PRT-209

Introducing RIFT to Protect Your Uncertain Schedule

Presented at the
International Cost Estimating and Analysis Association (ICEAA)
Professional and Training Workshop
Atlanta Georgia
7-10 June 2016

Nick DeTore
ndetore@tecolote.com

Peter Frederic
pfrederic@tecolote.com

23 June 2016

Preface

Several decades ago, cost estimators realized that a logical way to establish reserves was through detailed cost risk assessment via Monte Carlo simulation. This led immediately to a dilemma – people discovered that the numbers at any stated level of confidence did not sum. While mathematically correct, this was difficult to explain to management and not directly useful for reserve allocation.

Many cost allocation schemes have been developed to overcome this problem. However, recent emphasis on schedule risk assessment has exposed a similar issue with schedule reserve allocation. Durations at a stated confidence level along a critical path do not add up to the overall duration of the path, which can lead to hard-to-explain phenomena such as apparently negative reserve values at interim milestones along the critical path. In schedule networks, this problem is compound by the potential for alternate critical paths.

NASA funded Tecolote to conduct a study to define a methodology for allocating schedule reserves to lower-level activities in a schedule network such that reserves are available at the time they are most likely to be used. In the course of this study, numerous allocation approaches were considered. One approach was called Conditioned on Criticality (CoC) which involved performing an initial wave of allocation using statistics based only on Monte Carlo iterations in which a task was on the critical path. This seemed contradictory to our unstated goal which was to allocate "reserve" to tasks that were most likely to "need it”. CoC seemed to do the opposite, allocating more "reserve" to tasks that were less critical and therefore less likely to slip due to nodal bias.

In 2014, a new research effort was conducted on behalf of the ACEIT/JACS development team. Early in this effort, review of the CoC concept led to a realization: from the manager's perspective, tasks that are more critical, i.e. more likely to impact the ultimate project finish date, should be managed more closely, that is allowed less tolerance for slippage. Notice the use of the word "tolerance" instead of "reserve." Traditionally, we have confused ourselves by talking about schedule reserve as if it is some amount of time that we can hold in a special rainy day account to use when bad things happen. What we're really trying to do is establish finish dates that preserve the desired probability of success. The 2014 research effort has yielded a relatively straightforward methodology for doing that. The methodology works from existing simulation results produced by any Monte Carlo schedule risk tool and requires no explicit knowledge of schedule network logic.

INTRODUCTION

Our story introduces a new date. This is a unique date derived from a risk and uncertainty analysis on a schedule. To the best of our knowledge, this date has been calculated for the entire schedule, but neither calculated nor defined for each task. The desire to determine such dates at the task level has been an ongoing inquiry for some time. There are many heuristic approaches, but they all have serious drawbacks and none of them have gained traction as the consensus solution. Finding an efficient, meaningful, repeatable and defendable implementation has been elusive.

We believe we have found not only a practical solution to this common probabilistic schedule issue, but in the process, opened a new area of analytical reports that could (and should) become as common as reporting probability level results.

We call this metric the **Risk-Informed Finish Threshold (RIFT)**.

Where we begin:

- With the results of an schedule risk and uncertainty analysis
- With a date with a desired probability at the total project level as our finish date
- With a gap in time between the deterministic finish and this probabilistic finish

Where we go:

- Calculating the RIFT date to know when a task/milestone begins to adversely affect the probabilistic project finish date
- Managing tasks/milestones within our schedule to best protect our project from slipping past the selected probabilistic project finish
- Embracing a subtle paradigm shift from attempting schedule allocation to analyzing task tolerance within the schedule

WHAT THIS STORY IS ABOUT

This story is about a new concept to help manage your schedule. This story describes a new metric that opens the door for innovative analysis and practical management.

An uncertainty analysis provides a probabilistic finish date for the entire project. These projects are often scheduled to finish on a date at a selected probability on the project's total finish date s-curve. This is where the schedule analysis hands off the results to the managers, giving an overall probabilistic finish date and some gap, a "probabilistic contingency" of some duration amount. There exists little to no objective metrics that help managing the project on a task-by-task basis in a way that ensures this probabilistic finish is on track or in danger of falling behind. The RIFT dates provide the answer.

This story is about what to do with the uncertain results of a schedule analysis; it is about how to manage your schedule armed with meaningful dates, with RIFT dates.

WHAT THIS STORY IS NOT ABOUT

Before we dive into the main body of the story, we pause to preemptively remove misconceptions often caused by the cost estimating paradigm. Cost estimates and schedules have fundamental differences, the most prominent being that everything in a cost estimate adds up to the total; this is absolutely not the case in schedules.

In cost risk and uncertainty analysis, several well-known and well used implementations exist to allocate the difference between the point estimate and the probabilistic estimate (risk dollars) at the total level down through the work breakdown structure (WBS). This algorithm is referred to as “cost risk allocation”

Using terminology and concepts coined by decades of cost estimating, it is easy to get stuck thinking schedules are analogous to cost estimates. While the application of uncertainty might be similar, what to do with the uncertain schedule results can be quite different.

This story is not about “risk allocation”.

JUMPING TO THE END

The RIFT date provides a timeframe for an individual task or milestone at which the probabilistic project finish date will be maintained. A RIFT analysis provides insight into how individual activities affect the probabilistic finish of the entire project.

The conclusion of our story is just the beginning of the RIFT concept. With diligence, interest, inspiration and a bit of luck, RIFT analysis will become a fundamental aspect of the process of analyzing and managing schedules.

PROLOGUE

The question that began the exploration:
“How do we manage a schedule once we select the project’s probabilistic finish date?”

The idea that begins our story:
Conditional probability is a way to calculate the outcome of two related random variables once information about one of them is given.

Equations for the conditional distribution of two correlated normal distributions (two lognormal or a lognormal-normal combination for that matter) are simply statistical properties. These equations and derivations can be found in the appendix.

We start by defining two new distributions that act as our key players and end with a simple concept (along with tangible metrics and exciting questions) that may inspire a new generation of schedule analysis.

MEET THE KEY PLAYERS

THE SCHEDULE

A schedule is constructed for the purposes of this demonstration. This example schedule is illustrated in Figure 1 and is intended to be simple yet non-trivial. More aspects of the specification and inputs to the schedule will be revealed as we proceed, but for the introduction, just the most basic features are needed.

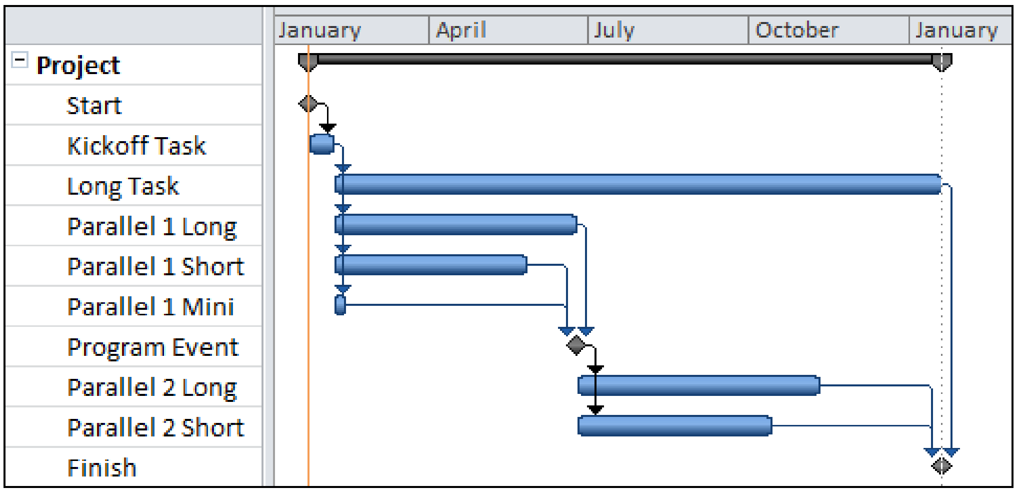

Figure 1: The Schedule

Specific dates and durations are unnecessary for now. The only other pertinent information is that there are probability distributions on the task durations enabling us to run a simulation and analyze the uncertainty results.

The analysis and results presented in this paper apply even if a different approach to apply uncertainty was used, for example, the risk factor approach or inserting risk events.

THE QUESTION, THEN AND NOW

Once we have modeled uncertainty and chosen a probabilistic project finish date (either chosen a date itself or chosen a probability from the finish date s-curve that then defines the date) we want to know when individual tasks or milestones slip to a point in the schedule that they endanger pushing out this selected finish, or “finish-no-later-than” date.

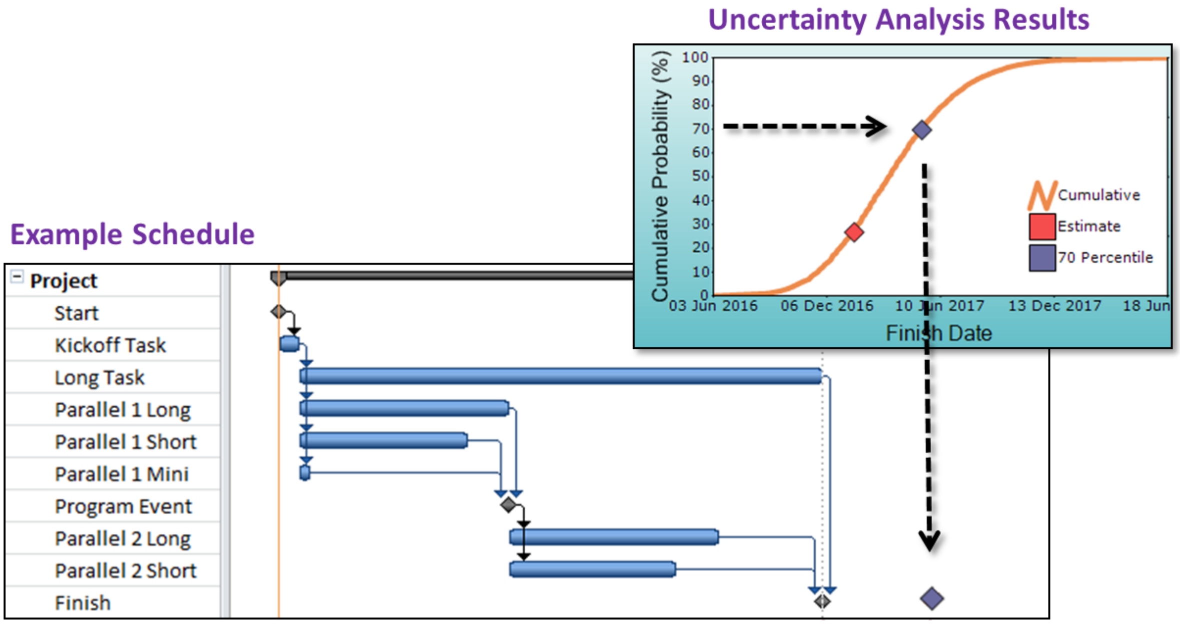

Figure 2: Uncertainty Results on the Schedule

The original question:
How do we spread this gap, this probabilistic contingency, between the plan and the selected project finish throughout our project in order to stay on schedule? (i.e., trying to draw an analogy to the cost risk and uncertainty model approach)

The new question:
At what point will each task's finish date cause the project to extend beyond the selected probabilistic finish date?

There are actually two variations on the new question that RIFT can answer:

1. Out of the entire schedule, which tasks are most likely to endanger my probabilistic finish?

2. Considering the tasks I am managing, when will they most likely endanger my schedule?

Using the schedule model introduced above, we first consider the second question in order to focus the explanation on a singular date and elaborate on the implications. Once the groundwork is laid, the first question can be addressed.

The conflict:
Here are just a few of the commonly reported metrics produced from a schedule uncertainty analysis:

- **Total Slack**: The amount of time before a task falls onto the critical path
- **Correlation**: Strength of the linear (or monotonic) relationship between two random variables
- **Criticality**: The average time on critical path over all possible outcomes

- **Duration Cruciality**: Combines criticality and correlation for a slightly more robust singular metric useful for overall ranking drivers
- **Sensitivity Index**: Combines the relative uncertainty with criticality for another useful metric for ranking overall drivers

All of these are useful but have the common downside that they are calculated using all the iterations from the simulation. These metrics inherently consider all possible outcomes, thus preventing a direct measurement of how parts of the schedule relate to a singular finish date. Even when conditioning on criticality (only calculating metrics with those iterations that fall on the critical path), there is still no direct measurement against the any singular finish date, just a measurement of the behavior of critical paths to all possible finish dates.

THE NEW DISTRIBUTIONS

Distribution B-LF ("B" for short) and A-LF ("A" for short) are distributions describing the entire duration before (B) and after (A) a task. By these definitions, the sum of B-LF and A-LF on any given iteration is the duration of the entire project, simply Distribution D (for project duration). Figure 3 outlines the general idea of these new distributions.

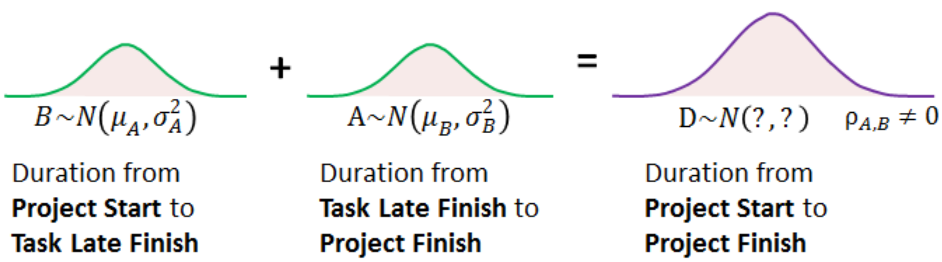

Figure 3: New Distributions

- **Distribution B-LF**– The entire duration from the start of the project to the late finish date[1] of any selected task or milestone on any given iteration (Duration "Before")
- **Distribution A-LF** – The duration from the task late finish date to the project finish date on any given iteration (Duration "After")
- **Distribution D** – The uncertainty distribution describing the duration of the entire project

To depict these distributions, seen in Figure 4 we have selected the Program Event in our schedule to define B-LF and A-LF in order to calculate the RIFT date for it. These definitions and analysis that follows is applicable to any selected task or milestone.

[1] The late finish date is the finish date plus total slack (if any). In other words, the latest date a task can finish without delaying the entire project.

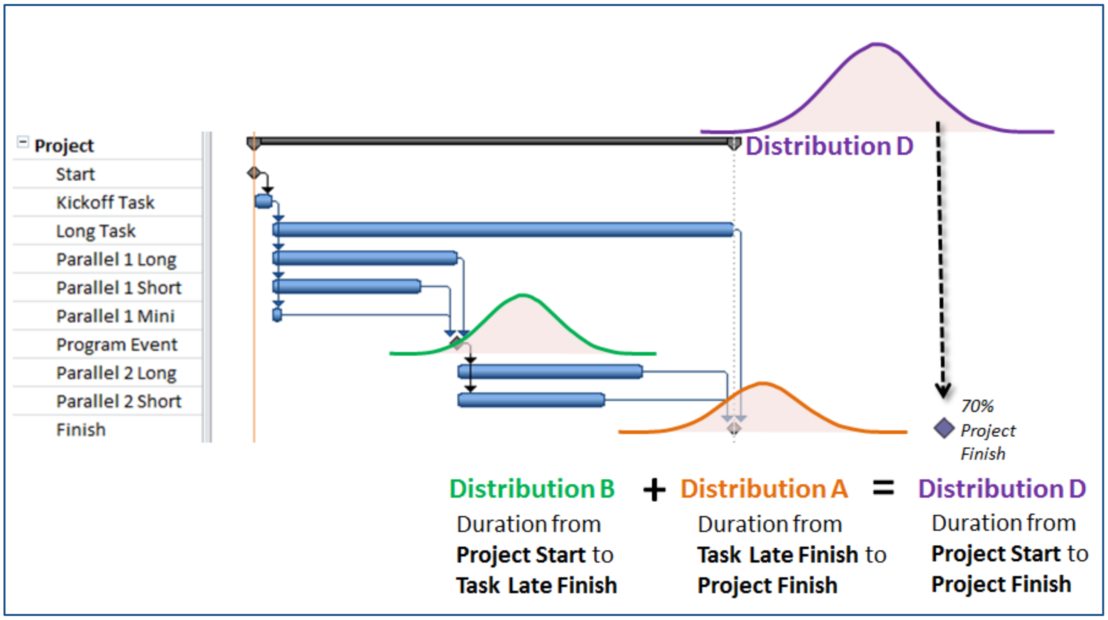

Figure 4: New Distributions in the Schedule

CONDITIONAL PROBABILITY IN A NEW SUIT

Deriving a formula for the conditional probability is a statistical matter-of-fact. What is new in this paper is the application of these conditional probability formulas to solve a widespread issue in dealing with uncertainty results of schedule analysis. There is also a new, innovative use of the Taylor series expansion to find a closed-form approximation to the solution of the more complicated conditional formulas, found in the appendix. Even more than a solution, this original application opens the doors for an entirely new dimension of analysis.

Taking into account the entirety of the uncertainty before and after the selected task/milestone, we take advantage of a conditional probability formula to find the RIFT date.

RESOLUTION IN SIGHT

We begin by describing the theoretical method of the solution, then show how it looks in practice, and finish with a single picture which we hope will be worth the thousand words leading up to it.

LAYING THE GROUNDWORK WITH GRAPHS

Imagine a scatter plot of task finish date and project finish date pairs, such as Figure 5 below. Clearly the correlation (Pearson or rank) strength is 1: there is a perfect one-to-one relationship between the task finish date and the project finish date.

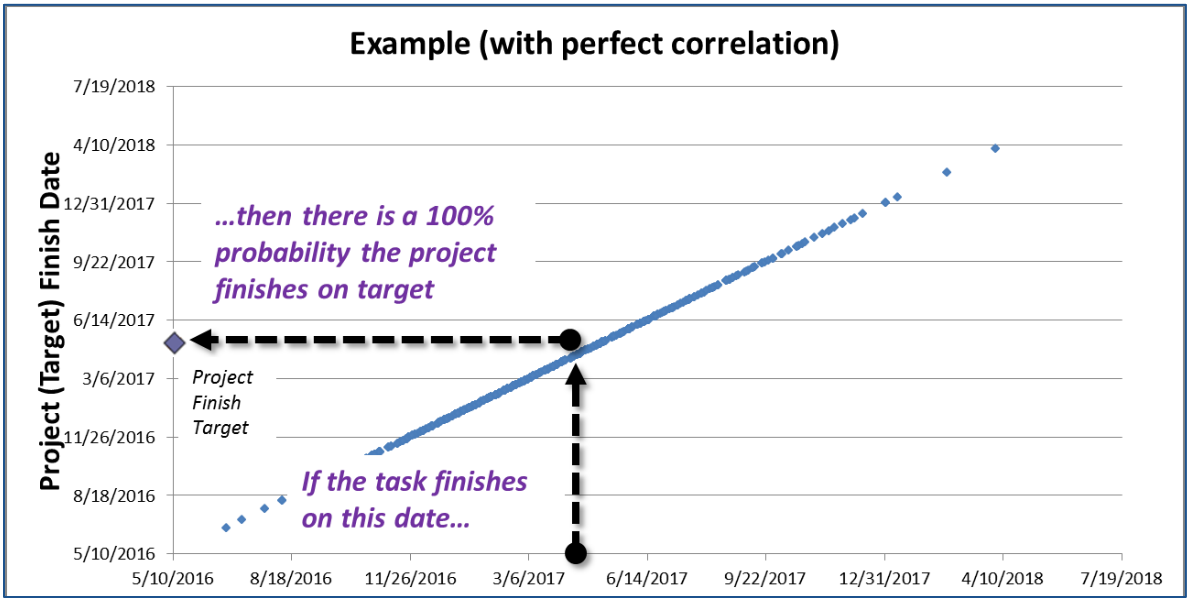

Figure 5: Perfectly Correlated Scatter

A scatter plot like the one above tells us that when we are given a date that the task finishes on, we know precisely and unambiguously when the project finishes.

Let's now consider a slightly less trivial example. The scatter in Figure 6 below comes from our example schedule, plotting the finish dates of the "Long Task" against the project finish date. This is quite an interesting example of how schedule scatter plots can take on interesting shapes, revealing insightful information about the structure of the schedule's uncertainty.

If "Long Task" finishes on the early side, we cannot then say with absolute certainty when the project finishes. Notably though, if "Long Task" finishes on the later side, it is always on the critical path, as seen when the scatter tightens into a straight line.

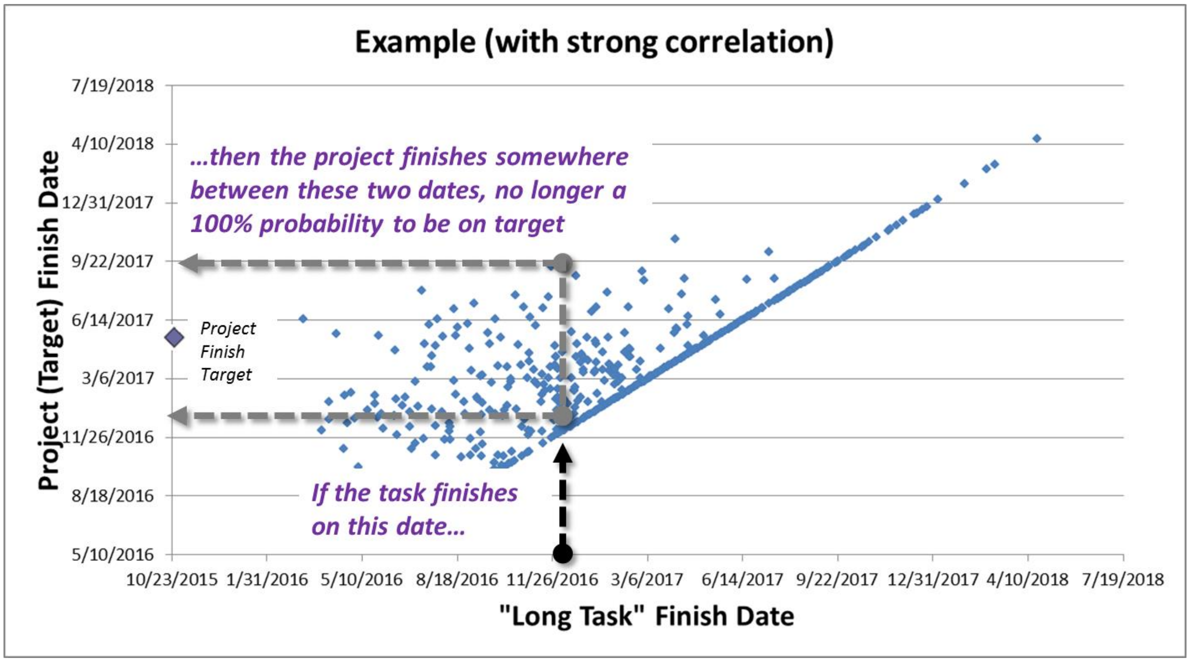

Figure 6: Strongly Correlated Scatter

What we would like to know is *what is the probability the project finishes on the selected finish date if "Long Task" finishes on the date marked with the dotted vertical black arrow*?

CONDITIONAL PROBABILITY TO THE RESCUE

To calculate the probability the project finishes on the selected date, we need to find where that date lies relative to all possible project finish dates given the "Long Task" finish date (the dotted vertical black arrow).

You will notice that the arrow on the above scatter is on a single date for the "Long Task". The vertical grey dotted line goes through what look like project finish date points that land on that exact "Long Task" finish date, but in reality, there might be only a couple points or none at all that land exactly on that line.

Zooming into Figure 6 we can see in Figure 7 where we might run into an issue in calculating the probability of finishing on the selected date given the "Long Task" finish date.

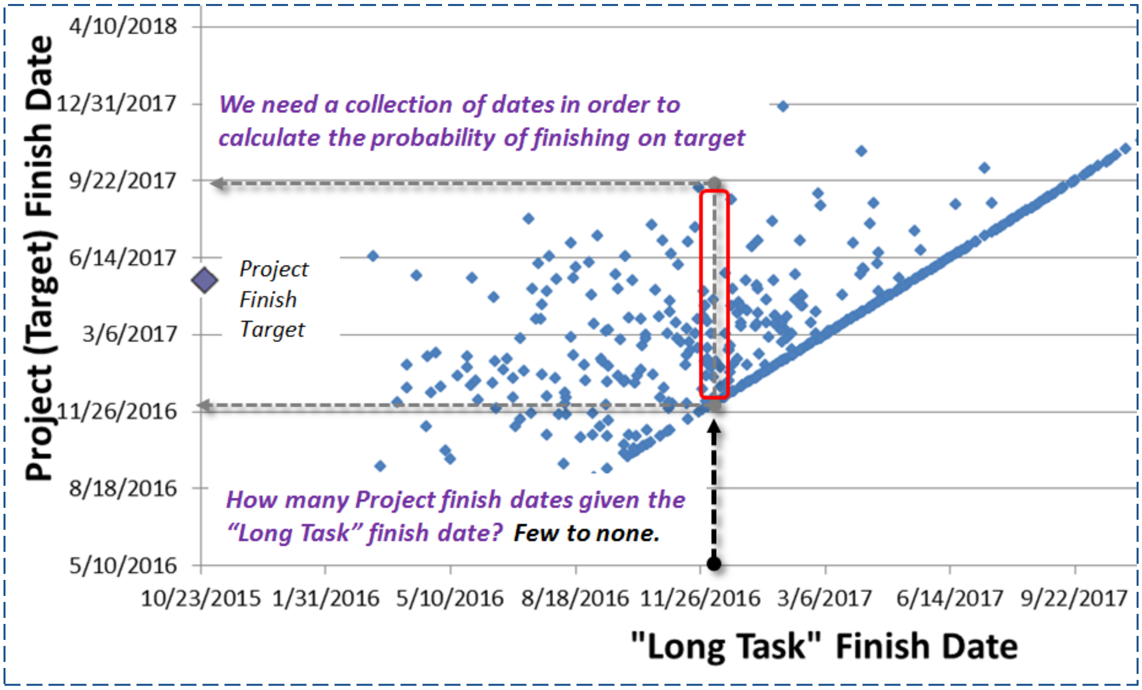

Figure 7: Strongly Correlated Scatter - Zoomed

The problem with finite data:
With few or no project finish dates landing *exactly* on any given task date, we cannot calculate *any* probabilities. This is why you see histograms generated from simulation data rather than continuous probability density functions (PDFs). For similar reasons, you never see the probability of finishing exactly on a date or at a specific cost, it is always "at or before" using cumulative probability (CDFs).

How we get around this problem:
While we cannot pick a singular date, we can select a range[2] of dates nearby our date of interest. When we look at the collection of results, call it a "**RIFT range**", we can calculate (using just this subset of results) the probability of our project finish date in order to ensure we are maintaining the original probability. This allows us to analyze what happens in our schedule around the given time frame.

Conditional probability enters:
This finally brings us back to the conditional probability of the selected finish date. The RIFT analysis finds the date for the selected task (in our scatter plot example this was the "Long Task" finish date) that if finished on, likely results in the project finishing on selected date. Say the original probability of finishing at or before was 70%, then in order to validate the RIFT date is accurate, we want to see that on the RIFT date the project finish date's probability is still 70%.

[2] We will ignore the issue of bin sizing for now; we have bigger fish to fry, but it is not forgotten.

Since we cannot just use the exact RIFT date, we use a range around it. Since the RIFT date is considered a “threshold”, we do not take an equal range on both sides but rather a little bit more before than after.

Looking at a zoomed in scatter plot we can see this range depicted in Figure 8.

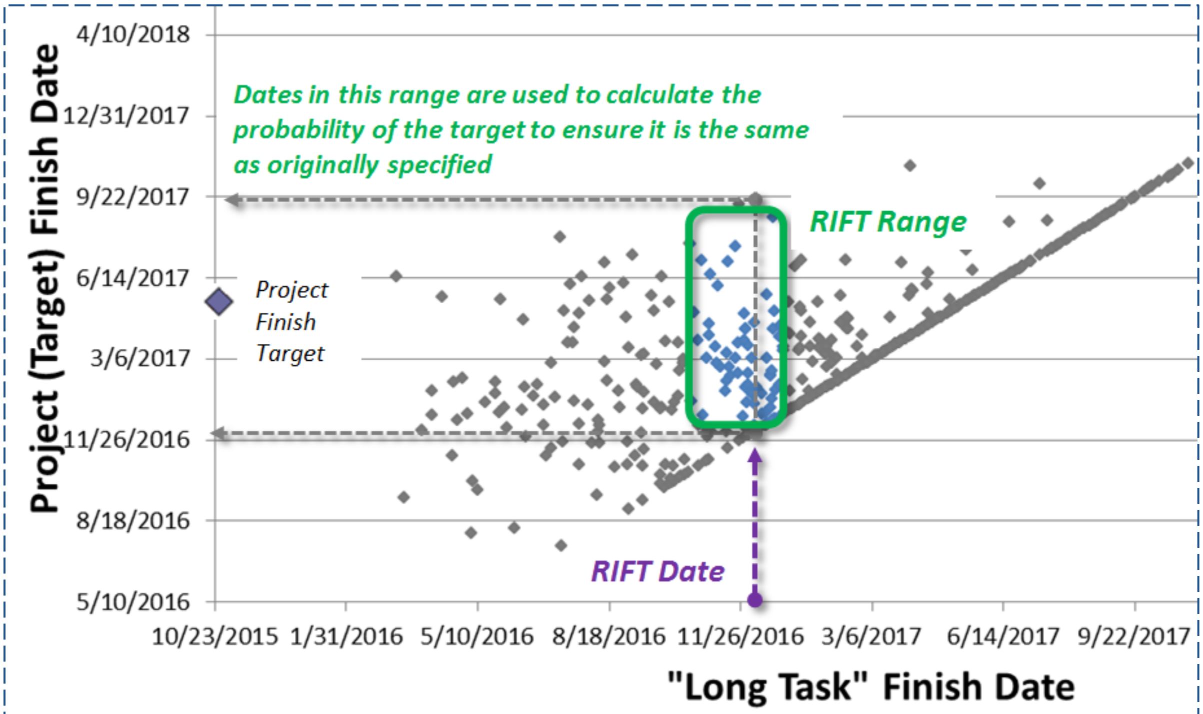

Figure 8: Strongly Correlated Scatter – RIFT Range

The conditional probability of the selected project finish date is the probability of the original selected project finish date calculated using only those dates that lie in the “RIFT range”[3].

CONDITIONAL PROBABILITY SERVING A FEW PURPOSES

The analysis described above illuminates a potential brute force method for determining the RIFT date if one chose to “search” the scatter range for the desired conditional probability of a project finish. However, we seek a way to work this process in reverse: we want to use a closed-form, one-step method for solving that problem.

However, the brute force analysis described in this section serves to illustrate the practical underpinnings and more importantly serves as a way to independently verify the results provided by the closed-form solution.

[3] The exact width of the window is still a point of analysis, currently a fixed percentage of the entire range of dates. It is possible to make the width a function of other metrics such as correlation, criticality, or a conditional metric based on concepts not yet discovered.

Additionally, we can calculate the brute force metric over time in order to graph the behavior of the relationship between the task/milestone finish date and the project finish date, opening up the door to a world of new analysis techniques and insights (more on that in the following chapters).

More pictures, less words, purpose number one:
The purpose is described in Figure 9 below.

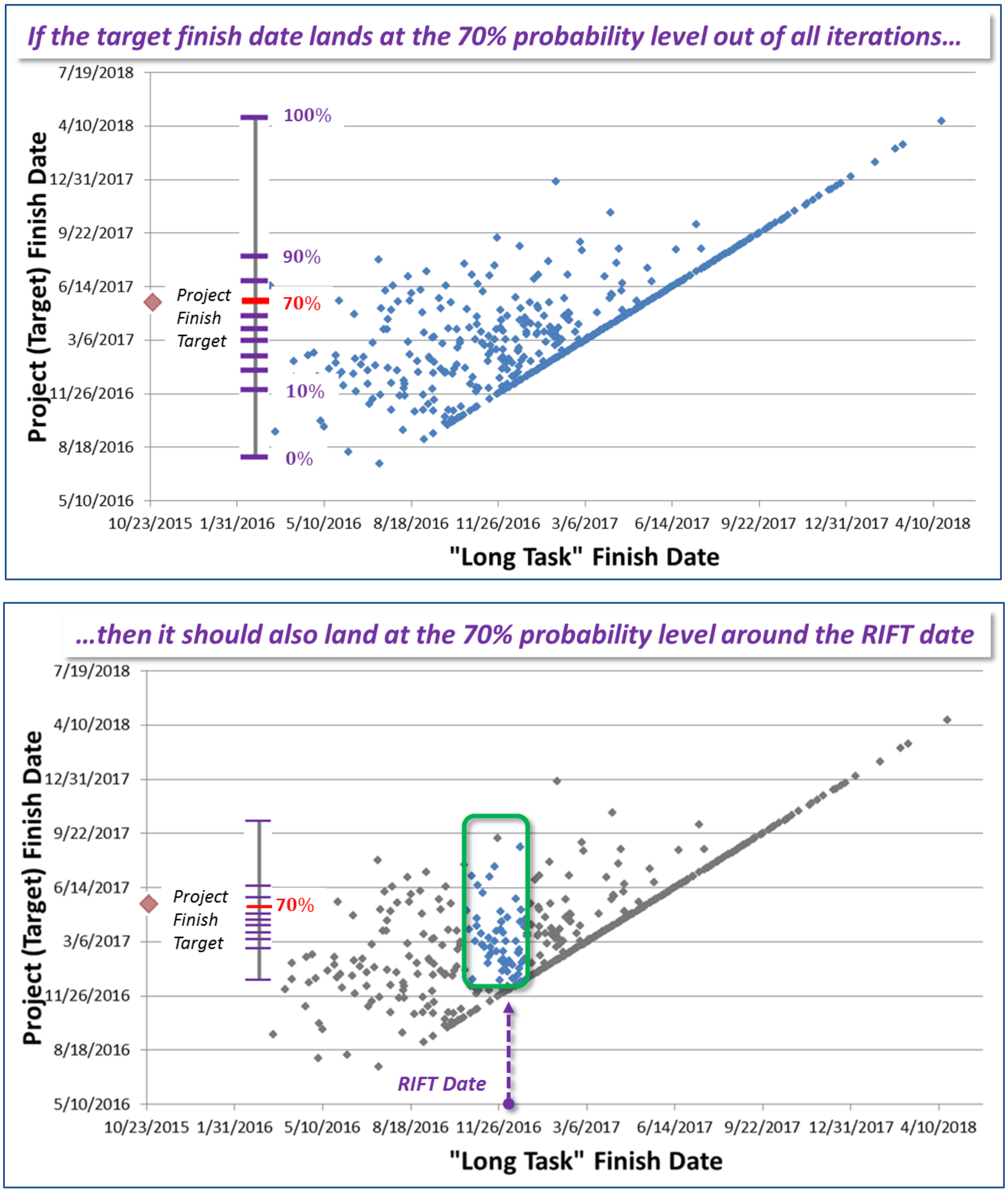

Figure 9: Comparing Results

If the difference between the original probability and the conditional probability of the selected project finish date is small, then that is a great indicator for the RIFT date being accurate.

More pictures, less words, purpose number two:
What else can be done with our RIFT range and the concept of binning our dates? Let's build one of the main outputs of the RIFT analysis, the **RIFT Chart**, seen in Figure 10.

ENTER THE RIFT CHART

- **Green outline**: Defines the range used to calculate the conditional probability
- **Horizontal line:** Marking the selected probabilistic project finish date
- **Vertical line:** Marking the RIFT date
- **Red label**: Conditional probability of the selected probabilistic project finish date
    - Labeled 70% as expected. Recall the premise was that the project's selected finish date would maintain the original probability of 70% when the task finishes around the RIFT date.
- **Red dot**: The *x-axis* value is the mid-point of the range (green outline). The *y-axis* value comes from the probability of the red label used to calculate the date at that probability (out of all iterations).
    - The red dot is near the crosshairs as expected since the conditional result (subset of iterations) and the overall result (entirety of iterations) should match near the RIFT date.

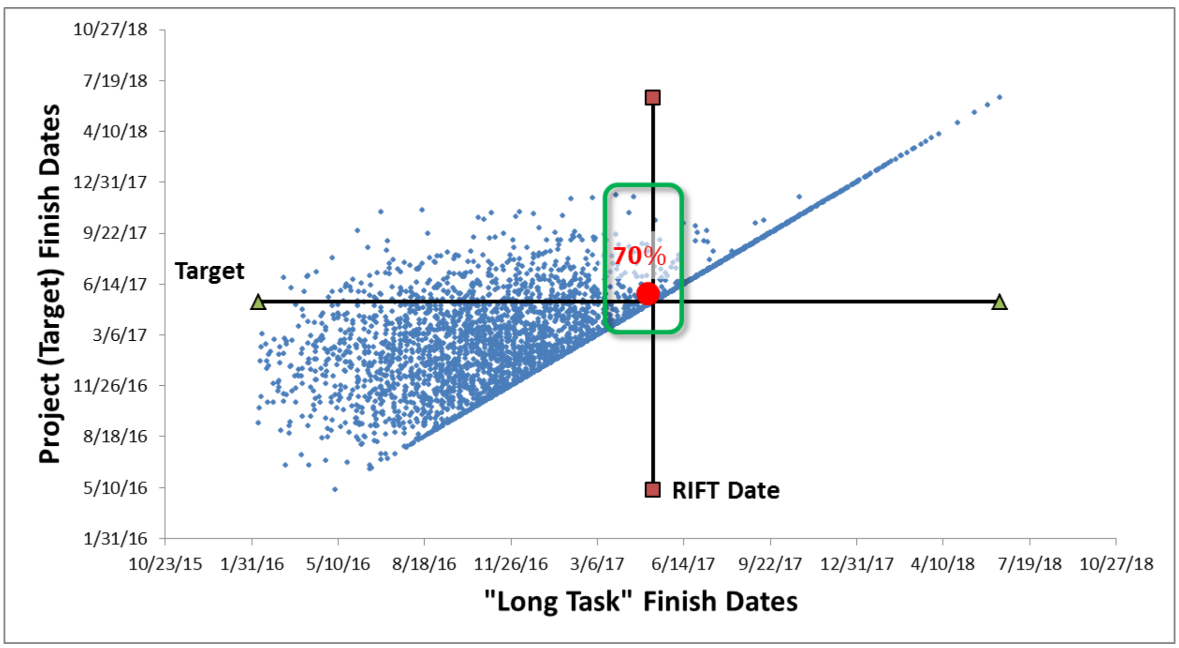

Figure 10: Introducing the RIFT Chart

Continuing the same idea, it is possible to continue binning the scatter and calculating the conditional probability, seen below in Figure 11.

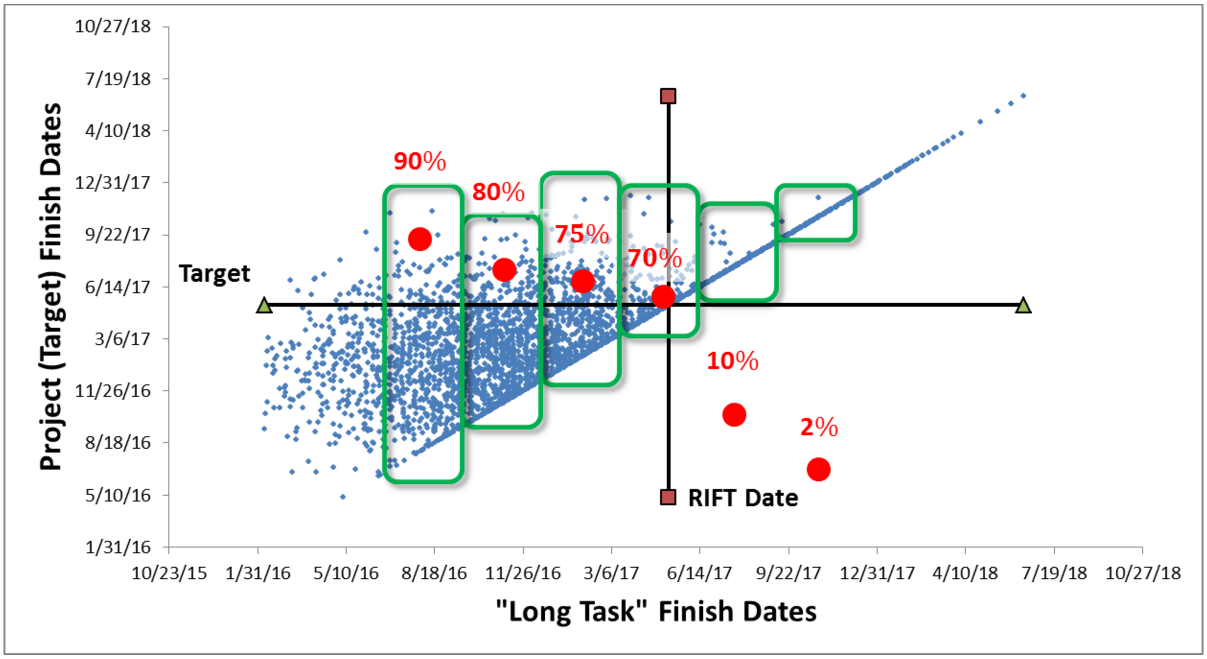

Figure 11: RIFT Chart – More Bins

Since the red dots are simply indicating a trend, instead of dots we can use a line to represent the values of the conditional probabilities over time. Similarly we do not need to outline the ranges used to calculate these probabilities, leaving us with the RIFT Chart as depicted in Figure 12, cropped and resized to enable clearer reading of each label.

The completed RIFT Chart:

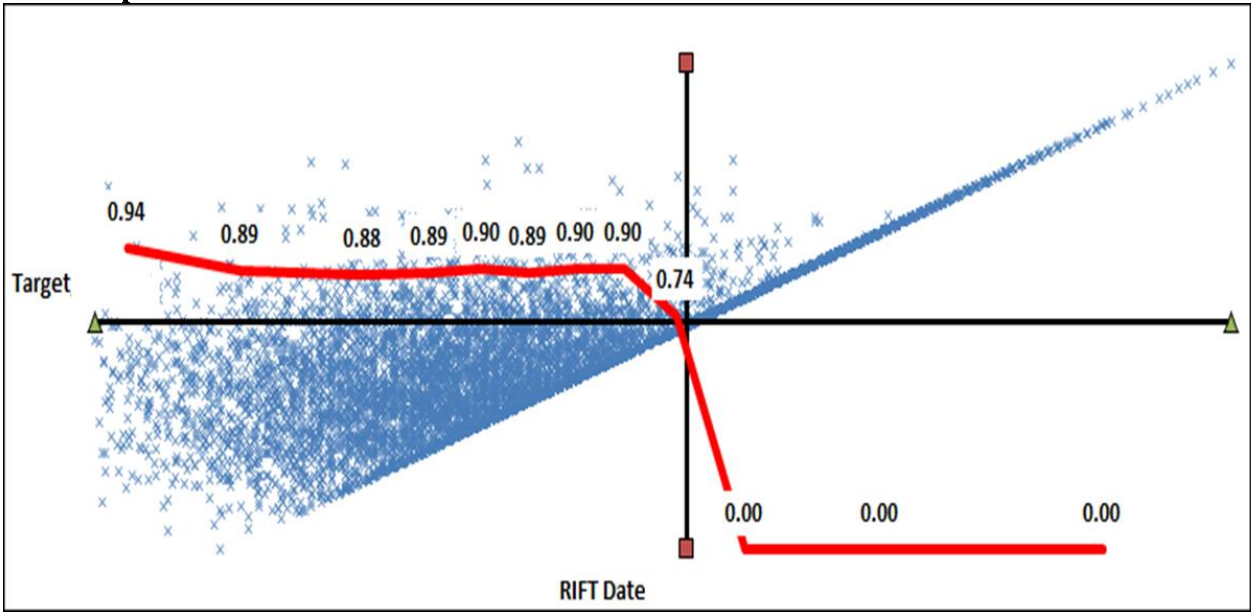

Figure 12: RIFT Chart – Completed[4]

WHAT MIGHT BE IN THE SEQUEL?

By plotting the trend of the conditional probability of the selected project finish date, we get not only a visual of our RIFT date's accuracy, but a tangible indicator of what happens if our selected task/milestone slips beyond the RIFT date, among other pieces of information.

Analyzing and plotting the conditional probabilities along with the RIFT date opens up an entirely new set of answerable questions:

- What is the probability of finishing at or before the RIFT date?
- How does the critical path relate to the trend of the red line?
- How dramatically does missing our RIFT date affect the schedule?
- What much time does it save to finish earlier than the RIFT date?
- Could there be something wrong in the schedule logic causing my project to behave this way?

This is just the beginning of the set of questions possible by calculating and visualizing the conditional relationship between the task/milestone and the selected probabilistic finish date.

[4] This chart is generated directly from the RIFT Report (in ACEIT's JACS tool) so it varies slightly from the charts prior that were slightly modified for demonstration purposes.

THE FULL RESOLUTION

This chapter discusses the closed-form solution based on statistical theory that allows us to avoid the computationally prohibitive brute force method alluded to in the previous chapter. We jump straight into the equation and then back up to explain where it came from.

THE RIFT EQUATION, SIMPLY PUT

Without further ado, the RIFT date is defined as follows:

$$\boldsymbol{b} = \frac{\boldsymbol{d} - \boldsymbol{\mu_a} + \boldsymbol{\rho}\frac{\boldsymbol{\sigma_a}}{\boldsymbol{\sigma_b}}\boldsymbol{\mu_b} - \boldsymbol{Z(c)\sigma_a}\sqrt{\boldsymbol{1-\rho^2}}}{\boldsymbol{1+\rho}\frac{\boldsymbol{\sigma_a}}{\boldsymbol{\sigma_b}}}$$

- **b** – duration from project start to the selected task's RIFT date
- **d** – duration from project start to selected project finish date
- $\boldsymbol{\mu_a}$ – mean of distribution A-LF
- $\boldsymbol{\mu_b}$ – mean of distribution B-LF
- $\boldsymbol{\sigma_a}$ – standard deviation of distribution A-LF
- $\boldsymbol{\sigma}_b$ – standard deviation of distribution B-LF
- $\boldsymbol{\rho}$ – correlation between distribution A-LF and B-LF
- $\boldsymbol{Z(c)}$ – quantile function of the standard normal distribution
- $\boldsymbol{c}$ **–** the cumulative probability at the selected project finish date

THE RIFT EQUATION, IN DEPTH

Consider any two correlated distributions that we are drawing values from at random. By the nature of their correlation, if we pick a value at random from one distribution, we then have some idea about what we will pick from the other distribution.

The theory behind the RIFT equation

Let's take this one step further. This time, we add together the two values drawn, keeping track of their sum. We now have correlation between each distribution and the sum (which is a distribution itself). Picking a value from one distribution tells us something about their sum.

> **Generally, our solution is:**
> Calculated a date, from all possible finish dates of our selected task, that when added to the uncertainty of the range of possible durations for the remainder of the project, most likely results in finishing on our project's selected finish date.

Statistically, our solution is:
Calculated a value b, from Distribution B-LF, that when added to Distribution A-LF, most likely equals our selected finish date d, from Distribution D.

The equation behind the RIFT theory
Now that we can specifically describe our solution and the tools we have, it is time to bring in conditional probability.

If X and Y are correlated, normal random variables, then this statement of conditional probability describes their relationship:

$$Y \sim N(\mu_Y, \sigma_Y^2)$$
$$X \sim N(\mu_X, \sigma_X^2)$$
$$\rho_{X,Y} \neq 0$$
$$(Y|X = x) \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X),\ \sigma_Y^2(1 - \rho^2)\right)$$

It reads "the distribution of Y, given a value x from distribution X, is normally distributed with a mean of $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ and variance of $\sigma_Y^2(1 - \rho^2)$ "

To simplify the mean and variance that defines the conditional normal distribution:

Let $U_{X,Y} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
Let $V_Y = \sigma_Y^2(1 - \rho^2)$

Therefore:

$$(Y|X = x) \sim N(U_{X,Y}, V_Y)$$

Using this idea, we can think of our goal as a statement of conditional probability, that is, we can replace X and Y with the variables from our story to be able to describe our scenario even more succinctly.

Recall our "statistically speaking" statement:
Calculate a value b, from Distribution B-LF, that when added to Distribution A-LF, most likely equals our selected finish date d, from Distribution D.

Literally translates into our conditional statement:

$$(A + \boldsymbol{b}|B = b) \sim N(U_{A,B} + \boldsymbol{b},\ V_A)$$

<u>BRINGING OUR KEY PLAYERS BACK INTO THE PICTURE</u>

The conditional distribution $(A + \boldsymbol{b})$ describes the total duration of the project, since **b** is a duration from project start to the late finish of a task we have chosen, and A is always the distribution of the remaining duration to the project finish.

But there's a catch! We are not given **b**, we want to know **b** because that defines the task finish leading to the selected project finish.

What we do know is the selected project finish date, which provides us the value for **d**, therefore we can use the inverse of the cumulative normal distribution function to solve for **b**.

Leaving the explicit derivation and algebra to the appendix we can rewrite the conditional probability taking the inverse as:

$$d = N^{-1}(c) = (U_{A,B} + b) + V_A * Z(c)$$

From here, we plug in the variables for V_A , $U_{A,B}$ and perform some algebra to arrive at the duration to the RIFT date:

$$\boldsymbol{b = \frac{d - \mu_a + \rho\frac{\sigma_a}{\sigma_b}\mu_b - Z(c)\sigma_a\sqrt{1-\rho^2}}{1 + \rho\frac{\sigma_a}{\sigma_b}}}$$

For any selected task[5]:

RIFT Date = Project Start Date + ***b***

Figure 13 depicts the solution.

[5] The RIFT date is calculated for an individual task or milestone, meaning the distributions B and A are recalculated for each and every selection. If we wanted to be extremely formal in our notation, we might add indices to A and B, to denote the selected task/milestone, but here we have omitted these for simplicity.

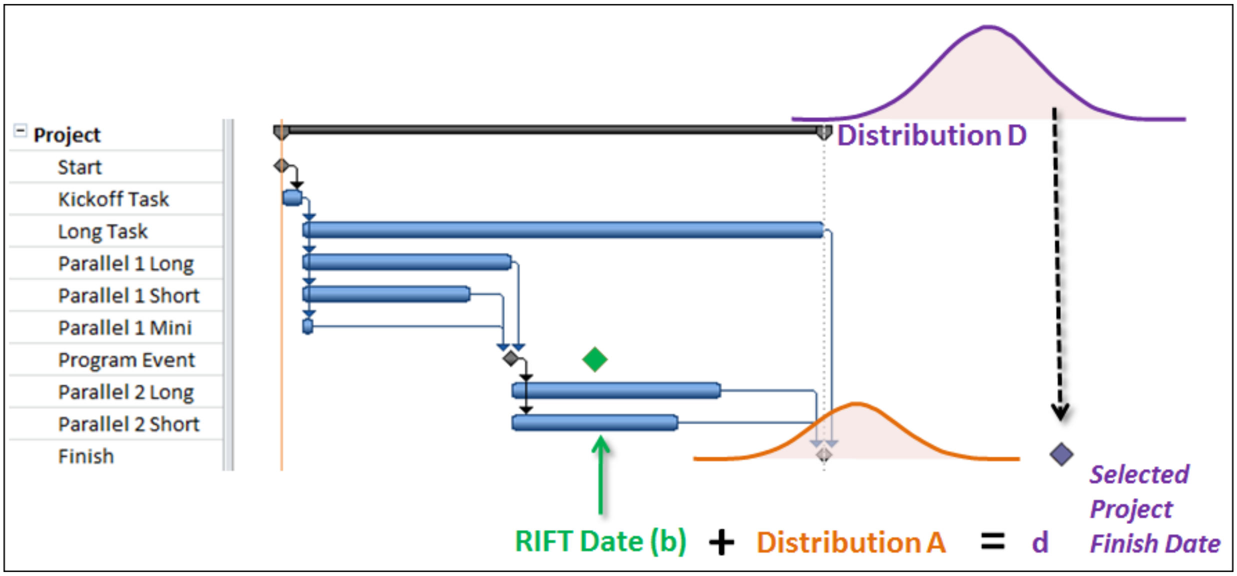

Figure 13: The Schedule and RIFT Date

Figure 13 does not show the distribution $(A + \boldsymbol{b})$ although it would look like a narrower version of distribution D, with the value at the selected probability $\boldsymbol{c}$ (70% in the schedule example depicted throughout this paper) lining up with each other. This concept is informally depicted in Figure 14 the specific formulation of the solution in the bottom half of the image.

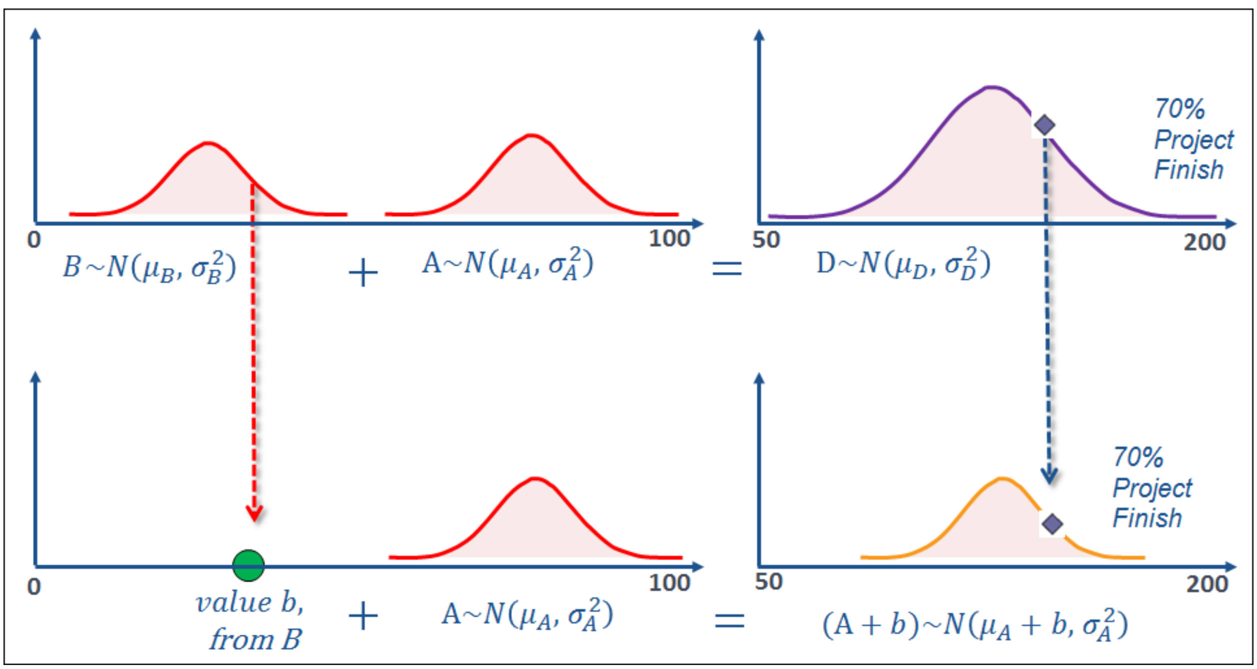

Figure 14: RIFT Date Solution- Informal Representation

CONCLUSION – THE RIFT ANALYSIS

With the variety of shapes and sizes schedules come in, they tend to find ways to add caveats to nearly any metrics attempting to describe them. The following sections briefly discuss general trends of how RIFT dates behave based on the structure of the schedule.

Correlation and criticality are common approaches to measuring how a task or milestone affects the project finish date. Correlation measures the strength of how much two finish dates move together. Criticality measures how often the task is on the critical path, i.e. the location of the task during the simulation. Both these metrics tell an important part of the story, but neither metric should be used on its own to pinpoint or categorize a schedule.

A task can be highly correlated to the project finish date but never actually critical. Conversely, a task might always be on the critical path but weakly correlated to the project finish. Additionally, a task may be highly correlated to an end date but in fact have zero causal influence – it shows up in the correlation report not because it is important, but because it is correlated to one that is. Care must be taken when ranking tasks based on a singular metric.

In similar fashion, the RIFT date for any selected task/milestone must be taken in context with the schedule and relevant metrics. This chapter discusses a few examples of how to perform and interpret a RIFT analysis.

Bringing back the Uncertain Schedule

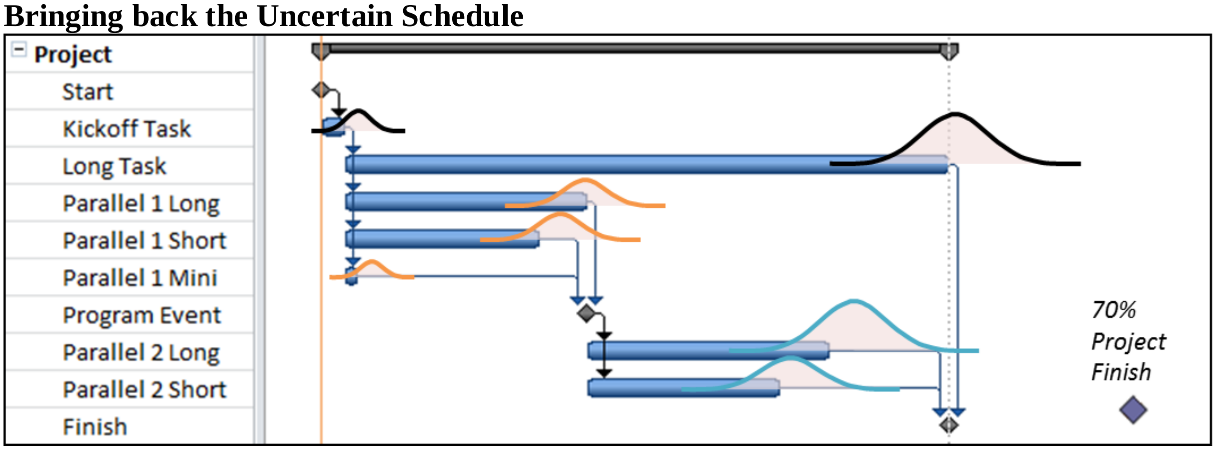

Figure 15: The Uncertain Schedule Returns

SAMPLE OF A BASIC RIFT ANALYSIS

Program Event is critical because of its pivotal location in the schedule, and correlated to the selected project finish since it is a gateway for the majority of the schedule needed to proceed.

Thus, the Program Event is a task for which a RIFT date is vital. Figure 16 shows the probability of achieving the selected project finish date given a task finish date as the red trend line, overlaid on a scatter plot of simulated task finish dates and corresponding project finish dates (i.e. the RIFT chart).

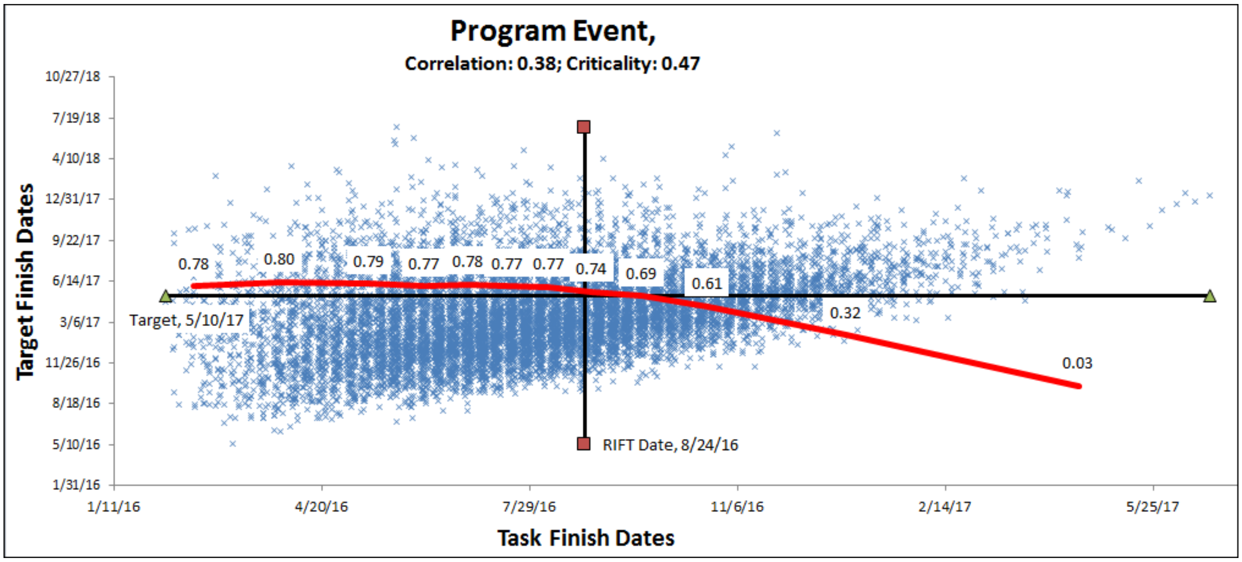

Figure 16: RIFT Chart – Program Event

Note the flat trend of the red line prior to the RIFT date, this indicates that the "Long Task" is the dominate force pushing out the selected project finish because the conditional probability is unchanged for a large portion of the Program Event's possible finish dates. It is not until the Program Event begins to finish later on (presumably after "Long Task") that it begins to have an influence on the selected project finish. Also note that the slope of the red line is not dramatic once passing the RIFT date, indicating only a slow decline in achieving the selected project finish rather than a dramatic delay in the entire schedule.

Kickoff Task is always critical because of its pivotal location in the schedule, highlighted in Figure 17, but due to the small amount of uncertainty and the overwhelming influence of the rest of the schedule ahead of it, its correlation to the project finish date is minimal. Thus a RIFT date might not serve much purpose on this task.

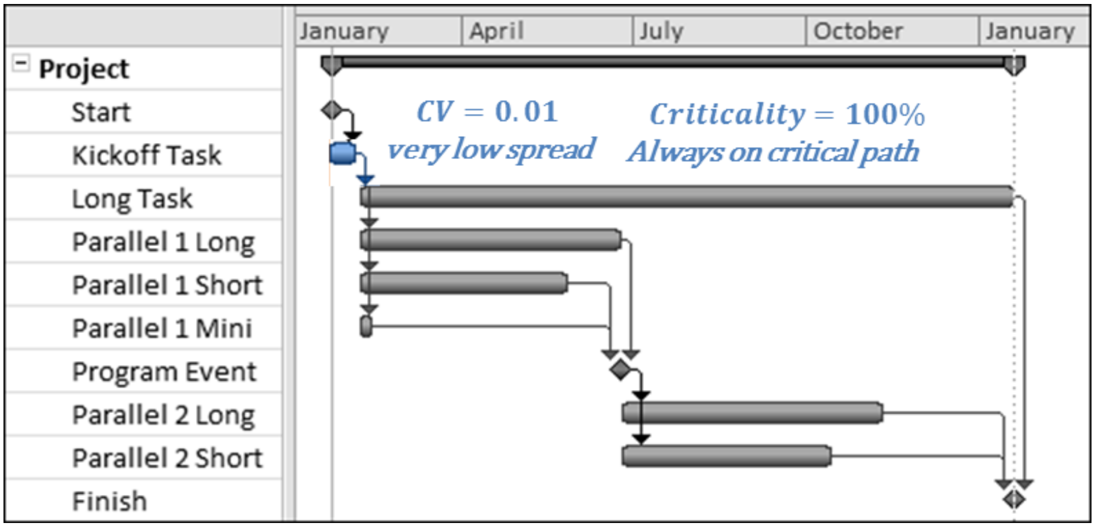

Figure 17: The Schedule – Kickoff Task

Long Task is highly critical because it often extends the entire length of the project, highlighted in Figure 18. The correlation comes from the direct influence on the critical path, not from any applied. Being highly correlated and fairly critical, a RIFT date is vital on this task.

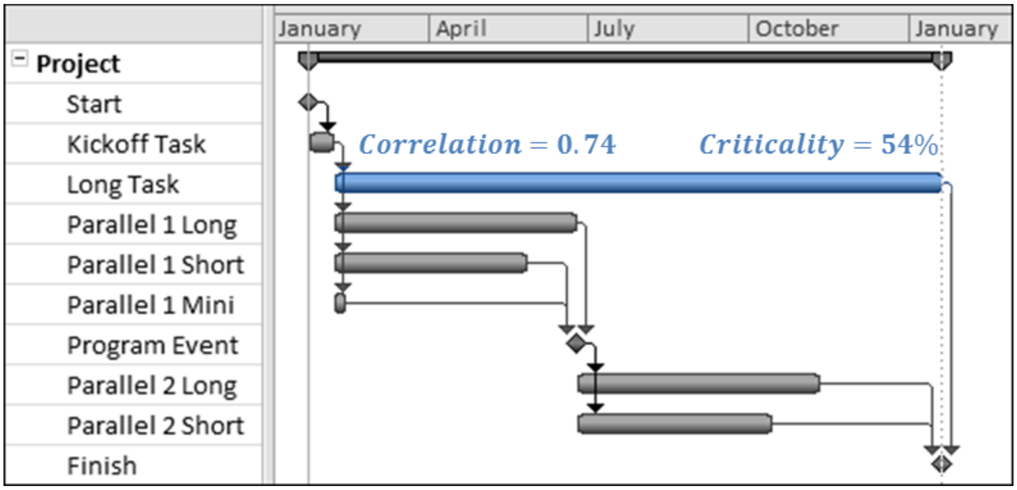

Figure 18: The Schedule – Long Task

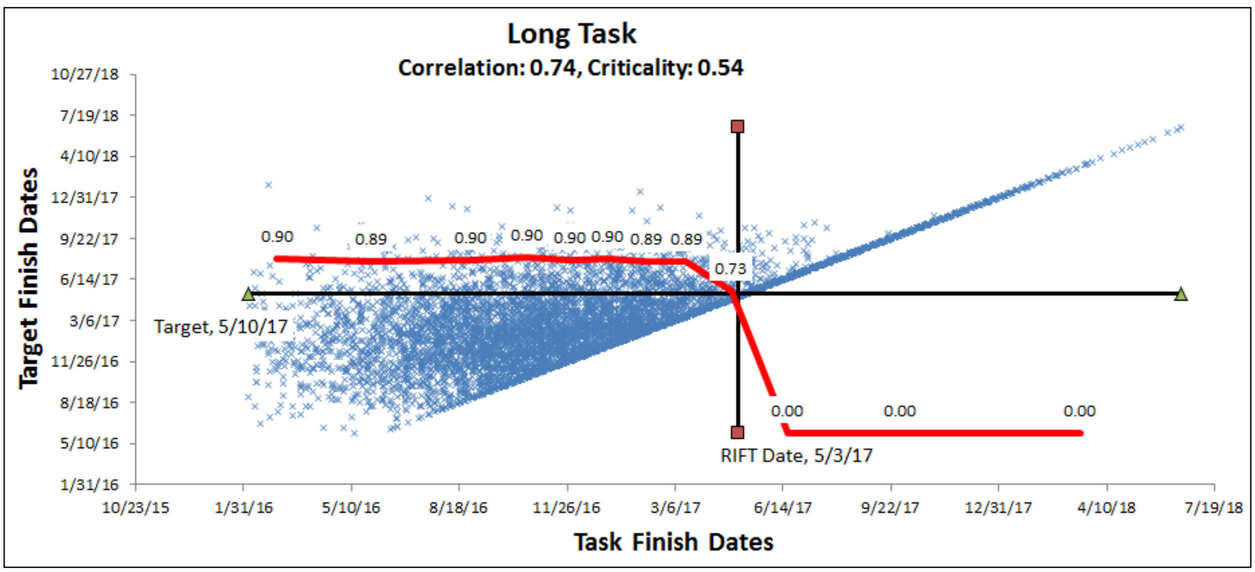

Figure 19: RIFT Chart – Long Task

Note the interesting shape of the scatter. "Long Task" shares the critical path with "Parallel 2 Long" and "Parallel 2 Short" when it is relatively short as seen in the wide scatter before the RIFT date. When the RIFT date is surpassed, the scatter begins to tighten because "Long Task" (aptly named) hogs the critical path and creates a one-to-one relationship with the project finish date.

The most interesting part of this scatter is the relationship between those points that lie on the critical path and the slope of the red line. Further research into creating a red line for "conditional criticality" might provide useful insight on top of the RIFT analysis.

Parallel 1 Mini is never critical, highlighted in Figure 20, because it is always shorter than the other two "Parallel 1" tasks, but due to the large amount of uncertainty and the applied correlation between the three "Parallel 1" tasks, "Parallel 1 Mini" is highly correlated to the project finish date. Thus a RIFT date would help a little, indirectly, but there must be better candidates in the schedule that more directly influence the project's outcome.

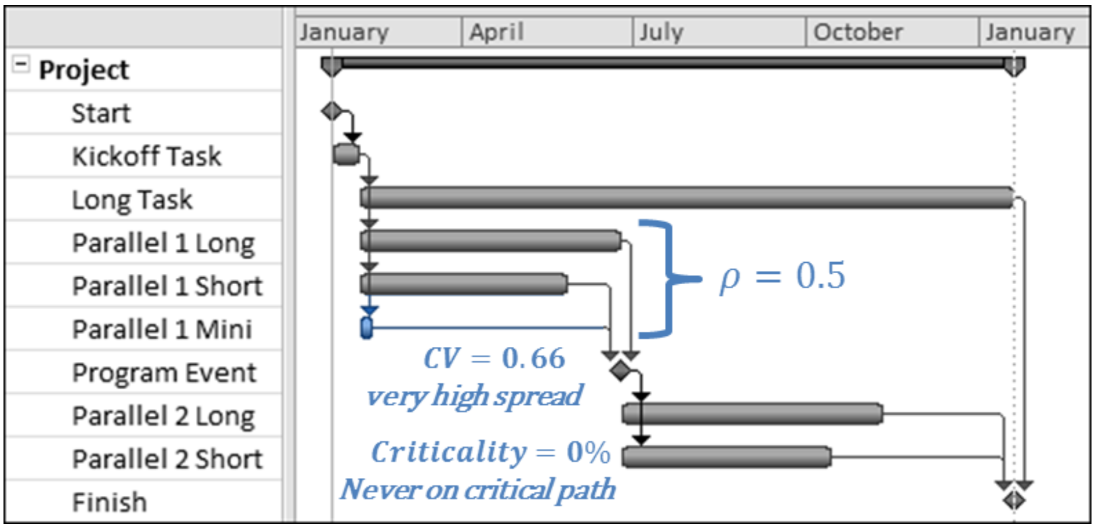

Figure 20: The Schedule – Parallel 1 Mini

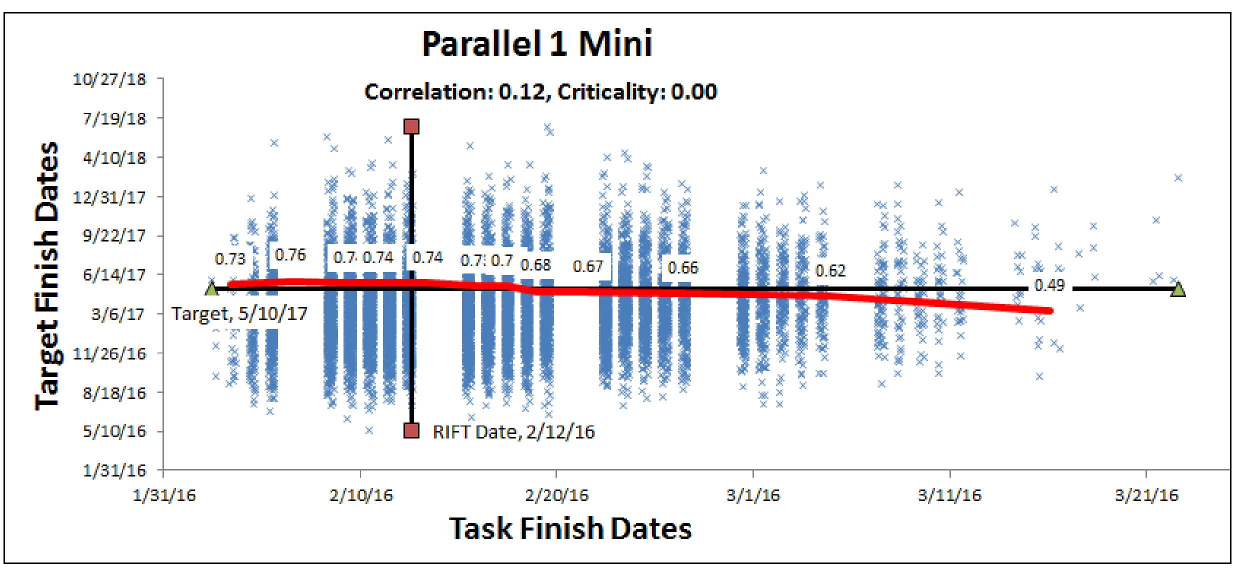

Figure 21: RIFT Chart – Parallel 1 Mini

Note how the RIFT date seems to be almost chosen at random along the red line in Figure 21. The flatness of the red line indicates the minimal influence on the project finish date regardless of the finish of “Parallel 1 Mini”.

CONCLUDING THOUGHTS ON THE EXAMPLE RIFT ANALYSIS

The moral of this chapter:
If a task is not related by any metrics to the project finish date, the RIFT date will be trivial since it is in essence a measurement of that relationship as well. Conversely, a task that is 100% critical or 100% correlated to the project finish will not need a RIFT date because every single day the task slips is one day the project slips, no matter what selected finish you are measuring to protect against.

A candidate for a “good” RIFT date is a task with at least some correlation and some criticality. A preliminary heuristic is that a task with at least 10% correlation and 10% criticality tend to generate meaningful RIFT dates. These dates have some amount of slack and some amount of influence, thus lend themselves to management.

Morals are relative:
Most confusingly, if a task is extremely correlated but never critical, or always critical but weakly correlated, the answer gets a little less obvious on how to treat a RIFT date. There is a grey area in which, unfortunately, human subjectivity must come into play.

A task that is highly critical but has weak or no correlation to the project finish date indicates a pivotal spot in the schedule logic but not much uncertainty, and thus the benefit from calculating the RIFT date would be minimal since the task has nearly a deterministic finish.

A task that is strongly correlated to the project finish but never critical indicates a very uncertain task influencing other tasks on the critical path but there is likely another task more critical that is directly influencing the project finish date. It would be more beneficial and efficient to find the task that is correlated and critical.

SAMPLE OF THE TABULAR OUTPUT

In addition to the graphs, a tabular report can be generated for any or all tasks in the schedule. The noteworthy item not yet discussed, left for the reader to ponder, is the additional information of the probability that the task actually finishes by the RIFT date, as seen in the 4th column below. This adds another consideration to the RIFT analysis. Additionally, the “buffer” in column 5 is the difference (in workdays) between the deterministic finish and RIFT date.

Task Name	Task Finish	Task RIFT	Probability Task Finishes before RIFT	Buffer	Conditional Target Finish Date %	Criticality	Task Correlation to Target
Kickoff Task	2/4/2016	2/9/2016	99%	2	67%	1.00	0.03
Long Task	1/19/2017	5/15/2017	79%	82	51%	0.54	0.74
Parallel 1 Long	6/23/2016	11/17/2016	93%	104	71%	0.32	0.34
Parallel 1 Short	5/26/2016	11/17/2016	99%	124	71%	0.15	0.19
Parallel 1 Mini	2/11/2016	11/17/2016	100%	199	71%	0.00	0.12
Program Event	6/23/2016	11/17/2016	91%	104	71%	0.47	0.37
Parallel 2 Long	11/10/2016	5/15/2017	91%	132	51%	0.32	0.45
Parallel 2 Short	10/13/2016	5/15/2017	96%	152	51%	0.16	0.39

Figure 22: RIFT Tabular Report

How would you rank the table below if searching for the “vital activities” to manage? Based on the probability of achieving the RIFT date? What if there is only a few days between the plan and the RIFT, i.e. the “buffer”? Maybe some combination of criticality and correlation?

It is clear that not every RIFT date is unique. One undeniably observation in this example is the ability for the consistency of the RIFT dates to reveal relationships within the network logic. The “Parallel 1” task network can be considered a singular predecessor of the Program Event. The late finish date of all these tasks drives the Program Event so they plausibly share a RIFT date. In a more complicated schedule than this example, a shared RIFT date might reveal a more complicated network of relationships that otherwise might be difficult to pinpoint.

What about cost? Can conditional analysis provide the means for a cost RIFT analysis? The short answer is: Yes. Equations have already been derived that utilize the same concepts discussed in this paper. Due to the added complexity of the time-dependent nature of cost-loaded schedules, research on this subject is still in the early stages.

RIFT analysis provides a new, unique, and powerful tool to add to the analysts and managers tool-kit. The method and purpose to which this tool is yielded has exciting potential.

A new world of analysis awaits.

THE BEGINNING

THE CONCLUSION TO OUR STORY IS JUST THE BEGINNING

The RIFT date is an output generated from an analysis that has enormous potential beyond a single result. Discussing the ideas and distributions that led us to the RIFT date is difficult because there is no common language yet for this type of analysis. This type of analysis describes uncertainty at a level that is not commonly (or ever) considered. Measuring the schedule using Distribution A-LF and B-LF along with potentially other conditional metrics opens the door for analyzing the schedule in ways not possible before.

UNRESOLVED PLOT – POINTS FOR FURTHER RESEARCH

The following questions are meant to illuminate and inspire the potential of further research. These are not drawbacks of the current capabilities of RIFT analysis; the goals set out in the initial implementation of RIFT have been achieved, these are the incredible new areas that can be addressed using the tools discussed in this paper.

The red line is a unique character with much more potential. How quickly does a task degrade your probability of the selected finish after the RIFT date? Can we create objective metrics for analyzing the trend of the red line? Measuring the slope before and after (creating a ratio) could be an objective metric to compare RIFT dates across different tasks/milestones to create a ranking scheme or a "RIFT Driver" chart.

RIFT dates are calculated individually, so how do we measure the probability of achieving the selected project finish if multiple RIFT dates are met?

Can the RIFT algorithm be modified to account for individual managers "tolerance" to slippage? Do we allow inputs prior to the RIFT report being run?

THE SEQUEL

If the concept of a RIFT analysis becomes a regular aspect of analyzing uncertain schedules it will add an entirely new dimension to our ability to manage schedules. Gathering RIFT dates in a tabular report is useful, but an extensive report answering the questions listed above (and more) would enhance the process of schedule management immensely.

There are difficulties in getting everyone to buy into a consistent process. Getting on the same page is one step, then translating the technical analysis into reports and actionable intelligence is an entirely separate beast. Nothing would be more remarkable than to look back years from now and see that the RIFT date was a fork in the road leading to a smarter way for everyone to conduct schedule uncertainty analysis.

To conclude, let it be teased that the theory is sound and the equations have recently been derived for a *cost* RIFT analysis, paving the way for a joint cost and schedule RIFT analysis.

APPENDIX A – DERIVATION OF STATISTICAL EQUATIONS

This appendix derives the conditional probabilities used to calculate the RIFT date. When those formulas require, a Taylor series expansion is used to solve the equation in a closed-formed single step. When the normal assumption of the schedule duration is not adequate, variations using log-normal and normal distribution combinations for A-LF and B-LF are formulated here.

NORMAL AND BIVARIATE NORMAL DISTRIBUTIONS

Normal Distribution. If a random variable X follows a normal distribution with a mean of μ and a variance of σ^2, i.e., X ~ N(μ, σ^2), then the probability density function (pdf) of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi x}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \quad -\infty < x < \infty \tag{1}$$

Note that if X is distributed as N(0,1), then X is the so-called standard normal distribution.

Cumulative Distribution Function (CDF). Any probability distribution can be expressed in cumulative form. The CDF completely describes the probability distribution of a real-valued random variable, just like the probability density/mass function. For every real number x, the CDF of a random variable X is given by

$$F(x) = P(X \le x) = \text{Probability of obtaining a value less than or equal to } x$$

$$= \begin{cases} \int_{-\infty}^{x} f(t)\,dt & \text{if X is continuous} \\ \sum_{t \le x} p(t) & \text{if X is discrete} \end{cases} \tag{2}$$

where f() represents the pdf of the random variable X if it is continuous while p() is the probability mass function if X is discrete. Therefore, the value of a CDF is bounded between 0 and 1, with 0.5 indicating the median of the population. The probability that X lies in the interval (a, b] is hence F(b) − F(a) if a < b.

Using the above definition, the CDF of a normal distribution X can be expressed as follows:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \phi((x-\mu)/\sigma) \quad \text{for } -\infty < x < \infty$$

where ϕ denotes the CDF of the standard normal distribution with a mean of 0 and variance 1, i.e., N(0,1). Note that N(0,1) is commonly denoted by the letter Z.

Bivariate Normal Distribution. If two random variables X and Y are assumed to follow a joint normal distribution, then the joint pdf of X and Y is given by

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\{\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y})+\frac{(y-\mu_y)^2}{\sigma_y^2}\}\right) \quad (3)$$

where:

X follows a normal distribution with a mean of μ_x and variance σ_x^2 ($X \sim N(\mu_x,\sigma_x^2)$)

Y follows a normal distribution with a mean of μ_y and variance σ_y^2 ($Y \sim N(\mu_y,\sigma_y^2)$)

ρ is the Pearson's correlation coefficient between X and Y

The joint distribution of X and Y is commonly denoted by Bivariate N((μ_x, μ_y), (σ_x^2, σ_y^2), ρ).

Conditional Distribution. The conditional pdf of Y given X = *x* is derived as

$$f(y\,|\,x)=f(x,y)/f(x)=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{[y-\mu_y-\rho\sigma_y/\sigma_x(x-\mu_x)]^2}{2\sigma_y^2(1-\rho^2)}\right) \quad (4)$$

As a result, the conditional mean of the distribution of Y at a given X value (X = *x*) is a linear function of *x*:

$$\begin{aligned}E(Y\,|\,X=x)&=\int_{-\infty}^{\infty} y\,f(y\,|\,X=x)dy=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\\&=(\mu_y-\rho\frac{\sigma_y}{\sigma_x}\mu_x)+\rho(\frac{\sigma_y}{\sigma_x})x=\alpha+\beta x\end{aligned} \quad (5)$$

However, the conditional variance of Y given X = *x* does **not** depend on *x* at all:

$$Var(Y\,|\,X=x)=\int_{-\infty}^{\infty} y^2\ f(y\,|\,X=x)dy-(E(Y\,|\,X=x))^2=\sigma_y^2(1-\rho^2) \quad (6)$$

Therefore, according to Equation 4, the conditional distribution of Y at a given *x* value is listed below (under the normality assumption):

$$Y \mid X = x \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma_y^2(1-\rho^2)\right) \tag{7}$$

Equation 5 is the population regression equation when regressing the dependent variable Y against the independent variable X.

The conditional distribution provides probability of the type $P(Y \leq y \mid X = x_0)$ or $P(y_1 \leq Y \leq y_2 \mid X = x_0)$. This is useful in cost analysis tradeoff studies when assessing the impact of a given schedule on the likelihood that the system cost will not exceed a certain threshold.

Using the notation in schedule analysis (Y = A and X = B), the conditional distribution of A at a given b value (i.e., Equation 7) can be expressed as

$$A \mid B = b \sim N\left(\mu_A + \rho\frac{\sigma_A}{\sigma_B}(b-\mu_B), \sigma_A^2(1-\rho^2)\right) \tag{8}$$

Therefore, the conditional distribution of A plus b (denoted by D) when B is at a given “b” value is then

$$A + b \mid B = b \sim N\left(b + \mu_A + \rho\frac{\sigma_A}{\sigma_B}(b-\mu_B), \sigma_A^2(1-\rho^2)\right) \tag{9}$$

With a chosen probability level c, if we obtain a value **d** from the distribution of D corresponding to this level c, the value **d** can be expressed as

$$\begin{aligned} d &= \text{Mean} + Z_c * \text{Standard Deviation} \\ &= b + \mu_A + \rho\frac{\sigma_A}{\sigma_B}(b-\mu_B) + Z_c * \sigma_A\sqrt{1-\rho^2} \\ &= b + \mu_A + \rho\frac{\sigma_A}{\sigma_B}(b-\mu_B) + \text{Normsinv}(c) * \sigma_A\sqrt{1-\rho^2} \end{aligned} \tag{10}$$

where:

$Z_c = \phi^{-1}(c) = c^{th}$ percentile of N(0,1), i.e., inverse CDF of N(0,1) at the probability level c

Normsinv() = inverse CDF of N(0,1) in Excel; for example, Normsinv(0.95) = 1.645 and $P(Z < 1.645) = 0.95$

Consequently, the value b can be expressed using d based upon Equation 10:

$$b = \frac{d - \mu_A + \rho \dfrac{\sigma_A}{\sigma_B}\mu_B - Z_c * \sigma_A\sqrt{1-\rho^2}}{1 + \rho \dfrac{\sigma_A}{\sigma_B}} \tag{11}$$

LOG-NORMAL AND BIVARIATE LOG-NORMAL DISTRIBUTIONS

Log-normal Distribution. If a random variable X is associated with a log-normal distribution with a mean of μ and variance σ^2 in **log** space, i.e., X ~ LN(μ, σ^2), then the pdf of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}x\sigma}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad \text{for } x > 0 \tag{12}$$

Listed below are the mean, median, mode, and standard deviation of X:

$$\text{Mean} = e^{\mu + \frac{\sigma^2}{2}} \tag{13}$$

$$\text{Median} = e^{\mu} \tag{14}$$

$$\text{Mode} = e^{\mu - \sigma^2} \tag{15}$$

$$\text{Stdev} = e^{\mu + \frac{\sigma^2}{2}}\sqrt{e^{\sigma^2} - 1} = E(X)\sqrt{e^{\sigma^2} - 1} \tag{16}$$

Note that the mean, median, and mode of a log-normal distribution are all different. Additionally, if the unit-space mean (Mean) and standard deviation (Stdev) are given, the log-space mean (μ) and standard deviation (σ) can be derived using Equations 13 and 16:

$$\sigma \text{ (in log space)} = \sqrt{\ln(1 + (Stdev / Mean)^2)} = \sqrt{\ln(1 + CV^2)} \tag{17}$$

$$\mu \text{ (in log space)} = \ln(\text{Mean}) - \sigma^2/2 \tag{18}$$

Cumulative Distribution Function (CDF). Based upon Equation 2, the CDF of a log-normal distribution X can be expressed as follows:

$$F(x) = P(X \le x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}} dt = \phi((\ln(x) - \mu)/\sigma) \quad \text{for } x > 0 \tag{19}$$

where ϕ denotes the CDF of N(0,1). We can also use the following Excel functions to describe Equation 19:

$$F(x) = P(X \le x) = \text{NormDist}(\ln(x), \mu, \sigma, \text{TRUE}) = \text{LognormDist}(x, \mu, \sigma) \quad \text{for } x > 0 \qquad (20)$$

Bivariate Log-normal Distribution. Now let us assume X_i follows a log-normal distribution, i.e., $X_i \sim LN(\mu_i, \sigma_i^2)$ for i = 1 and 2, and the correlation coefficient between $LN(X_1)$ and $LN(X_2)$ is ρ. X_1 and X_2 are then said to have a bivariate log-normal distribution with the following pdf:

$$f(x_1, x_2) = \frac{1}{2\pi x_1 x_2 \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \{ \frac{(\ln(x_1)-\mu_1)^2}{\sigma_1^2} - 2\rho(\frac{\ln(x_1)-\mu_1}{\sigma_1})(\frac{\ln(x_2)-\mu_2}{\sigma_2}) + \frac{(\ln(x_2)-\mu_2)^2}{\sigma_2^2} \} \right) \qquad (21)$$

where $x_1 > 0$, $x_2 > 0$, and $\rho = \rho_{\ln x1, \ln x2}$. The above bivariate distribution of X_1 and X_2 is commonly denoted by Bivariate $LN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho)$. Note that all the parameters (μ_i, σ_i, and ρ) are evaluated in **log** space.

Correlation coefficient. If two continuous random variables X_1 and X_2 follow a Bivariate $LN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho)$, then the correlation coefficient between X_1 and X_2 is given by

$$\rho_{x_1, x_2} = \frac{e^{\rho\sigma_1\sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1}} \qquad (22)$$

where ρ is the correlation coefficient in log space between $LN(X_1)$ and $LN(X_2)$ as mentioned above, i.e., $\rho = \rho_{\ln x1, \ln x2}$.

Based upon Equation 22, the log-space correlation ρ can be expressed by $\rho_{x1,x2}$ as follows:

$$\rho = \frac{1}{\sigma_1\sigma_2} \ln\left(1 + \rho_{x_1, x_2} \sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1} \right) \qquad (23)$$

Note: Given the unit-space mean and standard deviation, we can use Equations 17 and 18 to compute the respective log-space mean and standard deviation. The correlation coefficient $\rho_{x1,x2}$ in the unit space can be calculated accordingly using Equation 22.

It also follows from Equation 22 that the correlation coefficient $\rho_{x1,x2}$ is bounded between two numbers:

$$\frac{e^{-\sigma_1\sigma_2}-1}{\sqrt{e^{\sigma_1^2}-1}\sqrt{e^{\sigma_2^2}-1}} \le \rho_{x_1,x_2} \le \frac{e^{\sigma_1\sigma_2}-1}{\sqrt{e^{\sigma_1^2}-1}\sqrt{e^{\sigma_2^2}-1}} \qquad (24)$$

Conditional Distribution. If two random variables X_1 and X_2 follow a bivariate lognormal distribution $LN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho)$, the conditional pdf of X_1 given $X_2 = x_2$ is derived as

$$\begin{aligned} f(x_1 \mid X_2 = x_2) &= f(x_1, x_2) / f(x_2) \\ &= \frac{1}{\sqrt{2\pi}x_1\sigma_1\sqrt{1-\rho^2}} \exp\left(-\frac{[\ln x_1 - \mu_1 - \rho\sigma_1/\sigma_2(\ln x_2 - \mu_2)]^2}{2\sigma_1^2(1-\rho^2)} \right) \end{aligned} \qquad (25)$$

In other words, the conditional distribution of X_1 given $X_2 = x_2$ is given by

$$X_1 \mid X_2 = x_2 \sim LN(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho^2)) \quad \text{for } x_1 > 0 \text{ and } x_2 > 0 \qquad (26)$$

As a result, the conditional mean of the distribution X_1 given $X_2 = x_2$ is a function of x_2:

$$\begin{aligned} E(X_1 \mid X_2 = x_2) &= \exp(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\ln x_2 - \mu_2) + \sigma_1^2(1-\rho^2)/2) \\ &= x_2^{\rho\sigma_1/\sigma_2} \exp(\mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2 + \sigma_1^2(1-\rho^2)/2) \end{aligned} \qquad (27)$$

The conditional standard deviation of X_1 given $X_2 = x_2$ is also a function of x_2:

$$\begin{aligned} Stdev(X_1 \mid X_2 = x_2) &= \exp(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\ln x_2 - \mu_2) + \sigma_1^2(1-\rho^2)/2)\sqrt{e^{\sigma_1^2(1-\rho^2)}-1} \\ &= x_2^{\rho\sigma_1/\sigma_2} \exp(\mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2 + \sigma_1^2(1-\rho^2)/2)\sqrt{e^{\sigma_1^2(1-\rho^2)}-1} \end{aligned} \qquad (28)$$

Using the notation in schedule analysis (Y = A and X = B), the conditional distribution of A at a given b value (i.e., Equation 26) can be expressed as

$$A \mid B = b \sim LN\left(\mu_A + \rho\frac{\sigma_A}{\sigma_B}(\ln(b) - \mu_B), \sigma_A^2(1-\rho^2) \right) \text{ for } a > 0 \text{ and } b > 0 \qquad (29)$$

Therefore, the conditional distribution of A plus b (denoted by D) when B is at a given “b” value is given by

$$A + b \mid B = b \sim LN\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B} (\ln(b) - \mu_B), \sigma_A^2 (1 - \rho^2) \right) + b \tag{30}$$

With a chosen probability level c, if we obtain a value **d** from the distribution of D corresponding to this level c, the value d can be expressed as

$$\begin{aligned} d &= \exp\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B} (\ln(b) - \mu_B) + Z_c * \sigma_A \sqrt{1 - \rho^2} \right) + b \\ &= \exp\left(\mu_A - \rho \frac{\sigma_A}{\sigma_B} \mu_B + Z_c * \sigma_A \sqrt{1 - \rho^2} \right) \exp\left(\rho \frac{\sigma_A}{\sigma_B} \ln(b) \right) + b \\ &= K * \exp(\ln(b^{\rho \frac{\sigma_A}{\sigma_B}})) + b = K * b^{(\rho\sigma_A/\sigma_B)} + b = K * b^m + b \end{aligned} \tag{31}$$

where the letter K denotes the constant:

$$\mathrm{K} = \exp\left(\mu_A - \rho \frac{\sigma_A}{\sigma_B} \mu_B + Z_c * \sigma_A \sqrt{1 - \rho^2} \right) \tag{32}$$

The value b in Equation 31 cannot be solved by a closed-form formula. Therefore, we use Solver to derive a solution for b at a given d value and a chosen probability level c. Note: In Equation 31, the means (μ_A, μ_B), standard deviations (σ_A, σ_B), and correlation coefficient (ρ) are all evaluated in **log** space, while the values of b and d are in unit space. Note: Use Equations 17 and 18 to derive the log space mean and standard deviation and use Equation 23 to derive the correlation in log space.

Analytic solution:

We can use the Taylor series expansion to approximate the solution for b in Equation 31. A Taylor series is a series expansion of a function at a single point (which we termed an anchor point in this document). A Taylor series of a real function $f(x)$, which is infinitely differentiable at a real value x_o is given by

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \tag{33}$$

where f(n)(xo) is the nth derivative of f evaluated at the point xo and n! denotes the factorial of n. The derivative of order zero of f is f itself by definition and (x − xo)0 and 0! are both defined to be one. When xo is 0, the series (Equation 33) is also known as a Maclaurin series.

If we just use the first couple of terms of the Taylor Series expansion to approximate *f*(*x*):

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0) \tag{34}$$

The root of *f*(*x*) is then given by $\boldsymbol{x_o - f(x_o)/f'(x_o)}$.

Now let's assume $\boldsymbol{y} = f(x) = Kx^m + x - d$, where m denotes $\rho*(\sigma_A/\sigma_B)$. It follows from Equation 34 that the solution for the root of *y* (i.e., b value in Equation 31) can be approximated by the first order derivative of *f*(*x*) using the Taylor series expansion:

$$b\_approx = x_o - \frac{f(x_o)}{f'(x_o)} = x_o - \frac{K(x_o)^m + x_o - d}{Km(x_o)^{m-1} + 1} \tag{35}$$

where $\boldsymbol{x_o}$ **is chosen to be the solution for the normality case** (i.e., Equation 11). An anchor point is a crucial part in the Taylor series expansion for finding a good approximation for the root of *f*(*x*). We have explored several anchor points to approximate the root of *f*(*x*) and found that using this point $\boldsymbol{x_o}$ (the normality case) in the Taylor series expansion seems to be a very good approximation of the solution for b. In other words, Equation 35 satisfies Equation 31 at a given d value.

Validate the Threshold Equation using Crystal Ball

Given both A and B are log-normally distributed, we set up a forecast cell D in Crystal Ball as the sum of A and B. We took a slice of **d** (d +- 0.5%) and then a slice of **b** from Equation 35 at a given probability level. We found the respective "**a**" value is very close to the target percentile. This shows the threshold formula provides consistent results.

BIVARIATE NORMAL/LOGNORMAL DISTRIBUTION

Now let us assume X_1 follows a normal distribution, i.e., $X_1 \sim N(\mu_i, \sigma_i^2)$ while X_2 follows a log-normal distribution, i.e., $X_2 \sim LN(\mu_2, \sigma_2^2)$, and the correlation coefficient between X_1 and $LN(X_2)$ is ρ. X_1 and X_2 are then said to have a bivariate normal/log-normal distribution with the following pdf:

$$f(x_1, x_2) = \frac{1}{2\pi x_2 \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{\ln(x_2)-\mu_2}{\sigma_2}\right) + \frac{(\ln(x_2)-\mu_2)^2}{\sigma_2^2} \right\} \right)$$

(36)

where $-\infty < x_1 < \infty$, $0 < x_2 < \infty$, and $\rho = \rho_{x1,\ln x2}$. The above bivariate distribution of X_1 and X_2 is commonly denoted by Bivariate NLogN((μ_1, μ_2), (σ_1^2, σ_2^2), ρ). Note that the parameters, μ_1 and σ_1, are evaluated in unit space, while μ_2 and σ_2 are evaluated in **log** space.

In cost analysis, it is not uncommon for the cost to follow a normal-ish distribution while the system's total schedule is log-normally distributed because the total schedule is the sum of many correlated activities in the network.

Note also the correlation between X_1 and X_2 is driven by the correlation between X_1 and $\ln(X_2)$:

$$\rho_{x_1,x_2} = \rho_{x_1,\ln(x_2)} \frac{\sigma_2}{\sqrt{e^{\sigma_2^2}-1}} = \rho \frac{\sigma_2}{\sqrt{e^{\sigma_2^2}-1}} = \rho \frac{\sigma_2}{CV_{X_2}} \tag{37}$$

Based upon Equation 37, the correlation ρ between X_1 and $\ln(X_2)$ can be expressed by $\rho_{x1,x2}$ as follows:

$$\rho = \rho_{x_1,\ln(x_2)} = \rho_{x_1,x_2} \frac{\sqrt{e^{\sigma_2^2}-1}}{\sigma_2} = \rho_{x_1,x_2} \frac{CV_{X_2}}{\sigma_2} \tag{38}$$

Since the range for ρ is between -1 and 1, the range for $\rho_{x1,x2}$ is bounded by the following:

$$\frac{-\sigma_2}{\sqrt{e^{\sigma_2^2}-1}} \le \rho_{x_1,x_2} \le \frac{\sigma_2}{\sqrt{e^{\sigma_2^2}-1}} \tag{39}$$

where X_1 and X_2 follow a Bivariate NLogN((μ_1, μ_2), (σ_1^2, σ_2^2), ρ).

Conditional Distribution. If the random variables X_1 and X_2 follow a Bivariate **N**LogN((μ_1, μ_2), (σ_1^2, σ_2^2), ρ), then the conditional pdf of X_1 given $X_2 = x_2$ is given by

$$f(x_1 \mid X_2 = x_2) = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left(-\frac{[x_1 - \mu_1 - \rho(\sigma_1/\sigma_2)(\ln(x_2)-\mu_2)]^2}{2\sigma_1^2(1-\rho^2)}\right) \tag{40}$$

This conditional pdf is very similar to Equation 4 except that the value x_2 is replaced by $\ln(x_2)$. Equation 40 can also be described as follows:

$$X_1 \mid X_2 = x_2 \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (\ln(x_2) - \mu_2), \sigma_1^2 (1 - \rho^2) \right) \tag{41}$$

Using the notation in schedule analysis (X_1 = A and X_2 = B), the conditional distribution of A at a given b value (i.e., Equation 41) can be expressed as

$$A \mid B = b \sim N\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B} (\ln(b) - \mu_B), \sigma_A^2 (1 - \rho^2) \right) \text{ for } -\infty < a < \infty \text{ and } b > 0 \tag{42}$$

Therefore, the conditional distribution of A plus b (denoted by D) when B is at a given “b” value is given by

$$A + b \mid B = b \sim N\left(b + \mu_A + \rho \frac{\sigma_A}{\sigma_B} (\ln(b) - \mu_B), \sigma_A^2 (1 - \rho^2) \right) \text{ for } -\infty < a < \infty \text{ and } b > 0 \tag{43}$$

With a chosen probability level c, if we obtain a value **d** from the distribution of D corresponding to this level c, the value d can be expressed as

$$\begin{aligned} d &= b + \mu_A + \rho \frac{\sigma_A}{\sigma_B} (\ln(b) - \mu_B) + Z_c * \sigma_A \sqrt{1 - \rho^2} \\ &= b + \mu_A - \rho \frac{\sigma_A}{\sigma_B} \mu_B + Z_c * \sigma_A \sqrt{1 - \rho^2} + \rho \frac{\sigma_A}{\sigma_B} \ln(b) \\ &= b + K + \rho \frac{\sigma_A}{\sigma_B} \ln(b) \end{aligned} \tag{44}$$

where the letter K denotes the constant:

$$K = \mu_A - \rho \frac{\sigma_A}{\sigma_B} \mu_B + Z_c * \sigma_A \sqrt{1 - \rho^2} \tag{45}$$

Analytic solution:

The value of b in Equation 44 cannot be solved by a closed-form formula. However, it can be approximated by the Taylor series expansion. Let’s assume $y = f(x) = x + K + m*\ln(x) - d$, where m denotes $\rho*(\sigma_A/\sigma_B)$. The approximated solution for the root of y is derived by using the first order derivative of the Taylor series expansion:

$$b\_approx = x_o - \frac{f(x_o)}{f'(x_o)} = x_o - \frac{x_o + K + m * \ln(x_o) - d}{1 + m / x_o} \tag{46}$$

where x_o is the solution for the normality case (i.e., Equation 11). This solution b_approx appears to be a good approximation, which satisfies Equation 44. Note that in Equation 44, the mean μ_B and standard deviations σ_B are evaluated in **log** space, while the correlation coefficient ρ is equal to the correlation between A and ln(B); namely, $\rho = \rho_{A,\ln(B)}$. Note: Use Equation 38 to derive the correlation between A and ln(B).

Conditional Distribution. If the random variables X_1 and X_2 follow a Bivariate NLogN((μ_1, μ_2), (σ_1^2, σ_2^2), ρ), then the conditional pdf of X_2 given $X_1 = x_1$ is given by

$$f(x_2 \mid X_1 = x_1) = \frac{1}{\sqrt{2\pi} x_2 \sigma_2 \sqrt{1-\rho^2}} \exp\left(- \frac{[\ln(x_2) - \mu_2 - \rho\sigma_2 / \sigma_1 (x_1 - \mu_1)]^2}{2\sigma_2^2 (1-\rho^2)} \right) \tag{47}$$

Equation 47 can also be described as follows:

$$X_2 \mid X_1 = x_1 \sim LN\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), \sigma_2^2 (1-\rho^2) \right) \tag{48}$$

Using the notation in schedule analysis (X_2 = A and X_1 = B), the conditional distribution of A at a given b value (i.e., Equation 48) can be expressed as

$$A \mid B = b \sim LN\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B} (b - \mu_B), \sigma_A^2 (1-\rho^2) \right) \text{ for } a > 0 \text{ and } \infty > b > -\infty \tag{49}$$

Therefore, the conditional distribution of A plus b (denoted by D) when B is at a given "b" value is given by

$$A + b \mid B = b \sim LN\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B} (b - \mu_B), \sigma_A^2 (1-\rho^2) \right) + b \tag{50}$$

With a chosen probability level c, if we obtain a value **d** from the distribution of D corresponding to this level c, the value d can be expressed as

$$\begin{aligned} d &= \exp\left(\mu_A + \rho \frac{\sigma_A}{\sigma_B}(b - \mu_B) + Z_c * \sigma_A \sqrt{1-\rho^2} \right) + b \\ &= \exp\left(\mu_A - \rho \frac{\sigma_A}{\sigma_B}\mu_B + Z_c * \sigma_A \sqrt{1-\rho^2} \right) \exp\left(\rho \frac{\sigma_A}{\sigma_B}(b) \right) + b \\ &= K * \exp\left(\rho \frac{\sigma_A}{\sigma_B}(b) \right) + b \end{aligned} \tag{51}$$

where the letter K denotes the constant:

$$\mathrm{K} = \exp\left(\mu_A - \rho \frac{\sigma_A}{\sigma_B}\mu_B + Z_c * \sigma_A \sqrt{1-\rho^2} \right) \tag{52}$$

Analytic solution:

The value of b in Equation 51 cannot be solved by a closed-form formula. However, it can be approximated by the Taylor series expansion. Let's assume $y = f(x) = \mathrm{K}*\exp(\mathrm{m}*x) + x - \mathrm{d}$, where m denotes $\rho*(\sigma_A/\sigma_B)$. The approximated solution for the root of y is derived by using the first order derivative of the Taylor series expansion:

$$b\_approx = x_o - \frac{f(x_o)}{f'(x_o)} = x_o - \frac{K * \exp(m * x_o) + x_o - d}{Km * \exp(m * x_o) + 1} \tag{53}$$

As described above, x_o is the solution for the normality case (i.e., Equation 11). This solution b_approx appears to be a good approximation, which satisfies Equation 51. Note that in Equation 51, the mean μ_A and standard deviations σ_A are evaluated in **log** space, while the correlation coefficient ρ is equal to the correlation between ln(A) and B; namely, $\rho = \rho_{\ln(A),(B)}$.

APPENDIX B – GLOSSARY

- **Allocation** – The process of spreading a top-level result (cost or duration) to lower levels
- **Buffer** – In the context of RIFT analysis, the difference (in workdays) between the deterministic finish date and probabilistic finish date
- **Conditional Probability** – The probability of an event occurring given knowledge of the outcome of another related event
- **Correlation** – Strength of the linear (or monotonic) relationship between two random variables
- **Criticality** – The average time a task falls on the critical path during a simulation; the probability a task falls on the critical path on any given iteration
- **Distribution A-LF** – The duration from the task late finish date to the project finish date on any given iteration (Duration "After")
- **Distribution B-LF**– The entire duration from the start of the project to the late finish date of any selected task or milestone on any given iteration (Duration "Before")
- **Distribution D** – The uncertainty distribution describing the duration of the entire project
- **Late Finish Date –** Task finish date plus total slack, if any; the latest date a task can finish without delaying the entire project
- **Probabilistic Contingency** – The duration gap between the deterministic finish date and the probabilistic finish date
- **Red Line** – In the context of RIFT analysis, the trend of the conditional probability of the selected project finish date plotted over time by dividing/binning the scatter to calculate the probability of the selected project finish date using only those dates in the bin
- **RIFT Range** – A span of time around the RIFT date used to calculate the conditional probability
- **Taylor Series** - A representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point
- **Tolerance** – In the context of RIFT analysis, the duration considered acceptable for a task to slip
- **Total Slack** - The amount of time that the task can be delayed without delaying the project finish date; the duration from task finish date to the critical path