

The Best of Both Worlds: Combining Parametric Cost Risk Analysis with Earned Value Management Using Bayesian Parameter Learning

Murray Cantor, Ph.D.

Cantor Consulting, LLC

Christian Smart, Ph.D., C.C.E.A.

Jet Propulsion Laboratory/California Institute of Technology

*I need everything the data can give me...
If we could have the best of both worlds
A little bit of heaven right here on Earth
(with apologies to Van Halen)*

Introduction

Cost risk analysis and earned value data are used separately and independently to compute Estimates at Completion (EAC). However, combining the two has the potential to significantly improve the accuracy of EAC forecasting. This paper provides a rigorous method using Bayesian methods. We also provide some examples that illustrate the strengths of this method.

In earned value management (EVM), EAC is a critical metric. It is used to forecast the effort's total work cost as it progresses. In particular, the EAC is used to see if the work is running over or under its planned budget, which is specified as the budget at completion (BAC).

Figure 1 visualizes this progress. Each graph displays a probability density function (PDF) of the EAC at the onset of the effort and after some activities have been completed. The green area is the probability that EAC will fall within the BAC, and the red area is the probability that it will not. The left chart shows the status of an effort's onset: The budget is fixed, and the EAC is uncertain. The right diagram shows the status after some updates of the actual costs (AC) reported as effort is completed. As the work progresses, the EAC uncertainty should decrease, and the likelihood of meeting the budget should increase. This is the situation shown in the figure. If the area of the green does not increase as the work progresses, the budget is in jeopardy, and some management action is warranted.

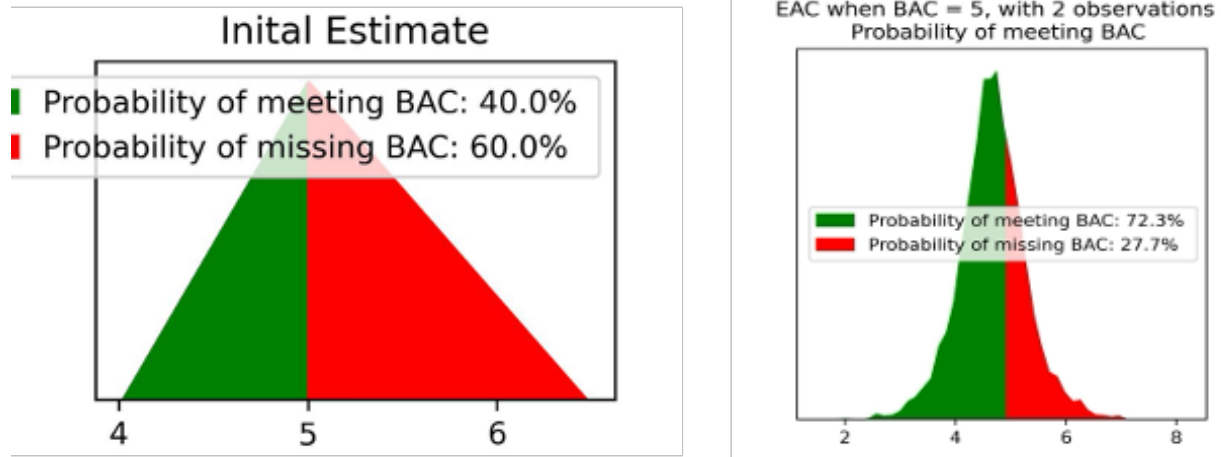


Figure 1

This paper will explain how to specify the initial PDF and learn the later PDFs from the data tracked in EVM. We describe the technique called Bayesian parameter learning (BPL). We chose this technique because it is the most robust for exploiting small sets of progress data and is easily used by practitioners.

There are several advantages to this approach:

- Provides more detail and insight than the conventional cost performance index (CPI) for estimating EAC.
- Leverages initial estimates, whether obtained by team consensus or sophisticated cost estimation techniques. (Smart, 2020)
- Updates the estimates using the observations of the actual spending rate.

Probabilistic cost estimates based on parametric models are often used early in a project's lifecycle, often up to the Critical Design Review milestone. After that point, earned value is often used for the EAC. In this paper, we show how to use these two key pieces of analysis in a unified and mathematically rigorous framework.

Why Bayes

The set of actual cost observations is generally very small. Bayesian analysis is based on applying the axioms of probability theory. It is more robust than frequency-based probability analysis because it applies to small data sets¹.

The Bayesian perspective is grounded in several fundamental principles:

- Prior knowledge, expert judgment, or beliefs are explicitly used. This is especially valuable in small data contexts, which is often the situation for NASA and Department of Defense projects. This approach allows the integration of expert opinion or historical data, enhancing the analysis when empirical data is sparse.
- Bayesian analysis treats all variables, including unknown parameters, as random variables, thereby explicitly including uncertainty. That is, a random variable's values are numerical outcomes of a random phenomenon, and each is associated with a probability density function (PDF). This PDF describes the random variable's probability of assuming specific values, allowing for a comprehensive expression of uncertainty and variability.

The power of the perspective follows from **Bayes' Theorem**. The theorem updates *prior* beliefs with new evidence to form *posterior* beliefs, reflecting a process of continuous learning from data.

¹ For a deeper discussion of the advantages of Bayesian methods, see (Clayton, 2021)

Overall, with its unique treatment of parameters as random variables accompanied by PDFs, the Bayesian approach is adept at evolving beliefs using data. This approach effectively provides meaningful insights to support decision-making when data is limited.

The perspective is appropriate in predicting the EAC:

- Given that the EAC is uncertain, we will treat it as a random variable and find its PDF, as shown in Figure 1.
- The estimation team has, or should have, a prior probabilistic estimate of the EAC. These are discussed below.
- A set of observations of (actual cost and percent complete) might be used in Bayes Theorem to update the prior as shown in the left-hand graph of Figure 1.

Bayesian Parameter Learning

We use the EVM information to apply the Bayes theorem to get a posterior PDF of the EAC. This is not entirely straightforward. For a variable of interest (Var) and a set of observations, the theorem states that it holds that the conditional probability, $P(\text{Var}|\text{Observations})$, is

$$P(\text{Var}|\text{observations}) = \frac{P(\text{Observations}|\text{Var})P(\text{Var})}{P(\text{Observations})}$$

Posterior
Likelihood
Prior
↓
↓
↓
 $P(\text{Var}|\text{observations})$
 $P(\text{Observations}|\text{Var})P(\text{Var})$
 $P(\text{Observations})$

↑

Marginal

To do Bayesian analysis, one needs:

1. A prior random variable.
2. A set of observations is used to iteratively apply Bayes theorem to get a set of posteriors.
3. A likelihood PDF for the observations.

Here, our observations are a sequence of observations provided by the cost center, i.e., the actual cost spent to date (AC) and completion percentage (%cost comp) of the effort. So, each observation is of the form $O_n = (AC_n, \%comp_n)$ where:

- AC_n = the actual cost of work performed (ACWP) to date
- $\%comp$ = percent of the work completed to date.

The challenge is to exploit the Bayes theorem to update the EAC using these observations. This is done by noting:

- We can convert the O_n sequence to a sequence of cost rates, cr , by taking the quotient of the observation elements so that $cr_n = AC_n / \%comp_n$. Note here that it is assumed that the percent complete is greater than zero after the project starts, so cr_n is well defined.
- The prior of the EAC is not only the original belief of the EAC but also the initial belief of the cost rate; that is, the prior is EAC at 100% completion, giving the slope of the EAC as a function. Note that the AC = 0 at 0% complete and is expected to be the EAC at 100% complete.

So, we have a prior and observations of the cost rate. We also need a likelihood². This is where Bayesian parameter learning comes into play. Without going too deep into the math, we use a logistic PDF with parameters μ and σ for the likelihood function. The logistic distribution is more suitable than the conventional choice of the

² We do not need the marginal as it can be inferred by using the constraint that the integral of the posterior must be 1.

normal distribution for modeling in our context. We often have very small sets of observations, and there is no reason they should be samples of independent, identically distributed processes (Wikipedia, 2023).

The parameters, mu and sigma, are uncertain, so they are treated as random variables with their own PDFs. These are learned using the Bayes Theorem. They are, in turn, used to obtain the posterior cost rate. As described in (Fenton & Neil, 2019), this technique is called *Bayesian parameter learning* (BPL).

In our implementation, we:

- Use Bayesian refinement, applying Bayes theorem iteratively for each observation and then using each of the posteriors as the priors for the next iteration. It is particularly useful in complex systems where the actual state of the world is uncertain and can only be estimated from incomplete or noisy data.
- Take a digitized approach (Kruschke, 2014). Here, the calculation is broken into three grids.
 - Prior grid: The joint probability of the priors. This is their product since they are independent.
 - Likelihood Grid: The likelihood of the observations given the priors. In our case, the logistic distribution for each pair of values of the digitized mu and sigma is applied.
 - Posterior Grid: The pointwise product of the prior and likelihood grids (ala Bayes theorem). The normalized margins give the density for the evolved mu and sigma.
- Use these PDFs to build the PDF of the cost rate using a Monte Carlo simulation.
- Learn the EAC by using the learned cost rate PDF, the sunk cost, and the remaining completion in a Monte Carlo simulation of:

$$EAC = Sunk\ Cost + CostRate * \%Remaining$$

With the posterior of the cost rate, we use it to compute the PDF of the EAC using a Monte Carlo simulation to do PDF arithmetic of the formula³.

Some readers may find the internal, intermediate PDFs interesting. An example of these is shown in Figure 2. The figure shows how the PDFs for the mu, sigma, and cost rate evolve over a sequence of Bayesian refinements. (Note that these data are used in the second example below.) Recall it is the cost rate PDF that is used for the EAC computation.

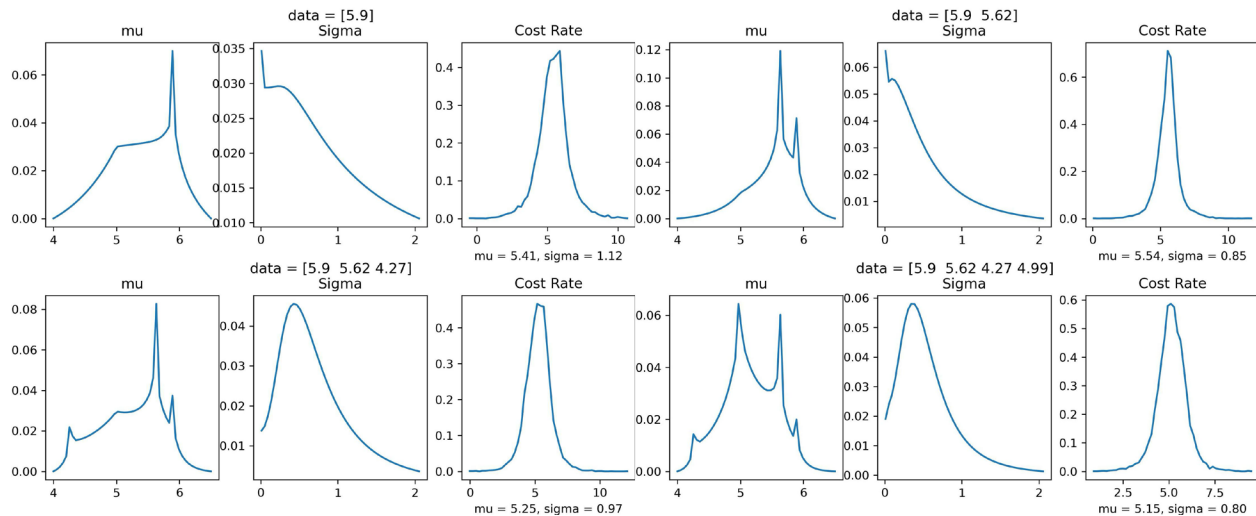


Figure 2

Note that these PDFs are sensible in that:

³ You can find the mathematical details in (Cantor, 2023).

- The mu PDF is multi-model and reflects the impact of the prior triangular by the Bayesian update using the data,
- The sigma PDF updates from the initial uniform peaks at zero for the first two similar data points and then changes shape to be mono-model and skewing, left with a dissimilar observation.
- Even with the strange shapes of the mu and sigma PDFs, the cost rate PDF is nicely mono-model and narrows with each observation.

Picking Priors

Bayesian parameter learning can be summed up as updating the expected EAC from an initial estimate, the so-called prior. There are two methods for setting the prior:

1. Using domain-based industry estimation tools to get a parameter-based distribution (Smart, 2020)
2. An empirical PDF based on industry studies.
3. Eliciting the subject matter experts' best-case, most likely, and worst-case values (Hubbard, 2014)⁴. The values can then specify a triangular distribution (Mun, 2003).

Each case yields probability distributions that are suitable priors. Methods 1 and 2 fully utilize estimation research and provide well-established distribution models. These are preferable when applicable. Otherwise, method 3 is a good substitute because it takes advantage of the experts' judgment and intuitions. It is commonly used for schedule estimation in more sophisticated project management tools.

Visualizing the Results

The output of the BPL is a PDF of the EAC. We are interested in taking action, so two views of the PDF are provided.

1. Target View: The probability of meeting or beating the target BAC.
2. Percent View: The percentiles of the EAC.

The first view gives a quick view of whether you will likely come in as planned. The second provides a more detailed view of how much you are off target.

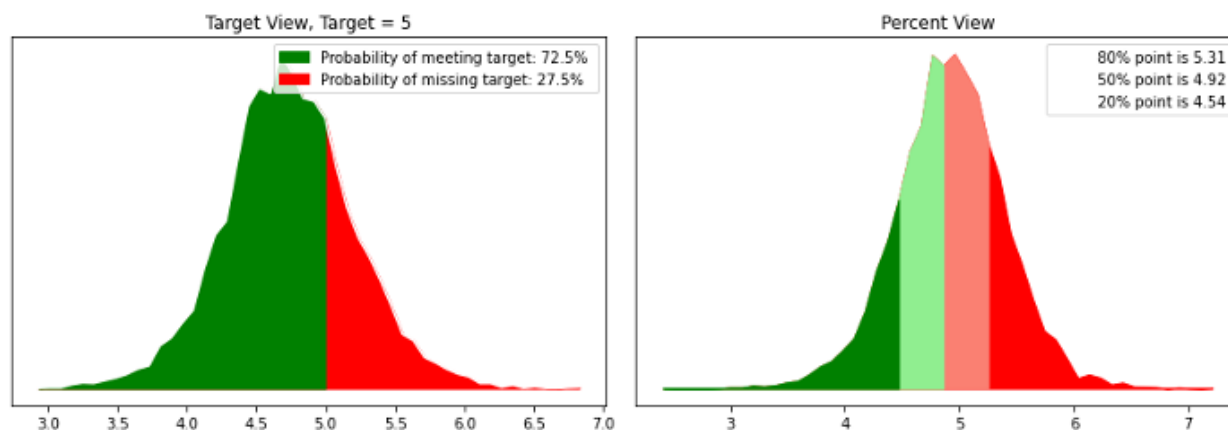


Figure 3

Examples

⁴ These are sometimes called Pareto estimates

We show two examples:

1. The “near-uniform spending” scenario is one in which spending goes close to the planned rate.
2. The “recovery” scenario involves spending initially at an above-budget rate but ultimately close to the planned budget.

In the examples, the BAC is five units (say \$M), but there is some uncertainty about the EAC, so the team used method 3 to provide a three-point estimate. To start the Bayesian learning, this is converted into a triangular distribution (Figure 4).

Note that from the choices, the model indicates the amount they might go above the budget is greater than the amount they might be under. This yields an asymmetric PDF with the mean larger than the mode.

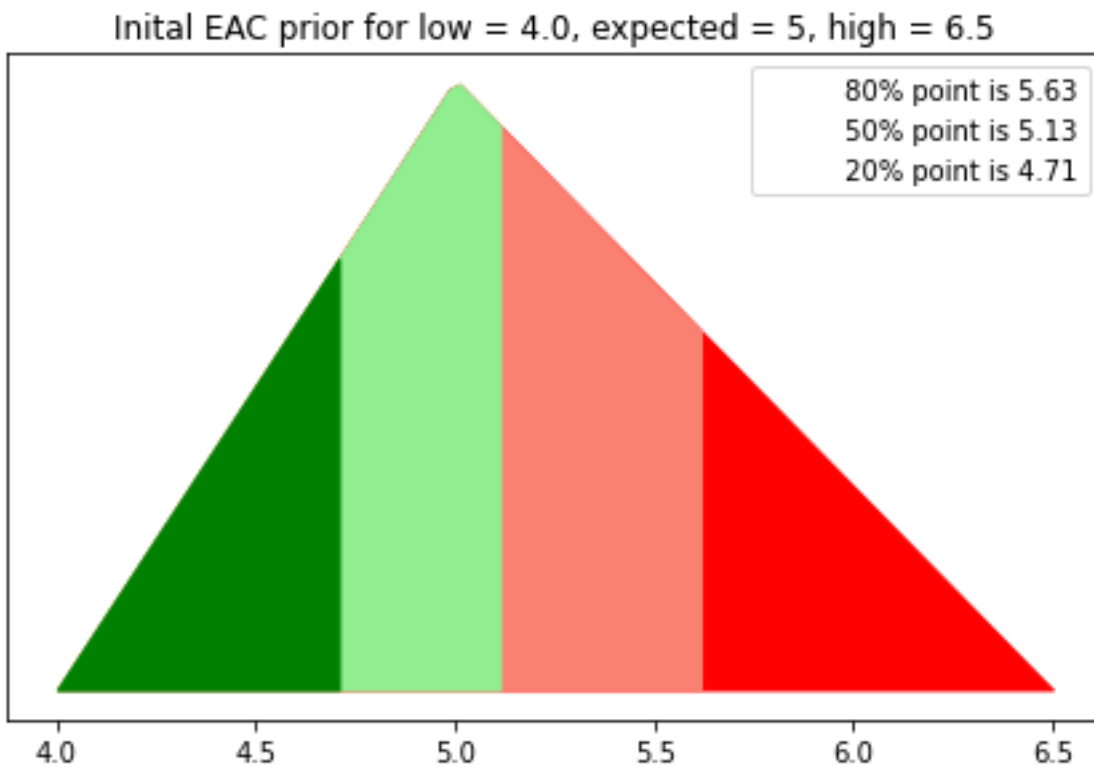


Figure 4

The area of each shaded region is the probability of the EAC being with the values at its base. Figure 4 is then interpreted as follows:

- There is an 80% chance that the cost at completion will be less than 5.63.
- There is a 50% chance that the cost at completion will be less than 5.13.
- There is a 20% chance that the cost at completion will be less than 4.71

The data is synthetically generated with four observations of the commonly tracked cost earned values parameters: actual cost and percent complete. It is generated with some random noise to reflect the real world better.

- The table of the observations with their associated cost rates.
- The output EAC PDFs after each observation.

Near Constant

Observation	AC	% Complete	Cost Rate
0	1.25	0.25	5.01
1	2.33	0.47	4.99
2	3.40	0.68	4.98
3	4.50	0.90	5.00

Table 1

The PDFs (Figure 4) show the ongoing BPL analysis for the observations. The first PDF is based on one observation, the second on the first two observations, etc.

Notice that the expected value, the 50% point, hovers around the BAC. However, the difference between the 20% and 80% points narrows considerably. Note at the 90% completion point, the task is 80% likely to miss the BAC by no more than 0.01 units.

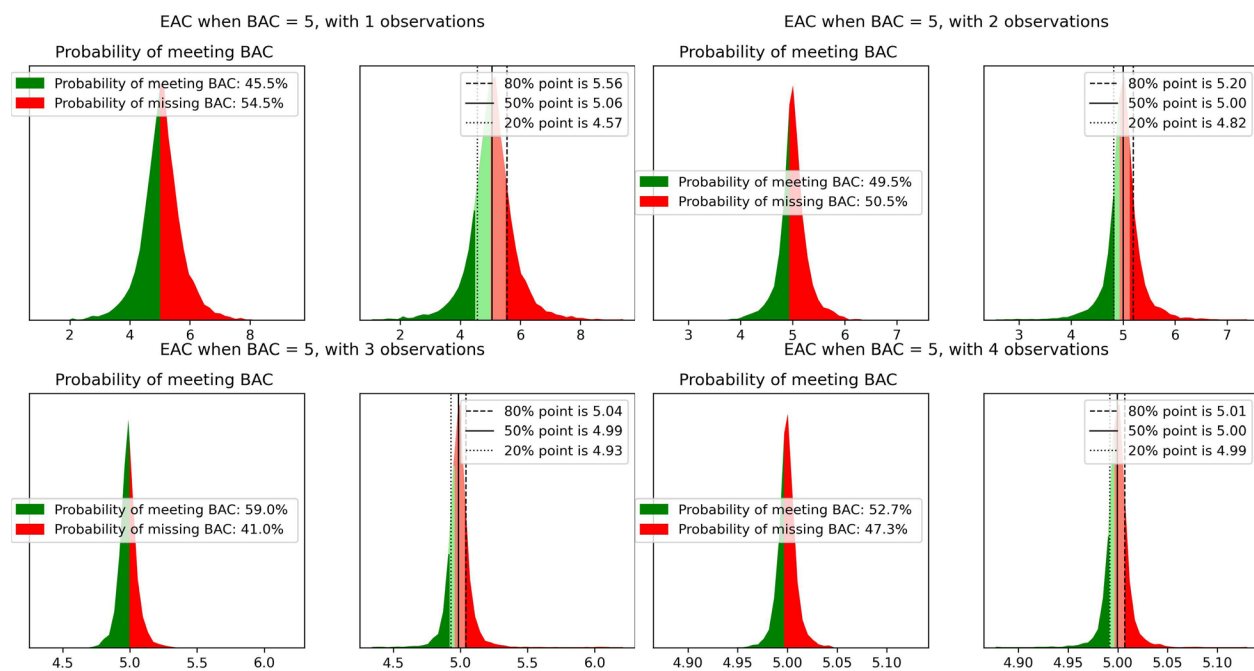


Figure 5

Recovery

This example is more realistic. The spending for the first two observations exceeds the average cost rate of 5. The team sees (Figure 6) that after two observations, they are very likely to come in over budget, so they complete the next by cutting spending. The final observation comes in at the expected rate (Table 2). So, at the 90% completion point, they are 48% likely to meet the target and 80% likely to miss at most 0.06 units.

I

Observation	AC	% Complete	Cost Rate
0	1.47	0.25	5.90
1	2.62	0.47	5.62
2	2.92	0.68	4.27
3	4.49	0.90	4.99

Table 2

Table 2

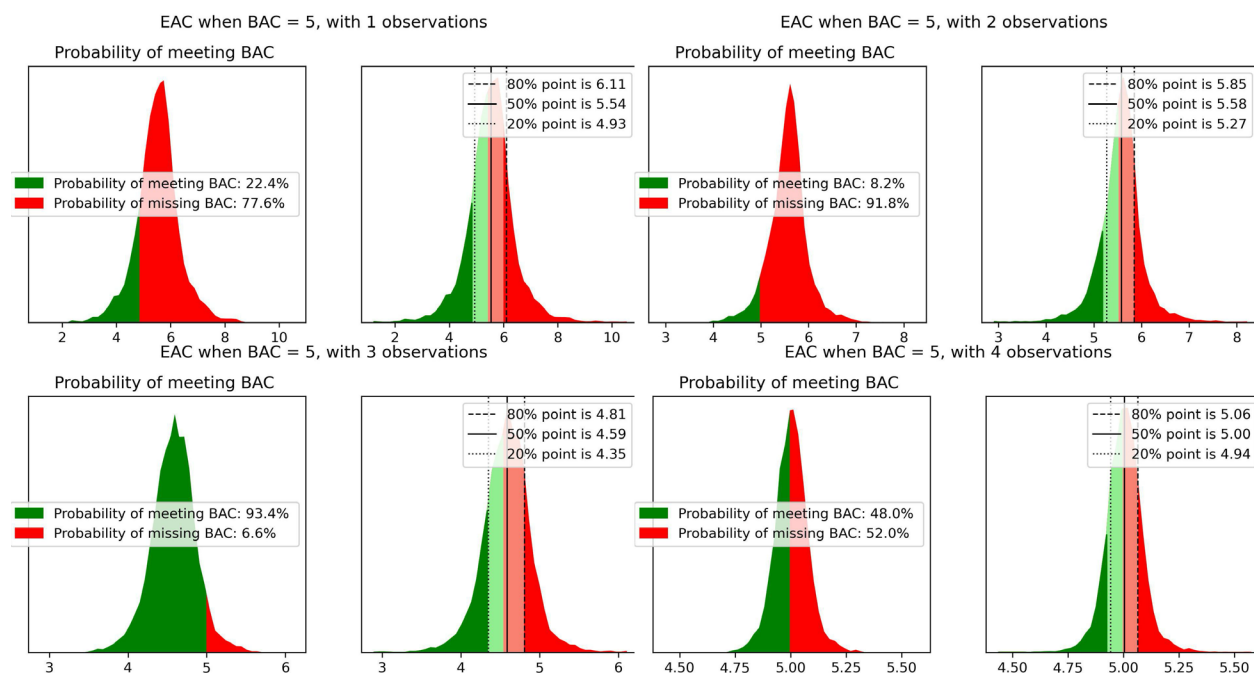


Figure 6

Next Steps

There are various ways this algorithm could be extended. These include:

- Nonlinear spending plans – Planned spending at internal minor milestones.
- Include both fixed and variable costs when variables are assigned independent of progress.
- Causal models – Bayesian nets to reflect how causal variance on one effort might affect another.

Some final points

Why this approach:

- BPL is mathematically sound, based on elementary probability theory.
- Makes the best use of the available small data.

- It gives early indications of the impact of spending trends.
- The PDFs are helpful for further analysis.
 - Roll-up of component and assembly costs.
 - Joint budget analysis.
 - Likelihood of desired underruns or acceptable overruns.

However:

- A reason that it has yet to be commonly deployed is that until recently, it was too computationally intensive.
- With the emergence of machine learning and data science techniques, these Bayesian techniques are becoming more mainstream (Clayton, 2021) (McElreath, 2020) (Ellenburg, 2014).

Acknowledgment

The authors thank Dr. James Cantor for his insightful reviews of an earlier draft of this paper.

Works Cited

- Cantor, M. (2023). *Bayesian Parameter Learning V2*. Retrieved from LinkedIn.com:
https://www.linkedin.com/posts/murraycantor_intro-to-bayesian-parameter-learning-v2-activity-7117523872907677698-4Dry?utm_source=share&utm_medium=member_desktop
- Clayton, A. (2021). *Bernoulli's Fallacy: Statistical Illogic and the Crisis of Modern Science*. Columbia University Press.
- Ellenburg, J. (2014). *How Not to Be Wrong: The Power of Mathematical Thinking*. Penguin Press.
- Fenton, N., & Neil, M. (2019). *Risk Assessment and Decision Analysis with Bayesian Networks, Second Edition*. CRC Press.
- Hubbard, D. (2014). *How to Measure Anything: Finding the Value of Intangibles in Business 3rd Edition*. Wiley.
- Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan 2nd Edition*. Academic Press.
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN (Chapman & Hall/CRC Texts in Statistical Science) 2nd Edition*. Chapman and Hall/CRC.
- Mun, J. (2003). *Applied Risk Analysis*. Hoboken, NJ: John Wiley & Sons.
- Smart, C. (2020). *Solving for Project Risk Management: Understanding the Critical Role of Uncertainty in Project Management*. McGraw Hill.
- Wikipedia. (2023, October). *Logistic Distribution*. Retrieved from Wikipedia:
https://en.wikipedia.org/wiki/Logistic_distribution