Kurtosis of the Topp-Leone distributions

Samuel Kotz
Department of Engineering Management and Systems Engineering, George Washington University, 1776 G Street, N.W. Washington D.C. 20052, USA
kotz@gwu.edu

Edith Seier
Department of Mathematics, P.O.Box 70663. East Tennessee State University, Johnson City, TN 37614, USA
seier@etsu.edu

Abstract

The kurtosis of the Topp-Leone (T-L) family of distributions is explored by means of the spread-spread function to compare it with the left triangular distribution that originates the T-L family. Based on the second derivative of the spread-spread function, intervals of values of the parameter $b$ are identified regarding the behavior of kurtosis.

Key Words: Triangular distribution, Spread function, Spread-spread plots, Convexity

1. Introduction

Kurtosis as a measure of ‘flat-toppedness’ of a probability density function of a continuous random variable was introduced by Pearson (1905) but it is currently understood as related to its center and tails. For a comprehensive discussion of kurtosis see Zenga (2006). A family of distributions, introduced by Topp and Leone (1955) over 50 years ago was recently resurrected by Nadarajah and Kotz (2003) and further studied by Kotz and Van Dorp (2004). The Topp-Leone (T-L) family of distributions is generated from the left triangular distribution by elevating the CDF to a power $b > 0$, not necessarily an integer, that becomes the parameter of the new distribution.
When new probability distributions are defined it is interesting to study their characteristics. The insight we can get about them can be of help to professionals looking to use those distributions as models. Kurtosis is also related to the performance of tests for variability and normality. In this note we shall explore the kurtosis of the T-L family of distributions, comparing the distribution with parameter $b$ with the left triangular distribution. Since the T-L distributions are skewed, the spread-spread plot introduced by Balanda and MacGillivray (1990) seems to be an appropriate tool for comparison. The convexity of the spread-spread function is studied by means of its second derivative and three intervals of values of the parameter $b$, (identified in a first exploration of the convexity of the function): $0 < b < 1$, $1 < b < 2$, and $b \geq 2$ are examined. In the final section an appraisal of the kurtosis is carried out by several alternative kurtosis measures.

2. The spread function and spread-spread plots in Topp - Leone distributions

The pdf of the Topp and Leone distributions is $f(x) = 2bx^{b-1}(1-x)(2-x)^{b-1}$ for $0 \leq x \leq 1$; with the cdf $F(x) = x^b(2-x)^b$, the inverse cdf being $1 - \sqrt{1 - \sqrt{y}}$. Evidently for $b = 1$, the T-L is reduced to the left triangular distribution. Balanda and MacGillivray (1990) defined the spread function to be $S_H = H^{-1}(0.5+u) - H^{-1}(0.5-u)$ for $0 \leq u \leq 0.5$, where $H^{-1}$ is the inverse cdf of the distribution. The spread-spread plot introduced by Balanda and MacGillivray (1990) assigns $S_H$ and $S_F$ for two different distributions $H$ and $F$ to the $y$ and $x$ axes respectively, the distribution $H$ is considered to have larger kurtosis than $F$ iff the curve in the graph $(S_HS_F^{-1})$ is convex. For the Topp and Leone distribution with parameter $b$, the spread function becomes:

$$S_b(u) = H^{-1}(0.5+u) - H^{-1}(0.5-u) = \sqrt{1 - \sqrt{0.5+u}} - \sqrt{1 - \sqrt{0.5+u}}$$  \hspace{1cm} (1)

Evidently for $b = 1$, the spread function becomes that of the left triangular distribution given by

$$S_1(u) = \sqrt{0.5+u} - \sqrt{0.5-u}$$  \hspace{1cm} (2)

It is of interest for applications to compare, in terms of kurtosis, the T-L distributions with parameter $b = 1$ with the left triangular distribution by
means of the spread-spread plot. Let \( y = S_1(u) \), from (1) and (2) we have \( u = S_1^{-1}(y) = 0.5y\sqrt{2-y^2} \) and the corresponding spread-spread function appearing in the spread-spread plot to compare a distribution with parameter \( b \) with the distribution with parameter \( b = 1 \) becomes (for \( 0 < y < 1 \)):

\[
S_b(S_1^{-1}(y)) = \sqrt{1 - (0.5(1 - y\sqrt{2-y^2}))^{1/b}} - \sqrt{1 - (0.5(1 + y\sqrt{2-y^2}))^{1/b}} 
\]

As it was already alluded, for a given value of \( b \neq 1 \), the convexity of the line \( S_b(S_1^{-1}(y)) \) indicates that the T-L distribution with parameter \( b \) possesses a larger kurtosis than the left triangular distribution; if the line is concave the opposite would be true. If \( S_b(S_1^{-1}(y)) \) is neither concave nor convex, the two distributions are not comparable in terms of kurtosis. To determine whether the \( S_b(S_1^{-1}(y)) \) is convex we calculate the second derivative of (3) with respect to \( y \). This results to be

\[
\Delta^2 = (S_b(S_1^{-1}(y)))'' = M(ABC' + AB'C + A'BC - (DEG' + DE'G + D'EG)) 
\]

where \( k = \sqrt{0.5} \), \( M = -k/(2b) \), \( r = y\sqrt{t} \), \( t = 2 - y^2 \), \( v = 1 - k\sqrt{1-r} \), \( u = 1 - k\sqrt{1+r} \), \( A = 1/\sqrt{v} \), \( B = (1-r)^{1/b-1} \), \( C = (y^2/\sqrt{v}) - \sqrt{t} \), \( D = 1/\sqrt{u} \), \( E = (1+r)^{1/b-1} \) and \( G = -C \). Thus, \( A' = -MBCv^{-3/2} \), \( G' = -C' \), \( B' = (1/b-1)C(1-r)^{1/b-2} \), \( C' = y^3/t^{3/2} + 3y/\sqrt{t} \), \( E' = (1/b-1)(1+r)^{1/b-2}G \) and \( D' = -MEGu^{-3/2} \).

We shall now analyze the behavior of the second derivative (4) for all \( 0 < y < 1 \), providing the spread and the spread-spread plots for \( b < 1 \), \( 1 < b < 2 \), and \( b \geq 2 \).

### 3. Topp-Leone distributions with \( b < 1 \)

For \( 0 < b < 1 \), the T-L distribution is a J-shape. These distributions were actually the original Topp and Leone distributions and have been found to have applications in reliability. Their authors were searching for distributions to model failure data of equipment and devices. Figure 1 displays the density function for a number of values of the parameter \( b = 0.1(0.2)1 \). The smaller is \( b \), the more negatively skewed the distribution becomes.

The surface in Figure 2 corresponds to the spread function (1) for \( 0.02 \leq b \leq 1 \) and \( 0 \leq u \leq 0.5 \). The sections of the surface obtained by fixing the
value of b show the manner in which the spread function varies as b increases. Figure 3 displays the spread-spread function (3) evaluated for $0.1 \leq b \leq 1$ and $0 \leq y \leq 1$ comparing each one of the corresponding T-L distributions with the left triangular distribution. The sections of the surface in Figure 3 obtained by fixing the value of b result in spread-spread plots such as those in Figure 4A. For low values of b, the spread-spread plot depicts a convex line, but as b approaches to 1, the spread-spread plot becomes almost linear (it is linear for b=1 because here we are comparing the left triangular distribution with itself.)

It is of course of interest to compare the Topp-Leone distributions among themselves and not only with the left triangular distribution. Figure 4b (where spread-spread plots compare $S_b$ and $S_{b+0.2}$), seem to suggest that the
Figure 3: Spread-Spread function $S_b(S_1^{-1}(y))$ to compare T-L distributions with $0.1 \leq b \leq 1$ with the left triangular distribution.

Topp-Leone distributions are indeed kurtosis ordered for the most part of the interval $0 < b < 1$, the smaller the value of the parameter $b$ the larger the kurtosis. The behavior in the vicinity of the extremes of the interval $[0,1]$ requires special attention. Observing the change in shape of the curve $S(u)$ along the values of $b$, we can conclude that kurtosis increases very fast as $b$ decreases in the interval $[0.1,0.3]$, gradually increases as $b$ decreases in the interval $[0.3,0.8]$, and seems to change very little after that. The issue of the kurtosis when $b$ is very close to 0 ($0 < b < 0.05$) seems to be quite delicate. Figure 2 displays the extreme form of the spread function for very low values of the parameter $b$. To study the convexity of the spread-spread function its second derivative is displayed in Figure 5. Observe that when $b$ is very close to 0, the second derivative takes value 0 for a substantial range of values of $y$. Figure 5 confirms the convexity $S_b(S_1^{-1}(y))$ when $0.1 \leq b \leq 0.88$. However rather unexpectedly the second derivative takes on values slightly below zero for a range of values of $y$ when $0.89 < b < 1$. The interval of values of $y$ for which the second derivative of $S_b(S_1^{-1}(y))$ is negative is located in the lower part of the interval $[0,1]$ and it prolongs as $b$ approaches 1 from the left. When $b = 0.95$ (last plot in Figure 5), the second derivative takes on a negative value for $y < 0.38$, however those negative values are very close to 0. For example, the minimum value attained by the second derivative for $b = 0.95$ is $-0.0008$. Since the second derivative (4) has a complicated expression and its value is expected to be close to zero in the vicinity of
Figure 4: Spread-spread plots for pairs of T-L distributions

$b = 1$, the negative value might be the consequence of the accumulation of rounding off errors. Due to this behavior of the spread-spread function it may be safer not to claim that the Topp-Leone distribution with $0.89 \leq b < 1$ has higher kurtosis than the left triangular distribution. For $b \leq 0.04$, the second derivative takes on value 0 for a substantial range of values of $y$. From all the figures in this section it is evident that in the interval $0.1 \leq b \leq 0.8$, the T-L distributions have a smooth spread function and strictly larger kurtosis than the left triangular distribution. Moreover the T-L distributions seem to be ordered with respect to kurtosis (the lower the value of the parameter $b$, the higher the kurtosis is), and small changes in $b$ can provide a very substantial change in kurtosis. The curvature of the spread-spread functions in Figure 4 seem to indicate that the change in kurtosis for equal increments in the parameter $b$ decreases as $b$ increases in $[0.1, 0.8]$.

4. **Topp-Leone distributions for** $1 < b < 2$

Figure 6 displays the density function of the Topp-Leone distributions for $b = 1(0.1)2$. Unlike the distributions in Figure 1 (for $0 < b < 1$), the mode here is not 0 and $f(\text{mode}) < 2$; $f(\text{mode})$ first decreases and then increases. As $b$ increases in this interval the distribution becomes less skewed.
Figure 5: Second derivative $\Delta^2$ of Spread-Spread function $S_b(S_1^{-1}(y))$ for $0 < b < 1$

Figure 6: Topp-Leone distributions for $1 \leq b \leq 2$

Figure 7 seems to indicate that the spread functions for the different values of b in the interval $[1, 2]$ are quite similar. This is confirmed by Figure 8 (to be compared with Figure 4a), which displays the spread-spread plot comparing the T-L distributions with $b = 1.1(0.1)2$ with the left triangular distribution. The second derivative in expression (4) is displayed in Figure 9 for $1 < b < 2$; it takes negative values for some values of y thus indicating that $S_b(S_1^{-1}(y))$ is not convex and hence one can not claim that the T-L distributions with the parameter in the interval $[1, 2]$ posses higher or lower kurtosis than the left triangular distribution. The values of y for which the second derivative takes negative values are at the right side of the interval $0 < y < 1$, in the vicinity of 1.
5. **Topp-Leone distributions for \( b \geq 2 \)**

The T-L densities for a number of values of \( b \geq 2 \) are displayed in Figure 10. Here both the mode and \( f(\text{mode}) \) increase when the parameter \( b \) increases. A comparison with figures 1 and 6 clearly indicates the change in shape from J shaped to distributions similar to the exponential and then unimodal Gamma type distributions. Skewness takes value 0 in the vicinity of \( b = 2.56 \), is positive for \( b < 2.56 \) and negative for \( b > 2.57 \).

The sharp turn in the spread function in Figure 11 becomes more noticeable as \( b \) increases. Figure 12 suggest that for \( b \geq 2 \), the Topp-Leone distribution has larger kurtosis than the left triangular distribution. This is confirmed by the second derivative of the spread-spread function in (4)
Figure 9: Second derivative $\Delta^2$ of Spread-Spread function $S_b(S_1^{-1}(y))$ for $1 < b < 2$

Figure 10: Topp and Leone distributions with $b \geq 2$

displayed in Figure 13 for $b \geq 2$. Note that for larger values of $b$, the second derivative takes rather high values for $y$ in the vicinity of $y=1$ and very low ones in the vicinity of $y=0$. Figures 12 and 14 also seem to indicate that as $b$ increases the difference in kurtosis between T-L distributions and the left triangular distribution is located mainly in the tails rather than at the peak since the curvature is larger at the right side of the spread-spread plot. The trend noticed in the shape of the spread and the spread-spread functions as $b$ increases in figures 11 and 12, continues to be valid when $b$ takes much larger values.

Figure 14 suggests that the distributions seem to be ordered according to the value of $b$, the larger the value the higher the kurtosis. Figure 14 is to be compared with Figure 4B; the change in the kurtosis as $b$ increases beyond 2 does not seem to be as prominent as when $b$ decreases below 0.8.
6. Kurtosis measures

Based on the analysis of the spread-spread function we know the ranges of values of the parameter \( b \) for which the T-L distribution has higher kurtosis than the left triangular distribution (\( b \leq 0.88 \) and \( b \geq 2 \)). Also as \( b \) decreases below 0.88 the kurtosis increases, and the behavior of the kurtosis for values of \( b \) in the vicinity of 0 and 1 may be somewhat delicate. Indeed, small changes in \( b \) in the interval (0.1, 0.5) may result in large changes in kurtosis. For \( b \geq 2 \) we would expect small increments in the kurtosis as \( b \) increases, while in the interval 1 < \( b < 2 \) the comparison of the T-L distributions with the left triangular distribution in terms of kurtosis is just meaningless since the corresponding spread-spread function is not convex. It maybe of interest to examine the behaviour of kurtosis measures for the T-L distributions.

By now it is well known that it is often misleading to represent kurtosis by a single number. Pearson’s kurtosis appears much more frequently in the literature dealing with statistical distributions; however other measures of kurtosis do exist. The distributions that we were studying are skewed so that the kurtosis measures defined for symmetric distributions are not applicable here. We shall use 3 kurtosis measures that are expected values of functions of the standardized variable \( z \), \( z = (x - \mu) / \sigma \), where \( \mu \) and \( \sigma \) are the mean and standard deviation of the distribution respectively. One of those measures is the well known Pearson’s coefficient \( \beta_2 = E[z^4] \) (Pearson, 1905). The other two measures to be considered are those proposed by Seier and Bonett (2003)
Figure 12: Spread-Spread plots to compare T-L \((b \geq 2)\) with the left triangular distribution.

: \(K_1(e) = E[5.7344\exp(-|z|)]\) and \(K_2(1) = E[14.84(1-|z|)]\), the latter being a scaled version of the robust kurtosis defined by Stavig(1982). These three measures are of the form \(E[g(z)]\) where \(g(z)\) is a continuous even function, strictly monotonic at each side of \(z=0\). However, \(g(z) = z^4\) (corresponding to Pearson’s \(\beta_2\)) increases as \(z\) departs from 0 giving more importance to the tails of the distribution while the \(g(z)\) appearing in the other two give more importance to the middle. All three measures attain value 3 for the normal (Gaussian) distribution, but the ranges of values are different. It should perhaps be noted that the L-kurtosis, a measure of kurtosis quite popular in hydrology, provides for a number of distributions values that tend to be highly correlated with the value of \(K_1(e)\); \(r = 0.992\) for a set of 21 symmetric distributions (Seier and Bonett, 2003) and \(r = 0.903\) for a set of 30 skewed distributions. Whether this is also valid for the T-L distributions is to the best of our knowledge an open problem. From Kotz and van Dorp (2004) the values of the mean and the variance are:

\[
\mu = 1 - 4^b \frac{\Gamma(b+1)\Gamma(b+1)}{\Gamma(2b+2)} \quad \text{and} \quad \sigma^2 = (2^{2b+1})B(b+1,b+1)[1-2^{2b-1}B(b+1,b+1)]-2^{2b+3}B(b+2,b+1)B(0.5|b+2,b+1),
\]

where \(B = \Gamma(a)\Gamma(b)/\Gamma(a+b)\) and \(B\) is the incomplete beta function.

Strictly speaking we shouldn’t quantify the T-L distributions kurtosis \(1 < b < 2\) because based on the distribution functions we can not really assert that they have higher or lower kurtosis than the left triangular, however we plot the values of \(\beta_2\), \(K_1(e)\) and \(K_2(1)\) for \(0.01 \leq b \leq 10\) in Figure 15. Each row of the figure corresponds to one of the 3 measures, each column...
Figure 13: Second derivative $\Delta^2$ of Spread-Spread function $S_b(S_1^{-1}(y))$ for $b \geq 2$

Figure 14: Spread-spread to compare pairs of T-L distributions with $b$ 2 units apart ($b \geq 2$)

corresponds to different intervals of values of the parameter $b$. Use $b = 1$ as a reference point since we are comparing all the distributions with the originating left triangular distribution. For values of $b$ in the vicinity of 0 we have noticed some problems with $K_1(e)$, and in a lesser extent with $K_2(1)$. A program to calculate these measures by means of numerical integration reveals some problems. The value for $K_1(e)$ peaks for $b=0.06$ and then decreases as $b$ goes (backwards) towards 0. It is only for $b=0.01$ that $K_1(e)$ attains a value lower than the one for the left triangular (2.8299). $K_2(1)$ peaks at $b=0.05$ and goes down as $b$ goes backwards towards 0, however its values are always higher than the one corresponding to the left triangular distribution. The coefficient $\beta_2$ does not experience problems in that region:
Figure 15: Values of the three kurtosis measures for different intervals for the parameter $b$

its value continues to increase as $b$ approaches 0. On the other hand, both $K_1(e)$ and $K_2(1)$ confirm our previous comments about kurtosis based on the spread-spread function for $b > 2$: here the kurtosis of the T-L distributions is greater than the kurtosis of the left triangular distribution, and seems to increase slightly as $b$ increases. However, $\beta_2$ takes values lower than 2.4 (the value corresponding to the left triangular distribution) for the whole range of values $2 < b < 4.28$. It is understandable that the measures would disagree in their behavior in the interval $1 < b < 2$ since the distributions here are not kurtosis ordered with respect to the left triangular distribution. Figures 1, 6 and 11 vividly demonstrate the movement of the T-L density encompassing a wide range of forms. Table 1 summarizes the values of Pearson's skewness and kurtosis coefficients and with parameter $b$ in the range $0 < b \leq 10$. Additional explorations indicate that the kurtosis coefficient $\beta_2$ converges to approximately 3.2 as the parameter $b$ increases indefinitely.

7. Conclusions

The Topp-Leone family of distributions originated from the left triangular distribution by elevating the cdf to the power $b > 0$. Kurtosis of this family of
Table 1: Pearson’s skewness and kurtosis

<table>
<thead>
<tr>
<th>b</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1000</td>
<td>3.3325</td>
<td>14.8985</td>
</tr>
<tr>
<td>0.2000</td>
<td>2.2124</td>
<td>7.6306</td>
</tr>
<tr>
<td>0.5000</td>
<td>1.1513</td>
<td>3.4940</td>
</tr>
<tr>
<td>0.8000</td>
<td>0.7373</td>
<td>2.6292</td>
</tr>
<tr>
<td>1.0000</td>
<td>0.5657</td>
<td>2.4000</td>
</tr>
<tr>
<td>1.5000</td>
<td>0.2916</td>
<td>2.1986</td>
</tr>
<tr>
<td>2.0000</td>
<td>0.1253</td>
<td>2.1800</td>
</tr>
<tr>
<td>2.5631</td>
<td>0.0000</td>
<td>2.2208</td>
</tr>
<tr>
<td>3.0000</td>
<td>-0.0711</td>
<td>2.2658</td>
</tr>
<tr>
<td>4.3000</td>
<td>-0.2106</td>
<td>2.4021</td>
</tr>
<tr>
<td>5.0000</td>
<td>-0.2600</td>
<td>2.4664</td>
</tr>
<tr>
<td>10.0000</td>
<td>-0.4287</td>
<td>2.7528</td>
</tr>
</tbody>
</table>

distributions is studied as a function of parameter $b$. Since the distributions are non-symmetric, the spread-spread function is an appropriate tool for this investigation. Small values of $b$, in the vicinity of 0, cause the spread function to be almost constant in some intervals, at the same time taking sharp turns. For $b < 0.05$ the second derivative of the spread-spread function takes value 0 for part of the domain of the variable $y$. The value of $b = 0$ presents the extreme situation in which all the mass is concentrated at one point. Values of $b$ close to 1 do not produce convex spread-spread functions when the distribution is compared with the left triangular. Neglecting the extreme and the non-comparable situations we feel comfortable to assert that for $0.1 \leq b \leq 0.8$ a Topp-Leone distribution has larger kurtosis than the left triangular distribution and that the lower the value of $b$ in that interval is, the higher the kurtosis is. Small changes in the parameter $b$ in the interval $(0.1, 0.5)$ produce noticeable changes in kurtosis. For the values of $b$ in the interval $(1, 2)$, the comparison between kurtosis of the T-L distribution and the left triangular distribution is non-feasible. For $b > 2$, the distributions have larger kurtosis than the left triangular, however in this range even large increments in the value of the parameter $b$ result in minor changes in kurtosis.
References


