AACE INTERNATIONAL RECOMMENDED PRACTICE

66R-11

SELECTING PROBABILITY DISTRIBUTION FUNCTIONS FOR USE IN COST AND SCHEDULE RISK SIMULATION MODELS

AACE INTERNATIONAL

AACE International Recommended Practice No. 66R-11

SELECTING PROBABILITY DISTRIBUTION FUNCTIONS FOR USE IN COST AND SCHEDULE RISK SIMULATION MODELS

TCM Framework: 7.6 – Risk Management

Rev. August 14, 2012

Note: As AACE International Recommended Practices evolve over time, please refer to www.aacei.org for the latest revisions.

Contributors:

Disclaimer: The opinions expressed by the authors and contributors to this recommended practice are their own and do not necessarily reflect those of their employers, unless otherwise stated.

Bruce E. Reynolds (Author)
Ricardo Accioly
James E. Arrow
Paul Bradley
Christopher P. Caddell, PE CCE
Dr. Ovidiu Cretu, PE
Michael W. Curran
Abhijnan Datta, CCE
Larry R. Dysert, CCC CEP
Dennis R. Hanks, PE CCE
John K. Hollmann, PE CCE CEP
Feng Jin
Matthew Schoenhardt
Robert F. Wells, CEP

AACE® International Recommended Practice No. 66R-11

SELECTING PROBABILITY DISTRIBUTION FUNCTIONS FOR USE IN COST AND SCHEDULE RISK SIMULATION MODELS

TCM Framework: 7.6 – Risk Management

August 24, 2012

INTRODUCTION

Scope

This recommended practice (RP) of AACE International provides guidance for selecting probability distribution functions (PDFs) for use in probabilistic decision and risk management (DRM) simulation models within cost engineering and total cost management (TCM)[1]. These DRM models are used to analyze asset and project cost, schedule, profitability and similar measures in consideration of risks. While AACE's definition of risk is fully defined elsewhere[2], this RP uses the term risk to address both threats and opportunities.

Purpose

This RP will describe the basic characteristics of commonly used PDFs and define their advantages and disadvantages for use in typical probabilistic modeling. It is intended to provide practical advice for non-statisticians who are applying risk analysis tools such as Monte Carlo modeling. In this area of practice, the data inputs are generally subjective; they are not based on sampling of data about a population or statistical analysis of empirical data. In that regard, this RP does not cover the use of curve fitting or statistical analysis of empirical data such as linear regression. The goals of this RP are to help users find PDFs: a) with an inferred "goodness of fit" to the opinions and perceptions of participants in risk analysis exercises; and b) that are reasonably simple to apply, understand and communicate. It does not cover methods to elicit these opinions and perceptions.

Background

The description of PDFs has traditionally resided in, and been somewhat obscured by, the domain of inferential statistics. It is hoped that this RP will provide some practical clarity.

For this RP, the PDFs can be classified in a number of ways that will assist the user in selection:

- First, this RP includes only those PDFs that can be assigned using subjective opinions from risk analysis participants and other subject matter experts. Typical applications are provided. Next, distributions can be either discrete or continuous. When modeling discrete variables (e.g., if a risk impact is to increase the number of sump pumps, it can be 1 or 2 pumps, but not 1.67), a discrete distribution should be used. Time and costs can be continuous or discrete variables. For example, depending on the risk event or condition, its cost impact may be any value within a continuous range or it may have only a few possible discrete values.

- Distributions can also be bounded or unbounded. In a bounded distribution, a minimum or maximum value can be stated for which there is a probability of 0% or 100% (e.g., as in triangle or uniform PDFs). In an unbounded distribution, the min/max may be infinity (e.g., such as in normal or the high side of lognormal PDFs). Because time and cost variables are typically bounded, the choice of unbounded PDFs requires special care in interpretation of the extremes of model outcomes (e.g., schedule durations in a CPM schedule cannot have negative values.)

August 24, 2012

Parameters

PDFs require parameters to define and describe them. The probability function with its parameters determines the probabilities for each model input value under consideration. Depending on the PDF used, the parameters can vary. Those parameters typically used in the practical PDFs for cost and schedule risk analysis are as follows:

- Low estimate: The lowest possible value as perceived by the team (0% probability of any value being lower). Note, the lowest value may be less than the team's perception.

- High estimate: The highest possible value as perceived by the team (100% probability of any value being equal to or lower than). Note that "optimism" bias is common; i.e., the highest value, including outliers, may be more than the team's perception.

- Most likely: Exactly the same as mode. This is the peak of the curve and represents a value that occurs most often.

- Mean: The central moment of the distribution. This value is also known as the as the arithmetic mean or expected value. In a skewed distribution the mean is distinct and different from the 50 value. This is the only measure of central tendency where the sum of the deviations of each value from the mean is zero.

- Standard deviation: A measure of dispersion of values from the mean. The square root of the variance.

The low estimate, high estimate, and most likely parameters are typically used in distributions such as Triangle, and PERT (a special form of a beta distribution, sometimes called BetaPERT). A Trigen distribution accommodates for optimism bias, when needed.

A key challenge of PDF selection is that the hard-and-fast rules set in the definitions above (e.g., highest value has 100% chance of being equaled or under-run) are commonly altered for practical considerations. For example, for range estimating and expected value contingency models[2,3], the high estimate assumed would exclude outcomes that approach absurdity (but are still possible) if for no other reason than human imagination has difficulty envisioning the true extremes. This is done in accordance with one of the principles of risk models which recommend that they be "fit for use" in the RCM process[4].

RECOMMENDED PRACTICE

PDF Selection

After one has come up with the base or point estimates of a critical cost item, activity, risk impact, or other uncertain variable in the model, users must select a PDF to apply to each variable. Other sources describe some of these models (as well as PDF selection)[5]. To aid this selection, Table 1 describes the basic characteristics of commonly used PDFs and defines their advantages and disadvantages for use in common probabilistic cost, schedule and similar risk modeling.

Traditionally, users have tended to select the triangular distribution because it is fairly easy to estimate the low, most likely, and high values. However, this distribution has significant potential disadvantages and this RP is intended to help users make better advised selections.