

# Project and Program Management Analytics

By Bharat Gera, PMP

“ Uncertainty in processes cannot be neglected, and a predictive approach is the rule of the day. ”

Any project executed in today's marketplace, whether a gaming project launch in Las Vegas, Nevada, USA; a healthcare project in Beijing, China; a government policy project in the Philippines; or a space launch project at NASA, is based on strong analytical foundations. Today, mankind has a steady influx of high technology, which can be used to harness any information source or intelligence, including our own mother nature, to gain deeper insights to model and predict underlying process behaviors and outcomes with a sheer statistical paradigm.

Uncertainty in processes cannot be neglected, and a predictive approach is the rule of the day. The CEOs, CIOs, and CTOs (chief technology officers) of today no longer accept uncertainty as the “ignore” factor when making decisions. As a project manager, I no longer report the estimated project schedule or cost with definitive numbers and make a disclaimer of project risks. Traditional project management envisages and supports the legacy paradigm. Today, project managers need to report the project metrics in terms of “analytical certainty,” as Project X has:

- Schedule of 3PW (person weeks) + 0.5 PW with a 95% confidence interval

- Cost of US\$3000 + US\$100 with a 99% confidence interval

Am I asking statisticians to be project managers? Not really! Rather, I propose the basics of statistical analytics, which can be learned by tenth-grade students and incorporated in *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* — Fourth edition (PMI, 2008).

Before I detail the enhancements to the project management principles (as per the *PMBOK® Guide*), I briefly introduce the analytical concepts required for this project management report.

## A Brief Introduction to Analytical Concepts

Numerical measures of processes and systems have an inherent characteristic, which follows the probability distributions determined by the underlying process type (e.g., a process measure relying on independent accumulated data values). Essentially, this concept means that any numerical measure of the system is likely to occur a certain number of times if the system process is repeated in the long term. The metrics of uncertainty is captured in a probability distribution.

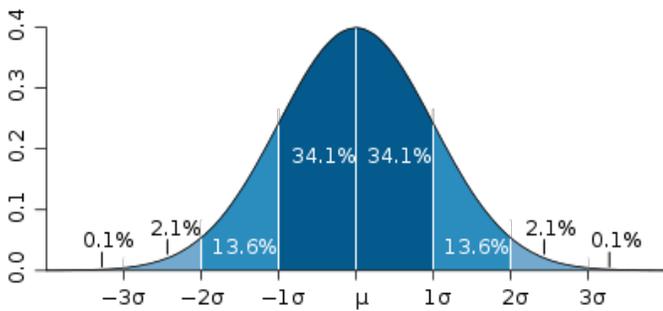
**Common Probability Distributions Observed in Project Management Scenarios (as per the PMBOK® Guide)**

In this section, I consider some common analytical probability distributions frequently encountered in project settings. The distributions are “analytical” because they are the products of analyses that mathematicians and statisticians have performed on the nature of the underlying processes that give rise to them. (They conducted these analyses precisely because the processes are frequently encountered.) The advantage of having these distributions is that once we determine that they are appropriate for a given situation, we can more accurately represent the uncertainty we face (e.g., in building a project risk model) and frequently gain important insight into a situation with much less effort.

**Uniform Distribution**

We start with perhaps the simplest of distributions in terms of the underlying process. If each value in a range has an equally likely chance of being generated by a process, then a uniform distribution accurately describes the likelihood of getting individual outcomes.

**Normal Distribution (A Fundamental Aggregation Distribution Observed in Nature):**

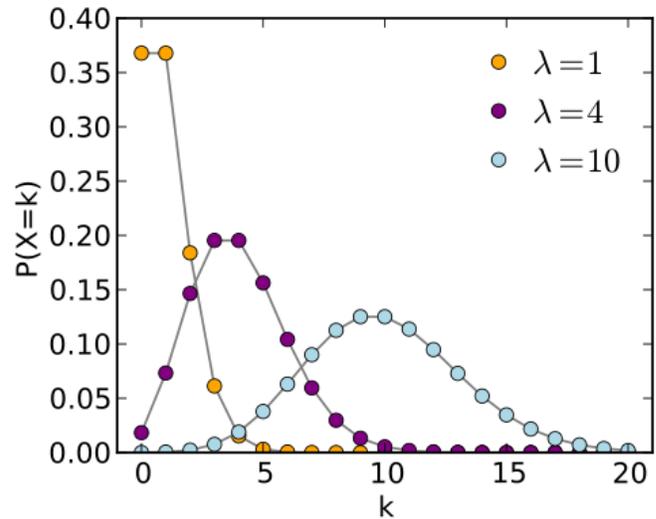


**Figure 1: Normal distribution.**

The familiar bell-shaped curve of the normal distribution (Figure 1) is undoubtedly the most recognizable of all probability distributions and also the most frequently misused. Although many processes in project management (as described in the *PMBOK® Guide*) and organizational management/businesses give rise to normal distributions, it is by no means ubiquitous. The underlying process that gives rise to a normal distribution is an accumulation process. Whenever an outcome is really the sum (or average) of the

outcomes of a number of uncertain quantities, different or the same, the probability distribution of the outcome is frequently a normal distribution.

**The Poisson Distribution**



**Figure 2: The Poisson distribution.**

The underlying process that gives rise to the Poisson distribution (Figure 2) is a counting process with an indefinite number of opportunities for an event to occur. Examples of such counting processes include the number of refrigerators sold in a department store in one week, the number of bonds issued on 1 December 2004, the number of people arriving at an ATM queue between 11:00 and 11:05 am on a given day, or the number of change requests (as defined in the *PMBOK® Guide*) that a project encounters per quarter. In general, counting processes or discrete countable independent event-generating systems are approximated to Poisson processes.

It is also the case that if the average number of occurrences per interval is large, the normal distribution gives us a good approximation of the Poisson probabilities. The deviation of a Poisson distribution is equal to the square root of  $m$ , the distribution mean. The example below is a demonstration of how analytical results can be practically applied in project executions.

For example, Project X is expected to hit 24 risks in the one-year schedule. Assuming a Poisson risk process, assess the level of risk preparedness required by the project team.

The above Poisson risk process can be approximated by a normal distribution of mean 24 and deviation  $\text{SQRT}(24)$ , assuming the project team is geared to cover risk outages up to the upper limit of 99% risk distribution:

Upper Risk Limit = NORMINV(0.99,24,SQRT[24]) ~35

The team has to be geared toward covering 35 risks during the year (average of ~10 day turn-around per risk), ensuring a 99% risk occurrence coverage in the project.

Also note that, only the mean value was used in the calculation, and the deviation was derived assuming an underlying Poisson process.

### Exponential Distribution (Decaying Systems)

Like the Poisson distribution, the underlying process that generates the exponential distribution (Figure 3) is concerned with the occurrence of an event. We are concerned with the time until the next event occurs, and therefore, the resulting probability distribution of the time until the next event or waiting time. The distribution is memory-less in the sense that the event waiting time is not dependent on when the last event occurred.

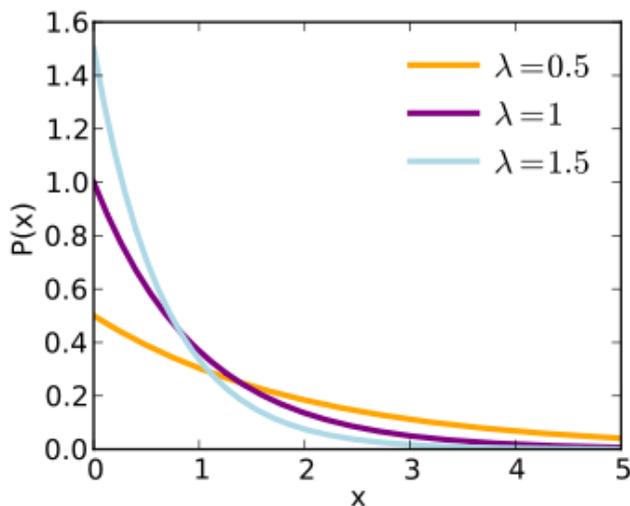


Figure 3: Exponential distribution.

### Queue Models — An Overview

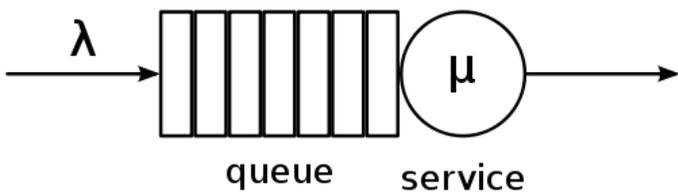


Figure 4: The basic M/M/1 queue model.

We will just touch upon the basic queue model (Figure 4), given the model’s relevance to analytically estimate project schedules. We assume we have a single task-processing system

and the task arrival is a Poisson process (rate λ). The time taken to complete the task is exponentially distributed (service rate μ); technically, this is an M/M/1 queue model.

The total expected waiting time (queue service) is,

$$T = \frac{1}{\mu - \lambda}$$

### Determining the Appropriate Probability Distribution

We must determine which analytical distribution to use for a given problem. There are two major methods we can use: The first method compares the processes that generate the data; the second compares the data distributions themselves. In the first approach, we must define the process that is generating the experienced uncertainty/variation and compare it with those known to generate particular analytical distributions. In the second approach, we generate a frequency distribution from historical data and compare it directly with known analytical probability distributions. The second method is the philosophy behind “best fit” software packages. If we find a significantly close match between our distribution and a particular analytical distribution, we can assume that the underlying process must be of the type that produces that particular analytical distribution, even if we do not have a precise definition of the process itself. It is highly recommended that project and/or program managers use the first method, given that the underlying processes are fairly known or provided in most circumstances.

Analytical probability distributions can be very useful, if and when they apply, and can be used only when the underlying processes satisfy the conditions necessary for their derivation, or when you have sufficient data you are comfortable with, assuming the distribution fits. When you do not have sufficient data, or cannot understand the underlying process sufficiently, you may want to fall back on simple distributions like the uniform distribution to represent uncertainty.

### Application of Analytics to Project Management Principles (as described in the PMBOK® Guide)

I used to skip management classes on analytical theory and, here I am, writing a three-page article on analytical theory! Well, I’m glad I just quantitatively used that statement of pun... Let’s begin with real-time analytical application in project management.

### Risk Management

Risks in project execution typically arrive at a Poisson process with the inter-risk time exponentially distributed.

For example: The oil & gas exploration industry has matured over the past century. Project data analyses over all initial project launch executions have revealed that these start-up projects encounter an average of approximately six risks within the first year. Ralance Industries, in conjunction with the State Government of Mumbai, has jump-started an oil exploration project in Mumbai's Southern Coast. The project manager of Ralance Industries is faced with the unique challenge of lack of government cooperation in risk management for the next 30 days. What should the project manager recommend to the CEO of Ralance Industries in his or her project status report?

These are real-time situations that global project managers will need to deal with. The project decisions made today must have strong analytical foundations. An analytical approach with this background is to estimate the chances of Ralance Industries being hit with a risk within the next 30 days.

Using the MS Excel® EXPONDIST formula to calculate the additive/cumulative probability of the occurrence of a risk event in the next one month, where

$$P(T \leq 30) = \text{EXPONDIST}(1/12, 6, 1) = 39\%$$

The distribution  $\lambda(\text{rate})$  is assumed to be 6/year in the above calculation.

The project manager should clearly indicate that the project is subject to an exposure of a 39% chance of being halted due to lack of government support. In practical terms, Ralance Industries may go with a risk of less than 40% against halting the project.

### Cost Management

Project cost is a vital piece in today's global project management. Analytics in cost optimization has become more vital because of the uncertainty of global economic equilibrium. Growth countries (e.g., India) have high inflation rates and face the heat of inflationary forces, whereas more mature markets (e.g., the United States) are being pulled into short-term imbalances of credit investments without underlying dollar generation.

Project management principles (as per the *PMBOK® Guide*) rightly advocate the principle of lowering project execution costs. Analytics brings in enormous power to this decision-making factor.

### Case Study I

The Philippine government passed the Biofuel Act in 2006, and the agriculture department was in charge of overseeing

its implementation. The law required specific proportions of bio-ethanol to be blended with gasoline. The prime intent was to boost the local production of biofuel feedstock and reduce imports from low-cost production economies (e.g., Brazil). The project management of the bio-ethanol supply chain network for local production was based on highly optimized linear analytical models calculating the production requirements of bio-ethanol in various locations across the Philippines. The complete details are beyond the scope of this article; nonetheless, a basic appreciation of optimization requirements and the importance of project management costing to policymaking is the intent of listing this case.

### Case Study 2

The World Health Organization (WHO) has decided to fast track its initiative to reduce the rapid spread of HIV in Kenya. The pilot rollout project aims to reduce spread of the disease and provide treatment to 60% of the infected population. The expected average cost of medication for those under 20 is US\$10,000, US\$25,000 for those in the 20- to 35-year-old age group, and US\$15,000 for those over the age of 35. The National Health Institute of Nairobi has reported that HIV cases are normally distributed with a mean age of 30 and a standard deviation of 7, and also estimates that there are six million people infected with the disease.

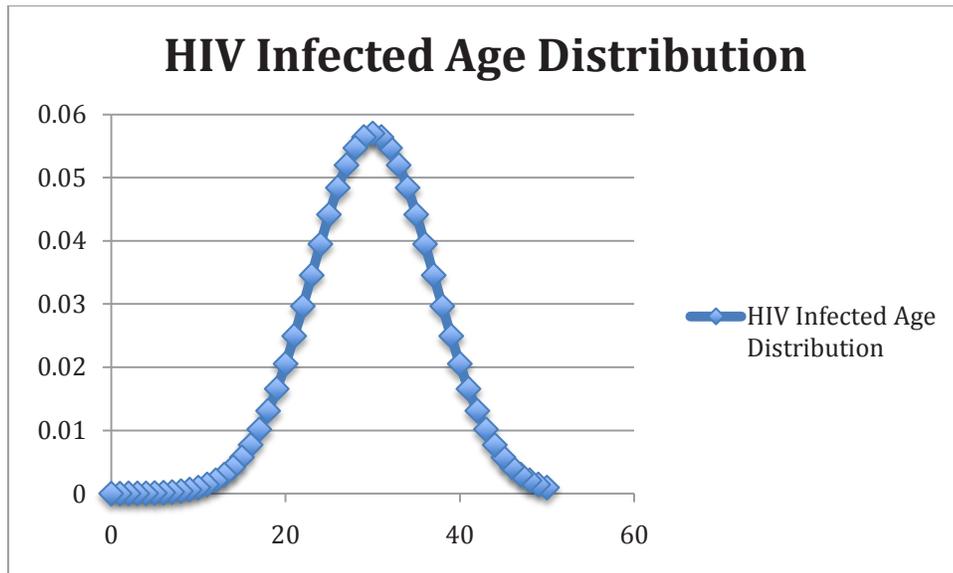
The project plan and budget details have to be submitted for the funding. How does the involved project office calculate the best execution plan and cost of the project?

The statistical information should be leveraged for this case. As per the normal distribution (Figure 5), approximately 68% of the population is covered in one standard deviation from the mean (either side).

The plot illustrates that approximately 68% of the infected population is between the age group of 30 +/-7 (i.e., [23-37]). This suggests that the project plan should target the age group of 23 to 37 for the initial project rollout. This will ensure that the WHO guideline of 60% is met and the highly dense HIV distribution is targeted.

The project costing has to be split for the age groups [23-35] and [35+]. The percentage of population covered in this age group can again be calculated from the distribution graph.

$$\begin{aligned} \text{Population covered under [23-35]} &= \\ &= \text{NORMDIST}(35, 30, 7, 1) - \text{NORMDIST}(23, 30, 7, 1) \\ &= 60.4\% \end{aligned}$$



**Figure 5: HIV-infected population age distribution.**

Population covered under  $[35, \infty]$  =  
 $1 - \text{NORMDIST}(35, 30, 7, 1)$   
 = 23.6%

Cost =  $[25K * 60.4% * 6 M] + [15K * 23.6% * 6 M] =$   
 \$US11.2 billion

### Schedule Management and Resource Allocation

One of the challenges faced in maintenance projects is the uncertainty associated with defect inflow from end consumers. Scheduling and budgeting for these processes are also prone to high inaccuracies in estimating the defect closure rate.

It is worthwhile modeling such projects as an M/M/1 queue to analytically estimate the characteristics of the system, such as the average system response delay to incoming requests or defects. In general, decaying systems and processes fall under the umbrella of **Poisson and Exponential Distributions**.

For example, consider a level 3 software maintenance project that has historically encountered ten defects from the market end customer per month. The arrival rate can be approximated to be a Poisson process with a rate  $\lambda = 10/\text{month}$ . The team working on the level 3 project has a collective average throughput of 24 defects per month, which is equivalent to a service rate of  $\mu = 24/\text{month}$ .

Modeling the system as an M/M/1 queue, the average response time of the system is  $1/(\mu - \lambda) = 1/(24 - 10) = 0.0714 \text{ month} \sim 2 \text{ days}$ . Any defect being logged will take an average of two days for closure.

The project service level agreement should be driven from the calculated response times to ensure customer commitments, in addition to avoid levying any penalty charges to the organization.

### Conclusion

I have made a sincere effort to demonstrate the modeling of project management principles to include statistical business analytics to enable projects to execute optimally in all subject areas. By no means do I expect project managers to be statisticians; but a culture of leveraging statistical know-how to processes can be a key factor these days. I hope the project management principles driven by Project Management Institute will be enhanced over time to add relevant analytics in key project calculations and evaluation procedures.

### About the Author

Bharat Gera is a line manager at IBM Corporation. He has been involved in the project management of enterprise software and hardware projects over the past decade and a half. He has worked primarily in storage, database, data warehouse, and analytics domains. He is a Project Management Professional (PMP)<sup>®</sup> and has an executive management degree with a specialization in business analytics from the Indian Institute of Management, Bangalore. He can be reached at [bharat.gera@alumni.iimb.ernet.in](mailto:bharat.gera@alumni.iimb.ernet.in)

Disclaimer: The views expressed here are strictly the personal thoughts of the author as deemed applicable to today's global project execution.