

## **CONSTRAINED BAYESIAN INFERENCE OF PROJECT PERFORMANCE MODELS**

Funlade T. Sunmola  
School of Engineering and Technology  
University of Hertfordshire  
College Lane Campus  
Hatfield, AL10 9AB, UK

### **ABSTRACT**

Project performance models play an important role in the management of project success. When used for monitoring projects, they can offer predictive ability such as indications of possible delivery problems. Approaches for monitoring project performance relies on available project information including restrictions imposed on the project, particularly the constraints of cost, quality, scope and time. We study in this paper a Bayesian inference methodology for project performance modelling in environments where information about project constraints is available and can be exploited for improved project performance. We apply the methodology to probabilistic modelling of project S-curves, a graphical representation of a project's cumulative progress. We show how the methodology could be used to improve confidence bounds on project performance predictions. We present results of a simulated process improvement project in agile setting to demonstrate our approach.

**Keywords:** Bayesian Inference, Project Performance Model, S-Curve, Project Constraints

### **1 INTRODUCTION**

Project performance models play an important role in the management of project success. They provide paths to discovering what one do not know. In manufacturing, performance models are integral part of projects, e.g. process improvement projects, development projects, product-based projects and manufacturing infrastructure projects, helping companies to consistently succeed in timely delivery of quality products and services (Lim and Yeo 1995, Yang 2013). When used for monitoring project progress, project performance models offer predictive ability by providing, for example, early warning signals of possible activity-coordination and delivery problems, resource conflicts, and risk of overruns. This is especially so in projects characterized by very high, difficult to quantify, levels of uncertainty created in part by the quality and completeness of information, diversity of interests and susceptibility to external influences (Atkinson, Crawford, and Ward 2006). The quality of the models depends largely on the available project information incorporating restrictions imposed on the project.

There are established techniques for monitoring and analyzing project performance including program evaluation and review technique (PERT), the critical path method (CPM), earned value management (EVM), project management control charts, and more recently burn-up, burn down and cumulative flow charts in agile environments (Aguanno 2009, Fleming and Koffleman 2010). The approaches are primarily deterministic and draw criticisms for their failure to account for the inherent uncertainty in project planning, execution, and performance measurement. The general criticism of the deterministic approaches centres on a) their inability to account for uncertainty, b) the erroneous mathematical assumption in techniques such as PERT and CPM that in any project network there is a unique, longest (i.e. critical) path, and c) the lack of confidence bounds on predictions which are essential to effective decision making under uncertainty (Pohl and Chapman 1987). These criticisms motivate the need for probabilistic approaches and new probabilistic performance modeling methods for project monitoring such as Stochastic S-Curves (Barraza, Back, and Mata 2004), control limit curves (Barraza and Bueno 2007), and Bayesian betaS-curve (Kim and Reinschmidt 2009) have been developed. We study in this paper a Bayesian inference methodology for project performance

modelling in environments where information about project constraints is available and can be exploited for improved project performance. We use the idea of truncated distributions. Truncated distributions arise naturally in many practical situations particularly when the variable of interest falls within certain tolerance limits. To truncate a distribution is to restrict its values to an interval and re-normalize the density so that the integral over that range is 1. We apply the methodology to probabilistic modelling of project S-curves and illustrate how the methodology could be used to improve bounds on project performance predictions, using a simulated process improvement project in agile setting.

We introduce project S-Curves in Section 2. In Section 3 we present our Bayesian approach and this is followed in Section 4 by an illustration of the approach using a simulated process improvement project in agile setting. We end the paper in Section 5 with concluding remarks and future research directions.

## 2 PROJECT S-CURVES

A project S-Curve is a graphical representation of the cumulative progress of a project from start to finish, with the horizontal and vertical axes showing time and cumulative project effort/progress respectively. The term derives from S-like shape of the curve, flatter at the beginning and end and steeper in the middle - a universal characteristic of all projects, regardless of the type, size, and complexity of a project. Since all projects are unique, the individual shapes for various projects vary according to the nature of the projects. Different methods can be used at different stages of a project to infer project S-curves.

At the initial stages, with only sketchy project definition, the general approach is to use empirical methods to derive baseline S-curves using historical data acquired from previous projects. As the project progresses and project-specific information becomes available, target S-curves based upon actual data can be developed for use in monitoring the project and refining target performance levels. Analysis of S-Curve permits a quick identification of project growth, slippage, and potential problems that could adversely impact the project if corrective action is not taken (Figure 1). Several alternative deterministic methods of inferring S-curves have been studied along with various mathematical

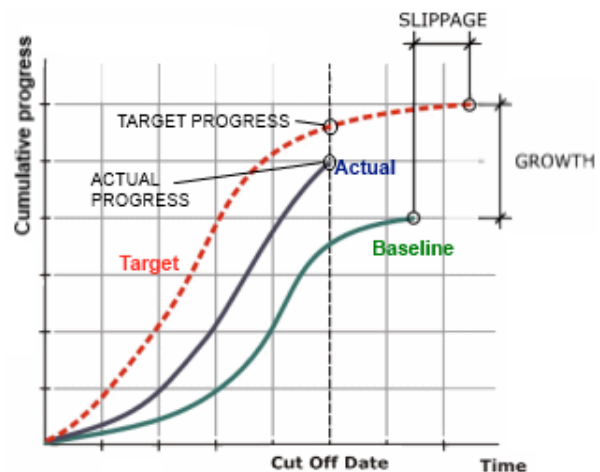


Figure 1: A Project S-Curve

formulas for generalizing cumulative project progress as a function of time e.g. Miskawi (1989), Murmis (1997) and Cioffi (2005). Deterministic inferences of project s-curves can be inadequate because projects are often subject to the influence of many factors and executes in uncertain environments. In contrast to deterministic approaches, researchers have suggested the use of probabilistic frameworks for project performance modelling using S-Curves. As an alternative to using deterministic S curves and traditional forecasting methods, Barraza, Back, and Mata (2004) introduced stochastic S curves (SS curves) to determine forecasted project estimates. Simulation is used to generate stochastic S-curves based on defined variability in duration and cost of the individual activities within the process. Stochastic S curves provide probability distributions for the budget and time values required to complete the project at every selected point of intermediate completion. Barraza and Bueno (2007) introduced an implementation of performance control limit curves for both actual cost and elapsed time, obtained using Stochastic S curves. Kim and Reinschmidt (2009) developed Bayesian BetaS-curve method (BBM) for schedule performance control and risk management of on-going projects. The BBM uses Bayesian principles and beta distribution in a probabilistic forecasting framework to drive continuous monitoring of project performance by revising beliefs about target project S-curve based on additional observations obtained during the project. The BBM is known to provide, early in the project, much more accurate forecasts than the earned value method or the earned schedule method and as accurate forecasts as the critical path

method without analyzing activity-level technical data. An advantage of the beta distribution is its ability in extensive distributions of forms only with two shape parameters. The probability density function (PDF) of the generalized beta distribution defined on the interval  $[A, B]$  with shape parameters  $\alpha$  and  $\beta$  is given by:

$$f(x) = \frac{1}{B(\alpha, \beta)} \frac{(x - A)^{\alpha-1} (B - x)^{\beta-1}}{(B - A)^{\alpha+\beta-1}} \quad A \leq x \leq B; A < B < \infty; \alpha, \beta > 0 \quad (1)$$

where  $B(\alpha, \beta)$  is the Beta function defined by:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1 - t)^{\beta-1} dt \quad (2)$$

The case where  $A = 0$  and  $B = 1$  is called the standard beta distribution. The formula for the cumulative distribution function (CDF) of the beta distribution, also called the incomplete beta function ratio, is defined by:

$$F(x) = \frac{\int_0^x t^{\alpha-1} (1 - t)^{\beta-1} dt}{B(\alpha, \beta)} \quad (3)$$

$$= \frac{B(x|\alpha, \beta)}{B(\alpha, \beta)}$$

The BetaS-curve function is defined as follows.

$$BetaS-Curve(x|\alpha, m) = \frac{B(x; T|\alpha, \beta)}{B(\alpha, \beta)} \quad (4)$$

where  $m$  is the location of the mode and  $T$  is project duration. The shape parameters are restricted to  $\alpha, \beta \geq 1$  in order to confine the plausible solution space to S-curves with unimodal PDF.

Bayesian paradigm can be applied from the onset of a project by integrating prior performance information (e.g. the prior probabilistic estimates of project duration and costs) with observations of new actual performance to arrive at posterior performance information. Bayesian inference offers a formal method to combine evidence external to a project, represented by a prior probability curve, with the evidence generated during the execution of the project, represented by a likelihood function. The attractiveness of Bayesian methods has also grown due to the availability of powerful MCMC simulation methods that provide the technology for sampling the posterior distribution of parameters and there are now more efficient techniques for handling constraints (Gelfand, Smith, and Lee 1992, Chen, Shao, and Ibrahim 2000).

### 3 BAYESIAN S-CURVES WITH PARAMETER CONSTRAINTS

In the Bayesian approach, we assume there is an S-Curve model  $S$  in a space  $\mathcal{S}$  of possible S-Curves and that  $S$  fits our project.  $S$  is parameterised by a set of random variables  $\theta$  that establish the geometric properties of the curve. In this paper, following Kim and Reinschmidt (2009), we specify  $S$  using a beta distribution and focus on three parameters - the shape parameters  $\alpha, \beta \geq 1$ , and a duration parameter  $T$ . The location of the mode  $m$  can be derived from  $\alpha$  and  $\beta$ . Bayesian inference concerns the estimation of the values of the unknown model parameters about which there may be some prior beliefs. These prior beliefs can be expressed as a probability density function  $f(\theta) = f(\alpha) f(\beta) f(T)$ , assuming the parameters are chosen independently, and may be interpreted as the probability placed on all possible parameter values before collecting any project-specific data. We assume constraints  $C$  are specified a-priori on the parameters.  $C$  truncates the prior distributions, restricting the domain of the probability distribution and allowing the parameters to only take permissible values. For example, for the random variable  $T$ ,  $T_l \leq T \leq T_u$  places lower  $T_l$  and upper  $T_u$  bounds on project duration. The PDF of  $T$  over the interval  $[T_l, T_u]$ , is given by:

$$f(T) = \begin{cases} \frac{g(T)}{G(T_T) - G(T_L)}, & \text{if } T_L \leq T \leq T_T \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $g(\bullet)$  and  $G(\bullet)$  are the PDF and CDF of the unconstrained distribution. Suppose we model  $T$  using a normal distribution  $N(\mu, \sigma^2)$ , then let the constrained duration parameter  $T_c$  be a truncated Normal  $TN(\mu, \sigma^2, T_L, T_T)$  random variable. The probability density function of  $T_c$  over interval  $[T_L, T_T]$  is given by:

$$f(T) = \begin{cases} \frac{\frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-(T-\mu)^2/2\sigma^2}}{\Phi\left(\frac{T_T-\mu}{\sigma}\right) - \Phi\left(\frac{T_L-\mu}{\sigma}\right)}, & \text{if } T_L \leq T \leq T_T \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where  $\Phi$  is the standard normal CDF. As the project progresses, we capture project-specific data  $D = (d_1, d_2, \dots, d_n)$  whose elements  $d_i$  consists of cumulative progress data  $w_i$  at time step  $t_i$ . We use the information to revise our prior beliefs. The dependence of observations  $D$  on the parameters  $\theta$  can be expressed as the likelihood function,  $f(D|\theta)$ .  $f(D|\theta)$  is calculated as the product of the likelihood of each observation. We make the deviations between actual times of performance reporting and the planned times determined by our model  $S(w_i|\theta)$  to be normally distributed with zero mean and standard deviation  $\sigma$ . That is,

$$\begin{aligned} f(D|\theta) &= \prod_{i=1}^N f(w_i, t_i|\theta) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp[-0.5((t_i - S(w_i|\theta))/\sigma)^2] \end{aligned} \quad (7)$$

The likelihood function is used to update the prior beliefs on  $\theta$  to account for the new data. This updating is performed in the constrained space  $\theta_c$  using Bayes theorem which can be expressed:

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int_{\theta_c} f(D|\theta)f(\theta) d\theta} \quad (8)$$

where  $f(\theta|D)$  is called the posterior distribution and expresses the probability of the parameter values after observing the new data. Once the posterior distribution is available, any features of  $\theta$ , such as the marginal distributions or means and variances of the individual parameters, as well as the predictive distribution of future observations, require integrating over the posterior distribution. For example, the marginal posterior distribution of an individual parameter e.g.  $T$  can be calculated as

$$f(T|D) = \int_{\theta_c, -T} f(\theta|D) d_{\theta_c, -T} \quad (9)$$

using Gibbs Sampling (Gelfand, Smith, and Lee 1992).  $\theta_c, -T$  represents all the parameters except  $T$ .

#### 4 ILLUSTRATION

We illustrate the Bayesian approach using a simulated process improvement project in agile setting. For our purpose we focus on requirements evolution in a project whose dynamics is illustrated in Figure 2. The duration of the example project is between 42 and 55 time steps.

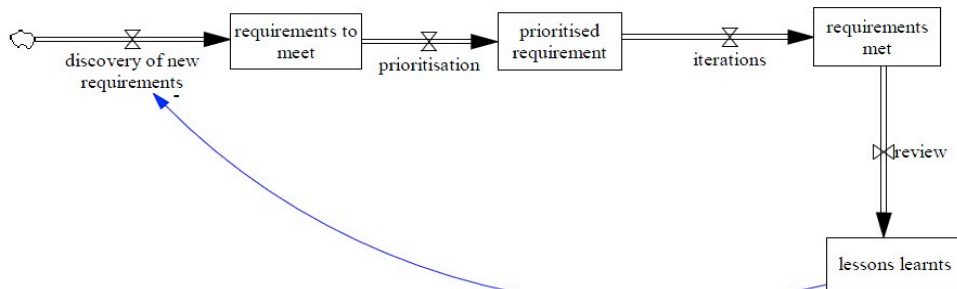


Figure 2: System Dynamics of the Project

Table 1: List of Simulation Parameters and their Settings

Parameter	Description	Setting
baseReqt:	Set of requirements initially specified to be met	Uniform (20,30)
discoveredReqtOrd:	Set of requirements routinely discovered	Uniform (10, 25)
	Arrival rate of the discovered requirements	Poisson (1.4)
discoveredReqtLearn:	Set of requirements discovered through learning	$0.4 \times \text{baseReqt} \times \exp(3.0)$
	Acceptance probability of the discovered	Uniform (0.6,1)
valueReqt:	Value of requirement	Uniform (100, 200)
	Volatility of requirements value	Normal $\mu = 0$ $\sigma = 0.15 \times \text{valueReqt}$
effortReqt:	Effort/Cost required to deliver a requirement	Uniform (10, 100)
PrioritisationStrategy:	Requirements are prioritized according to highest value first	

The process improvement project is initiated with a base set of requirements. A requirement is an ordered pair (cost, value). The requirement values are volatile and non-negative. As the project progresses new requirements are discovered and added to a prioritized list of requirements to meet. The project is executed iteratively. In each iteration, a requirement is selected from the prioritized requirements list and processed. When the selected requirement is fulfilled, a lessons-learnt review is conducted and work moves to the next iteration. New requirements may be discovered routinely and may also emerge as a result of the lessons-learnt review. The discovered requirements are added to the requirements list at the start of a new iteration. The execution process continues in cycles until all the requirements are fulfilled. The simulation run ends when all the requirements are fulfilled. To control how the simulation runs, we set the parameters of the simulation to values listed in Table 1.

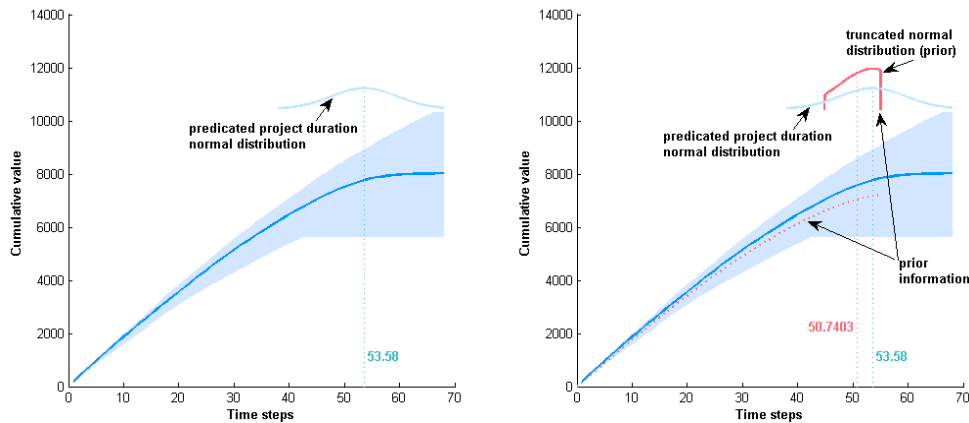


Figure 3: Predicted Performance – a) obtained from simulation and b) the prior information.

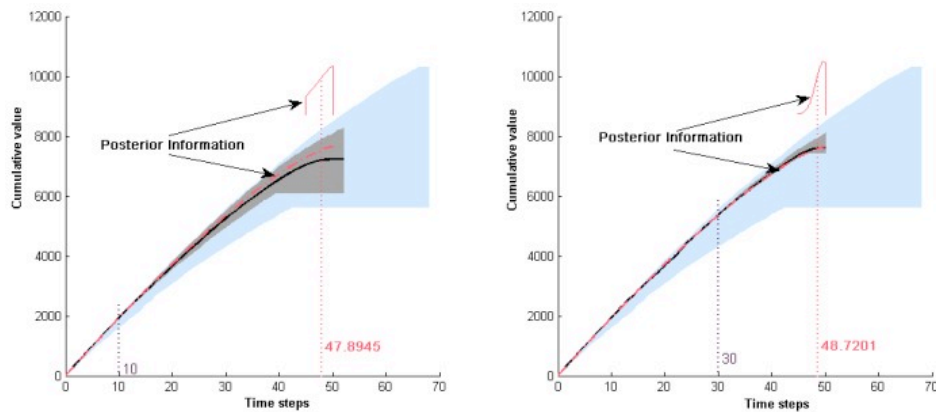


Figure 4: Predicted Project Performance at intermediate points – a) 10 time steps and b) 30 time steps.

The results of the simulation are averaged over 1000 runs. Figure 3a shows the predicted project duration obtained from simulation. The predicted duration is normally distributed with mean 53.58 and standard deviation of 6.04. Figure 3b shows the predicted normal distribution truncated to 45 and 55 time steps, below and above respectively. We use the truncated normal distribution alongside a Beta (1.01, 1.32) as prior information (Figure 3b). The posterior distributions for actual project data at time steps 10 and 30 are shown in Figures 4a and 4b respectively. The truncated distributions provide improved confidence bounds on project duration and the resulting models respond quickly to early project reports.

## 5 CONCLUSIONS AND FUTURE WORK

The focus in this paper has been on Bayesian inference for project performance modelling in environments where information about project constraints is available and can be exploited for improved project performance. We adopted the idea of truncated distributions and use a simulated process improvement project in an agile setting to illustrate the approach. Project constraints were assumed to be valid and known a-priori. Capturing the project constraint information and accounting for uncertainty in the constraints as part of the Bayesian framework are some areas of future work.

## REFERENCES

- Aguanno, K. 2009. *Agile Project Performance Management: Using Agile Metrics to Better Monitor and Control Agile Development Projects*. Multi-Media Publications Inc.
- Atkinson, R., L. Crawford, and S. Ward. 2006. Fundamental uncertainties in projects and the scope of project management. *International Journal of Project Management* 24, (8): 687 – 698.
- Barraza, G., W. Back, and F. Mata. 2004. Probabilistic Forecasting of Project Performance Using Stochastic S Curves. *Journal of Construction Engineering and Management* 130 (1): 25–32.
- Barraza, G., and R. Bueno. 2007. Probabilistic Control of Project Performance Using Control Limit Curves. *Journal of Construction Engineering and Management* 133 (12): 957–965.
- Chen, M.-H., Q.-M. Shao, and J. Ibrahim. 2000. Monte Carlo Methods for Constrained Parameter Problems. In *Monte Carlo Methods in Bayesian Computation*, Springer Series in Statistics, 191–212. Springer New York.
- Cioffi, D. F. 2005. A tool for managing projects: an analytic parameterization of the S-curve. *International Journal of Project Management* 23 (3): 215 – 222.
- Fleming, Q. W., and J. M. Koppelman. 2010. *Earned Value Project Management*. Fourth Edition. Project Management Institute.
- Gelfand, A. E., A. F. M. Smith, and T. -M. Lee. 1992. Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling. *Journal of the American Statistical Association* 87 (418): 523–532.
- Kim, B., and K. Reinschmidt. 2009. Probabilistic Forecasting of Project Duration Using Bayesian Inference and the Beta Distribution. *Journal of Construction Engineering and Management* 135 (3): 178–186.
- Lim, K. K., and K. T. Yeo. 1995. Management by projects in manufacturing industry In *Proceedings of 1995 IEEE Annual International Engineering Management Conference, 1995 Global Engineering Management: Emerging Trends in the Asia Pacific.*, 402–407.
- Miskawi, Z. 1989. An S-curve equation for project control. *Construction Management and Economics* 7 (2): 115–124.
- Murmis, G. M. 1997. ‘S’ curves for monitoring project progress. *Project Management Journal* 28 (3):29–35.
- Pohl, J., and A. Chapman. 1987. Probabilistic Project Management. *Building and Environment* 22 (3): 209–214.
- Yang, L.-R. 2013. Key practices, manufacturing capability and attainment of manufacturing goals: The perspective of project/engineer-to-order manufacturing. *International Journal of Project Management* 31 (1): 109 – 125.