The Statistician (2002)
51, *Part* 1, *pp.* 63–79

A novel extension of the triangular distribution and its parameter estimation

J. René van Dorp and Samuel Kotz

George Washington University, Washington DC, USA

[Received January 2001. Final revision November 2001]

Summary. An extension of the three-parameter triangular distribution utilized in risk analysis is discussed. Special cases of the resulting four-parameter family include the triangular distribution, the power function distribution and the uniform distribution. Expert judgment elicitation of its parameters is discussed as well as moment estimation and maximum likelihood estimation of its parameters by using sample data.

Keywords: Expert judgment; Maximum likelihood; Moment estimation; Two-sided power distribution

1. Introduction

In recent years two papers dealing with triangular distributions and their extensions have appeared in *The Statistician*. Johnson (1997) and Johnson and Kotz (1999) dealt with neglected applications of this distribution as an alternative to the beta distribution which suffers from difficulties involved in its maximum likelihood parameter estimation and whose parameters do not have a clear-cut meaning. Johnson (1997) is particularly relevant to the current work. Johnson used triangular distributions as a proxy to the beta distribution, specifically in problems of assessment of risk and uncertainty, such as the project evaluation and review technique (PERT). The parameters of a triangular distribution have a one-to-one correspondence with an optimistic estimate a, most likely estimate m and pessimistic estimate b of a quantity under consideration, providing to the triangular distribution its intuitive appeal (see, for example Williams (1992)). Similarly to the beta distribution, the triangular distribution can be positively or negatively skewed (or symmetrical) but must remain unimodal. Johnson (1997) pointed out that there is no triangular distribution which would reasonably approximate uniform, J-shaped or U-shaped distributions.

In this paper we investigate an extension of the three-parameter triangular distribution, to be called the two-sided power (TSP) distribution, as a meaningful alternative to the beta distribution. The four-parameter distribution proposed herein does allow for J-shaped and U-shaped forms. Since the TSP distribution extends the triangular distribution it should inherit its intuitive appeal and meaningful parameters. In addition, TSP distributions have the attractive property that their maximum likelihood estimation (MLE), although ingenious, is computationally straightforward and apparently robust, involving only elementary functions. Also, the cumulative distribution function of a TSP variable and its inverse can be derived in closed form,

Address for correspondence: J. René van Dorp, Department of Engineering Management and System Engineering, School of Engineering and Applied Science, George Washington University, 707 22nd Street North West, Washington DC 20052, USA.
E-mail: dorpjr@seas.gwu.edu

 0039–0526/02/51063

allowing straightforward and efficient use of this distribution in Monte-Carlo-type uncertainty or risk analyses. It has been brought to our attention by one of the referees that because of its flexibility the TSP family may also serve as a rich family of prior distributions in a Bayesian analysis. In such analyses parameters need to be assessed as well, either by using expert judgment or data. This paper concentrates on the estimation of the parameters of the TSP distribution by using a variety of methods. Some properties of the two-parameter TSP distribution with support $[0, 1]$ have been discussed in van Dorp and Kotz (2002).

In Section 2, the TSP distribution is briefly introduced. We discuss an expert judgment elicitation method for its four parameters in Section 3 followed by the method of moments for a TSP distribution with known support in Section 4. The method of moments discussed herein uses a classical method for solving cubic equations known as Cardano's method. This method was first discovered by Tartaglia in 1539 and later published by Cardano in 1545 (see Cardano (1993) for full details). In Section 5, the two-parameter MLE procedure for TSP distributions with known support is briefly reviewed. To the best of our knowledge MLE of the three-parameter triangular distribution has not been investigated, but it follows naturally from the two-parameter MLE procedure described in Section 5 and will be discussed in Section 6. In Section 7, we expand the three-parameter MLE procedure in Section 6 to an MLE procedure for the four-parameter TSP distribution. Finally, we provide some concluding remarks in Section 8. To the best of our knowledge the TSP family is mentioned only in passing in Nadarajah (1999). We could not locate other literature citations.

2. Two-sided power distributions

Let X be a random variable with probability density function given by

$$f(x|a, m, b, n) = \begin{cases} \dfrac{n}{b-a}\left(\dfrac{x-a}{m-a}\right)^{n-1} & a < x \leqslant m, \\ \dfrac{n}{b-a}\left(\dfrac{b-x}{b-m}\right)^{n-1} & m \leqslant x < b. \end{cases} \tag{1}$$

The random variable X is said to follow a TSP distribution, $\mathrm{TSP}(a, m, b, n)$, $a \leqslant m \leqslant b, n > 0$. For $n > 1$, the mode of the density function is at m and the value of the probability density function at the mode is always $n/(b-a)$. For $0 \leqslant n < 1$ and $a < m < b$ the mode of the density function is at a or b and $f(\cdot|a, m, b, n) \to \infty$ at its modes. For $n = 1$, $f(\cdot|a, m, b, n)$ simplifies to a uniform$[a, b]$ distribution. For $n = 2$, $f(\cdot|a, m, b, n)$ reduces to a triangular distribution triang(a, m, b). Finally, for $a = 0$ and $m = b = 1$, $f(\cdot|a, m, b, n)$ corresponds to a power function distribution and for $a = m = 0$ and $b = 1$ to its reflection. Fig. 1 provides examples of symmetric $\mathrm{TSP}(0, m, 1, n)$ distributions, i.e. $m = 0.5$, including uniform, triangular and some U-shaped distributions. Fig. 2 presents examples of positively and negatively skewed $\mathrm{TSP}(0, m, 1, n)$ distributions, including examples of triangular distributions. Finally, Fig. 3 provides examples of J-shaped $\mathrm{TSP}(0, m, 1, n)$ distributions. van Dorp and Kotz (2002) refer to TSP distributions on $[0, 1]$ as standard TSP distributions.

The cumulative distribution function of a $\mathrm{TSP}(a, m, b, n)$ distribution follows from expression (1) as

$$F(x|a, m, b, n) = \begin{cases} \dfrac{m-a}{b-a}\left(\dfrac{x-a}{m-a}\right)^{n} & a \leqslant x \leqslant m, \\ 1 - \dfrac{(b-m)}{(b-a)}\left(\dfrac{b-x}{b-m}\right)^{n} & m \leqslant x \leqslant b. \end{cases} \tag{2}$$

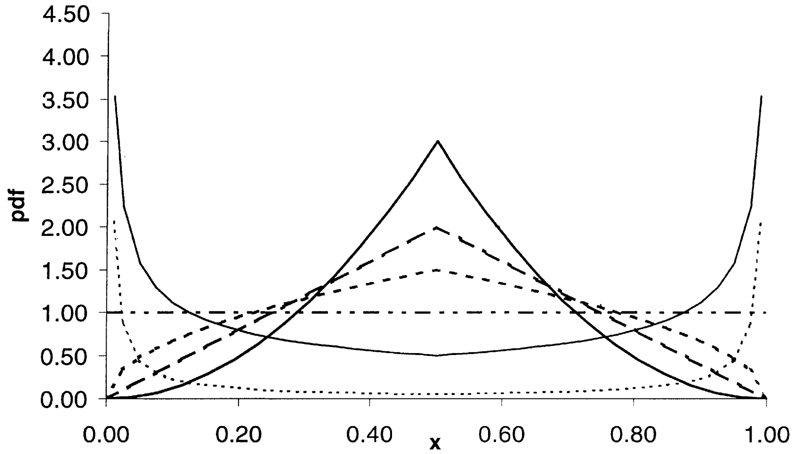

Fig. 1. Symmetric TSP$(0, m, 1, n)$ distributions $(m=0.5)$: ······, $n=0.05$; ———, $n=0.5$; · · — · · , $n=1$; - - - - - -, $n=1.5$; — —, $n=2$; ———, $n=3$

The expressions for the mean and the variance can be obtained from expression (1) and simplify to

$$E(X) = \frac{a + (n-1)m + b}{n+1} \tag{3}$$

and

$$\operatorname{var}(X) = (b-a)^2 \frac{n - 2(n-1)(m-a)/(b-a) \times (b-m)/(b-a)}{(n+2)(n+1)^2}. \tag{4}$$

The meaning of the parameters is as follows: a and b are the end points of the support, n is the shape parameter and m is the threshold parameter for a change in the form of the probability density function. The parameters a and b may be related to pessimistic and optimistic estimates of the associated TSP(a, m, b, n) variable. For $n > 1$, m coincides with the most likely estimate (the mode) of a TSP(a, m, b, n) variable. These interpretations mirror those for the three-parameter triang(a, m, b) distribution. Regardless of the value of n, the parameter m identifies the $100(m-a)/(b-a)$th percentile of the TSP distribution.

3. Elicitation of two-sided power parameters

Johnson (1997) proposed the triangular distribution as an alternative to the beta distribution (see, for example, Johnson *et al.* (1995)) as its parameters have a one-to-one correspondence with an optimistic estimate a, a most likely estimate m and a pessimistic estimate b of an activity duration T in a PERT network. Over 40 years ago, Malcolm *et al.* (1959) fitted a beta distribution to such estimates a, m and b by using the method of moments to overcome the difficulty with interpreting the beta parameters by setting

$$\begin{aligned} E(T) &= \frac{a + 4m + b}{6}, \\ \operatorname{var}(T) &= \frac{1}{36}(b-a)^2. \end{aligned} \tag{5}$$

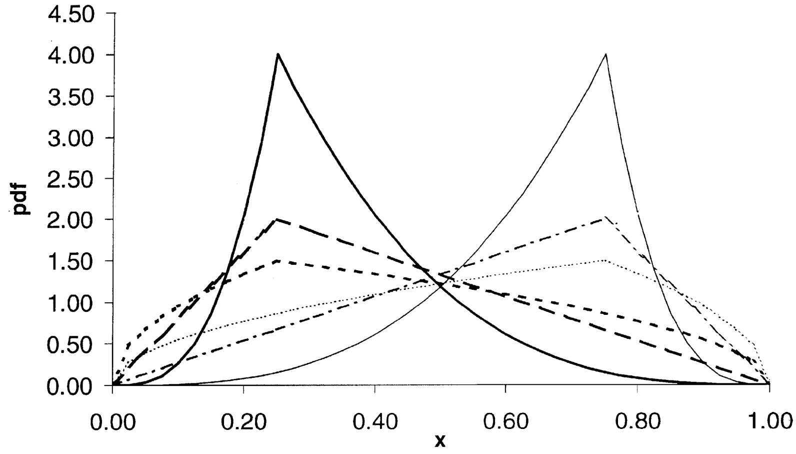

Fig. 2. Positively skewed TSP(0, m, 1, n) distributions ($m > 0.5$ and $n > 1$) and negatively skewed TSP(0, m, 1, n) distributions ($m < 0.5$ and $n > 1$): - - - - -, $m = 0.25$, $n = 1.5$; — —, $m = 0.25$, $n = 2$; ———, $m = 0.25$, $n = 4$; ······, $m = 0.75$, $n = 1.5$ · — · — ·, $m = 0.75$, $n = 2$; ———, $m = 0.75$, $n = 4$

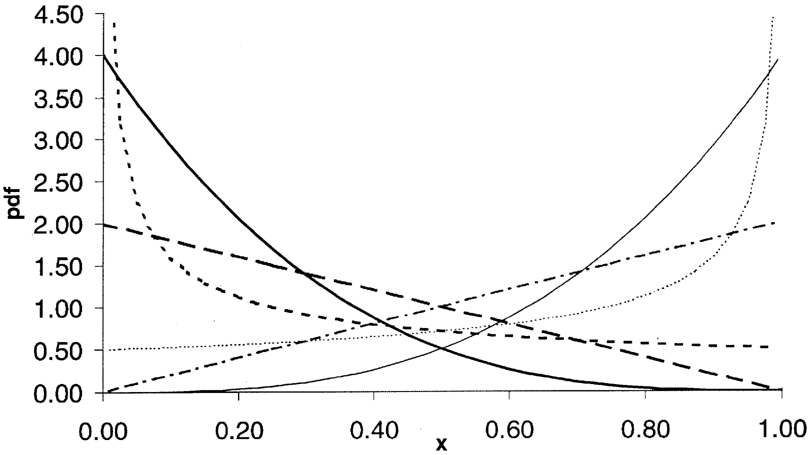

Fig. 3. Some J-shaped TSP(0, m, 1, n) distributions: ······, $m = 0$, $n = 0.5$; — —, $m = 0$, $n = 2$; ———, $m = 0$, $n = 5$; - - - - -, $m = 1$, $n = 0.5$;· — · — ·, $m = 1$, $n = 2$; ———, $m = 1$, $n = 5$

Solving for the beta parameters using expression (5) is somewhat controversial (see for example Clark (1962) and Grubbs (1962)) and its use is still (see for example Kamburowksi (1997)) subject to discussion. From equations (3) and (4) it follows that for a triangular distribution ($n = 2$)

$$E(X) = \frac{a+m+b}{3}, \tag{6}$$

$$\frac{3}{72}(b-a)^2 \leqslant \operatorname{var}(X) \leqslant \frac{1}{18}(b-a)^2. \tag{7}$$

From equation (6) it follows that $E(X)$ may overestimate or underestimate $E(T)$ in expression (5) depending on whether m is less or greater than the midpoint $(a+b)/2$. More importantly, it follows from inequality (7) that $\operatorname{var}(X)$ of a triangular distribution is always larger than $\operatorname{var}(T)$ in expression (5), regardless of the values for a, m and b. This may partially explain the existing controversy around using expression (5) to fit the beta parameters.

Instead of using a triangular distribution we may also use the more general TSP(a, m, b, n) distribution as an alternative to the beta distribution. Specifically, setting $n = 5$ and utilizing equations (3) and (4) we have

$$E(X) = \frac{a + 4m + b}{6}, \tag{8}$$

$$\frac{1}{84}(b - a)^2 \leqslant \operatorname{var}(X) \leqslant \frac{5}{252}(b - a)^2. \tag{9}$$

Hence, in the latter case the mean values $E(T)$ in expression (5) and $E(X)$ in equation (8) agree, but now $\operatorname{var}(X)$ in inequality (9) is always less than $\operatorname{var}(T)$ in expression (5), regardless of the values for a, m and b, perhaps adding additional fuel to the controversy surrounding the use of expression (5).

In place of using comparisons with expression (5) to specify the parameter n of a TSP(a, m, b, n) distribution, we may indirectly elicit n by asking an expert for the relative importance of the already elicited most likely value m compared with the bounds a or b. From equation (3) it follows that $n + 1$ may be interpreted as the sample size of a virtual sample with $n - 1$ observations m, with one additional observation a and one additional observation b. Suppose that an expert assigns the relative importance of the most likely value m to be $y = n - 1$; it then follows from the interpretation above that

$$n = y + 1. \tag{10}$$

Hence, if an expert responds that the most likely estimate m is as important as the bounds a or b (i.e. $y = 1$), it follows from equation (10) that a triangular distribution ($n = 2$) models the expert's uncertainty. If an expert responds that the most likely estimate m is more (or, correspondingly, less) important than the bounds a or b, the elicitation will yield a TSP distribution with variance smaller (or, correspondingly, larger) than that of the triangular distribution. An expert would have to assign relative importance $y = 4$ for the mean of the resulting TSP variable $E(X)$ to agree with $E(T)$ in expression (5).

In the following sections, more classical estimation procedures for the TSP(a, m, b, n) distribution are discussed using data.

4. Two-parameter moment estimation

Consider a hypothetical example in Table 1 of eight completed replications $x_i, i = 1, \ldots, 8$, of an activity duration in a PERT context known to take at least 2 hours and not more than 12 hours. The third column of Table 1 contains standardized replications $x_i' = (x_i - 2)/10$.

Without loss of generality, we may apply the method of moments to a TSP$(0, m', 1, n)$ variable using the x_i'-values in Table 1 (instead of to a TSP(a, m, b, n) variable using the original data) with $m' = (m - a)/(b - a)$ and fixed $a = 2$ and $b = 12$. The advantage is that population quantities $E(X)$ and $\operatorname{var}(X)$ (see equations (3) and (4)) simplify to

$$\begin{aligned} E(X) &= \frac{(n - 1)m' + 1}{n + 1}, \\ \operatorname{var}(X) &= \frac{n - 2(n - 1)\, m'(1 - m')}{(n + 2)(n + 1)^2}. \end{aligned} \tag{11}$$

Table 1. Replications of an activity duration known to be completed within 2–12 hours

i	*Original data* x_i	*Standardized or scaled data* x'_i
1	3	0.10
2	4.5	0.25
3	5	0.30
4	6	0.40
5	6.5	0.45
6	8	0.60
7	9.5	0.75
8	10	0.80

Equating sample quantities

$$\bar{x}' = \frac{1}{8}\sum_{i=1}^{8} x'_i,$$

$$\hat{\sigma}^2 = \frac{1}{7}\sum_{i=1}^{8}(x'_i - \bar{x}')^2$$

with their population equivalents (see expression (11)), we obtain the following cubic equation in parameter n:

$$cn^3 + dn^2 + en + f = 0, \tag{12}$$

where the coefficients are given by

$$\left.\begin{array}{r} c = \frac{1}{2}\hat{\sigma}^2, \\ d = \hat{\sigma}^2, \\ e = -\{(\bar{x}' - \frac{1}{2})^2 + \frac{1}{2}\hat{\sigma}^2 + \frac{1}{4}\}, \\ f = \frac{1}{4} - (\bar{x}' - \frac{1}{2})^2 - \hat{\sigma}^2. \end{array}\right\} \tag{13}$$

After solving the cubic equation for n, the estimate of m' follows from

$$\hat{m}' = \frac{1}{2} + \frac{n+1}{n-1}\left(\bar{x}' - \frac{1}{2}\right) \tag{14}$$

and that of m is given by

$$\hat{m} = \hat{m}'(b-a) + a. \tag{15}$$

Expression (12) may be solved by using Cardano's method briefly described in Appendix A. Utilizing the standardized data in Table 1 we have $\bar{x}' = 0.456\,25$ and $\hat{\sigma}^2 = 0.060\,313$. Substituting these values in equations (13) yields $c = 0.030\,156$, $d = 0.060\,3125$, $e = -0.282\,070$ and $f = 0.187\,773$. Using the values for c, d, e and f in equation (44) (Appendix A) and solving for

p and q yields $p = -3.562320$ and $q = 6.527513$. With $q^2 + p^3 = -2.597847$, the following three real-valued solutions are calculated (see equation (43) and the left-hand side column of Table 4 in Appendix A):

$$\hat{y}_1 = -3.762539, \qquad \hat{y}_2 = 2.144776, \qquad \hat{y}_3 = 1.617762.$$

Utilizing the values for c, d and equation (43) yields the following three solutions for equation (12):

$$\hat{n}_1 = -4.429201, \qquad \hat{n}_2 = 1.478109, \qquad \hat{n}_3 = 0.951095.$$

The first solution $\hat{n}_1$ is unacceptable since we must have $n > 0$. Hence, we solve for $\hat{m}'_2$ and $\hat{m}'_3$ using $\hat{n}_2$, $\hat{n}_3$ and equation (14), yielding

$$\hat{m}'_2 = 0.273, \qquad \hat{m}'_3 = 2.245.$$

The third solution is disqualified since we must have $0 < m' < 1$. Substituting $a = 2$, $b = 12$ and $\hat{m}'_2$ in equation (15) the unique moment estimators are

$$\begin{aligned} \hat{m}_2 &= 4.7308, \\ \hat{n}_2 &= 1.478109. \end{aligned} \tag{16}$$

In the next section, we shall briefly review two-parameter MLE for a TSP(a, m, b, n) distribution with fixed a and b using the original data in Table 1.

5. Maximum likelihood estimation procedure for two-parameter two-sided power distribution

For a random sample $\mathbf{X} = (X_1, \ldots, X_s)$ with size s from a TSP(a, m, b, n) distribution, let the order statistics be $X_{(1)} < X_{(2)} \ldots < X_{(s)}$. Utilizing expression (1), the likelihood for $\mathbf{X}$ is by definition

$$L(\mathbf{X}; a, m, b, n) = \left(\frac{n}{b-a}\right)^s H(\mathbf{X}; a, m, b)^{n-1}, \tag{17}$$

where

$$H(\mathbf{X}; a, m, b) = \frac{\prod_{i=1}^{r} (X_{(i)} - a) \prod_{i=r+1}^{s} (b - X_{(i)})}{(m-a)^r (b-m)^{s-r}}, \tag{18}$$

$X_{(0)} \equiv a$, $X_{(s+1)} \equiv b$ and r is implicitly defined by $X_{(r)} \leqslant m < X_{(r+1)}$. From the structure of equation (17) it follows that the two-parameter MLE procedure maximizing equation (17) as a function of m and n (with a and b fixed) is a two-stage optimization problem, namely we may first determine $\hat{m}$ at which equation (18) attains its maximum as a function of m. Next, utilizing $\hat{m}$, we may calculate $\hat{n}$ maximizing $L(\mathbf{X}; a, \hat{m}, b, n)$ (see equation (17)) as a function of n. Fig. 4 displays equation (18) as a function of m for the original data in Table 1 with fixed $a = 2$ and $b = 12$. (We are indebted to the referee for providing a draft of Fig. 4.)

van Dorp and Kotz (2001) proved that equation (18) attains its maximum, for fixed a and b, at one of the order statistics $(X_{(1)}, \ldots, X_{(s)})$. Specifically,

$$\hat{m}(a, b) = X_{(\hat{r}(a,b))} \tag{19}$$

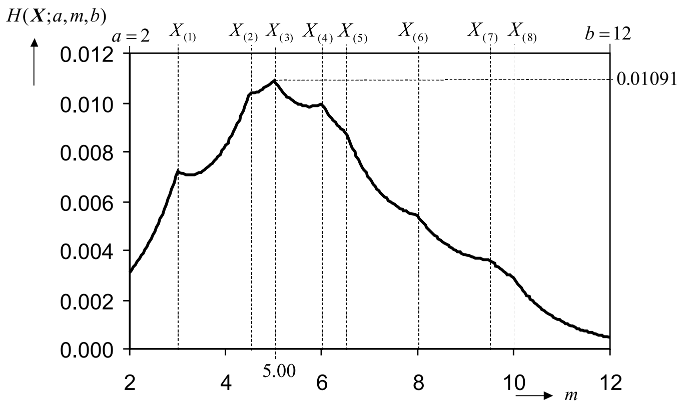

Fig. 4. Graph of $H(\mathbf{X}; a, m, b)$ (see equation (18)) for the original data in Table 1 with fixed $a = 2$ and $b = 12$

where

$$\hat{r}(a, b) = \underset{r \in \{1,...,s\}}{\arg\max} \{M(a, b, r)\}, \tag{20}$$

and

$$M(a, b, r) = \prod_{i=1}^{r-1} \frac{X_{(i)} - a}{X_{(r)} - a} \prod_{i=r+1}^{s} \frac{b - X_{(i)}}{b - X_{(r)}}. \tag{21}$$

(For notational convenience the arguments a and b will occasionally be suppressed in $\hat{r}(a,b)$.) Next, utilizing $H\{\mathbf{X}; a, \hat{m}(a, b), b\} = M\{a, b, \hat{r}(a, b)\}$, the maximum likelihood estimator of n, $\hat{n}(a, b)$ say, maximizing equation (17) together with the maximum likelihood estimator of m, $\hat{m}(a, b)$ (see equation (19)), is

$$\hat{n}(a, b) = -\frac{s}{\log[M\{a, b, \hat{r}(a, b)\}]}. \tag{22}$$

van Dorp and Kotz (2002) applied the two-parameter MLE procedure above (with fixed a and b) to estimate uncertainty in 294 observed monthly differences of interest data on 30-year Treasury bonds over the period 1977–2001. The procedure for $\hat{m}(a, b)$ (see equation (19)) with fixed a and b is almost identical with the MLE procedure for a one-parameter triang$(0, m, 1)$ distribution discussed in Johnson and Kotz (1999). The standardized data in Table 1 coincide with the example in Johnson and Kotz (1999) and for comparison we demonstrate the procedure using the latter example. Consider the matrix $C = (c_{i,r})$ in Table 2, where

$$c_{i,r} = \begin{cases} \dfrac{X_{(i)} - a}{X_{(r)} - a} & i < r, \\ \dfrac{b - X_{(i)}}{b - X_{(r)}} & i \geqslant r \end{cases}$$

is calculated utilizing the original data in Table 1 with $a = 2$ and $b = 12$. The last row in Table 2 contains products of the matrix elements in the rth column which are equal to the values of

Table 2. Example of MLE for TSP(*a, m, b, n*) with fixed *a* and *b*†

i	$X_{(i)}$	*Estimates for the following values of r and* $X_{(r)}$:							
		r							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
		$X_{(r)}$							
		3.00	*4.50*	*5.00*	*6.00*	*6.50*	*8.00*	*9.50*	*10.00*
1	3.00	1.00000	0.40000	0.33333	0.25000	0.22222	0.16667	0.13333	0.12500
2	4.50	0.83333	1.00000	0.83333	0.62500	0.55556	0.41667	0.33333	0.31250
3	5.00	0.77778	0.93333	1.00000	0.75000	0.66667	0.50000	0.40000	0.37500
4	6.00	0.66667	0.80000	0.85714	1.00000	0.88889	0.66667	0.53333	0.50000
5	6.50	0.61111	0.73333	0.78571	0.91667	1.00000	0.75000	0.60000	0.56250
6	8.00	0.44444	0.53333	0.57143	0.66667	0.72727	1.00000	0.80000	0.75000
7	9.50	0.27778	0.33333	0.35714	0.41667	0.45455	0.62500	1.00000	0.93750
8	10.00	0.22222	0.26667	0.28571	0.33333	0.36364	0.50000	0.80000	1.00000
$M(a,b,r)$		0.00725	0.01038	0.01091	0.00995	0.00879	0.00543	0.00364	0.00290

†The values in Johnson and Kotz (1999) corresponding to the last three entries in the last row contain the following typographical errors: 0.00547 should read 0.00543, 0.00137 should replace 0.00364 and 0.00029 should be 0.00290.

$M(a, b, r)$ (see equation (21)). It follows immediately that

$$\left.\begin{aligned} \hat{r}(2, 12) &= 3, \\ \hat{M}(2, 12, \hat{r}) &= 0.01091, \\ \hat{m}(2, 12) &= X_{(3)} = 5.00, \\ \hat{n}(2, 12) &= \frac{-8}{\log\{\hat{M}(2, 12, \hat{r})\}} = 1.77060. \end{aligned}\right\} \tag{23}$$

Compare these values with the global maximum of $H(\mathbf{X}; a, m, b)$ depicted in Fig. 4 attained at $X_{(3)}$. Note that $\hat{n}(2, 12) < 2$, the value of n for which a TSP(a, m, b, n) distribution reduces to a triang(a, m, b) distribution. This illustrates the added flexibility of the two-parameter MLE procedure in our case. Finally, the value of $\hat{n}(2, 12) < 2$ is consistent with the moment estimates in Section 4 (see equation (16)).

Before considering the MLE procedure for the four-parameter TSP(a, m, b, n) distribution, it is helpful to discuss the MLE procedure for the three-parameter triang(a, m, b) distribution which, to the best of our knowledge, has not been tackled before.

6. Maximum likelihood estimation procedure for three-parameter triangular distribution

A TSP$(a, m, b, 2)$ distribution reduces to a triang(a, m, b) distribution. Thus, analogously to Section 5 for given values of $a < X_{(1)}$ and $b > X_{(s)}$ of a triang(a, m, b) variable X, we have

$$\max_{a<m<b}\{L(\mathbf{X}; a, m, b, 2)\} = \left(\frac{2}{b-a}\right)^s M\{a, b, \hat{r}(a, b)\}, \tag{24}$$

where $L(\mathbf{X}; a, m, b, 2)$, $M(a, b, r)$ and $\hat{r}(a, b)$ are given by equations (17), (21) and (20) respectively. It thus follows with equation (24) that

$$\max_{S(a,m,b)} \left[\log\{L(\mathbf{X}; a, m, b)\}\right] = \max_{a<X_{(1)},\, b>X_{(s)}} \{s \log(2) + G(a, b)\}, \tag{25}$$

where $S(a, m, b) = \{(a, m, b) | a < X_{(1)}, b > X_{(s)}, a < m < b\}$ and

$$G(a, b) = \log[M\{a, b, \hat{r}(a, b)\}] - s \log(b - a). \tag{26}$$

Note that the function $G(a, b)$ is defined for values of a and b such that $a < X_{(1)}$ and $b > X_{(s)}$.

Fig. 5 shows the form of the function $G(a, b)$ given by equation (26) for the original data in Table 1. $G(a, b)$ is continuous, but partial derivatives with respect to a or b may not be defined at a finite number of points (specifically $s - 1$) since $\hat{r}(a, b)$ is given by equation (20). The following properties can be derived for $\hat{r}(a, b)$ (see equation (20)) as a function of b, keeping $a < X_{(1)}$ fixed:

(a) $\hat{r}(a, b)$ is decreasing in b;
(b) $\lim_{b \downarrow X_{(s)}}\{\hat{r}(a, b)\} = s$;
(c) $\lim_{b \to \infty}\{\hat{r}(a, b)\} = 1$;
(d) $\hat{r}(a, b)$ has $s - 1$ discontinuities at

$$f_b(a, r) = \frac{X_{(r+1)} - X_{(r)} \sqrt[s-r]{\{(X_{(r)} - a)/(X_{(r+1)} - a)\}^r}}{1 - \sqrt[s-r]{\{(X_{(r)} - a)/(X_{(r+1)} - a)\}^r}}, \qquad r \in \{1, \ldots, s - 1\}. \tag{27}$$

Similar properties can be derived for $\hat{r}(a, b)$ as a function of a, keeping $b > X_{(s)}$ fixed. Fig. 6 gives the form of the function $\hat{r}(a, b)$ for the original data in Table 1. The function $\hat{r}(a, b)$ may be viewed as a bivariate step function or a *winding staircase function,* which could serve as a promising concept for studying non-differentiable bivariate distributions.

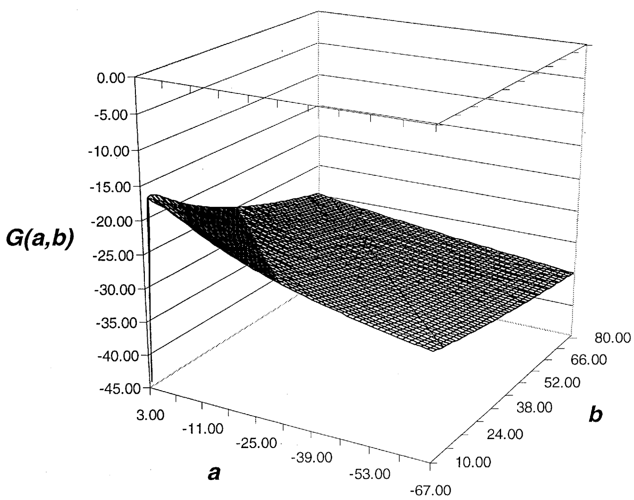

Fig. 5. Function $G(a, b)$ given by equation (26) for the original data in Table 1

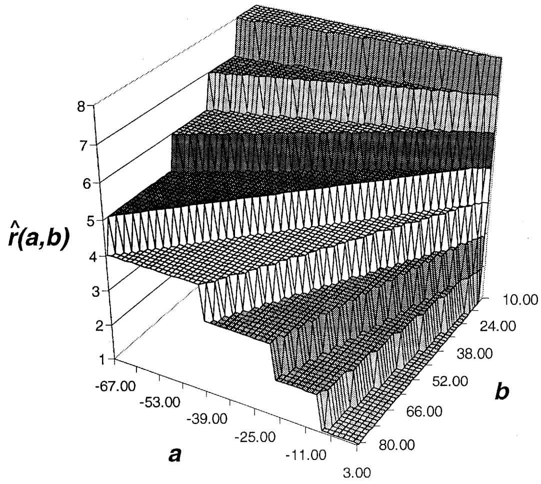

Fig. 6. Function $\hat{r}(a,b)$ given by equation (20) for the original data in Table 1

Barring the points of discontinuity of the function $\hat{r}(a, b)$, the function $G(a, b)$ is differentiable with respect to a and b. From equation (26) we obtain

$$\frac{\partial}{\partial a}G(a, b) = \frac{\partial M\{a, b, \hat{r}(a, b)\}/\partial a}{M\{a, b, \hat{r}(a, b)\}} + \frac{s}{b - a}, \tag{28}$$

$$\frac{\partial}{\partial b}G(a, b) = \frac{\partial M\{a, b, \hat{r}(a, b)\}/\partial b}{M\{a, b, \hat{r}(a, b)\}} - \frac{s}{b - a}, \tag{29}$$

where

$$\frac{\partial}{\partial a}M\{a, b, \hat{r}(a, b)\} = M\{a, b, \hat{r}(a, b)\} \sum_{j=1}^{\hat{r}-1} \frac{X_{(j)} - X_{(\hat{r})}}{(X_{(\hat{r})} - a)(X_{(j)} - a)} < 0 \tag{30}$$

and

$$\frac{\partial}{\partial b}M\{a, b, \hat{r}(a, b)\} = M\{a, b, \hat{r}(a, b)\} \sum_{j=\hat{r}+1}^{s} \frac{X_{(j)} - X_{(\hat{r})}}{(b - X_{(\hat{r})})(b - X_{(j)})} > 0. \tag{31}$$

A routine `BSearch` has been developed utilizing equations (27), (29) and (31) to determine $\hat{b}(a)$ for fixed a, where

$$\hat{b}(a) = \arg\max_{b > X_{(s)}}\{G(a, b)\}. \tag{32}$$

This routine follows a bisection approach (see, for example Press *et al.* (1989)) and is described in Appendix B. With `BSearch` to determine $\hat{b}(a)$ for fixed a, a routine `ABSearch` determines $\hat{a}$ and $\hat{b}(\hat{a})$ such that

$$\hat{a} = \arg\max_{a < X_{(1)}}[G\{a, \hat{b}(a)\}].$$

The latter routine utilizes equations (28) and (30), also follows a bisection approach and is described in Appendix B as well. The routine `ABSearch` evaluates equation (25) by successively utilizing `BSearch` and yields maximum likelihood estimators

$$\hat{a}, \qquad \hat{b} = \hat{b}(\hat{a}), \qquad \hat{m}(\hat{a}, \hat{b}) = X_{(\hat{r}(\hat{a}, \hat{b}))},$$

where $\hat{b}(\cdot)$ and $\hat{r}(a, b)$ are given by equations (32) and (20) respectively. For the original data in Table 1 `ABSearch` calculates the following maximum likelihood estimates:

$$\left.\begin{aligned} \hat{a} &= 2.0762; \\ \hat{b} &= 11.9393; \\ \hat{m}(\hat{a}, \hat{b}) &= 5.00000. \end{aligned}\right\} \tag{33}$$

Note that $\hat{a} < X_{(1)} = 3.0$ and $\hat{b} > X_{(s)} = 10.00$. Observe that $\hat{m}(\hat{a}, \hat{b})$ in equation (33) for a triang(a, m, b) distribution coincides with the maximum likelihood estimator $\hat{m}(2, 12)$ in equation (23) for the two-parameter TSP$(2, m, 12, n)$ distribution with fixed $a = 2$ and $b = 12$.

7. Maximum likelihood estimation procedure for four-parameter two-sided power distribution

Only minor modifications are needed in the procedure for the three-parameter triang(a, m, b) distribution established above to derive the MLE procedure for the four-parameter TSP(a, m, b, n) distribution. Analogously to Section 5, it follows that, for given values of $a < X_{(1)}$ and $b > X_{(s)}$ of a TSP(a, m, b, n) variable X, we have

$$\max_{n>0, a<m<b} \{L(\mathbf{X}; a, m, b, n)\} = \max_{n>0} \left(\frac{n}{b-a}\right)^s \left[M\{a, b, \hat{r}(a, b)\}\right]^{n-1}, \tag{34}$$

where, as above, $L(\mathbf{X}; a, m, b, n)$, $M(a, b, r)$ and $\hat{r}(a, b)$ are given by equations (17), (21) and (20) respectively. The global maximum in equation (34) is attained at

$$\hat{n}(a, b) = -\frac{s}{\log[M\{a, b, \hat{r}(a, b)\}]} \tag{35}$$

for given values of a and b (see equation (22)). It thus follows from equations (34) and (35) that

$$\max_{S(a,m,b,n)} \left[\log\{L(\mathbf{X}; a, m, b, n)\}\right] = \max_{a<X_{(1)}, b>X_{(s)}} \{G(a, b)\}, \tag{36}$$

where $S(a, m, b, n) = \{(a, m, b, n) | \, a < X_{(1)}, b > X_{(s)}, a < m < b, n > 0\}$ and

$$G(a, b) = s\left[\log\left\{\frac{\hat{n}(a, b)}{b-a}\right\} - \frac{\hat{n}(a, b) - 1}{\hat{n}(a, b)}\right]. \tag{37}$$

We are using the same notation for equation (37) as in equation (26) since both functions play a similar role in the MLE routine `ABSearch` described in Appendix B. Barring the points of discontinuity of $\hat{r}(a, b)$ (see equation (20)), it follows that

$$\frac{\partial}{\partial a} G(a, b) = \frac{\partial M(a, b)/\partial a}{M(a, b)} \left[\frac{-s}{\log\{M(a, b)\}} - 1\right] + \frac{s}{b-a}, \tag{38}$$

$$\frac{\partial}{\partial b} G(a, b) = \frac{\partial M(a, b)/\partial b}{M(a, b)} \left[\frac{-s}{\log\{M(a, b)\}} - 1\right] - \frac{s}{(b-a)}. \tag{39}$$

where $\partial M(a, b)/\partial a$ and $\partial M(a, b)/\partial b$ are given by equations (30) and (31) respectively. The routine to evaluate equation (36) is analogous to the routine `ABSearch` presented in the previous section to evaluate equation (25) and is described in Appendix B by making appropriate changes with respect to $G(a, b)$, $\partial G(a, b)/\partial a$ and $\partial G(a, b)/\partial b$.

Fig. 7 provides the form of the function $G(a, b)$ given by equation (37) for the original data in Table 1. The peak in Fig. 7 coincides with the point $(a, b) = (X_{(1)}, X_{(s)})$ and for these data `ABSearch` calculates the following maximum likelihood estimates:

$$\left.\begin{aligned} \hat{a} &= 3.0000; \\ \hat{b} &= 10.0000; \\ \hat{m}(\hat{a}, \hat{b}) &= 10.0000; \\ \hat{n}(\hat{a}, \hat{b}) &= 0.2632. \end{aligned}\right\} \tag{40}$$

In equations (40), $\hat{m}(\hat{a}, \hat{b}) = X_{(s)}$, which is perhaps a surprising result comparing the numerical values for $\hat{m}(\cdot, \cdot)$ in equations (23) and (33). This discrepancy may be attributed to using a small sample size of 8 to estimate four parameters.

To test the four-parameter MLE procedure for a larger sample, consider the sample of size 50 in Table 3 from a TSP(2, 7, 12, 2.5) distribution. Fig. 8 provides the form of the function $G(a, b)$ given by equation (37) for the data in Table 3.

For these values the routine `ABSearch` in Appendix B calculates the following maximum likelihood estimates:

$$\left.\begin{aligned} \hat{a} &= 2.7694; \\ \hat{b} &= 11.3742; \\ \hat{m}(\hat{a}, \hat{b}) &= 7.100; \\ \hat{n}(\hat{a}, \hat{b}) &= 2.5630. \end{aligned}\right\} \tag{41}$$

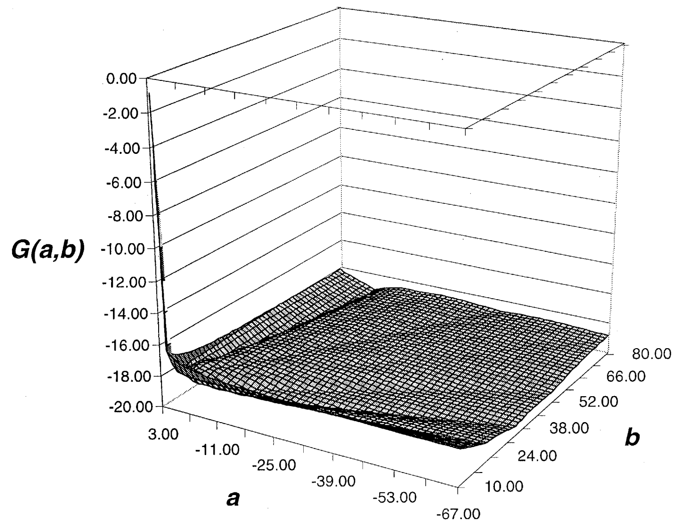

Fig. 7. Function $G(a,b)$ given by equation (37) for the original data in Table 1

Table 3. Ordered sample (of size 50) from a TSP(2,7,12, 2.5) distribution

3.469	5.820	6.846	7.323	8.615
4.366	6.040	6.871	7.469	8.893
4.639	6.210	6.964	7.541	8.991
4.770	6.422	7.001	7.687	9.035
5.025	6.448	7.064	7.714	9.265
5.040	6.515	7.100	7.720	9.280
5.132	6.537	7.129	7.743	9.340
5.317	6.681	7.204	8.076	9.350
5.326	6.721	7.270	8.347	9.651
5.625	6.743	7.300	8.459	10.662

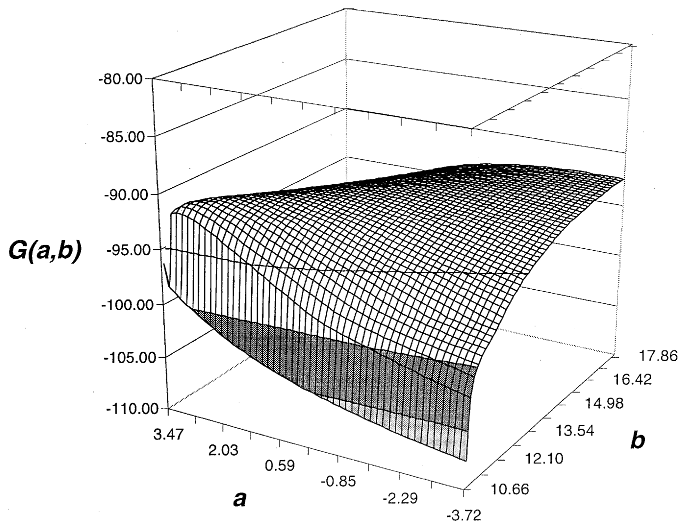

Fig. 8. Function $G(a,b)$ given by equation (37) for the data in Table 3

Note that $\hat{a} < X_{(1)} = 3.469$ and $\hat{b} > X_{(s)} = 10.662$. Compare this with the estimators of these two parameters for a smaller sample of size 8 given in equation (33). In the present case the estimators $\hat{a}$ and $\hat{b}$ are somewhat closer to the corresponding extreme order statistics. In addition, it may be observed that the maximum likelihood estimates in equation (41) are quite consistent with the parameters of the TSP(2, 7, 12, 2.5) distribution from which the sample in Table 3 was generated.

8. Concluding remarks

A new four-parameter family of TSP(a, m, b, n) distributions was proposed which has some attractive properties, especially those related to the meaning of the parameters involved and the structure of its expected value as a function of its parameters, as well as an instructive and algorithmically quite straightforward new MLE procedure. The family of TSP distributions naturally extends the three-parameter triangular distributions. The new four-parameter TSP(a, m, b, n) distribution seems to be a useful and a more flexible competitor to the four-parameter beta distribution than the triangular distribution, specifically in, but not limited to, PERT applications emphasized in this paper. It is our hope that the introduction of the proposed distribution into statistical practice will assist in the basic goals of applied statistical work.

Acknowledgements

The authors are indebted to Professor N. L. Johnson, to the Joint Editor and the referees for their most valuable comments and suggestions which improved both the content and the presentation of the first and second versions.

Appendix A: Cardano's method

We shall briefly discuss a version of Cardano's method for solving the cubic equation

$$cn^3 + dn^2 + en + f = 0. \tag{42}$$

After dividing equation (42) by c and introducing

$$y = n + \frac{d}{3c} \tag{43}$$

we have $y^3 + 3py + 2q = 0$, where

$$\begin{aligned} 3p &= \frac{3ce - d^2}{3c^2}, \\ 2q &= \frac{2d^3}{27c^3} - \frac{de}{3c^2} + \frac{f}{c}. \end{aligned} \tag{44}$$

Table 4 summarizes the solution method for solving equation (42).

Appendix B: Maximum likelihood estimation procedure in pseudo-Pascal

The numerical routines below in pseudo-Pascal require separate algorithms to evaluate $M(a_k, b_k, r_k)$ (see equation (21)) $G(a_k, b_k, r_k)$ (see equation (26)) for triang(a, m, b) MLE or $G(a_k, b_k, r_k)$ (see equation (37)) for TSP(a, m, b, n) MLE, $\partial G(a_k, b_k, r_k)/\partial a$ (see equation (28)) for triang(a, m, b) MLE or $\partial G(a_k, b_k, r_k)/\partial a$ (see equation (38)) for TSP(a, m, b, n) MLE, $\partial G(a_k, b_k, r_k)/\partial b$ (see equation (29)) triang(a, m, b) MLE or $\partial G(a_k, b_k, r_k)/\partial b$ (see equation (39)) for TSP(a, m, b, n) MLE, $\partial M(a_k, b_k, r_k)/\partial a$ (see equation (30)) and finally $\partial M(a_k, b_k, r_k)/\partial b$ (see equation (31)). The output parameters of routines are indicated in bold italics.

B.1. BSearch

Let $G(a, b)$ be the function defined by equation (26) (or equation (37)). For any given a the set of discontinuities in b of the function $G(a, b)$ is a null set and we could utilize equation (29) (or equation (39)) and

Table 4. Version of Cardano's method for solving a cubic equation

$r = sgn(q)\sqrt{\lvert p\rvert}$		
$p < 0$		$p > 0$
$q^2 + p^3 \leqslant 0$	$q^2 + p^3 > 0$	
$\cos(\varphi) = q/r^3$	$\cosh(\varphi) = q/r^3$	$\sinh(\varphi) = q/r^3$
$y_1 = -2r\cos(\varphi/3)$	$y_1 = -2r\cosh(\varphi/3)$	$y_1 = -2r\sinh(\varphi/3)$
$y_2 = 2r\cos(60^\circ - \varphi/3)$	$y_2 = r\cosh(\varphi/3) + \mathrm{i}\sqrt{3}r\sinh(\varphi/3)$	$y_2 = r\sinh(\varphi/3) + \mathrm{i}\sqrt{3}r\cosh(\varphi/3)$
$y_2 = 2r\cos(60^\circ + \varphi/3)$	$y_3 = r\cosh(\varphi/3) - \mathrm{i}\sqrt{3}r\sinh(\varphi/3)$	$y_3 = r\sinh(\varphi/3) - \mathrm{i}\sqrt{3}r\cosh(\varphi/3)$

equation (31) to determine an ascending search direction with respect to $G(a,b)$ for b. Define

$$B(a) = \max_{r\in\{1,\ldots,s-1\}} \{f_b(a,r)\}, \tag{45}$$

where $f_b(a,r)$ is given by equation (27). From the properties of $\hat{r}(a,b)$ (see equation (20)) mentioned in Section 6 it follows that for $b > B(a)$

$$G(a,b) = \log\left(\prod_{i=2}^{s} \frac{b - X_{(i)}}{b - X_{(1)}}\right) - s\log(b-a)$$

and

$$\frac{\partial}{\partial a} G(a,b) = \frac{s}{b-a} > 0. \tag{46}$$

Hence, from equation (46) it follows that necessary conditions for a local maximum of $G(a,b)$ with a fixed cannot be satisfied for $b > B(a)$. Thus, `BSearch` maximizing $G(a,b)$ as a function of b with a fixed may be confined to the interval $(X_{(s)}, B(a))$ only. The routine `BSearch` below evaluates equation (32), follows a bisection approach (see, for example, Press *et al.* (1989)) and requires a separate algorithm to evaluate equation (45).

B.1.1. BSearch$(a_k, \mathbf{X}, \hat{\boldsymbol{b}}_{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{k}}, \boldsymbol{r}_{\boldsymbol{k}})$

Step1: $l_k^b = X_{(s)}$.
Step2: $u_k^b = B(a_k)$; $\hat{b}_k = (l_k^b + u_k^b)/2$; $M_k = M(a_k, b_k, r_k)$; $G_k = \partial G(a_k, b_k, M_k, r_k)/\partial b$.
Step3: if $\mathrm{abs}(G_k) \geqslant \delta$ then
    if $G_k < 0$ then $u_k^b = \hat{b}_k$
    else $l_k^b = \hat{b}_k$
    else stop.
Step4: if $(u_k^b - l_k^b) \geqslant \delta$ go to step 2
    else stop.

B.2. ABSearch

Let $G(a,b)$ be the function defined by equation (26) (or equation (37)). For any given b the set of discontinuities in a of the function $G(a,b)$ is a null set and we could utilize equation (28) (or equation (38)) and equation (30) to determine an ascending search direction with respect to $G(a,b)$ for a. The routine `ABSearch` starts by establishing an interval $[A, X_{(1)}]$ such that

$$\frac{\partial}{\partial a} G\left\{A, \hat{b}(A)\right\} > 0. \tag{47}$$

To determine A in equation (47) we may utilize equation (28) (or equation (38)) and equation (30). From

$$\lim_{a\to-\infty} \{\hat{r}(a,b)\} = s,$$

where $\hat{r}(a,b)$ is given by equation (20), it follows that for any given b there is a sufficiently small a such that

$$G(a,b) = \log\left(\prod_{i=1}^{s-1} \frac{X_{(i)} - a}{X_{(s)} - a}\right) - s\log(b-a)$$

and

$$\frac{\partial}{\partial a} G(a,b) = \sum_{i=1}^{s-1} \frac{X_{(i)} - X_{(s)}}{(X_{(i)} - a)(X_{(s)} - a)} + \frac{s}{b-a}. \tag{48}$$

From equation (48) it follows that for any given b there is an a sufficiently small such that $\partial G(a,b)/\partial a > 0$. It is conjectured that an A given by equation (47) does exist. A numerical analysis supports this conjecture. Having established $[A, X_{(1)}]$, `ABSearch` follows a bisection approach (see, for example Press *et al.* (1989)) and evaluates equation (25) (or equation (36)) by successively utilizing the routine `BSearch`.

B.2.1. *ABSearch(*$\mathbf{X}, \hat{\boldsymbol{a}}_{\boldsymbol{k}}, \hat{\boldsymbol{b}}_{\boldsymbol{k}}, \boldsymbol{m}_{\boldsymbol{k}}, \boldsymbol{n}_{\boldsymbol{k}}$*)*

Step1: $u_k^a = X_{(1)}$; $l_k^a = X_{(1)} - (X_{(s)} - X_{(1)})$.
Step2: BSearch($l_k^a, \mathbf{X}, \hat{\boldsymbol{b}}_{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{k}}, \boldsymbol{r}_{\boldsymbol{k}}$); $G_k = \partial G(l_k^a, \hat{b}_k, M_k, r_k)/\partial a$.
Step3: if $G_k < 0$ then
    $u_k^a = l_k^a$; $l_k^a = l_k^a - (X_{(s)} - X_{(1)})$
    go to step 2.
Step4: $\hat{a}_k = (l_k^a + u_k^a)/2$; BSearch($\hat{a}_k, \mathbf{X}, \boldsymbol{b}_{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{k}}, \boldsymbol{r}_{\boldsymbol{k}}$); $G_k = \partial G(\hat{a}_k, \hat{b}_k, M_k, r_k)/\partial a$.
Step5: if abs$(G_k) \geqslant \delta$ then
    if $G_k < 0$ then $u_k^a = \hat{a}_k$
    else $l_k^a = \hat{a}_k$
  else go to step 7.
Step6: if $(u_k^a - l_k^a) \geqslant \delta$ go to step 4
  else go to step 7.
Step7: $m_k = X_{(r_k)}$.
Step8: $n_k = -s/\log(M_k)$

(Step 8 should be omitted for triang(a, m, b) MLE.)

References

Cardano, G. (1993) *Ars Magna or the Rules of Algebra* (Engl. transl.). New York: Dover Publications.
Clark, C. E. (1962) The PERT model for the distribution of an activity. *Ops Res.*, **10**, 405–406.
van Dorp, J. R. and Kotz, S. (2002) The standard two sided power distribution and its properties: with applications in financial engineering. *Am. Statistn*, **56**, in the press.
Grubbs, F. E. (1962) Attempts to validate certain PERT statistics or a 'Picking on PERT'. *Ops Res.*, **10**, 912–915.
Johnson, D. (1997) The triangular distribution as a proxy for the beta distribution in risk analysis. *Statistician*, **46**, 387–398.
Johnson, N. L. and Kotz, S. (1999) Non-smooth sailing or triangular distributions revisited after some 50 years. *Statistician*, **48**, 179–187.
Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions*, 2nd edn, vol. 2. New York: Wiley.
Kamburowski, J. (1997) New validations of PERT times. *Omega*, **25**, 323–328.
Malcolm, D. G., Roseboom, C. E., Clark, C. E. and Fazar, W. (1959) Application of a technique for research and development program evaluation. *Ops Res.*, **7**, 646–649.
Nadarajah, S. (1999) A polynomial model for bivariate extreme value distributions. *Statist. Probab. Lett.*, **42**, 15–25.
Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1989) *Numerical Recipes in Pascal.* Cambridge: Cambridge University Press.
Williams, T. M. (1992) Practical use of distribution in network analysis. *J. Ops Res. Soc.*, **43**, 265–270.