Presented at the 2013 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

Tecolote Research, Inc.
Bridging Engineering and Economics
Since 1973

# Top-level Schedule Distribution Models

Peter Frederic

June 2013

# Agenda

- Introduction
- The Process
- Simulation Data
- Additional Research Required
- Conclusions

# Introduction

- In initial stages of a development project, sometimes necessary to build summary-level schedule for planning and budgeting
- When uncertainty assessments performed on schedule networks containing few activities, the duration distribution forms can have significant impact on overall results
- Important to choose uncertainty distribution forms that accurately represent behavior of sub-network of activities represented by each summary activity
- Various distribution forms have been evaluated against a sample sub-network to identify ideal distribution form

# The Process

- **Focused on finding distributions suited to modeling the completion of multiple parallel activities**
- **Developed an Excel™/@Risk ™ tool to compare how various distributions behave vs. simulated data from simplified schedule network**
- **50 parallel activities feeding into one**
- **Each activity modeled as lognormal distribution with a mean duration of ten days and standard deviation of four days**
- **Test process:**
  1. Use @Risk to run 10,000 iterations of the sample network
  2. Capture completion date in all 10,000 iterations
  3. Bin the simulation results into a histogram
  4. Use the Excel Solver to fit PDFs for various distributions to the simulated data histogram

# Log-normal Distribution

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:

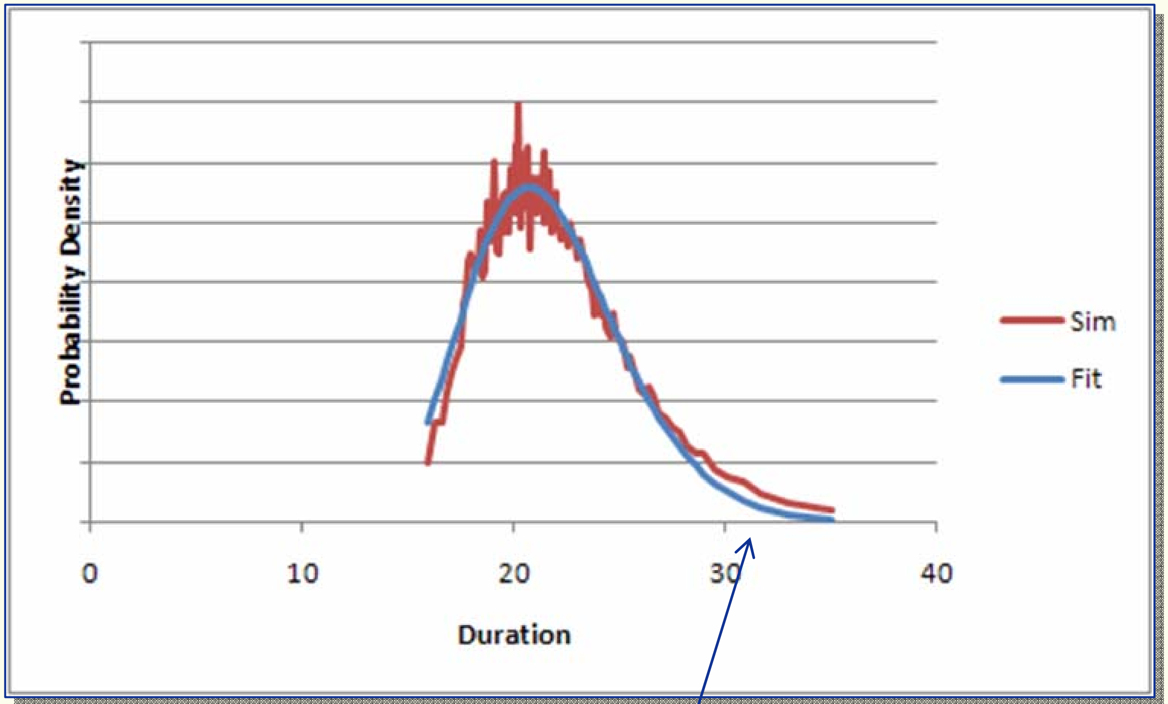
$$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$

Where:

x = Duration

μ = Mean ln(duration)

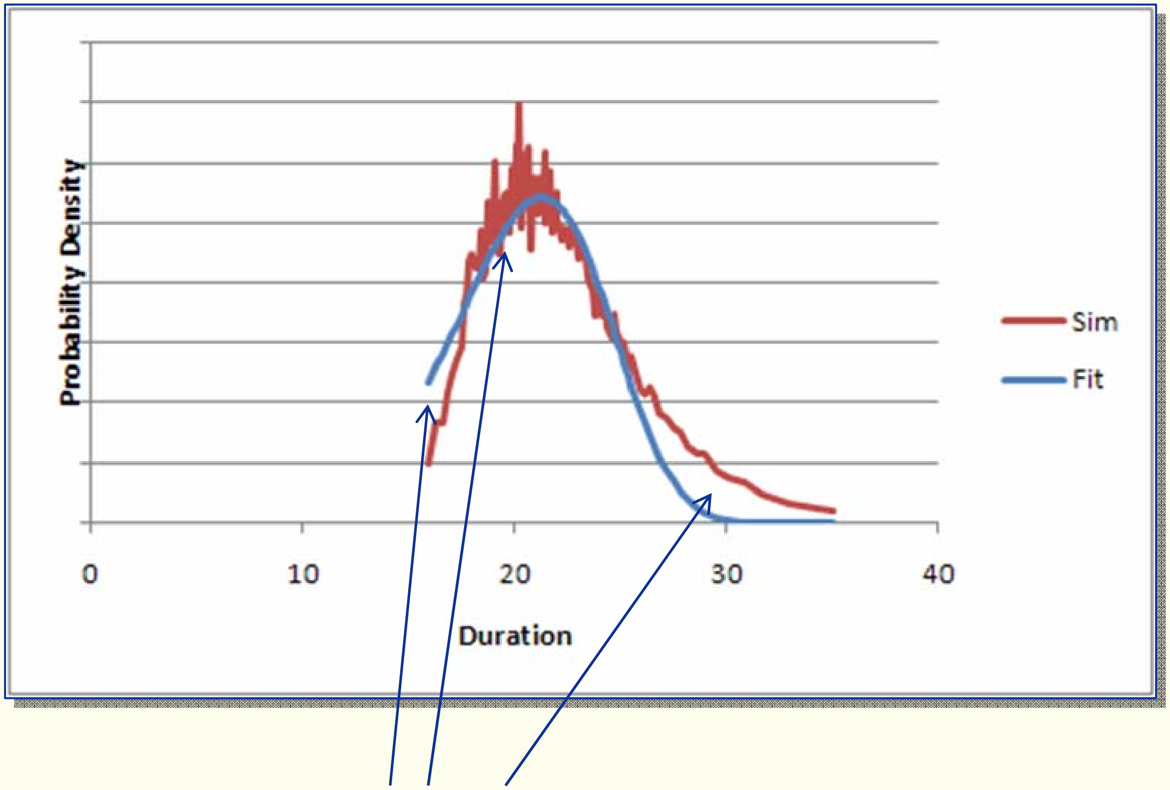
α = Standard deviation ln(duration)




Matches surprisingly well, but falls significantly short at the high tail of the distribution

# Weibull Distribution

Wikipedia: "The Weibull distribution is used:
•In survival analysis
•In reliability engineering and failure analysis
•In industrial engineering **to represent manufacturing and delivery times**"




Does not match well

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

# Erlang Distribution

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:

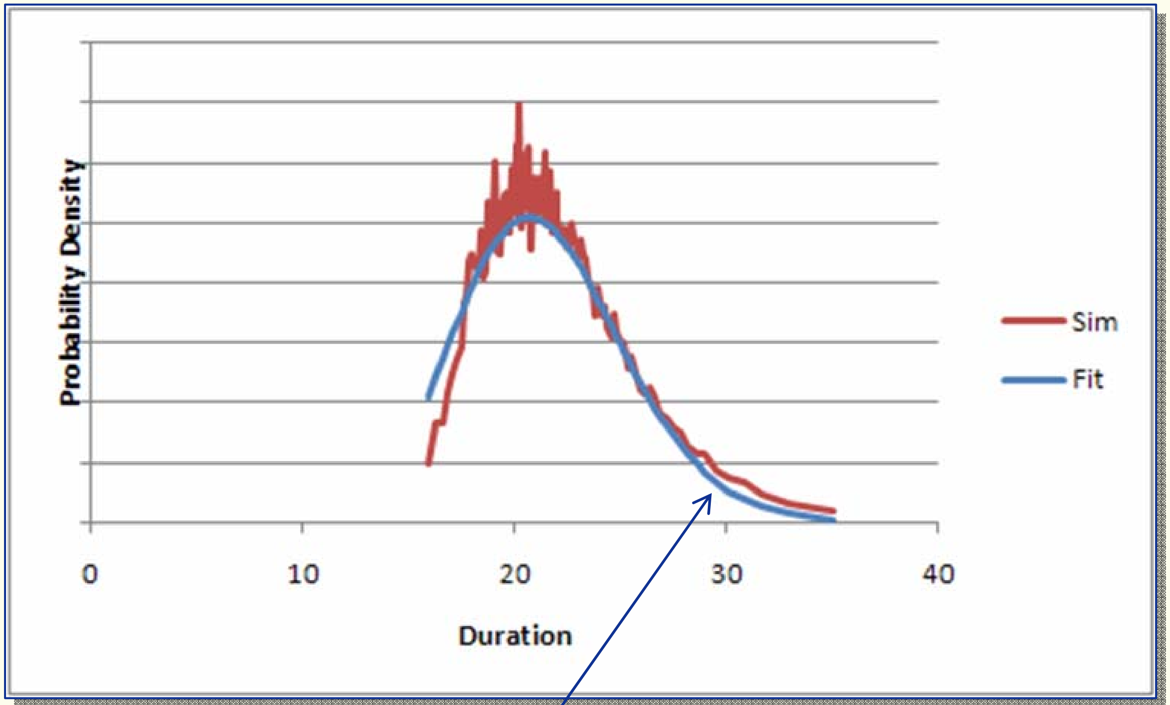
$$\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

Where:

x = Duration

$k$ = Shape parameter

$\lambda$ = Rate parameter




Matches fairly well, but falls significantly short at the high tail of the distribution

# Poisson Distribution

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**

# PERT Beta Distribution

@Risk Help: "The PERT distribution 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 * (μ - a) / (b - a)

β = Shape parameter = 6 * (b - μ) / (b - a)

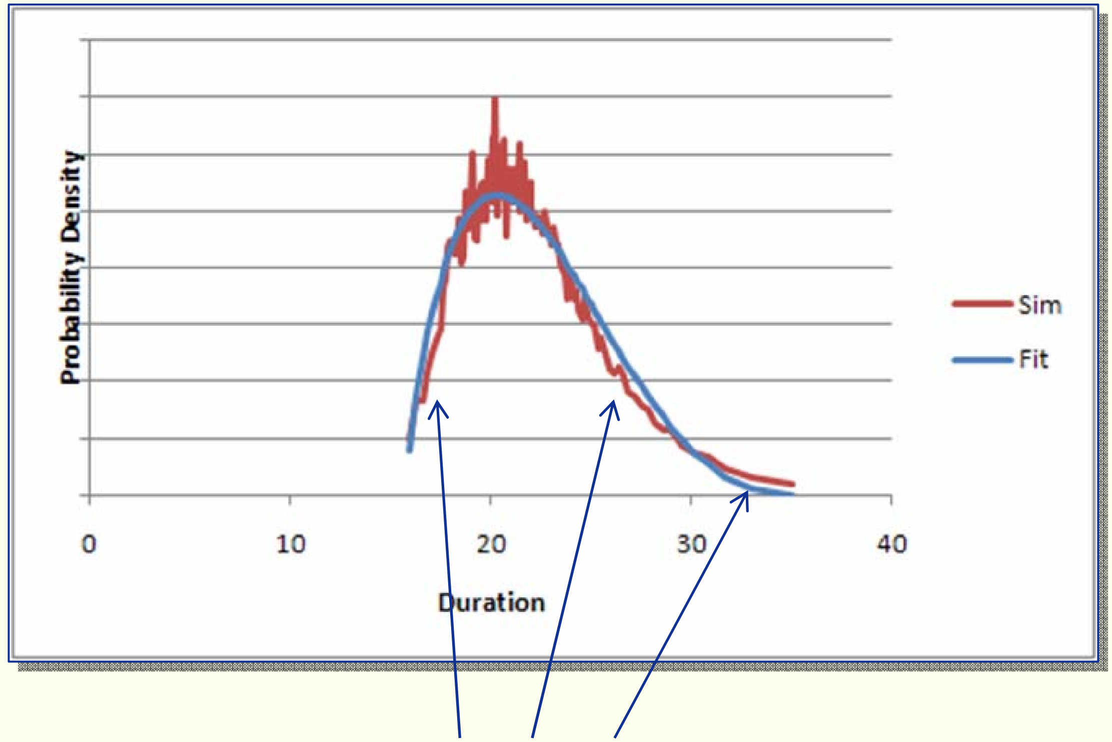
a = Absolute minimum y

b = Absolute maximum y

m = Most likely value of y

and:

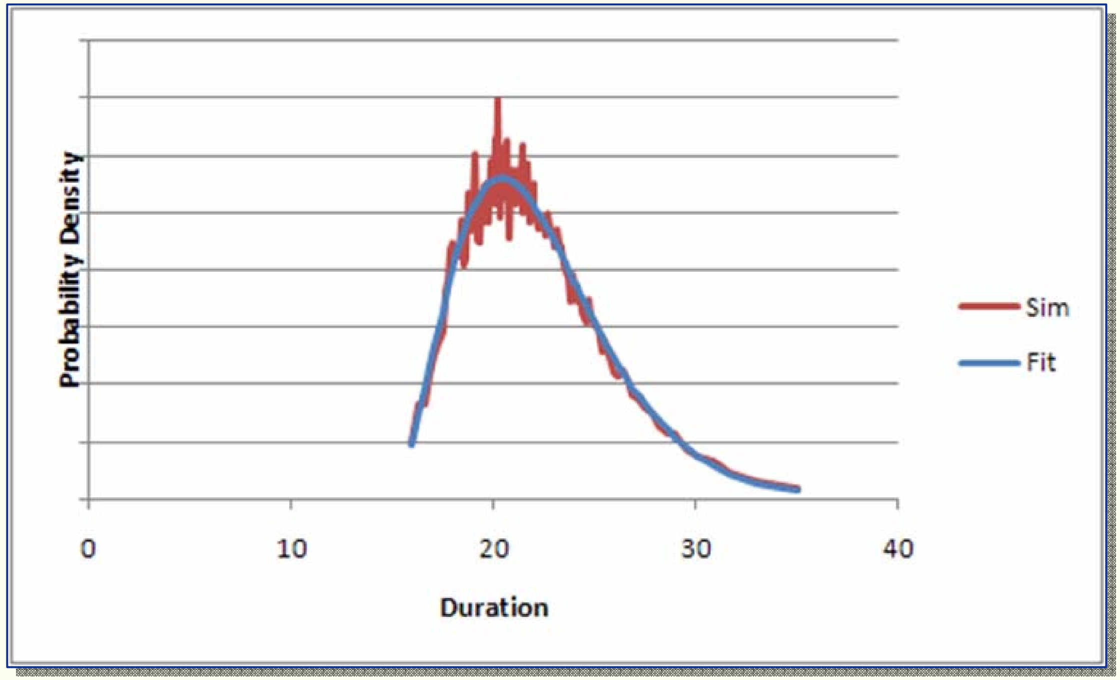
$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$$




Does not match particularly well

# General Beta Distribution

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… is **used extensively in… project management/ control systems to describe the time to completion of a task.**"




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:

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

**Matches almost exactly!**

# Additional Research Required

- **Broader set of sample cases**
  - 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
- **Develop a catalog of Beta distributions to address the sub-network types described above**
- **More statistical rigor?**
  - Characterize the error inherent in summarizing sub-networks

# 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.
- Additional research is required to develop a catalog of Beta distributions to address a broad range of schedule modeling situations