Volume 9, Issue 9, September– 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

Statistically Mitigating Subjective Estimates with PERT and Montecarlo

A Method for Combining PERT with Monte-Carlo Simulation

[1]Abel J. Stephen

Abstract:- This paper introduces a methodology for advanced project duration estimation, integrating the Program Evaluation and Review Technique (PERT) with Monte Carlo simulation. It employs various distributions — normal and beta — to enhance the accuracy of task duration modeling based on initial three-point estimates. This aproach refines these distributions, establishing a robust mode while maintaining a consistent 90th-percentile confidence level. The study illustrates the feasibility of the implementation using accessible tools i.e., Google Sheets and Power BI, ensuring practicality in project management. The conclusion underscores improved accuracy and reliability in project duration estimates, enhancing risk management and decision-making throughout project execution.

Keywords:- *Monte Carlo Simulation, PERT, Machine Learning, Artificial Intelligence, Probabilistic Modelling, Project Duration Estimation, Project Management.*

I. INTRODUCTION

Technical Program Managers (TPMs) often rely on Subject Matter Experts (SMEs) to provide duration estimates in planning timelines and levels of effort. It is crucial to statistically mitigate the subjectivity of these estimates to arrive at a probabilistic project completion duration. Accurate project duration estimation is the bedrock of effective project management. Traditionally, practitioners have relied on deterministic single-point estimates, which often fail to account for the uncertainties inherent in complex projects.

Integrating PERT and Monte Carlo simulations offers a sophisticated, probabilistic approach to project duration estimation that simplifies the recommendation process while increasing accuracy and reliability. (Ballesteros-Pérez et al., 2020) This methodology effectively transforms subjective expert judgments into robust data points, capturing project timelines' inherent variability and uncertainty. For instance, empowering a project manager to quickly and easily assert:

"***Simulations indicate that a specific task is most likely to take around 29.68 days, which is 97.87% longer than the initial estimate of 15.00 days. The recommended value is between 29.68 days (66.60% confidence) and 31.69 days (90% confidence)***. "

This is a valuable outcome when under the duress of planning. This combined approach elevates project duration estimation to a new level of precision and realism. It is highly relevant for project managers, as the ability to accurately forecast project duration provides a strategic advantage, enabling cost savings and quicker project completion. Transitioning from using single-point estimates or even employing PERT and Monte Carlo independently to now fusing PERT and Monte Carlo affords a probabilistic model that marks a significant step forward in project management rigor and ease.

This methodology is accessible, especially for projects that can be readily captured in Google Sheets as a simple task list. Adopting this paper's novel methodology offers a more sophisticated and realistic framework for project duration estimation.

➢ *Why Start with the Three-Point Estimates?*

Starting with three-point estimates — optimistic, most likely, and pessimistic — provides a structured way to capture the range of possible outcomes for project tasks. This method addresses the inherent uncertainty and variability in project estimation, offering a more comprehensive understanding than single-point estimates. Depending on the estimator's subjective view, these estimates can be overly optimistic or pessimistic.

In contrast, using three-point estimates allows for a more balanced and realistic assessment of potential outcomes. This approach incorporates the best-case scenario (O), the most likely scenario (M), and the worst-case scenario (P), providing a broader perspective on possible durations.

➢ *Triangular Distribution Overview*

Research suggests that a triangular distribution is defined by the following three parameters, which are used in distributions where there is a "lack of knowledge" of distribution (Kissell & Poserina, 2017).

- O: Optimistic (best-case) estimate
- M: Most likely estimate
- P: Pessimistic (worst-case) estimate

The probability density function (PDF) of a triangular distribution is piecewise linear and forms a triangle with a peak at M. (P. Bertsekas & N. Tsitsiklis., 2002, p. 140)

➢ *Variance of the Triangular Distribution*

The variance of a distribution is a measure of the spread of its values. For a triangular distribution, the variance (σ^2) is given by:

$$\sigma^2 = \frac{(P - O)^2}{24}$$

This formula can be derived by calculating the second moment about the mean. The following is a step-by-step explanation:

➢ ***Piecewise PDF**: The triangular distribution's PDF can be split into two linear segments:*

- From O to M:

$$f(x) = \frac{2(x - O)}{(P - O)(M - O)}$$

- From M to P:

$$f(x) = \frac{2(P - x)}{(P - O)(P - M)}$$

➢ ***Mean Calculation**: The mean (expected value, or the 'μ') of the triangular distribution is:*

$$\mu = \frac{O + M + P}{3}$$

➢ ***Variance Calculation**: The variance is the expected value of the squared deviations from the mean:*

$$\sigma^2 = E[(X - \mu)^2] = \int_O^P (x - \mu)^2 f(x)\ dx$$

We need to compute this integral in two parts, corresponding to the two segments of the PDF.

- First Segment (from O to M):

$$\int_O^M (x - \mu)^2 \cdot \frac{2(x - O)}{(P - O)(M - O)}\ dx$$

- Second Segment (from M to P):

$$\int_M^P (x - \mu)^2 \cdot \frac{2(P - x)}{(P - O)(P - M)}\ dx$$

➢ ***Combining the Integrals**: After computing the integrals separately, we sum them up, leading to the triangle distribution's variance as below:*

$$\sigma^2 = \frac{(P - O)^2}{24}$$

➢ *Explanation of the Factor of 24*

The factor of 24 in the denominator comes from the integration process and normalization for the variance. (P. Bertsekas & N. Tsitsiklis., 2002, p. 140)

➢ *Here is why:*

- *Normalization:*

The triangular distribution's PDF is normalized so that the total area under the curve equals 1. This normalization involves factors of (P−O), (M−O), and (P−M).

- *Integration and Squaring:*

When calculating the second moment about the mean, we integrate the squared deviations over the entire range [O, P], where OOO is the optimistic estimate, and MMM is the most likely estimate, and PPP is the pessimistic estimate. The process involves squaring the terms (x−μ) and integrating the resulting polynomial expressions.

- *Combination of Terms:*

The combined effect of these integrations and the structure of the triangular distribution's PDF leads to the specific factor of 1/24 in the variance formula.

This captures the spread of the distribution and is derived from the properties of the triangular distribution, specifically the piecewise linear PDF and the range (P−O). The factor of 24 results from the integration and normalization processes involved in computing the variance for this particular distribution. This detailed mathematical derivation ensures that the variability and uncertainty in task durations are accurately represented. Using three-point estimates and the triangular distribution allows project managers to make more informed decisions and better manage risks in project timelines.

➢ *Why PERT?*

Transitioning from three-point estimates to the Program Evaluation and Review Technique (PERT) builds on the foundation of capturing variability. It provides a more structured and probabilistic approach to project duration estimation. PERT refines the three-point estimate by applying a weighted average that emphasizes the most likely estimate while still considering the optimistic and pessimistic outcomes.

PERT is an essential tool for project management, particularly in task planning and scheduling. (Hernandez, 2021) It provides a weighted average that reflects a more realistic scenario by utilizing three estimates for each task—the best case, the most likely, and the worst case. This technique recognizes the inherent uncertainties in project timelines and helps mitigate risk.

PERT's strength lies in its ability to capture the most likely scenario while considering optimistic and pessimistic outcomes. This balanced approach helps project managers identify potential delays and allocate resources more effectively. Managers can better understand the variability and probability of completing the project within a given

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

timeframe by calculating the PERT mean and standard deviation.

Moreover, PERT aids decision-making by offering insights into the expected project duration and associated risks. This predictive capability ensures that project plans are grounded in realistic expectations, facilitating more accurate scheduling and efficient risk management.

PERT uses three estimates to capture a range of possible outcomes: the optimistic (O), pessimistic (P), and most likely (M) times. The following estimates are used to calculate the expected time (T.E.) for project tasks using the formula:

- *The Formula for PERT Mean (μ):*

$$\mu = \frac{\mathrm{O} + 4\mathrm{M} + \mathrm{P}}{6}$$

- *The Formula for PERT Standard Deviation (σ):*

$$\sigma = \frac{\mathrm{P} - \mathrm{O}}{6}$$

This approach allows PERT to account for variability by considering the potential spread of outcomes. Additionally, PERT calculates the task durations' standard deviation (σ) to quantify the uncertainty. By doing so, PERT provides a probabilistic estimate incorporating a range of possible durations, reducing the reliance on a single-point estimate and thereby addressing uncertainty through variability.

The PERT method can be understood through two types of distributions: the triangular distribution and the beta distribution. The PERT is based on a beta distribution but is often simplified into a triangular distribution for ease of use. (Broadleaf Capital International Pty Ltd, 2014) When simplified into a triangular distribution, PERT presents a simple shape that is easy to visualize and understand.

To understand this, let us consider an analogy of a triangle with three key points: the best case (optimistic), the worst case (pessimistic), and the most likely time for a task. The triangular shape makes it straightforward, as the triangle peak represents the most likely time, and the sides slope down to the best and worst-case times. This approach is practical for rough, quick estimates because you draw straight lines connecting these points.

Thus, in the triangular distribution, the mode is nothing more than the "most likely" time. Its core benefits pertain to practical estimation due to its simplicity and realistic evaluation. PERT is useful in project management and risk assessment because it considers best-case, most likely, and worst-case scenarios. Calculating the mean and standard deviation makes this distribution method easy to understand and implement.

Lastly, it can also provide more realistic estimates for project durations and costs because it incorporates expert judgment about uncertainty and variability. The drawbacks are mainly related to subjectivity and non-normality. The fundamental reason is that the accuracy of the PERT distribution depends on the subjective estimates of the optimistic, most likely, and pessimistic values.

The limitation in flexibility is due to its nature, which assumes a specific beta distribution form. This form may not be the right fit for all types of data, making some distributions less accurate. In such a case, the data may be unreliable since the PERT distribution may not approximate a normal distribution well, especially if the real distribution is heavily skewed or has fat tails. In contrast, the beta distribution offers a more flexible and accurate shape.

Instead of a simple triangle, imagine a curve that can bend and flex to fit real-world data better. This curve still relies on the best, worst, and most likely times, but it can model more complex situations. The beta distribution is more accurate because it can show that most tasks are likely to finish within the most likely time while also accounting for the possibility of delays.

It uses sophisticated calculations to give more weight to the most likely time but also considers the best and worst cases nuancedly. The mode in the beta distribution is derived from its parameters and is not necessarily the same as the "most likely" time, which is used in the triangular distribution. The critical difference between these two distributions lies in their shape and calculation. The triangular distribution is more straightforward to calculate and visualize, making it suitable for quick, rough estimates. It is practical for initial planning and simple projects. On the other hand, the beta distribution provides a detailed and accurate model by using advanced calculations. It is better suited for detailed planning when it is crucial to understand the variability and risks.

The PERT method is considered to be based on a beta distribution due to its flexibility and characteristics, which make it well-suited for modeling uncertainty and variability in project durations. The beta distribution is highly flexible and can take on various shapes depending on its parameters, allowing it to accurately model different possible distributions of project completion times. This is defined over a finite interval, typically between the optimistic and pessimistic estimates, making it ideal for modeling activities with clear minimum and maximum duration limits.

PERT calculates the expected duration using a weighted average, giving more weight to the most likely estimate, similar to the mean of a beta distribution. Additionally, the beta distribution allows for calculating variance and standard deviation, which is essential for understanding the dispersion or spread of possible project completion times. These characteristics make the beta distribution a realistic choice for representing project timelines.

The mean and variance are calculated similarly in both the triangular and beta distributions. However, the mode calculation differs. For the triangular distribution, the mode is simply the "most likely" time — $\boldsymbol{m}$. For the beta distribution, the mode is derived from its parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which are calculated based on the mean and variance.

This difference highlights why the beta distribution can provide a more nuanced and accurate representation of project durations, as it accounts for the probability distribution more flexibly and realistically than the triangular distribution. The mode is a critical statistical measure representing a distribution's most frequently occurring value. While the mean gives an average estimate and the median provides a middle value, the mode pinpoints the exact duration most likely to occur based on historical data or expert judgment.

If the mode significantly differs from the mean, it may indicate higher variability or risk in the project duration. Additionally, if the mode is closer to the optimistic or pessimistic estimates, it suggests a skewed distribution, indicating potential biases or risks in the estimates.

➢ *PERT Triangular Distribution*

- ***Mean (Expected Value)****: The mean of the Triangular Distribution in PERT is Calculated as:*

$$\text{Mean} = \frac{a + 4m + b}{6}$$

Where:

a is the optimistic estimate (best case),

m is the most likely estimate,

b is the pessimistic estimate (worst case).

- ***Standard Deviation****: The standard deviation for the triangular distribution is:*

$$\text{Standard Deviation} = \frac{b - a}{6}$$

- ***Mode****: The mode for the triangular distribution is simply the most likely estimate m.*

- *90th Percentile*

To determine the 90th percentile for a PERT distribution, you can use both linear interpolation and the z-score method. However, it's important to note that the z-score method assumes a normal distribution, which might not be as accurate for a PERT distribution, as it is actually a form of the Beta distribution. The linear interpolation method, on the other hand, leverages the nature of the Beta distribution while simplifying the process by assuming a piecewise linear cumulative distribution function (CDF) based on the best-case, most likely, and worst-case estimates. (P. Bertsekas & N. Tsitsiklis., 2002, p. 140) Here are the main steps:

- *Calculate Cumulative Probabilities at Critical Points:*

✓ Cumulative probability at the best case (C.D.F. $_{\text{best\_case}}$) = 0
✓ Cumulative probability at the most likely case (C.D.F. $_{\text{most\_case}}$) = 0.5
✓ Cumulative probability at the worst case (C.D.F. $_{\text{worst\_case}}$) = 1

- *Identify the Range where the Cumulative Probability Surpasses 0.9 (90%)*

✓ The range of interest is between the most likely and worst case, where the cumulative probability goes from 0.5 to 1.
✓ Use linear interpolation within this range:
✓ Linear interpolation is used to estimate values within a range based on known values at the endpoints.

We often use linear interpolation to find the 90th percentile in a triangular distribution. This involves solving for x, where the cumulative distribution function (CDF) equals 0.9. The CDF for a triangular distribution is piecewise and involves the following steps:

- *For x between a and m:*

$$F(x) = \frac{(x - a)^2}{(b - a)(m - a)}$$

- *For x between m and b:*

$$F(x) = 1 - \frac{(b - x)^2}{(b - a)(b - m)}$$

This approach provides a practical way to estimate the 90th percentile for a PERT distribution. It acknowledges its basis in the Beta distribution while utilizing a simpler, piecewise linear approximation.

➢ *PERT Beta Distribution*

The transition from PERT to the beta distribution represents a move towards even greater precision and flexibility in modeling project durations. While PERT often simplifies its calculations using the triangular distribution, it conceptually aligns with the beta distribution, which offers a more sophisticated representation of uncertainty and variability.

The beta distribution is defined over a finite interval, typically between the optimistic and pessimistic estimates, making it ideal for modeling activities with clear duration limits. It is highly flexible and can take on various shapes depending on its parameters, accurately reflecting different possible distributions of project completion times.

- **Mean (Expected Value)**: The mean for the beta distribution in PERT is the same as the triangular distribution:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

$$\text{Mean} = \frac{a + 4m + b}{6}$$

- **Standard Deviation**: The standard deviation for the beta distribution in PERT is:

$$\text{Standard Deviation} = \frac{b - a}{6}$$

- **Mode**: The beta distribution's mode is calculated differently and is more complex than the triangular distribution. However, for the sake of PERT simplification, the most likely estimate m is often used similarly to the triangular distribution.

$$Mean(\mu)\ \ and. Variance(\sigma^2)\ for\ PERT\ distribution$$

$$\mu = \frac{a + 4m + b}{6}$$

$$\sigma^2 = \left(\frac{b - a}{6}\right)^2$$

$$Parameters(\alpha\ and\ \beta) for\ Beta\ distribution$$

$$\alpha = \left(\frac{\mu - a}{b - a}\right)\left(\left(\frac{\mu - a}{b - a}\right)\left(1 - \frac{\mu - a}{b - a}\right)/\sigma^2 - 1\right)$$

$$\beta = \left(1 - \frac{\mu - a}{b - a}\right)\left(\left(\frac{\mu - a}{b - a}\right)\left(1 - \frac{\mu - a}{b - a}\right)/\sigma^2 - 1\right)$$

Where:

a is the optimistic estimate,

m is the most likely estimate,

b is the pessimistic estimate.

$$\text{Mode} = \frac{\alpha - 1}{\alpha + \beta - 2} \quad \text{for} \quad \alpha > 1 \quad \text{and} \quad \beta > 1$$

- **90th Percentile**: The 90th percentile for the beta distribution in PERT requires numerical methods or simulation because the beta distribution CDF does not have a simple closed form. Typically, this is achieved through:

$$\text{P}(X \leq x) = 0.90$$

The beta distribution's strength lies in its ability to model the probability distribution more flexibly and realistically than the triangular distribution, capturing the inherent uncertainty in project durations more accurately. This sophisticated modeling gives project managers a deeper understanding of potential risks and variability, enabling better decision-making and more effective project management.

- *Acknowledging Prior Work Around PERT*

Significant prior work has demonstrated the effectiveness of PERT in estimating project durations and managing project risks. One notable example is Malcolm, Roseboom, Clark, and Fazar's (1959) work, which introduced the PERT methodology for planning and controlling the Polaris missile project. This seminal work underscored the importance of incorporating uncertainties in project planning, leading to more realistic and achievable project schedules.

Acknowledging that projects rarely follow a single deterministic path, PERT provides a nuanced understanding of ambiguities by incorporating three-point estimates (optimistic, most likely, and pessimistic) for each task. This method yields a weighted average and a standard deviation, offering a balanced approach to estimating project durations. By doing so, PERT enhances the precision of project timelines and helps project managers identify potential delays early in the planning phase.

Kerzner (2017) and Meredith and Mantel (2019) explore PERT's application in various fields. These authors highlight how PERT, combined with other project management techniques, can improve the accuracy and reliability of project schedules. They emphasize that PERT's structured approach to uncertainty is invaluable in managing complex projects with multiple interdependent tasks.

The practicality of PERT, supported by extensive literature, makes it a fundamental tool in project management. Its ability to accommodate uncertainties and provide realistic estimates has been proven effective across various industries, from construction to software development. As we refine and integrate PERT with advanced simulation techniques, its relevance and applicability in project management remain robust and indispensable.

- *Why Monte Carlo Simulation?*

Monte Carlo Simulation is a powerful tool that mirrors the randomness inherent in real-world events. From the probability distributions defined by PERT's mean and standard deviation, Monte Carlo simulation generates many possible project durations. Each simulation represents a distinct realization, capturing the project task's unpredictability and risk. This computational prowess allows project managers to explore the full spectrum of outcomes, from best-case scenarios to worst-case contingencies. Project managers can use Monte Carlo simulation to make informed decisions based on a comprehensive view of potential outcomes. The method provides a rich backdrop of potential project trajectories, accounting for the improbabilities that pervade complex endeavors. Simulating thousands or even millions of scenarios reveals various outcomes' likelihood.

- *Acknowledging Prior Work Around Monte Carlo*

Significant prior work has demonstrated the effectiveness of Monte Carlo simulation in estimating project durations. One notable example is the study by Musa and Okumoto (1989), which pioneered Monte Carlo simulation

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

for software reliability. This study underscored the versatility of Monte Carlo simulation, extending its application beyond software to broader project management contexts.

Acknowledging that projects rarely unfold along a single deterministic path, PERT provides a nuanced understanding of ambiguities. However, the integration of Monte Carlo simulation propels this approach into a new echelon of precision, harnessing the power of computational algorithms to forecast a spectrum of possible outcomes.

Montgomery and Runger (2018) further explore statistical methods, including Monte Carlo simulation, offering practical examples adaptable to project management. Their work highlights how Monte Carlo simulation transforms subjective expert estimates into a robust statistical distribution. This method significantly diminishes human bias by generating thousands of potential outcomes through a Gaussian distribution, offering a granular perspective on project duration uncertainties.

Other literature further supports Monte Carlo's practicality. Karabulut's (2017) investigation of construction project planning exemplifies the effectiveness of using Monte Carlo simulation over the Critical Path Methodology (CPM). The study illustrates the relevance and significance of Monte Carlo in planning construction projects. Karabulut's study reveals that while CPM optimistically estimated completion within 186 working days, the Monte Carlo simulation suggested a 50% chance of completion within 205 days, highlighting the balanced perspective provided by the Monte Carlo simulation.

Our approach does not consider CPM at the moment, as we aim for a more generic solution that does not require the identification of task dependencies. This makes our method applicable even in cases where the initial project tasks are presented in a simple table with no dependencies outlined. This is particularly useful for straightforward initial estimations or smaller projects or even initial discovery efforts for larger projects, where dependencies are not yet necessary to outline.

Shaping Different Distributions with PERT and then Sampling with Monte Carlo to Get the 90th Percentile and Mode PERT provides a structured method to account for uncertainty using three-point estimates, yielding a weighted average and a realistic standard deviation. Monte Carlo simulation, on the other hand, leverages computational power to generate thousands of possible outcomes based on these PERT-derived parameters. This fusion allows project managers to visualize various potential project durations, offering a more comprehensive understanding of risks and uncertainties.

By integrating these two methodologies, we can predict not just the most likely outcomes but also the full spectrum of possibilities, thereby enabling more informed decision-making and robust risk management. Let us begin by examining the use of PERT (Program Evaluation Review Technique) as a starting point for obtaining 90% confidence levels in project duration estimation. We will explore PERT in the following contexts:

- PERT as a Standalone Method and 90th Percentile
- PERT with a Box-Muller Transform and 90th Percentile
- PERT-Informed Gaussian Distribution and 90th Percentile
- PERT-Informed Gaussian for Monte Carlo Simulation and 90th Percentile
- PERT-Informed Beta Distribution and 90th Percentile
- PERT-Informed Beta Distribution for Monte Carlo Simulation and 90th Percentile

By evaluating these methods, we aim to understand their individual strengths and weaknesses, which will help us achieve reliable confidence levels in project management.

II. SHAPING DIFFERENT DISTRIBUTIONS

This section delves into using PERT to shape various distributions. Then, it progresses to how Monte Carlo sampling generates a refined model while maintaining the 90th percentile confidence level. By leveraging different distributions, we can better capture project tasks' inherent variability and complexity, leading to more accurate and reliable project duration estimates.

➢ *PERT as a Standalone Method and 90th Percentile*

PERT (Program Evaluation Review Technique) uses optimistic, most likely, and pessimistic estimates to calculate the mean and standard deviation for project tasks. As a standalone method, PERT provides a straightforward calculation of the 90th percentile, offering a deterministic output. The mean and standard deviation derived from PERT remain the same because they are directly calculated from the three estimates, and the 90th percentile is determined through these fixed values.

The mode is simply the most likely estimate. Consider the following example scenario that has been given as follows, as part of a project management case to assess the estimated duration required for the project to be complete:

- Optimistic (a) = 10 days
- Most likely (m) = 20 days
- Pessimistic (b) = 40 days

➢ *PERT Formulae*

Mean (μ): $\mu = \frac{\mathrm{O+4M+P}}{6}$

Standard Deviation (σ): $\sigma = \frac{\mathrm{P-O}}{6}$

90th Percentile (Y): $Y = \mathrm{Y1} + \frac{(\mathrm{Y2-Y1})}{(X2-X1)} \times (\mathrm{X} - \mathrm{X1})$

Where;

Y is the value at the target percentile.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

Y1 is the value at the lower bound of the range containing the target percentile.

Y2 is the value at the upper bound of the range containing the target percentile.

X is the target percentile.

X1 is the cumulative probability at the lower bound.

X2 is the cumulative probability at the upper bound.

➢ *Applying PERT Formulae*

Mean (u):

$$\mu = \frac{\mathrm{a} + 4\mathrm{m} + \mathrm{b}}{6}$$

$$\mu = \frac{10 + 4 \times (20 + 40)}{6}$$

$$\mu = \frac{10 + 4 \times (20 + 40)}{6}$$

$$\approx 21.67 \text{ days}$$

Standard Deviation (σ):

$$\sigma = \frac{\mathrm{b} - \mathrm{a}}{6}$$

$$\sigma = \frac{40 - 10}{6}$$

$$\sigma = 5 \text{ Days}$$

➢ *Linear Interpolation for the 90th Percentile*

$$Value\ at\ 90th\ Percentile = \mathrm{m} + \frac{(\mathrm{b} - \mathrm{m})}{(1 - 0.5)} \times (0.9 - 0.5)$$

$$= 20 + \frac{(40 - 20)}{(1 - 0.5)} \times (0.4)$$

$$= 20 + \frac{(40 - 20)}{(0.5)} \times (0.4)$$

$$= 20 + \frac{(20)}{(0.5)} \times (0.4)$$

$$= 20 + 40 \times 0.4$$

$$= 20 + 16$$

$$= 36$$

The Box-Muller transform, as explained above through the formulae, is used to generate normal distribution samples efficiently. When combined with PERT, it transforms uniformly distributed random numbers into normally distributed ones, aligning the PERT mean and standard deviation with the generated samples, as explained by the usage of the formulae below. This method maintains the same 90th percentile confidence level from the initial PERT estimates to the transformed distribution because the Box-Muller transform retains the underlying statistical properties (mean, standard deviation) of the PERT estimates, ensuring that the transformed distribution reflects the same overall data characteristics.

The Box Muller transformation generates random pairs of independent, standard, normally distributed (zero mean, unit variance) random numbers given a source of uniformly distributed numbers. Box Muller Formulas use two means, '$\mu 1$'and '$\mu 2$', in the interval (0,1), in which the Standard Deviation and Normal Variables are 'Z_0' and 'Z_1',

Where:

$$Z_o = \sqrt{-2 \ln\ (\mu 1)} \times \cos(2.\pi.\mu 2)$$

$$Z_1 = \sqrt{-2 \ln\ (\mu 1)} \times \sin(2.\pi.\mu 2)$$

To scale these to the formula to a normal distribution, we scale the PERT mean and Standard Deviation to those of the Box Muller Approach:

$$X_o = Z_0 \times \sigma + \mu$$

$$X_1 = Z_1 \times \sigma + \mu$$

➢ *The following Methods should be Followed to Conduct the Box Muller Method:*

- Generate '$\mu 1$'and '$\mu 2$'two uniformly distributed random numbers: '$\mu 1$' and '$\mu 2$' are generated using 'Math.random()', which produces values between 0 (inclusive) and 1 (exclusive).
- Apply the Box-Muller Transform to Calculate z0 and z1, where the formula for the Box-Muller transform is as follows:

$$Z_o = \sqrt{-2 \ln(\mu 1)} \times \cos(2.\pi.\mu 2)$$

$$Z_1 = \sqrt{-2 \ln(\mu 1)} \times \sin(2.\pi.\mu 2)$$

This formula shows that the values ('z0' and 'z1') are normally distributed with a mean of 0 and a standard deviation of 1. Scale Z_0 (μ)and Z_1 (σ) to shift to the desired distribution using the PERT mean and standard deviation. '$\mu 1$' and '$\mu 2$' are the uniform random numbers. To transform `z0` to have the desired mean (μ) and standard deviation (σ), we use:

$$X_o = Z_0 \times \sigma + \mu$$

$$X_1 = Z_1 \times \sigma + \mu$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

$$\text{randomValue} = \text{z0} \times \text{stddev} + \text{mean}$$

'0' is the normally distributed random number with a mean of 0 and a standard deviation of 1.

'Z_0' * stddev + mean' scales 'Z_0' to have the desired mean and standard deviation.

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

- *Example:*

Assume we have $\mu 1$ = 0.5 and $\mu 2 = 0.7$. To calculate the z values, we will use the formula above from the Box-Muller transform method. First, we calculate Zo and Z1:

$Z_o = \sqrt{-2 \ln 0.5} \times \cos(2.\pi.0.5) \approx -0.30$

$Z_1 = \sqrt{-2 \ln(0.7)} \times \sin(2.\pi.0.7)\ ) \approx 1.11$

Next, we will transform Zo and Z1 to match our mean and standard deviation:

X_0 = 21.67 + 5 X (-0.364) ≈ 21.67-1.82 ≈ 19.85 days

X_1 = 21.67 +5x (-1.119) ≈ 21.67-5.595 ≈ 16.075 days

To find the 90th percentile using Box-Muller, we need to find the value of X such that 90% of the values are below it. The 90th percentile of the standard normal distribution is approximately 1.28 (Z-value). Transforming this to our distribution:

$X_{90} = \mu + \sigma \times Z_{90}$

X_{90} = 21.67 +5 X 1.28 21.67 +6.4

X_{90} = 28.07 days

Using Box-Muller, we transformed uniformly distributed random numbers into normally distributed ones and found the 90th percentile by transforming the Z-value 1.28 to our distribution. The example values are 19.85 days, 16.075 days, and 28.07 days (90th percentile). Using the Gaussian Formula, we calculated the probability density for a specific value and found the 90th percentile similarly by transforming the Z-value 1.28 to our distribution.

Our probability density examples were 25 days: 0.0637 and 90th percentile: 28.07 days. Both methods give us the same 90th percentile value when using the same mean and standard deviation derived from PERT.

- *Pros and Cons of Box-Muller Transform*

The Box-Muller transform has both advantages and disadvantages. One significant advantage is that it generates samples from an exact normal distribution, making it useful for simulations requiring normally distributed random variables. It is also relatively simple to implement, does not require complex mathematical functions, and is widely used, ensuring broad understanding and support within the field.

However, there are some drawbacks. The method requires the computation of logarithms and trigonometric functions, which can be computationally expensive compared to other methods. Additionally, the Box-Muller transform generates pairs of random variables, which might be unnecessary for applications requiring only one random variable. Furthermore, uniformly distributed random numbers are required as input, adding another step if such numbers are not readily available.

- *PERT-Informed Gaussian Distribution and 90th Percentile*

Using PERT estimates to inform a Gaussian distribution allows for directly applying the mean and standard deviation in the Gaussian probability density function. The 90th percentile confidence level remains the same from the initial PERT estimates through the Gaussian distribution because the mean and standard deviation align closely. The Gaussian distribution is symmetric around the mean, and since the mean and standard deviation are derived directly from the PERT estimates, the 90th percentile, which is a function of these parameters, remains consistent. The mode will differ as it reflects the symmetry of the Gaussian distribution. Using the PERT mean and standard deviation in Gaussian Distribution, which is defined by its probability density function (PDF):

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

Where Mean (μ) and Standard Deviation (σ).

This is evident through the processes above; in this case, the PERT Mean and Standard Deviation are supplied directly to the Gaussian formula. The following are some of the benefits and drawbacks of utilizing the Gaussian Distribution to convert non-normal data into normal data.

- *Pros and Cons of the Gaussian (Normal) Distribution Formula*

The Gaussian (Normal) distribution formula has several advantages and disadvantages. (P. Bertsekas & N. Tsitsiklis., 2002, p. 140) Among the pros, it features an exact analytical form, which is beneficial for theoretical analysis and applications requiring exact probabilities. It is widely applicable due to the central limit theorem, which states that the sum of a large number of random variables will be approximately normally distributed. (P. Bertsekas & N. Tsitsiklis., 2002, p. 140)Additionally, the properties of the Gaussian distribution are well-known and extensively studied, making it a reliable choice for many applications.

However, there are also cons to consider. One significant drawback is the assumption of normality, as many real-world phenomena are not normally distributed. Relying on the Gaussian distribution can lead to incorrect conclusions if the data does not fit this model. It is also not suitable for all

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

data, particularly for data with significant skewness or kurtosis, which requires alternative distributions for accurate modeling. Furthermore, parameter estimation, such as estimating the mean and standard deviation from data, requires a sufficient sample size and can be influenced by outliers.

➢ *PERT-Informed Gaussian for Monte Carlo Simulation and 90th Percentile*

Monte Carlo simulation, combined with a Gaussian distribution shaped by PERT estimates, involves generating numerous random samples from the normal distribution. This method captures the inherent variability in task durations and provides a distribution of possible outcomes. The 90th percentile confidence level remains the same due to the Central Limit Theorem, which ensures that the mean and standard deviation of the distribution of sample means will approximate the mean and standard deviation of the population from which the samples are drawn.

This means that the 90th percentile, based on the mean and standard deviation, will remain stable. However, the model is refined through empirical data. The PERT mean and standard deviation define the distribution of task durations in the Monte Carlo simulation. Typically, a normal (Gaussian) distribution is assumed, as the Central Limit Theorem suggests that the distribution of sample means tends to be normal with a sufficiently large sample size.

Monte Carlo simulation generates many random samples from the specified normal distribution, with the mean and standard deviation derived from the PERT estimates. Each simulation represents a possible outcome or realization of task durations. Monte Carlo simulation handles uncertainty by using random sampling to generate a range of possible outcomes based on defined probability distributions for input variables.

When applied to project estimation, Monte Carlo simulation repeatedly samples task durations from their probability distributions, often derived from PERT's expected time and standard deviation. Each simulation iteration generates a possible project duration based on these sampled values, creating a distribution of potential outcomes. This distribution provides insights into the likelihood of different project durations, enabling a more nuanced understanding of risk and uncertainty.

Monte Carlo simulation samples from a normal distribution shaped by the PERT mean and standard deviation. The range of values typically spans from the mean minus three standard deviations to the mean plus three standard deviations, covering approximately 99.7% of possible outcomes. This method captures the inherent variability in task durations and comprehensively analyzes potential project timelines.

Monte Carlo simulation iteratively generates numerous random samples (iterations) from the normal distribution. Each sample represents a possible outcome or realization of task duration, considering both the variability defined by the normal distribution and any additional uncertainty introduced through the simulation process. While the PERT estimates define the shape of the normal distribution, Monte Carlo simulation allows for the incorporation of additional uncertainty or variability beyond what is explicitly defined by these estimates.

This additional uncertainty can be represented by introducing randomness into the simulation process, such as sampling from distributions with broader ranges or different shapes, to account for unforeseen factors. The benefit of utilizing a Gaussian distribution in Monte Carlo Simulation is that it allows PERT to be properly utilized. The following are some of the key pros and cons of our approach:

➢ *Pros and Cons of Using PERT-Informed Gaussian for Monte Carlo Simulation*

The Monte Carlo simulation, when combined with PERT estimates to shape the Gaussian distribution, offers several advantages. Firstly, it provides improved accuracy, generating more realistic project duration estimates by capturing the inherent variability in task durations and offering a detailed analysis of potential project timelines (Kerzner, 2017). Secondly, it enhances risk management by incorporating additional uncertainty beyond the PERT estimates, which provides a more comprehensive picture of project risks, aiding better-informed decision-making and contingency planning (Montgomery & Runger, 2018).

Thirdly, the flexibility of Monte Carlo simulation allows it to account for many possible outcomes, making it adaptable to various project scenarios and complexities. This adaptability makes it a versatile tool in project management, capable of handling diverse project tasks and timelines (Karabulut, 2017). Additionally, the simulation offers insightful statistical measures, such as percentiles, that help project managers understand the likelihood of different outcomes.

This detailed insight aids in setting realistic expectations and planning buffers, allowing project managers to better gauge the probability of meeting specific project milestones and deadlines (Musa & Okumoto, 1989). However, there are also cons associated with this approach. The combination of PERT with Monte Carlo simulation introduces complexity in terms of calculations and interpretation, requiring project managers to understand both methods to use the combined approach effectively.

This complexity can be a barrier to adoption, particularly for teams lacking expertise in statistical analysis (Meredith & Mantel, 2019). Additionally, Monte Carlo simulation requires significant computational resources to run numerous iterations and generate a comprehensive range of outcomes, making it resource-intensive and time-consuming (Montgomery & Runger, 2018). Another downside is its dependency on the quality and accuracy of the initial PERT estimates.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

Inaccurate or biased input data can lead to misleading results, emphasizing the importance of gathering reliable data and leveraging expert judgment for the initial estimates (Kerzner, 2017). Finally, while the Central Limit Theorem supports using a normal distribution for large sample sizes, this assumption may not always hold true for all projects.

If the actual distribution of task durations deviates significantly from normality, it can lead to potential inaccuracies. Validating the normality assumption in the context of specific project data is crucial to ensure reliable outcomes (Meredith & Mantel, 2019). By leveraging many possibilities, Monte Carlo simulation creates a comprehensive picture of potential project timelines, capturing the inherent variability and providing statistical measures (such as percentiles) to inform decision-making.

Hence, combining PERT with Monte Carlo simulation optimizes the handling of uncertainty by leveraging the strengths of both methods. Leveraging the initial estimates from a PERT analysis to inform a Monte Carlo simulation is a valid project management and risk analysis approach.

➢ *PERT-Informed Beta Distribution and 90th Percentile*

Shaping a Beta distribution using PERT estimates allows for a more flexible representation of skewed data. The Beta distribution, defined by shape parameters alpha and beta, can accommodate various data patterns. The 90th percentile confidence level remains the same from the initial PERT estimates through the Beta distribution because the shape parameters (alpha and beta) are calculated based on the mean and variance derived from the PERT estimates.

The mode, however, will differ and is calculated based on the Beta distribution parameters discussed above, offering a more accurate reflection of the most probable project duration. The PERT mean and standard deviation are used to shape a beta distribution, which is defined by its probability density function (PDF). In this case, the PERT Mean and Standard Deviation are supplied to the Beta distribution formula. The following are some of the benefits and drawbacks of utilizing the Beta Distribution for converting non-normal data.

➢ *Pros and Cons of the Beta Distribution Formula*

The Beta distribution offers several advantages due to its flexibility and adaptability. Firstly, it can accommodate various shapes, including skewed distributions, making it more adaptable to real-world data with non-symmetrical characteristics. This flexibility allows for more accurate modeling of project durations. Secondly, the Beta distribution features customizable parameters, α (alpha) and β (beta), which can be tailored to fit specific data patterns, enhancing the accuracy of estimates.

Lastly, the Beta distribution is defined over a finite interval [0, 1], which can be scaled to fit any range, providing a realistic representation of task durations that cannot be negative. However, there are also drawbacks to consider. Determining the parameters alpha and beta from PERT estimates can be complex and requires careful scaling and transformation, introducing complexity in parameter estimation. Additionally, the calculations involved in shaping a Beta distribution are more complex than those for a Gaussian distribution, requiring more computational resources.

This computational intensity can be a limitation in some scenarios. Moreover, the accuracy of the Beta distribution relies heavily on the quality of the PERT input estimates. Inaccurate or biased data can lead to misleading results, making the distribution sensitive to input estimates.

➢ *PERT-Informed Beta Distribution for Monte Carlo Simulation and 90th Percentile*

Combining PERT-informed Beta distribution with Monte Carlo simulation involves generating random samples from the Beta distribution based on PERT-derived parameters. This method effectively handles skewness and provides a comprehensive analysis of potential project timelines. The 90th percentile confidence level remains the same due to the Central Limit Theorem, which ensures that the mean and standard deviation of the distribution of sample means will approximate the mean and standard deviation of the population from which the samples are drawn.

This ensures that the 90th percentile, based on these parameters, remains stable. Empirical data further refines the model, providing a more nuanced understanding of potential project durations. The PERT mean and standard deviation define the Beta distribution of task durations in Monte Carlo simulation.

This approach is used to generate many random samples from the specified Beta distribution, providing a robust framework for duration estimation. Monte Carlo simulation handles uncertainty by using random sampling to generate a range of possible outcomes based on defined probability distributions for input variables. When applied to project estimation, Monte Carlo simulation repeatedly samples task durations from their probability distributions, often derived from PERT's expected time and standard deviation.

Each simulation iteration generates a possible project duration based on these sampled values, creating a distribution of potential outcomes. This distribution provides insights into the likelihood of different project durations, enabling a more nuanced understanding of risk and uncertainty.

➢ *Shaping the Beta Distribution with PERT Parameters*

Monte Carlo simulation samples from a Beta distribution shaped by the PERT mean and standard deviation. The range of values typically spans from the best-case to the worst-case scenarios, capturing approximately all possible outcomes. This method comprehensively analyzes potential project timelines, considering inherent variability.
Iterative Simulation

Monte Carlo simulation iteratively generates numerous random samples (iterations) from the Beta distribution. Each sample represents a possible outcome or realization of task duration, considering both the variability defined by the Beta distribution and any additional uncertainty introduced through the simulation process.

➢ *Incorporating Additional Uncertainty*

While the PERT estimates define the shape of the Beta distribution, Monte Carlo simulation allows for the incorporation of additional uncertainty or variability beyond what is explicitly defined by these estimates. This additional uncertainty can be represented by introducing randomness into the simulation process, such as sampling from distributions with broader ranges or different shapes, to account for unforeseen factors. Using PERT-informed Beta for Monte Carlo simulation offers several significant advantages.

➢ *Pros and Cons of Using PERT-Informed Beta for Monte Carlo Simulation*

Firstly, the Beta distribution's flexibility allows for more accurate modeling of task durations, especially when they are not symmetrically distributed. This combined approach captures the inherent variability in task durations and provides a detailed analysis of potential project timelines, leading to improved accuracy. Secondly, the Monte Carlo simulation incorporates the beta distribution, which provides a comprehensive picture of project risks, allowing for better-informed decision-making and contingency planning.

This enhanced risk management helps project managers prepare more effectively for potential project delays and cost overruns. Additionally, the Beta distribution can model various possible outcomes, making it adaptable to various project scenarios and complexities. This flexibility is particularly useful in handling diverse project tasks and timelines.

Furthermore, the simulation offers valuable statistical measures, such as percentiles, that help project managers understand the likelihood of different outcomes. This detailed insight aids in setting realistic expectations and planning buffers, providing insightful statistical measures**.** However, there are also several drawbacks to consider.

Combining PERT with Monte Carlo simulation using the Beta distribution introduces complexity in terms of calculations and interpretation. Project managers need to understand both methods to use the combined approach effectively, which can be a barrier to adoption. Additionally, Monte Carlo simulation requires significant computational power to run numerous iterations and generate a comprehensive range of outcomes. This can be resource-intensive and time-consuming, posing challenges related to computational resources.

The accuracy of the simulation depends heavily on the quality and accuracy of the initial PERT estimates. Inaccurate or biased input data can lead to misleading results, highlighting a dependency on data quality. Lastly, estimating the parameters for the Beta distribution requires a careful scaling of PERT estimates, which can be complex and prone to error if not done correctly, presenting challenges in parameter estimation.

By leveraging many possibilities, Monte Carlo simulation creates a comprehensive picture of potential project timelines, capturing the inherent variability and providing statistical measures (such as percentiles) to inform decision-making. Combining PERT with Monte Carlo simulation using Beta distribution optimizes the handling of uncertainty by leveraging the strengths of both methods. This approach provides a sophisticated and realistic framework for managing uncertainty in project timelines.

➢ *Using the 90th Percentile and Mode*

The 90th percentile is a crucial metric for project managers because it captures project uncertainties while aligning with stakeholders' risk tolerance levels. It represents a value below which 90% of the data falls, providing a realistic view of project timelines that balances ambition with prudence. By focusing on the 90th percentile, project managers adopt a pragmatic benchmark that acknowledges both the best-case and worst-case scenarios. The 90th percentile confidence level remains consistent with the initial PERT estimates for the PERT-shaped distribution, even after Monte Carlo sampling of the PERT-shaped distribution. This stability is due to the fact that PERT-derived distributions preserve the mean and standard deviation of the initial estimates.

These statistical properties ensure that the 90th percentile, which depends on the mean and standard deviation, remains unchanged throughout the transformations. For instance, whether using a Gaussian or Beta distribution and even when applying Monte Carlo simulations, the 90th percentile confidence level stays the same due to the preservation of these fundamental statistical characteristics.

➢ *Importance of the Mode*

The mode, which represents the most frequently occurring value in a distribution, is critical for pinpointing the exact duration most likely to occur based on historical data or expert judgment. While the mean provides an average estimate and the median gives a middle value, the mode offers a precise indication of the most probable outcome. If the mode significantly deviates from the mean, it may indicate higher variability or risk in the project duration. A mode closer to the optimistic or pessimistic estimates suggests a skewed distribution, highlighting potential biases or risks in the estimates.

➢ *Why the Mode Changes*

Through the various methods and distributions applied, the mean and standard deviation typically align closely with the initial PERT estimates, maintaining a stable 90th percentile. However, the mode can change significantly. In the Beta distribution and its subsequent Monte Carlo simulations, the mode is not merely the "most likely" value, as in the triangular distribution.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

Instead, it reflects empirical outcomes more accurately. This refinement through Monte Carlo simulations enhances the precision of the mode, capturing real-world complexities and uncertainties more effectively. By leveraging both the 90th percentile and the mode, project managers gain a comprehensive understanding of project durations.

- *Refining the 90th Percentile and Mode*

The 90th percentile offers a conservative estimate that accounts for potential delays and uncertainties, while the mode provides the most likely project duration. Combining these two metrics allows for a balanced view that integrates optimism with caution, leading to more informed and realistic project planning and decision-making.

- *Monte Carlo – 90th Percentile and Mode*

The Monte Carlo simulation method captures a significant portion of project uncertainties and provides a realistic view of project timelines. By maintaining the 90th percentile confidence level and refining the mode through empirical data, Monte Carlo simulations enhance the robustness of project duration estimates. This dual perspective ensures that project planning and execution are informed by both theoretical models and real-world data, leading to more accurate and reliable estimates.

- *All Methods Considered*

Fusing PERT to shape distributions and then using Monte Carlo sampling from that shaped distribution to arrive at an empirical mode significantly enhances the precision and reliability of project duration estimations. This approach leverages the strengths of both methods to offer more accurate and nuanced insights into project timelines. Using PERT to shape different distributions, followed by Monte Carlo sampling, demonstrates that the 90th percentile confidence level remains consistent with the initial PERT estimates through the PERT-shaped distribution and into the distribution derived from Monte Carlo sampling.

This consistency is due to the preservation of underlying statistical properties such as the mean and standard deviation. For instance, the Box-Muller transform retains the statistical characteristics of the PERT estimates, ensuring the transformed distribution reflects the same overall data properties. Additionally, the Central Limit Theorem ensures that the empirical distribution approximates the true underlying distribution with enough simulations.

However, while the 90th percentile remains consistent, the mode often changes between the theoretical and empirical distributions. The mode, which represents the most frequently occurring value, is critical because it highlights the most likely project duration based on the data. In the Beta distribution, the mode is a theoretical estimate, whereas, in the Monte Carlo simulations, the mode reflects empirical outcomes.

This empirical mode can provide deeper insights into potential project durations, capturing real-world variability and complexities that theoretical models might miss. Considering the mode alongside the 90th percentile allows project managers to balance optimism and caution, providing a more robust framework for estimating project durations and planning resources effectively. Considering the PERT distribution approach, the method calculates the mean by heavily weighting the most likely estimate.

This provides a balanced approach with a realistic standard deviation reflecting the actual spread of data. However, it does not account for the full range of possible variability as Monte Carlo simulations do. For the Triangular Distribution shaped by PERT, this approach uses best case, most likely, and worst case to create a distribution with the most likely value at the peak. It is intuitive and easy to apply, giving some weight to the most likely estimate.

However, it is less accurate as it treats the range with equal weight on tails, potentially less realistic than the PERT method. The Normal Distribution shaped by PERT assumes a normal distribution using the most likely value as the mean, requiring a standard deviation. This method is familiar and straightforward, assuming a symmetrical distribution.

However, it may not accurately represent skewed data and requires additional information like the standard deviation. The Beta Distribution shaped by PERT uses shape parameters α (alpha) and β (beta) derived from best-case, most likely, and worst-case estimates. This allows for skewness and flexibility, representing a more realistic distribution.

The complexity in calculation and implementation is a downside. In the PERT Shaped Normal Distribution + Monte Carlo approach, PERT estimates are used to derive a normal distribution, followed by Monte Carlo simulations. This method is familiar and straightforward if data follows a normal distribution but assumes a symmetrical distribution, which may not capture skewness. Finally, the PERT Shaped Beta Distribution + Monte Carlo approach uses PERT estimates to derive shape parameters α (alpha) and β (beta) for a beta distribution, followed by Monte Carlo simulations. This method accurately represents skewed data, provides realistic estimates, and handles complexity and non-linearity effectively. However, it is computationally intensive and requires an understanding of beta distribution and Monte Carlo simulations. Overall, integrating PERT with Monte Carlo simulations, particularly through a beta distribution, provides a robust method for project duration estimation, accommodating variability and skewness while ensuring reliable and informed decision-making.

- *Final Approaches Selected*

Our methodology explored various distributions and how PERT can shape them, progressing to Monte Carlo sampling for enhanced estimation accuracy. Notably, there is no significant difference between PERT-informed Gaussian and PERT with Box-Muller Transform for our purposes, as both methods generate normal distribution samples efficiently. Our evaluation identified two robust approaches:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

- "*PERT Shaped Normal Distribution + Monte Carlo" and "PERT Shaped Beta Distribution + Monte Carlo.*"

These methods were chosen for their ability to balance simplicity with accuracy, providing realistic estimates that account for variability and skewness—crucial factors for effective project management. The PERT Shaped Normal Distribution + Monte Carlo approach assumes symmetry in the data, as represented by the Gaussian (normal) distribution. However, this assumption makes it less suitable for capturing skewed data accurately.

When PERT estimates indicate skewness, the Gaussian distribution might not reflect the true variability. On the other hand, the PERT Shaped Beta Distribution + Monte Carlo approach accommodates skewness effectively. Using shape parameters alpha (α) and beta (β) derived from PERT estimates, the Beta distribution offers a more accurate representation of the inherent skewness in the data.

Regarding the calculation and interpretation of standard deviation, the PERT Shaped Normal Distribution + Monte Carlo method uses the standard deviation derived from PERT estimates. However, since the Gaussian distribution assumes symmetry, it might not align with the actual data distribution, especially in skewed cases. Conversely, the PERT Shaped Beta Distribution + Monte Carlo method derives standard deviation from realistic PERT estimates (best case, most likely, worst case), accurately reflecting the actual data spread and contributing to a more precise representation of the data's variability.

In terms of robustness and reliability, Monte Carlo simulations on a Gaussian distribution may overlook significant characteristics of the data's distribution, such as skewness and heavy tails, due to the assumption of symmetry. In contrast, Monte Carlo simulations on the Beta distribution capture a wider range of possible outcomes, including skewness, resulting in a distribution more representative of real-world data. This provides a more reliable 90th percentile confidence level.

Bias reduction is another crucial aspect where these methods differ.

The Gaussian distribution's assumption of symmetry may not mitigate all biases, particularly if the underlying data is skewed, potentially leading to over- or under-estimation of the 90th percentile confidence level. The Beta distribution, however, minimizes the influence of extreme values by appropriately weighting the most likely estimate and accurately reflecting the data spread through its shape parameters, thereby reducing bias in the resulting distribution. From a practical application standpoint, the Normal Distribution Shaped by PERT is easier to implement and familiar to many practitioners.

However, it might not offer the same accuracy in representing skewed data, especially when the data does not follow a symmetric pattern. The PERT Shaped Beta Distribution + Monte Carlo method, though more complex to implement, provides valuable flexibility in representing skewed data and delivering realistic estimates. This makes it particularly useful for project management scenarios where an accurate representation of variability is crucial.

The 90th percentile confidence level remains consistent between the Beta distribution and the Monte Carlo histogram because they are derived from the same underlying Beta distribution parameters. This percentile represents the value below, at which 90% of the data falls, and remains a critical measure for understanding potential project duration. However, the mode, representing the most frequently occurring value, can differ between the theoretical Beta distribution and the Monte Carlo simulation results.

The mode from the Beta distribution provides a theoretical estimate, while the mode from the Monte Carlo simulation reflects empirical data from numerous simulated outcomes. The Central Limit Theorem supports that with enough simulations, the empirical distribution approximates the true underlying distribution. A higher empirical (Monte Carlo) mode may indicate a longer time to complete the project, capturing real-world variability more accurately.

Conversely, a lower empirical mode suggests faster completion, indicating that the project might proceed more quickly than anticipated. Even though the 90th percentile confidence interval remains consistent across both distributions, it plays a crucial role in project management by providing a conservative estimate that accounts for potential risks and uncertainties. This measure ensures a 90% confidence level that the project will not exceed this duration, offering a buffer against unforeseen delays.

Additionally, considering the mode provides further insights into the most probable outcomes. The mode helps to pinpoint the most likely project duration, offering a more refined understanding of expected timelines. This dual perspective—incorporating both the 90th percentile for risk management and the mode for pinpointing the most probable duration—enables a more nuanced and accurate approach to project planning and execution.

By integrating PERT with Monte Carlo simulations, project managers gain a dual perspective that combines theoretical models with an empirical representation of real-world data, leading to more accurate and reliable project duration estimates. This method allows for a comprehensive understanding of potential project outcomes, improving decision-making and project planning. Our estimation process begins with a simple three-point estimate consisting of the best-case, most likely, and worst-case scenarios.

These initial estimates provide a basic framework for the potential outcomes. To refine these estimates, we apply the PERT method, which gives more weight to the most likely outcome. This weighted approach provides a more balanced mean estimate by considering the likelihood of various scenarios. Next, we model these PERT estimates using a Beta distribution.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

This step accounts for any skewness in the data, allowing us to better represent the uncertainty and variability in the estimates. To refine our estimates further, we will perform a Monte Carlo simulation on the beta distribution. By running numerous simulations, we generate a large number of random samples.

This approach leverages the Central Limit Theorem, providing an empirical, real-world simulation of the potential outcomes. From the Monte Carlo simulation results, we calculate the mode (the most frequently occurring value) and the 90th percentile. The mode gives us an idea of the most likely duration based on the simulations, while the 90th percentile provides a high-confidence estimate that accounts for potential uncertainties.

Also, it is important to note that the mode represents the most frequently occurring value in a distribution, making it a crucial measure for identifying a project's most probable completion time. This statistical measure aids in setting realistic timelines and expectations, ensuring that project plans are grounded in the most likely outcomes. However, the mode can change based on the shape of the distribution and the influence of the optimistic, most likely, and pessimistic estimates.

Monte Carlo simulations, which account for variability and uncertainty, often result in a different mode than initial estimates, providing a more accurate reflection of potential project durations. In the context of project duration estimates, the mode should not be negative; accurate and reasonable estimates inherently prevent such anomalies.

The 90th percentile is another critical measure that complements the mode. It represents the value below, at which 90% of the observations fall, offering a high confidence interval for project completion times. By considering both the mode and the 90th percentile, project managers can develop a robust understanding of potential project outcomes. The mode provides the most likely completion time, while the 90th percentile offers a buffer for uncertainty, ensuring that timelines are not overly optimistic.

This combination helps identify and mitigate risks, set realistic schedules, and communicate effectively with stakeholders. For instance, if the mode indicates a completion time of 29.68 days and the 90th percentile suggests 31.69 days, project managers can confidently plan for a duration within this range, accounting for most potential delays and ensuring a higher probability of meeting deadlines. This integrated approach leads to more accurate and reliable project estimations, ultimately benefiting overall project management.

This step-by-step progression—from simple estimates to PERT, Beta distribution, and Monte Carlo simulation—allows us to refine our initial estimates into a statistically informed recommendation. The final recommendation includes both the mode and the 90th percentile values, offering a comprehensive and reliable estimate for project planning and risk management. In project management, understanding the difference between the initial most likely estimate (mode) and the mode obtained after Monte Carlo simulations on a beta distribution is crucial for making informed decisions.

The initial most likely estimate is derived from expert judgment and represents the duration that is most likely to occur based on initial assessments. It is part of the three-point estimation (optimistic, most likely, and pessimistic) used in PERT analysis. After shaping the PERT estimates into a beta distribution, Monte Carlo simulations are performed to account for variability and uncertainty. These simulations generate a large number of potential outcomes for the duration of the project. The mode from these simulations is identified as the most frequently occurring value among the simulated results, providing a more empirically driven estimate that reflects real-world variability.

The difference between the initial most likely estimate and the mode after Monte Carlo simulations is calculated to quantify the most probable completion time shift due to the additional variability captured in the simulations. This difference is given by:

$$\textbf{Difference} = \textbf{Mode Monte Carlo} - \textbf{Most Likely Estimate}$$

If the mode after Monte Carlo simulations is higher than the initial most likely estimate, it indicates potential delays or underestimation of the project duration. This requires careful attention to risks and contingency planning. Conversely, suppose the mode after Monte Carlo simulations is lower than the initial most likely estimate. In that case, it suggests possible efficiencies or overestimates the project duration, allowing for potential resource reallocation or schedule optimization.

We calculate the percentage difference between the initial most likely estimate and the Monte Carlo mode:

$$\text{Difference Percentage} = \left(\frac{\text{Mode Monte Carlo} - \text{Most Likely Estimate}}{\text{Most Likely Estimate}}\right) \times 100$$

This comprehensive approach, progressing from simple estimates to PERT, Beta distribution, and Monte Carlo simulation, refines initial estimates into statistically informed recommendations. The final recommendation includes both the mode and the 90th percentile values, offering a reliable estimate for project planning and risk management. Now, let us consider the real implementation of our approach to evaluate its efficacy.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

III. REAL WORLD IMPLEMENTATION

➢ *Diagrams*

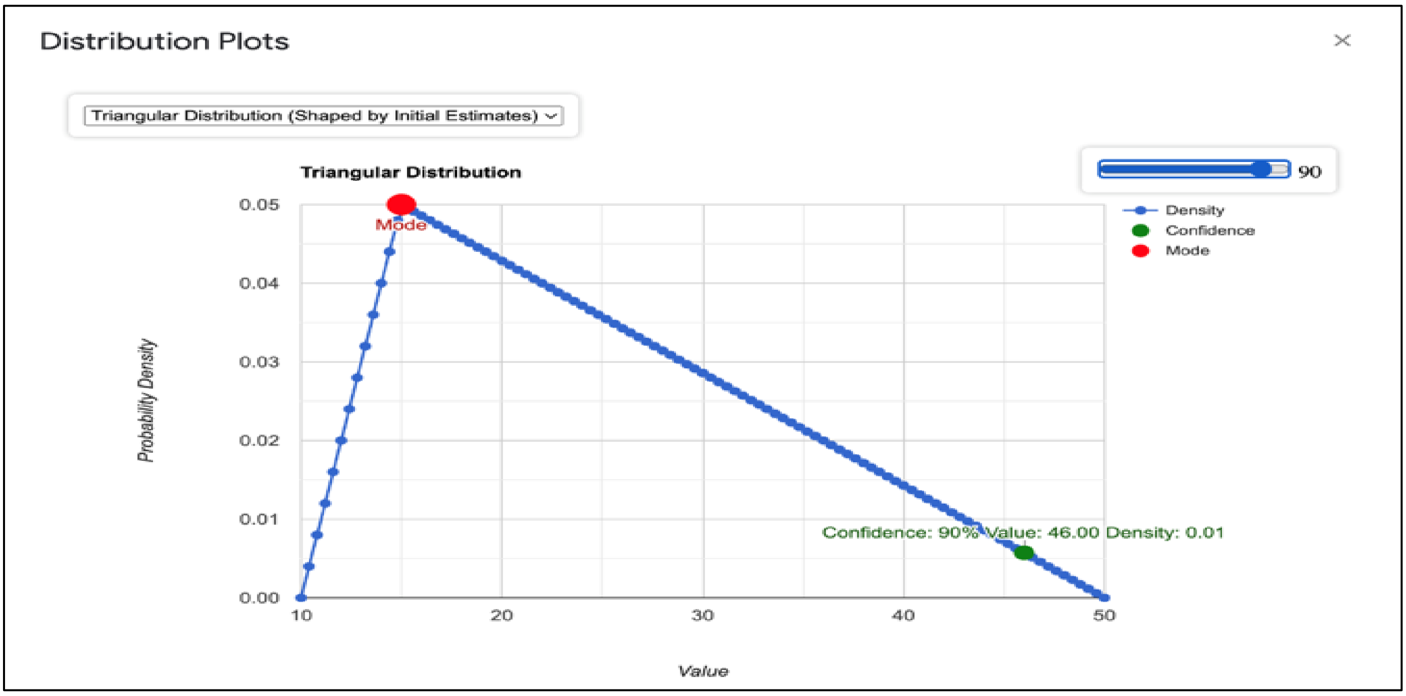

Fig 1 PERT Shaped Gaussian in Power B.I. with Co-Pilot Feature Enabled

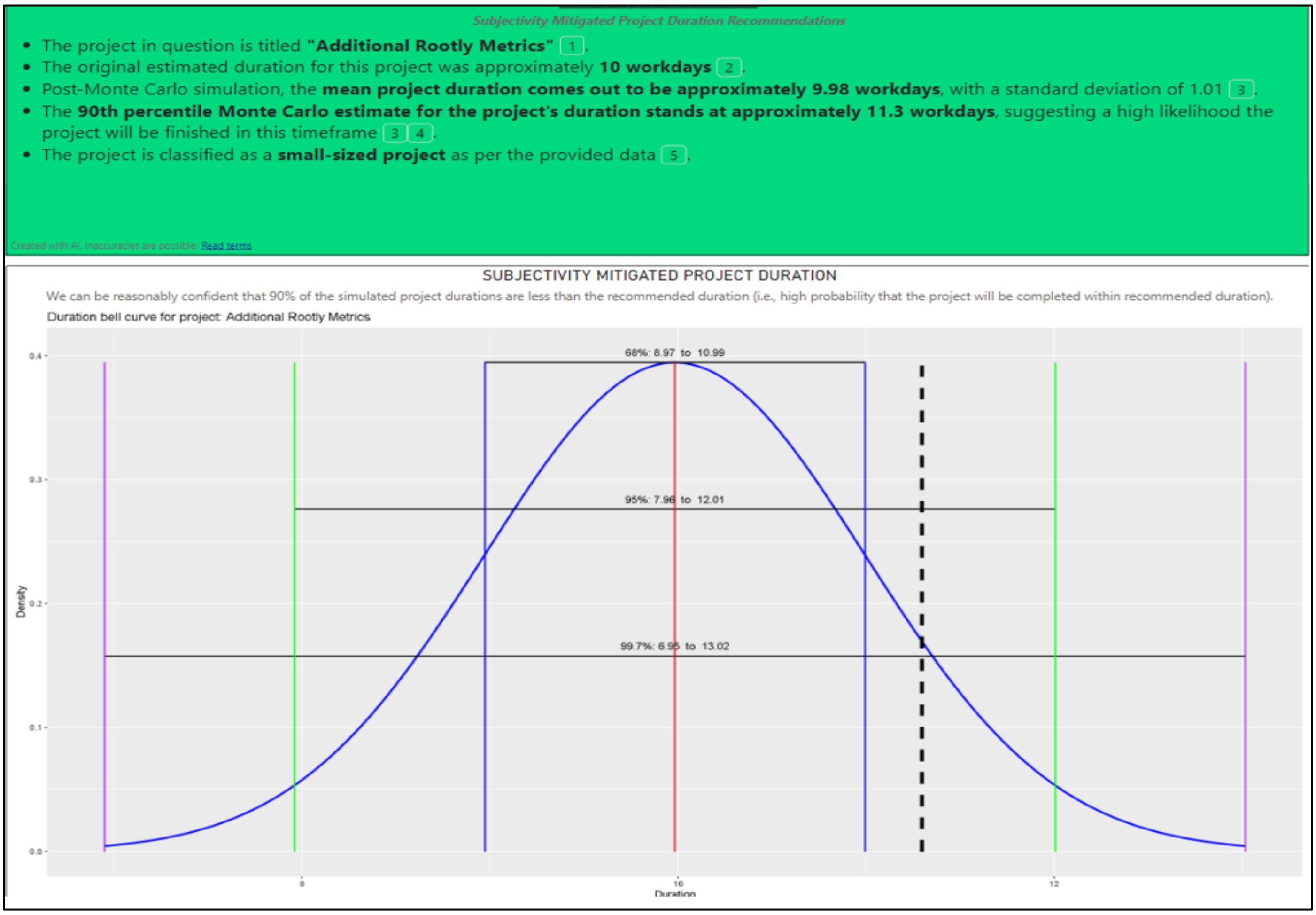

Fig 2 Sample of Our Approach in Power BI. with the Co-Pilot Feature

The screenshots above in Figures I and II show how our mathematical approach of PERT + Monte Carlo can be leveraged in Power BI. (using R script to do calculations) and also incorporates Power BI.'s co-pilot feature to auto-summarize the interpretation of the calculations in the context of projects. Having A.I. summarize such results dynamically opens a new world of continuous inquiry where the user can ask the AI to explain/interpret the results for self-serve understanding!

➢ *Google Sheet Plugin*

- We start with the Triangular distribution
- We use PERT to shape a distribution (either normal or beta), then sample from that with Monte Carlo to generate a histogram, and then smooth it.

✓ The mean, standard deviation, and 90th percentile remain the same before and after Monte Carlo.
✓ The mode, however, changes.
✓ Examples below:

➢ *Before Monte Carlo*

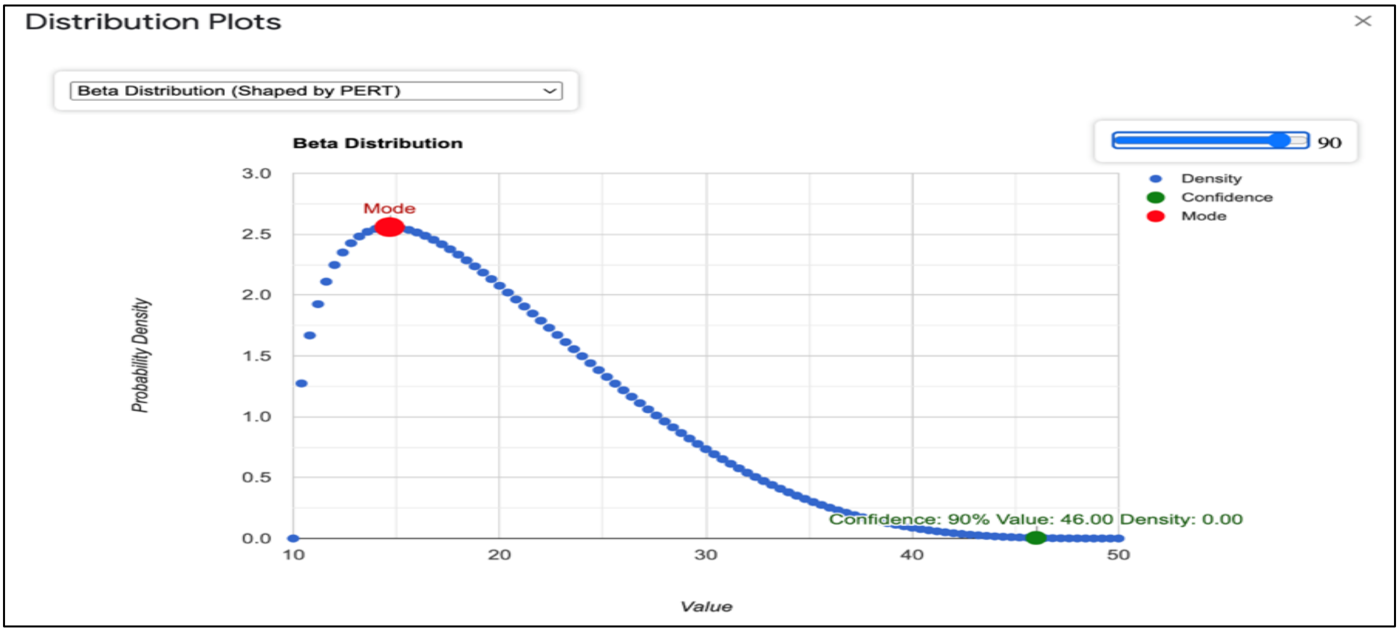

Fig 3 Distribution Plots before Monte-Carlo Method

➢ *After Monte Carlo*

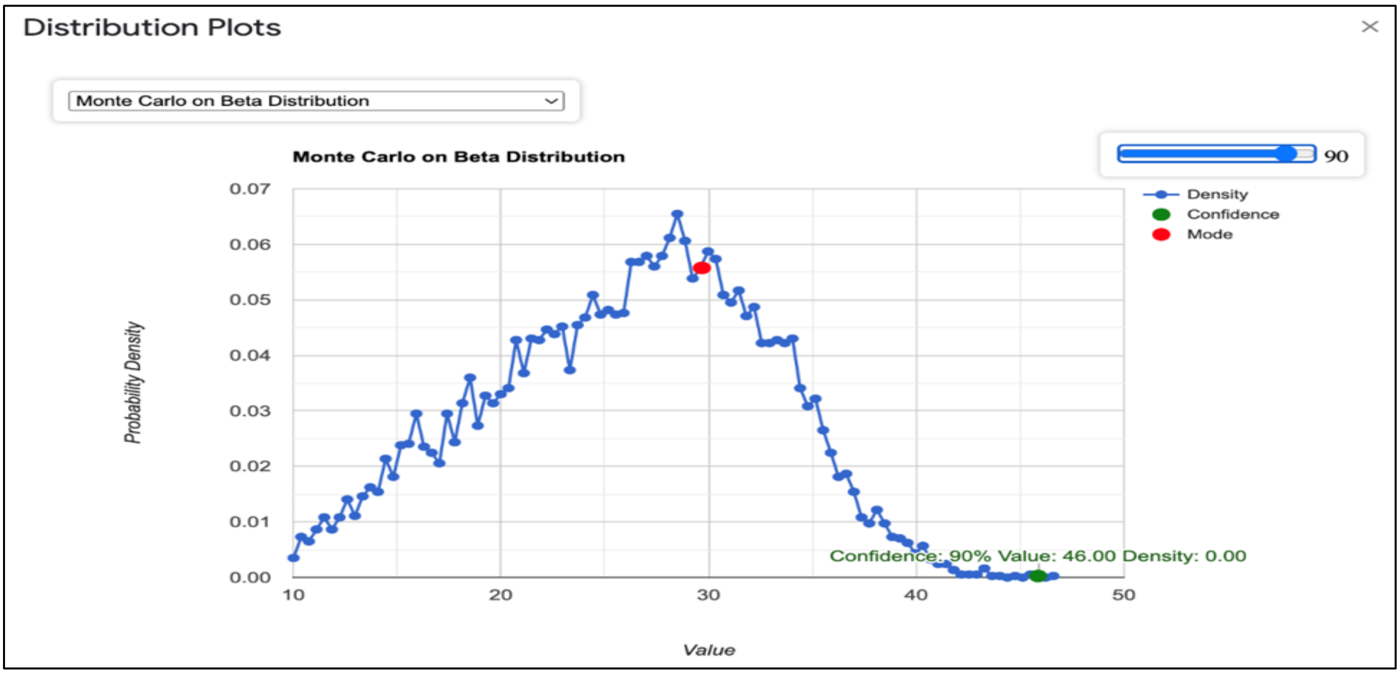

Fig 4 Google Sheets's Application of PERT + Beta Distribution + Monte Carlo

- Note: We have also created a video of how our solution works on our website.

The plugin solution is simple to follow and install on any computer with access to Google Sheets.

The video is available at the following website: https://icarenow.io/pmc-estimator-2.

The Monte Carlo simulation, before and after, as per Figures III and IV, generates a large number of random samples to estimate the distribution. The mode, shown as the red circle in the figure above, represents these samples' most frequently occurring value. However, the highest peak in the histogram may sometimes differ slightly from the actual mode due to how the samples are grouped into bins.

This is a normal variable in statistical visualization. The recommendation is that the simulations indicate that this item is likely to take around 29.68 days, which is 97.87% longer than the initial estimate of 15.00 days. The recommended value, therefore, we can assume is anywhere between 29.68 days (with a 66% confidence) and 31.69 days (with a 90% confidence).

IV. CONCLUSION

Our exploration demonstrates that integrating PERT and Monte Carlo simulation significantly advances project duration estimation beyond traditional methods. This approach balances theoretical models with empirical data, providing a comprehensive risk management and decision-making framework.

➢ *Our Methodology offers Flexibility by Accommodating Both Normal and Beta Distributions:*

- We use PERT values in the Gaussian equation for normal distributions, perform Monte Carlo simulations, and derive the 90th percentile and mode.
- For skewed data, we shape a Beta distribution using PERT values, run Monte Carlo simulations, and extract the 90th percentile and mode.

This dual approach allows project managers to capture variability and uncertainty in project tasks effectively. By leveraging initial PERT estimates to inform Monte Carlo simulation parameters, we model a larger sample size, resulting in a more nuanced understanding of potential project durations and enhancing the reliability of both 90th-percentile confidence estimations and the mode.

Implemented through accessible tools like Google Sheets and Power B.I., our methodology demonstrates the practical application of advanced statistical techniques in everyday project management. It empowers project managers to refine initial human estimates with statistical rigor, providing a powerful tool for accurate and reliable duration estimation.

Future work should explore "what-if" scenarios based on probabilistic outcomes for duration and cost estimates. This approach transforms duration estimation from a static exercise to a dynamic, data-driven process, ensuring decisions are grounded in statistical rigor.

To expand the methodology's applicability, exploring additional probabilistic models such as Poisson and Exponential distributions is essential for capturing distinct characteristics of different project tasks. Automating the reporting of calculation outcomes, as demonstrated in our screenshot of an R-version integration with Co-Pilot and Power BI, can provide added value through dynamic summarization of findings.

Emerging technologies like machine learning and artificial intelligence promise to further enhance project duration estimation. Algorithms that learn from historical data adapt to project dynamics and predict outcomes with unprecedented accuracy could offer corrective measures, paving the way for even more sophisticated and reliable project management tools.

In conclusion, we have demonstrated a successful framework for integrating PERT and Monte Carlo simulation to yield robust duration recommendations. This method offers a significant improvement in managing the uncertainties and complexities of modern project management, ultimately leading to more successful project outcomes. By embracing the inherent randomness in project management and grounding decisions in statistical rigor, project managers can make more informed and reliable recommendations, enhancing overall project success.

ACKNOWLEDGMENTS

➢ *I Favor Erwin Schrödinger's Sentiment:*
"***The task is not so much to see what no one has yet seen, but to think what nobody has thought about that which everybody sees.***"

I sincerely thank the Technical Program Management Leadership at LinkedIn, including but not limited to Gary Tauscher, Richard Chen, Shwetha Umesh, Pooja Singh, and Nidhi Mehta. They have all embodied a demanding management excellence standard, allowing room for exploration that has a practical, implementable impact. Many thanks to Daniel Tsang (MS in Statistics) for generously giving his time and statistical expertise to review this paper and drive clarity.

REFERENCES

[1]. Ballesteros-Pérez, P., Cerezo-Narváez, A., Otero-Mateo, M., Pastor-Fernández, A., Zhang, J., & Vanhoucke, M. (2020). Forecasting the project duration average and standard deviation from deterministic schedule information. *Applied Sciences, 10*(2), 654. DOI: 10.3390/app10020654

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP164

[2]. Broadleaf Capital International Pty Ltd. (2014). Creating value from uncertainty. In Broadleaf (pp. 1–7). Broadleaf Capital International Pty Ltd. https://broadleaf.com.au/wp-content/uploads/2014/07/Beta-PERT-origins-2014-v2.pdf
[3]. Datawithzon. (2023). How to connect Google Sheets with Power BI. *Medium.*
[4]. Enterprise D.N.A. (2023). Power B.I. with R and RStudio: How to get started.
[5]. Hernandez, F. (2021, January 5). Triangular Distribution vs Pert: Which is Best for Project Management? www.safran.com; Safran. https://www.safran.com/blog/triangular-distribution-vs-pert
[6]. Johnson, D. (1997). The Triangular Distribution as a Proxy for Beta Distribution in Risk Analysis. The Statistician, 46(1997), 387–398. https://doi.org/10.1111/1467-9884.00091
[7]. Karabulut, M. (2017). Application of Monte Carlo simulation and PERT/CPM techniques in the planning of construction projects: A case study. *Periodicals of Engineering and Natural Sciences*, *5*(3), 408-420.
[8]. Kerzner, H. (2017). *Project management: A systems approach to planning, scheduling, and controlling.* Wiley.
[9]. Kissell, R., & Poserina, J. (2017). Chapter 4 - Advanced Math and Statistics. In Optimal Sports Math, Statistics, and Fantasy (pp. 103–135). Elsevier. https://www.sciencedirect.com/science/article/abs/pii/B9780128051634000049
[10]. Malcolm, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W. (1959). Application of a technique for research and development program evaluation. *Operations Research*, *7*(5), 646-669.
[11]. Meredith, J. R., & Mantel, S. J. (2019). *Project management: A managerial approach.* Wiley.
[12]. Microsoft Fabric Community. (2023). How to connect Google Sheets to Power BI.
[13]. Musa, J. D., Iannino, A., & Okumoto, K. (1989). Software reliability: Measurement, prediction, application. *McGraw-Hill.*
[14]. Nieto-Rodriguez, A., & Vargas, R. V. (2023). How AI will transform project management. *Harvard Business Review.*
[15]. P. Bertsekas, D., & N. Tsitsiklis., J. (2002). Introduction to Probability (Online, 2002, p. 140). Athena Scientific. https://archive.org/details/introductiontopr0000bert/page/n7/mode/2up
[16]. Springer. (2023). Hybrid human-AI forecasting for task duration estimation in project management.
[17]. Theocharis, S. (2023). Dynamic AI project estimation. *Towards AI.*