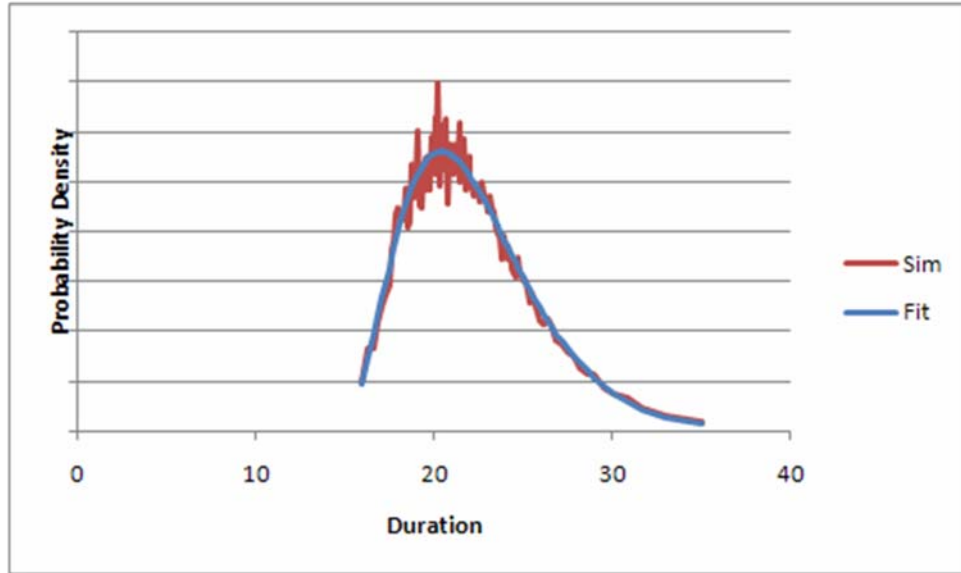
a = Absolute minimum y

b = Absolute maximum y

and:

$$B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt$$

Because of this four-parameter flexibility, the general Beta distribution is able to model the “50 parallel activities feeding into one” simulation data almost exactly:



3 ADDITIONAL RESEARCH REQUIRED

To develop a truly useful capability for developing summary-level analysis schedules, we should attempt to fit Beta curves to sample cases with a broad range of attributes, including the following:

- Various sub-network sizes (10 tasks, 100 tasks, etc.)
- Highly-serial versus highly-parallel sub-networks
- High-risk versus low-risk sub-network tasks
- Highly-skewed versus symmetrical task duration distributions
- Different distribution forms for individual task durations
- Highly correlated versus uncorrelated sub-network task durations
- Networks with many external constraints
- Discrete risk impacts
- Real-world networks with a mixture of the features above

We could develop a catalog of Beta distribution parameters to address a variety of schedule network topologies that would provide the capability to accommodate varying levels of available planning detail.

It might also be useful to apply more statistical rigor to the process. It should be possible to quantify how well the Beta curves fit their underlying sample cases. Furthermore, it would be useful to characterize the error inherent in summarizing sub-networks. Presumably, the catalog of Beta distributions will have a finite number of entries, so each Beta curve would have to address a range of conditions, and at the outer edges of the range, the distribution would be less accurate than at the center of the range. A sharp decision-maker would want to know how much of the uncertainty in the schedule model's predictions are due to the fact that the analysis was done at a summary level, and how much uncertainty would be eliminated by building a more detailed schedule model. It would be nice to have a quantitative answer to that question.

4 CONCLUSIONS

The general Beta distribution was the only distribution with sufficient degrees of freedom to fit the simulated completion-of-multiple-parallel-activities data well. If we ran the test model using various levels of uncertainty, degrees of correlation, and network topologies, we could develop a catalogue of Beta distribution shape parameters that would be useful in cases where a detailed schedule network and schedule risk analysis are not available. These could be used to develop simple analysis schedules in early-planning situations where a detailed schedule network does not exist, but a general understanding of the nature of the plan, whether conservative and sequential, or aggressive and concurrent, does exist.