I^2SDS

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2009-11

August 30, 2009

On Some Elicitation Procedures for Distributions with Bounded Support with Applications in PERT

J. R. van Dorp

Department of Engineering Management and

Systems Engineering

The George Washington University

In: Risk and Decision Analysis in Maintenance Optimization and Flood Management, M.J. Kallen and S.P. Kuniewski (Eds.), IOS Press, The Netherlands, November 2009, pp. 21 - 34

On Some Elicitation Procedures for Distributions with Bounded Support with Applications in PERT

Johan René van Dorp[1], The George Washington University, USA

Abstract: The introduction of the Project Evaluation and Review Technique (PERT) dates back to the 1960's and has found wide application since then in the planning of construction projects. Difficulties with the interpretation of the parameters of the beta distribution let Malcolm et al. (1959) to suggest the classical expressions for the PERT mean and variance for activity completion that follow from lower and upper bound estimates a and b and a most likely estimate θ thereof. The parameters of the beta distribution are next estimated via the method of moments technique. Despite more recent papers still questioning the PERT mean and variance approach, their use is still prevalent in operations research and industrial engineering text books that discuss these methods. In this paper an overview is presented of some alternative approaches that have been suggested, including a recent approach that allows for a direct model range estimation combined with an indirect elicitation of bound and tail parameters of generalized trapezoidal uniform distributions describing activity uncertainty. Utilizing an illustrative Monte Carlo Analysis for the completion time of an 18 node activity network, we shall demonstrate a difference between project completion times that could result when requiring experts to specify a single most likely estimate rather than allowing for a modal range specification.

1. Introduction

The three parameter triangular distribution $Triang(a, \theta, b)$, with lower and upper bounds a and b and most likely value θ, is one of the first continuous distributions on the bounded range proposed back in 1755 by English mathematician Thomas Simpson (1755, 1757). It received special attention

[1] Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, 1776 G Street, N.W., Washington D.C. 20052. Phone: 202-994-6638, Fax: 202-994-0245, E-mail: dorpjr@gwu.edu.

as late as in the 1960's, in the context of the PERT (see, e.g., Winston 1993) as an alternative to the four-parameter beta distribution

$$f_T(t|a,b;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{(t-a)^{\alpha-1}(b-t)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}, \quad a \le t \le b,\ \alpha > 0,\ \beta > 0. \tag{1}$$

which involves some difficulties regarding the interpretation of its parameters α and β. As a result Malcolm et al. (1959)[2] suggested the following PERT mean and variance expressions

$$E[T] = \frac{a + 4\theta + b}{6}, Var[T] = \frac{1}{36}(b-a)^2 \tag{2}$$

where T is a random variable modeling activity completion time, a and b being the lower and upper bound estimates and θ being a most like estimate for T. The remaining beta parameters α and β in (1) are next obtained from (2) utilizing the method of moments.

The somewhat non-rigorous proposition (2) resulted in a vigorous debate over 40-years ago (Clark 1962, Grubbs 1962, Moder and Rodgers 1968) regarding its appropriateness and even serves as the topic of more recent papers (see, e.g., Herrerías 1989, Kamburowski 1997, Herrerías et al. 2003). In a further response to the criticism of (2), Herrerías (1989) suggested substitution of

$$\alpha = 1 + s\,(\theta - a)/(b-a),\ \beta = 1 + s\,(b-\theta)/(b-a), \tag{3}$$

in (1) instead, where $s > -1$, $a < \theta < b$. This yields

$$E[T] = \frac{a + s\,\theta + b}{s+2}, Var[T] = \frac{(s+1)(b-a)^2 + s^2(b-\theta)(\theta-a)}{(s+3)(s+2)^2}. \tag{4}$$

Essentially, Herrerías (1989) reparameterizes the beta pdf (1) by managing to express α and β in terms of new parameters θ and s while retaining the lower and upper bounds a and b. For $s > 0$ the beta pdf (1) is unimodal and for $s = 0$ it reduces to a uniform distribution. Hence, Herrerías (1989) designated s to be a confidence parameter in the mode location θ such that higher values of s indicate a higher confidence. Indeed, for $s \to \infty$, the beta pdf converges to a single point mass at

[2]Kamburowski (1997) notes that: "*Despite the criticisms and the abundance of new estimates, the PERT mean and variance* [given by (2) in this paper] *can be found in almost every textbook on OR/MS and P/OM, and are employed in much project management software.*"

θ. For $-1 < s < 0$, the beta pdf (1) is U-shaped which is not consistent with θ being a most likely value.

As a further alternative to the beta pdf (1), Van Dorp and Kotz (2002) generalized the $Triang(a, \theta, b)$ distribution to a two sided power $TSP(a, \theta, b, n)$ distribution

$$f_X(x|a, \theta, b, n) = \frac{n}{b-a} \times \begin{cases} \left(\frac{x-a}{\theta-a}\right)^{n-1}, & a < x \leq \theta, \\ \left(\frac{b-x}{b-\theta}\right)^{n-1}, & \theta \leq x < b, \end{cases} \tag{5}$$

by the inclusion of an additional parameter $n > 0$ describing a power-law behavior in both tails. For $n = 2$ and $n = 1$ the distribution (5) reduces to the $Triang(a, \theta, b)$ and Uniform$[a, b]$ distributions, respectively. The following expressions for the mean and the variance follow from (5):

$$E[X] = \frac{a + (n-1)\theta + b}{n+1}, Var[X] = \frac{n(b-a)^2 - 2(n-1)(b-\theta)(\theta-a)}{(n+2)(n+1)^2}. \tag{6}$$

Interestingly, one immediately observes that by substituting $n = s + 1$ in (6), the beta mean value (4) and TSP mean value in (6) coincide. Moreover, recalling that $T \sim Beta(a, b, \alpha, \beta)$ given by (1) and $X \sim TSP(a, m, b, n)$ given by (5), and observing that for $s = 4$ or $n = 5$ the mean values in (4) and (6) agree and reduce to the PERT mean $E[T]$ in (2) as suggested by Malcolm et al. back in 1959, one might indeed conclude that they were lucky in this respect. However, observing that the variance in (4) for $s = 4$ is quite different from the PERT variance in (2), Malcolm et al. (1959) were after all not so lucky. Moreover, after some algebraic manipulations using variances in (4) and (6) it follows that:

$$Var[T] - Var[X] = \frac{(n-1)(b-\theta)(\theta-a)}{(n+2)(n+1)} = \begin{cases} \leq 0, & 0 \leq n < 1, \\ > 0, & n > 1. \end{cases} \tag{7}$$

Hence, in the unimodal domains of the TSP distribution (5), $n > 1$, and the beta distributions (1), $s > 0$, with parameterization (3), the variance of the TSP distribution is strictly less than the PERT variance modification of Herrerías (1989) given by (4). The result (7) is consistent with the TSP distributions being more "peaked" than the beta distribution (see, e.g. Kotz and Van Dorp 2004). Summarizing, given that an expert only provides lower bounds a and b and most likely value m, additional alternatives are provided in terms of the TSP(n) pdf's (5), $n \neq 2$, besides the existing beta and triangular pdf options, and one is left to wonder which one of these to use, perhaps extending the 50-year old controversy surrounding the use of (2).

The context of the controversy alluded to above deals with the larger domain of distribution selection and parameter elicitation via expert judgment, in particular those distributions with bounded support. In a recent survey paper, a leading Bayesian statistician O'Hagan (2006) explicitly mentions a need for advances in elicitation techniques for prior distributions in Bayesian Analyses, but also acknowledges the importance of their development for those areas where the elicited distribution cannot be combined with evidence from data, because the expert opinion is essentially all the available knowledge. Garthwaite, Kadana and O'Hagan (2005) provide a comprehensive review on the topic of eliciting probability distributions dealing with a wide variety of topics, such as e.g. the elicitation process, heuristics and biases, fitting distributions to an expert's summaries, expert calibration and group elicitation methods. Experts are, as a rule, classified into two, usually unrelated, groups: 1) *substantive* experts (also known as technical experts or domain experts) who are knowledgeable about the subject matter at hand and 2) *normative* experts possessing knowledge of the appropriate quantitative analysis techniques (see, e.g., De Wispelare et al. (1995) and Pulkkinen and Simola (2000)). In the absence of data and in the context of decision/simulation and uncertainty analyses, substantive experts are used (often by necessity) to specify input distributions albeit directly or indirectly with the aid of a normative expert. The topic of this paper deals with fitting specific parametric distributions to a set of summaries elicited from an expert.

In Section 2, we provide an overview of indirect elicitation procedures for TSP pdf (5) parameters and their generalizations developed in Kotz and Van Dorp (2006), Van Dorp et al. (2007) and Herrerías et al. (2009). Firstly, we shall present an indirect elicitation procedure for the bound parameters a, b and tail parameter n of TSP pdf's (5). It has the specific advantage of not requiring bounds elicitation whom may not fall within the realm of expertise of a substantive expert. Next, we present the indirect elicitation of both tail parameter of a generalization of TSP distribution allowing for separate power law behavior in both tails. This procedure was presented in Herrerías et al. (2009), but does require the bounds a and b to be available. We return to indirect bounds and power tail parameter elicitation for generalized trapezoidal uniform (GTU) distributions given lower and upper quantile estimates and a modal range specification. A substantive expert may be more comfortable with specifying a modal range rather than having to specify a single point estimate as required in (2), (3) and (5). The GTU elicitation procedure was developed in detail in Van Dorp et al. (2007). Finally, in Section 3, we shall demonstrate via an illustrative Monte Carlo analysis for the completion time of an 18 node activity network a potential difference between project completion times that could result when requiring experts to specify a single most likely estimate rather than allowing for a modal range specification.

2. Parameter elicitation algorithms for TSP distributions and some generalizations.

Let $X \sim TSP(\Theta)$ with pdf (5), where $\Theta = \{a, \theta, b, n\}$. The main advantage of the pdf (5) over the beta pdf (1) is that it has a closed form cdf expressible using only elementary functions:

$$F_X(x|\Theta) = \begin{cases} \frac{\theta-a}{b-a}\left(\frac{x-a}{\theta-a}\right)^n, & \text{for } a < x < \theta, \\ 1 - \frac{b-\theta}{b-a}\left(\frac{b-x}{b-\theta}\right)^n, & \text{for } \theta \leq x < b. \end{cases} \tag{8}$$

Suppose a lower and upper percentiles a_p, b_r and most likely value θ for X are pre-specified in a manner such that $a_p < \theta < b_r$. Kotz and Van Dorp (2006) showed that a unique bounds a and b solution

$$a \equiv a\{q(n)\} = \frac{a_p - \theta\sqrt[n]{p/q(n)}}{1 - \sqrt[n]{p/q(n)}},\ b \equiv b\{q(n)\} = \frac{b_r - \theta\sqrt[n]{\frac{1-r}{1-q(n)}}}{1 - \sqrt[n]{\frac{1-r}{1-q(n)}}}. \tag{9}$$

exists given a value for the parameter $n > 0$, where $q(n) = Pr(X < \theta)$[3]. The unique value for $Pr(X < \theta)$ follows by solving for $q(n)$ from the equation

[3]Herein we shall use the notation $\sqrt[n]{x} = x^{1/n}$ even when $n > 0$ is non-integer valued.

$$q(n) = \frac{(\theta - a_p)\Big(1 - \sqrt[n]{\frac{1-r}{1-q(n)}}\Big)}{(b_r - \theta)\Big(1 - \sqrt[n]{\frac{p}{q(n)}}\Big) + (\theta - a_p)\Big(1 - \sqrt[n]{\frac{1-r}{1-q(n)}}\Big)}, \tag{10}$$

using a bisection method with starting interval $[p, r]$. When $n \downarrow 0$,

$$q(n) \to q(0) = (\theta - a_p)/(b_r - a_p) \tag{11}$$

and when $n \to \infty$, $q(n)$ converges to the unique solution $q(\infty)$ of the equation

$$\frac{q(\infty)}{q(0)} Log\Big\{\frac{q(\infty)}{p}\Big\} = \frac{1-q(\infty)}{1-q(0)} Log\Big\{\frac{1-q(\infty)}{1-r}\Big\} \tag{12}$$

that, similar to $q(n)$ in (10), may be solved for using a bisection method with starting interval $[p, r]$. The pdf (5) itself, satisfying $a_p < \theta < b_r$, converges to a Bernoulli distribution with point mass $q(0)$ at a_p when $n \downarrow 0$ and when $n \to \infty$ converges to an asymmetric Laplace distribution

$$f_X(x|a_p, \theta, b_r) = \begin{cases} q(\infty)\mathcal{A}Exp\Big\{-\mathcal{A}(\theta - x)\Big\}, & x \le \theta, \\ \{1 - q(\infty)\}\mathcal{B}Exp\Big\{-\mathcal{B}(x - \theta)\Big\}, & x > \theta, \end{cases} \tag{13}$$

where the coefficients $\mathcal{A}$ and $\mathcal{B}$ are

$$\mathcal{A} = \frac{Log\Big\{\frac{q(\infty)}{p}\Big\}}{\theta - a_p} \text{ and } \mathcal{B} = \frac{Log\Big\{\frac{1-q(\infty)}{1-r}\Big\}}{b_r - \theta}. \tag{14}$$

(See also Kotz and Van Dorp (2005).)

Summarizing, the information $a_p < \theta < b_r$ does not uniquely specify a member within the TSP family. Kotz and Van Dorp (2006) suggest the elicitation of an additional quantile $a_p < x_s < b_r$ to indirectly elicit the remaining parameter n. They solve for a, b and n via an eight step algorithm. Its details are provided in Kotz and Van Dorp (2006) and a software implementation of this algorithm is available from the author upon request. Setting, $a_{0.10} = 6.5$, $x_{0.80} = 10\frac{1}{4}$, $b_{0.90} = 11\frac{1}{2}$ and $\theta = 7$ we have:

$$n \approx 3.873, q(n) = 0.209,\ a\{q(n)|n\} \approx 4.120, b\{q(n)|n\} \approx 17.878. \tag{15}$$

Figure 1 displays the TSP distribution with most likely value $\theta = 7$ and parameter values (15).

In: Risk and Decision Analysis in Maintenance Optimization and Flood Management, M.J. Kallen and S.P. Kuniewski (Eds.), IOS Press, The Netherlands, November 2009, pp. 21 - 34

2.1. GTSP parameter elicitation algorithm

Kotz and Van Dorp (2004) briefly mentioned generalized $GTSP(\Theta)$ distributions with pdf

$$f_X(x|\Theta) = \mathcal{C}(\Theta) \times \begin{cases} \left(\frac{x-a}{\theta-a}\right)^{m-1}, & \text{for } a < x < \theta, \\ \left(\frac{b-x}{b-\theta}\right)^{n-1}, & \text{for } \theta \le x < b, \end{cases} \tag{16}$$

where $\Theta = \{a, \theta, b, m, n\}$ and

$$\mathcal{C}(\Theta) = \frac{mn}{(\theta-a)n + (b-\theta)m}. \tag{17}$$

They reduce to $TSP(\Theta)$ pdf's (5) when $m = n$ and were studied in more detail by Herrerías et al. (2009). Their cdf's follow from (16) as:

$$F_X(x|\Theta) = \begin{cases} \pi(\Theta)\left(\frac{x-a}{\theta-a}\right)^{m}, & \text{for } a < x < \theta, \\ 1-[1-\pi(\Theta)]\left(\frac{b-x}{b-\theta}\right)^{n}, & \text{for } \theta \le x < b. \end{cases} \tag{18}$$

where

$$\pi(\Theta) = (\theta - a)\mathcal{C}(\Theta)/m. \tag{19}$$

To indirectly elicit the power parameters m and n, Herrerias et al. (2009) also suggest eliciting a lower quantile $a_p < \theta$ and an upper quantile $b_r > \theta$. Similar to the PERT mean and variance (2), however, lower and upper bounds a, b and a most likely estimate θ must have been directly pre-elicited. The parameters m and n are next solved from the following set of non-linear equations (the quantile constraints) :

$$\begin{cases} F(a_p|\Theta) = \pi(\Theta)\left(\frac{a_p-a}{\theta-a}\right)^{m} = p, \\ F(b_r|\Theta) = 1-[1-\pi(\Theta)]\left(\frac{b-b_r}{b-\theta}\right)^{n} = r. \end{cases} \tag{20}$$

Herrerias et al. (2009) showed that the first (second) equation in (20) has a unique solution $m^\bullet$ for every fixed value of $n > 0$ and thus it defines an implicit continuous function $\xi(n)$ such that the parameter combination $\{\theta, m^\bullet = \xi(n),\ n\}$ satisfies the first quantile constraint for all $n > 0$. This unique solution $m^\bullet$ may be solved for by employing a standard root finding algorithm such as, e.g., the Newton-Raphson method (Press et al., 1989) or a commercially available one such as, e.g., GoalSeek in Microsoft Excel. Analogously, the second equation defines an implicit continuous function $\zeta(m)$ such that the parameter combination $(\theta, m,\, n^{\bullet} = \zeta(m))$ satisfies the second quantile constraint for all $m > 0$. By successively solving for the lower and upper quantile constraint given a value for n or m, respectively, an algorithm can be formulated that solves (20). Details are provided in Herrerias et al. (2009). Setting $a = 2,\ \theta = 7,\ b = 15,\ a_{0.10} = 4\frac{1}{4},\ b_{0.90} = 11$ in (20) yields the power parameters

$$m \approx 1.883 \text{ and } n \approx 2.460. \tag{21}$$

Figure 1 displays the GTSP distribution with lower and upper bounds $a = 2,\ b = 15$, most likely value θ and the power parameter values (21).

2.2. GTU parameter elicitation procedure

Van Dorp et al. (2007) considered *Generalized Trapezoidal Uniform* (GTU) distributions. Letting $X \sim GTU(\Theta)$, where $\Theta = \{a, \theta_1, \theta_2, b, m, n\}$, they have for its pdf:

$$f_X(x|\Phi) = \mathcal{C}(\Theta) \times \begin{cases} \left(\frac{x-a}{\theta_1-a}\right)^{m-1}, & \text{for } a \le x < \theta_1, \\ 1, & \text{for } \theta_1 \le x < \theta_2, \\ \left(\frac{b-x}{b-\theta_2}\right)^{n-1}, & \text{for } \theta_2 \le x < b, \end{cases} \tag{22}$$

where the normalizing constant $\mathcal{C}(\Phi)$ is given by

$$\mathcal{C}(\Theta) = \frac{mn}{(\theta_1 - a)n + (\theta_2 - \theta_1)mn + (b - \theta_2)m}. \tag{23}$$

Defining stage probabilities $\pi_1 = Pr(X \le \theta_1),\ \pi_2 = Pr(\theta_1 < X \le \theta_2), \pi_3 = Pr(X > \theta_1)$, one obtains from (22) and (23):

$$\begin{cases} \pi_1(\Theta) = \mathcal{C}(\Theta)(\theta_1 - a)/m, \\ \pi_2(\Theta) = \mathcal{C}(\Theta)(\theta_2 - \theta_1), \\ \pi_3(\Theta) = \mathcal{C}(\Theta)(b - \theta_2)/n. \end{cases} \tag{24}$$

Utilizing the stage probabilities $\pi_i(\Theta), i = 1, \ldots, 3,$ one obtains the following convenient form for the cdf of (22)

$$F_X(x|\Theta) = \begin{cases} \pi_1(\Theta)\left(\frac{x-a}{\theta_1-a}\right)^m, & a \leq x \leq \theta_1, \\ \pi_1(\Theta) + \pi_2(\Theta)\frac{x-\theta_1}{\theta_2-\theta_1}, & \theta_1 < x \leq \theta_2\,, \\ 1-\pi_3(\Theta)\left(\frac{b-x}{b-\theta_2}\right)^n, & \theta_2 < x \leq b, \end{cases} \tag{25}$$

and for its quantile function

$$F_X^{-1}(y|\Theta) = \begin{cases} a + (\theta_1 - a)\sqrt[m]{\frac{y}{\pi_1(\Theta)}}, & 0 \leq y \leq \pi_1(\Theta), \\ \theta_1 + (\theta_2 - \theta_1)\frac{y-\pi_1(\Theta)}{\pi_2(\Theta)}, & \pi_1(\Theta) < y \leq 1-\pi_3(\Theta), \\ b - (b-\theta_2)\sqrt[n]{\frac{1-y}{\pi_3(\Theta)}}, & 1-\pi_3(\Theta) < y \leq 1. \end{cases} \tag{26}$$

The $GTU(\Theta)$ distributions reduce to trapezoidal distributions studied by Pouliquen (1970) by setting $m = n = 2$, to $GTSP(\Theta)$ distributions given by (16) and (17) by setting $\theta_1 = \theta_2$, and to $TSP(\Theta)$ distributions given by (5) by setting $\theta_1 = \theta_2 = \theta$ and $m = n$ in (22) and (23).

It shall be assumed here that the lower and upper bound parameters a and b and tail parameters m and n are unknown and that they need to be determined from (i) a directly elicited modal range $[\theta_1, \theta_2]$, (ii) the relative likelihoods π_2/π_1 and π_2/π_3 (or their reciprocals), and (iii) a lower $a_p < \theta_1$ and upper $b_r > \theta_2$ quantiles. The first (second) relative likelihood may be elicited by asking how much more likely it is for X to be within its modal range $[\theta_1, \theta_2]$ than being less (larger) than it. Stage probabilities (24) $\pi_i, i = 1, \ldots, 3,$ next follow with the restriction they must sum to 1. This manner of elicitating of $\pi_i, i = 1, \ldots, 3$ is analogous to the fixed interval elicitation method mentioned in Garthwaite, Kadane and O'Hagan (2005).

Van Dorp et al. (2007) showed that a unique solution for the power parameters m and n may be obtained from the equations $a^*(m) = \widetilde{a}(m),\ b^*(n) = \widetilde{b}(n)$, respectively, where

$$\begin{cases} a^*(m) \equiv \theta_1 - m\frac{\pi_1}{\pi_2}(\theta_2 - \theta_1), & \widetilde{a}(m) \equiv a_p - \frac{\sqrt[m]{p/\pi_1}}{1-\sqrt[m]{p/\pi_1}}(\theta_1 - a_p), \\ b^*(n) \equiv \theta_2 + n\frac{\pi_3}{\pi_2}(\theta_2 - \theta_1), & \widetilde{b}(n) \equiv b_r + \frac{\sqrt[n]{(1-r)/\pi_3}}{1-\sqrt[n]{(1-r)/\pi_3}}(b_r - \theta_2), \end{cases} \tag{27}$$

$\pi_i, i = 1, \ldots, 3,$ are given by (25), and provided

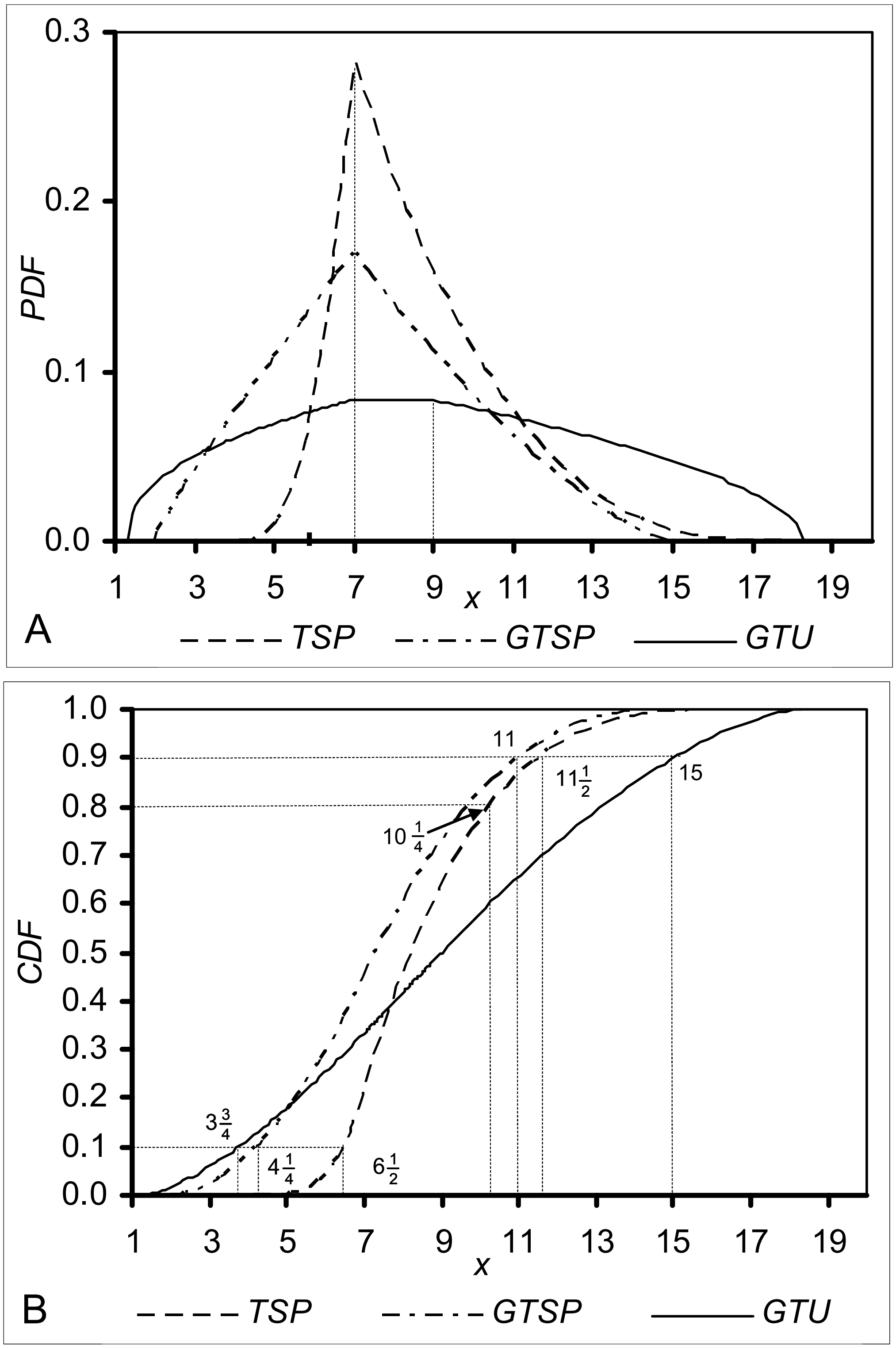

Figure 1. Pdf's (A) and cdf's (B) of TSP, GTSP and GTU distributions with parameter settings (15), (21) and (29). Elicited modal (quantile) values are indicated in Figure 1A (1B).

$$\begin{cases} a_p > \theta_1 - \xi(\theta_2 - \theta_1), & \text{where } \xi = \frac{\pi_1}{\pi_2}\log\left(\frac{\pi_1}{p}\right) > 0, \\ b_r < \theta_2 + \psi(\theta_2 - \theta_1), & \text{where } \psi = \frac{\pi_3}{\pi_2}\log\left(\frac{\pi_3}{1-r}\right) > 0. \end{cases} \tag{28}$$

Equations $a^*(m) = \widetilde{a}(m)$, $b^*(n) = \widetilde{b}(n)$ may be solved for using a standard root finding algorithm such as, e.g., the Newton-Raphson method (Press et al., 1989) or a commercially available one such as, e.g., GoalSeek in Microsoft Excel. No solution exist for power parameters m and n exist when conditions in (28) are not met. After solving for m the lower bound a follows by substitution of m in $a^*(m)$ or $\widetilde{a}(m)$. Solving for the upperbound b is analogous, but utilizes the expressions for $b^*(n)$ or $\widetilde{b}(n)$. Setting $[\theta_1, \theta_2] = [7, 9]$, $\pi_2/\pi_1 = 1/2$, $\pi_2/\pi_3 = 1/3$, $a_{0.10} = 3\frac{3}{4}$ and $b_{0.90} = 15$ in (27) yields the tail and lower and upper bound parameters

$$m \approx 1.423,\ n \approx 1.546,\ a \approx 1.306 \text{ and } b \approx 18.273. \tag{29}$$

Figure 1 displays the GTU distribution with modal range $[\theta_1, \theta_2] = [7, 9]$ and parameter values (29). Please observe in Figure 1 that both TSP and GTSP distributions posses mode $\theta = 7$, whereas the GTU distribution has a modal range $[7, 9]$. Quantile values for the TSP, GTSP and GTU examples in this section are indicated in Figure 1B.

3. An illustrative activity network example

We shall demonstrate via an illustrative Monte Carlo analysis for the completion time of an 18 node activity network from Taggart (1980), depicted in Figure 2, a potential difference between project completion times that could result when requiring experts to specify a single most likely estimate rather than allowing for a modal range specification. We shall assume that lower and upper quantiles $a_{0.10}$ and $b_{0.90}$ in Table 1 have been elicited via an expert judgment for each activity in the project network. We shall investigate four scenarios of mode specification for the activity durations in the project network, keeping their lower and upper quantiles $a_{0.10}$ and $b_{0.90}$ fixed. In the first scenario "GTU" activity duration uncertainty is modeled using a GTU distribution. The modal range $[\theta_1, \theta_2]$ is specified in Table 1. For all activities, a relative likelihood of 2.75 (1.25) is specified for the right

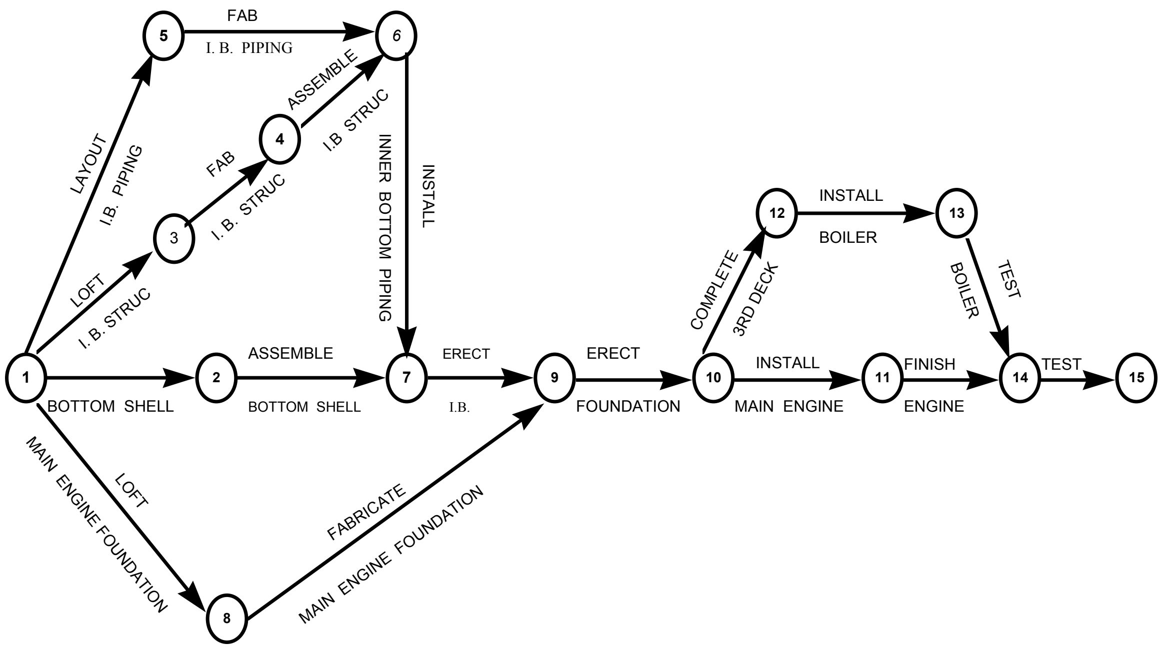

Figure 2. Example project network from Taggart (1980).

Table 1. Data for modeling the uncertainty in activity durations for the project network presented in Figure 5 for the Scenarios "GTU", "Uniform", "Laplace 1", "Laplace 2".

Activity Name	$a_{0.10}$	θ_1	θ_2	$b_{0.90}$
Shell: Loft	22	25	28	41
Shell: Assemble	35	38	41	54
I.B.Piping: Layout	22	25	28	41
I.B.Piping: Fab.	6	8	10	19
I.B.Structure: Layout	22	25	28	41
I.B.Structure: Fab.	16	18	20	29
I.B.Structure: Assemb.	11	13	15	24
I.B.Structure: Install	6	8	10	19
Mach Fdn. Loft	26	29	32	45
Mach Fdn. Fabricate	31	34	37	50
Erect I.B.	28	31	34	47
Erect Foundation	6	8	10	19
Complete 3rd DK	4	6	8	17
Boiler:Install	7	9	11	20
Boiler:Test	9	11	13	22
Engine: Install	6	8	10	19
Engine: Finish	18	21	24	37
FINAL Test	14	17	20	33

tail (left tail) as compared to the modal range $[\theta_1, \theta_2]$. From the relative likelihoods it immediately follows that the lower bounds θ_1 of the modal ranges in Table 1 equal the first quartile (probability $\frac{1}{4}$) of the activities, whereas a $\frac{1}{5}$ probability is specified throughout for the modal range $[\theta_1, \theta_2]$. Hence, the upper bounds θ_2 of the modal ranges are the 45-th percentiles of the activity durations and thus are strictly less than their median values. Moreover, all activity durations are right skewed (having a longer tail towards the right). We solve for the lower and upper bounds a and b using the procedure described in Section 2.2.

The next three scenarios involve limiting cases when activity duration uncertainties are distributed as a two-sided power (TSP) distribution with the pdf (5). Recall from Section 2 that Kotz and Van Dorp (2006) have shown that for every $n > 1$ in (5), a unique unimodal TSP distribution can be fitted given a lower quantile $a_{0.10}$, an upper quantile $b_{0.90}$ and a most likely value θ such that $a_{0.10} < \theta < b_{0.90}$. For $n \downarrow 1$, the fitted TSP distribution reduces to a uniform distribution with the bounds

$$a = \frac{0.90a_{0.10} - 0.10b_{0.90}}{0.80} \text{ and } b = \frac{0.90b_{0.90} - 0.10a_{0.10}}{0.80}. \tag{30}$$

We shall use bounds (30) for the <u>second</u> scenario designated "Uniform" combined with the values for $a_{0.10}$ and $b_{0.90}$ in Table 1. The uniform distribution with bounds (29) actually has the smallest variance amongst pdf's (5) given the constraint set by $a_{0.10} < \theta < b_{0.90}$ and their fixed values.

For $n \to \infty$ and with specified values $a_{0.10} < \theta < b_{0.90}$, the TSP distribution (5) converges to an asymmetric Laplace distribution (13) with parameters $a_{0.10}, \theta, b_{0.90}$ and $q(\infty)$, where $q(\infty)$ is the limiting probability of being less than the mode θ and the unique solution to Equation (12). This asymmetric Laplace distribution has the largest variance amongst the TSP distributions (5) given the constraint $a_{0.10} < \theta < d_{0.90}$ and their preset values. Hence, for our <u>third</u> scenario "Laplace 1" we set $\theta = \theta_1$, specified in Table 1, and use the values $a_{0.10}$ and $b_{0.90}$ in Table 1 to determine the remaining parameter $q(\infty)$. Similarly, we obtain the <u>fourth</u> scenario "Laplace 2" by setting $\theta = \theta_2$. Note that our first two scenarios "GTU" and "Uniform" <u>are consistent</u> with the mode specifications $a_{0.10} < \theta < b_{0.90}$ in the third and fourth scenarios "Laplace 1" and "Laplace 2", respectively. That is, in all the scenarios the activity durations have the lower and upper quantiles $a_{0.10}$ and $b_{0.90}$ in common and a mode at $\theta = \theta_1$ ($\theta = \theta_2$) for the third (fourth) scenario.

Now we shall generate the cdf of the completion time distribution of the project presented in Figure 2 for each of these scenarios "GTU", "Uniform", "Laplace 1" and "Laplace 2" by employing the Monte Carlo technique (Vose 1996) involving 25,000 independent samples from the activity durations and subsequently applying the critical path method (CPM) (see e.g. Winston 1993)[4]. Consequently, for each scenario we obtain an output sample of size 25000 for the completion time of the project network in Figure 2 from which one can empirically estimates its completion time distribution. The resulting cdf's for the four scenarios described above are depicted in Figure 3. Among the scenario's in Figure 3 only the scenario "Uniform" has symmetric activity duration distributions. The activity durations of all other scenarios are all right skewed with a mean value less than that of the same activity in the "Uniform" scenario. This explains why the completion time distribution of the "Uniform" scenario is located substantially to the right of all the other scenarios. Moreover, as explained above, the variances of activity durations in the "Uniform" scenario are smaller than those of the activities in the other one. Thus it explains why its project completion time cdf is the steepest.

The largest discrepancy between the cdf's in Figure 3 occurs between the "Uniform" and "Laplace 1" and equals ≈ 0.24 observed at ≈ 194 days. Hence, certainly the specification of lower and upper quantiles $a_{0.10}$ and $b_{0.90}$ and a most likely value θ seems to be insufficient to determine a pdf in the family (5). Note that the project completion time cdf of the "GTU" scenario in Figure 3 for the most part is sandwiched between those of the "Laplace 1" and "Laplace 2" scenarios with a maximal difference of ≈ 0.04 (≈ 0.07) between its cdf and the "Laplace 1" ("Laplace 2") cdfs observed at approximately 187 days (197 days).

[4]To avoid the occurence of negative activity durations in the sampling routine as a result of the infinite support of the Laplace distributions, a negative sampled activity duration is set to be equal to zero.

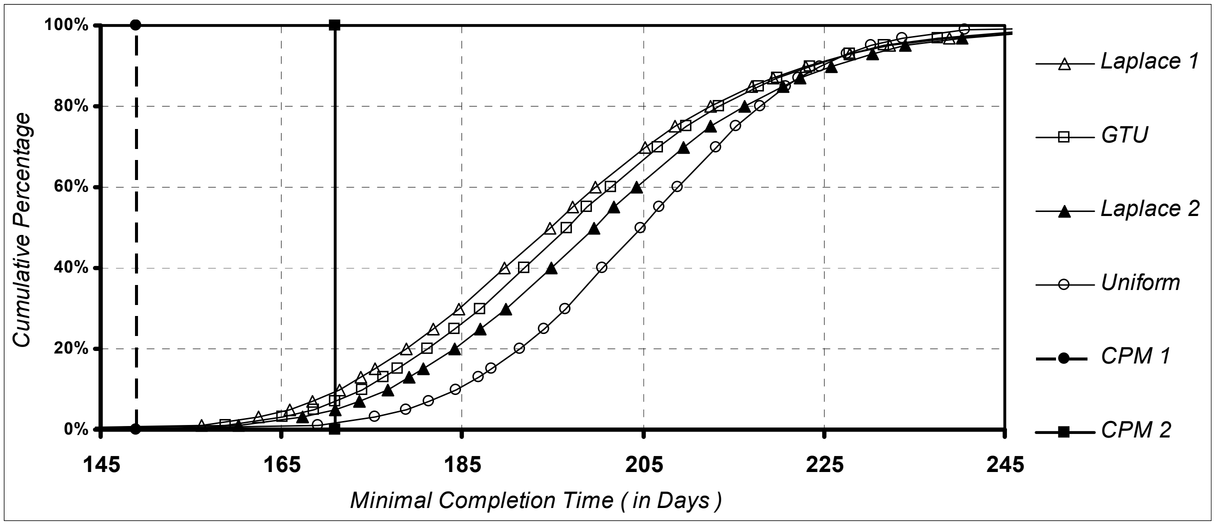

Figure 3. Comparison of CDF's of the completion times for the Project in Figure 2 for the scenarios "GTU", "Uniform", "Laplace 1" and "Laplace 2".

Finally, note that in Figure 3 the project completion time of 149 days following from the CPM using <u>only</u> the most likely values of θ_1 in Table 1, is represented by the bold vertical dotted line "CPM 1". Similarly, a completion time of 171 days follows using <u>only</u> the most likely values of θ_2 in Table 1 is indicated by the bold "CPM 2" line. Since the values of θ_1 are less than the median for all 18 activities in Table 1 (in addition to having right skewness), we observe from Figure 3 that the probability of achieving the "CPM 1" completion time of 149 days is negligible. For the "CPM 2" completion time of 171 days these probabilities are less than $\approx 10\%$ for all four scenarios. Although the skewness of the activity distributions in Table 1 may perhaps be somewhat inflated, a case could definitely be made that a skewness towards the <u>lower</u> bound may appear in assessed activity time distributions in view of a potential motivational bias of the substantive expert. These CPM results further reinforce the observation that in applications uncertainty results ought to be communicated to decision makers.

Concluding Remarks

A discussion some 50 years ago about the appropriateness of using the PERT mean and variance (2) utilizing either beta or triangular pdfs, was followed by a concern by others some 20 years later or more (e.g. Selvidge, 1980 and Keefer and Verdini, 1993) regarding the elicitation of lower and upper bounds a, b of a bounded uncertain phenomenon, since these typically do not fall within the realm of experience of an substantive expert. When instead eliciting a lower and upper quantiles a_p and b_r and a most likely value θ, however, even within the two-sided power (TSP) family of distribution with bounded support, infinitely many options exist that match these constraints. Hence, one arrives at the conclusion that additional information needs to elicited from the substantive expert for further uncertainty distribution specification. In case of the TSP family of distributions, Kotz and Van Dorp (2006) suggested the elicitation of an additional quantile to uniquely identify its lower and upper bounds a and b and power parameter n. Even when relaxing the TSP pdf or PERT requirement of specifying a single mode θ to allow for a modal range specification $[\theta_1, \theta_2]$ of a generalized trapezoidal uniform (GTU) distributions, a lower quantile $a_p < \theta_1$ and upper quantile $b_r > \theta_2$ specification is not a sufficient information to determine its lower and upper bounds $a < a_p$ and $b > b_r$ and its power parameters $m, n > 0$. Van Dorp et al. (2007) suggest to elicit in addition two relative likelihoods regarding the three stages of the GTU distribution to solve for these parameters.

Summarizing, lower and upper bounds specification or lower and upper quantiles specification combined with providing a single modal value, or even a modal range, does not uniquely determine an uncertainty distribution. In my opinion, this lack of specificity is one of the root causes regarding the controversy alluded to in the introduction of this paper surrounding the continued use of the PERT mean and variance (2) or other common arguments amongst practitioners regarding whether to use beta, triangular (or TSP) distributions to describe a bounded uncertain phenomena.

Acknowledgement

I am indebted to Samuel Kotz who has been gracious in donating his time to provide comments and suggestions in the development of various sections presented in this paper. I am also thankful to the referee whose comments improved the presentation of an earlier version.

References

Clark C.E. (1962). The PERT model for the distribution of an activity. *Operations Research*, 10, pp. 405-406.

DeWispelare A., Herren L., Clemen R.T. (1995). The use of probability elicitation in the high-level nuclear waste recognition program. *International Journal of Forecasting*, 11 (1), pp. 5 - 24.

Elmaghraby S.E. (1977). *Activity Networks: Project Planning and Control by Network Models*. Wiley, New York.

Garthwaite P.H., Kadane J.B and O'Hagan A. (2005). Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association*, 100 (470), 680-700.

Grubbs F.E. (1962). Attempts to validate certain PERT statistics or a 'Picking on PERT'. *Operations Research*, 10, 912-915.

Herrerías R. (1989). Utilización de Modelos Probabilísticos Alternativas para el Métedo PERT. Applicación al Análisis de Inversiones. *Estudios de Economía Aplicada,* Secretariado de Publicaciones de la Universidad de Valladolid, 89-112.

Herrerías R., García J. and Cruz S. (2003). A note on the reasonableness of PERT hypotheses. *Operations Research Letters*, 31, 60-62.

Herrerías J.M., R. Herrerías and Van Dorp J.R. (2009), "The Generalized Two-Sided Power Distribution", *Journal of Applied Statistics*, 36 (5), 573-587.

Kamburowski, J. (1997). New validations of PERT times. *Omega, International Journal of Management Science,* 25 (3) 323-328.

Keefer D.L. and Verdini A.V. (1993). Better estimation of PERT activity time parameters. *Management Science*, 39 (9), 1086 - 1091.

Kotz S. and Van Dorp J.R. (2004). *Beyond Beta, Other Continuous Families of Distributions with Bounded Support and Applications*. World Scientific Press, Singapore.

Kotz S. and Van Dorp J.R. (2005). A link between two-sided power and asymmetric Laplace distributions: with applications to mean and variance Approximations. *Statistics and Probability Letters*, 71, 382-394.

Kotz S. and Van Dorp J.R. (2006). A novel method for fitting unimodal continuous distributions on a bounded domain. *IIE Transactions*, 38, 421-436.

Malcolm D.G., Roseboom C.E., Clark C.E. and Fazar W. (1959). Application of a technique for research and development program evaluation. *Operations Research*, 7, 646-649.

Moder J.J. and Rodgers E.G.(1968). Judgment estimate of the moments of PERT type distributions. *Management Science*, 15 (2), B76-B83.

O'Hagan A (2006). Research in Elicitation. In *Bayesian Statistics and its Applications*, S. K. Upadhyay, U. Singh and D. K. Dey (eds.), Anamaya: New Delhi, 375-382.

Pouliquen L.Y. (1970). Risk analysis in project appraisal. *World Bank Staff Occasional Papers, 1*. John Hopkins University Press, Baltimore, MD.

Press W.H., Flannery B.P., Teukolsky S.A. and Vettering W.T. (1989). *Numerical Recipes in Pascal*. Cambridge, UK: Cambridge University Press.

Pulkkinen U. and Simola K. (2000). *An Expert Panel Approach to Support Risk-Informed Decision Making,* Sateiluturvakeskus (Radiation and Nuclear Safety Authority of Finland STUK), Technical report STUK-YTO-TR 129, Helsinki, Finland.

Selvidge, J.E. (1980). Assessing the extremes of probability distributions by the fractile method. *Decision Sciences,* 11, 493-502.

Simpson T. (1755). A letter to the Right Honourable George Earls of Maclesfield. President of the Royal Society, on the advantage of taking the mean of a number of observations in practical astronomy. *Philosophical Transactions*, 49 (1), 82-93.

Simpson T. (1757). An attempt to show the advantage∏ arising by taking the mean of a number of observations in practical astronomy. *Miscellaneous Tracts on some curious and very interesting Subjects in Mechanics, Physical Astronomy and Speculative Mathematics*, 64-75.

Taggart R. (1980). *Ship Design and Construction,* The Society of Naval Architects and Marine Engineers (SNAME), New York.

Van Dorp J.R. and Kotz S. (2002). A novel extension of the triangular distribution and its parameter estimation. *The Statistician,* 51 (1), 63-79.

Van Dorp J.R., Cruz S., García J. and Herrerías R. (2007). An Elicitation Procedure for the Generalized Trapezoidal Distribution with a Uniform Central Stage, *Decision Analysis*, 4, 156 - 166.

Vose D. (1996). *Quantitative Risk Analysis, A Guide to Monte Carlo Simulation Modeling.* Wiley*,* New York.

Winston, W.L. (1993). *Operations Research, Applications and Algorithms*. Duxbury Press, Pacific Grove, CA.