The Institute for Integrating Statistics in Decision Sciences

**On a Bounded Bimodal Two-sided Distribution
Fitted to the Old-Faithful Geyser Data**

Donatella Vicari
Department of Statistical Sciences
Sapienza University of Rome, Italy

Johan Rene van Dorp
Department of Engineering Management and Systems Engineering
The George Washington University, USA

On a bounded bimodal two-sided distribution fitted to the Old-Faithful Geyser Data

Donatella Vicari[1] and Johan Rene van Dorp[2]

Abstract — In this paper we shall develop a novel family of bimodal univariate distributions (also allowing for unimodal shapes) and demonstrate its use using the well known and almost classical data set involving durations and waiting times of eruptions of the Old-Faithful Geyser in Yellowstone park. Specifically, we shall utilize the Old-Faithful data set with 272 data points provided in Dekking et. al (2005). In the process, we develop a bivariate distribution using a copula technique and compare its fit to a mixture of bivariate normal distributions also fitted to the same bivariate data set. We believe the fit-analysis and comparison is primarily illustrative from an educational perspective for distribution theory modelers, since in the process a variety of statistical techniques are demonstrated. We do not claim one model as preferred over the other.

1. Introduction

The Old-Faithful data set has been a popular data set to demonstrate a variety of statistical techniques in particular kernel density estimation (e.g. Silverman, 1986), time-series analysis (e.g. Azzalini and Bowman, 1990), clustering (e.g. Atkinson and Riani, 2006), Hidden Markov Models, and distribution theory (e.g. Eilers and Borgdorff, 2007), to name a few. This paper falls in the latter category. Specifically, we shall illustrate a novel univariate parametric family of distributions allowing for bimodal shapes by developing a bivariate distribution for the Old-Faithful duration and waiting time data set provided in Dekking et. al (2005) consisting of 272 data points (d_i, w_i), $i = 1, \ldots, 272$. The bivariate distribution shall be constructed using a copula technique involving diagonal band copulas (see, Kotz and van Dorp, 2010).

Figures 1A and 1B presents an exploratory data analysis of the Old-Faithful Geyser data. Figure 1A depicts a scatter plot of the duration and waiting time data. One observes a clear clustering and a strong statistical dependence between the two random variables duration

---

[1] Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
[2] Corresponding Author, Department of Engineering Management and Systems Engineering, The George Washington University, Washington D.C., USA

(D) and waiting time (W). Transforming the 272 durations d_i using their empirical cumulative distribution function

$$F_\eta(d_i) = \frac{1}{\eta}\sum_{j=1}^{\eta} 1_{[0,d_i]}(d_j),\ i = 1,\ldots,\eta,\ \eta = 272$$

and waiting times w_i using their empirical distribution function

$$G_\eta(w_i) = \frac{1}{\eta}\sum_{j=1}^{\eta} 1_{[0,w_i]}(w_j),\ i = 1,\ldots,\eta,\ \eta = 272$$

yields a bivariate data set ($F_\eta(d_i), G_\eta(w_i)$) on the unit-square $[0,1]^2$. A scatter plot of this bivariate data set is presented in Figure 1B.

Observe from Figure 1B that while univariate data sets $F_\eta(d_i)$ and $G_\eta(w_i)$ naturally follow a uniform behavior on $[0,1]$, the bivariate data set ($F_\eta(d_i), G_\eta(w_i)$) does not follow this behavior on the unit square $[0,1]^2$. Firstly, introducing the parameter δ and estimating

$$\widehat{P}(\delta) = \frac{\#\ \text{data points in Figure 1B}\ \in [0,\delta]^2 \cup (\delta,1]^2}{\eta}$$

as a function of $\delta \in [0,1]$, we observe from Figure 1C that over 98.5% of the data points ($F_\eta(d_i), G_\eta(w_i)$) $\in [0,\delta]^2 \cup (\delta,1]^2$ for $\delta \in (0.355,\ 0.358)$. In fact, only 4 out of the 272 data points do not fall within this area. Secondly, evaluating lower correlation of lower data points

$$(F_\eta(d_i), G_\eta(w_i)) \in [0,\delta]^2$$

we obtain a value of approximately 0.214. Evaluating correlation of upper data points

$$(F_\eta(d_i), G_\eta(w_i)) \in [\delta,1]^2$$

we obtain a value of approximately 0.278.

A challenge in modeling a bivariate copula distribution (Nelsen, 1999) for the data in Figure 1B is the lack of observations in the $[0,\delta] \times [\delta,1]$ and $[\delta,1] \times [0,\delta]$ areas. One possibility is to use a two-dimensional mixture technique reminiscent of the univariate mixture technique of Two-sided distributions (Vicari et al, 2008) that define separate generating densities for its two branches. Specifically we suggest the copula mixture

$$C(u,v|\delta) = \delta \times S(\frac{u}{\delta},\frac{v}{\delta}) + (1-\delta)T(\frac{u-\delta}{1-\delta},\frac{v-\delta}{1-\delta}),\ (u,v) \in [0,1]^2,\ \delta \in [0,1]. \qquad (1)$$

where $S(\,\cdot\,,\,\cdot\,)$ and $T(\,\cdot\,,\,\cdot\,)$ are its generating copulas. It is not difficult to show that from $S(\,\cdot\,,\,\cdot\,)$ and $T(\,\cdot\,,\,\cdot\,)$ being copulas on $[0,1]^2$ it follows that $C(u,v|\delta)$ is a copula on $[0,1]^2$.
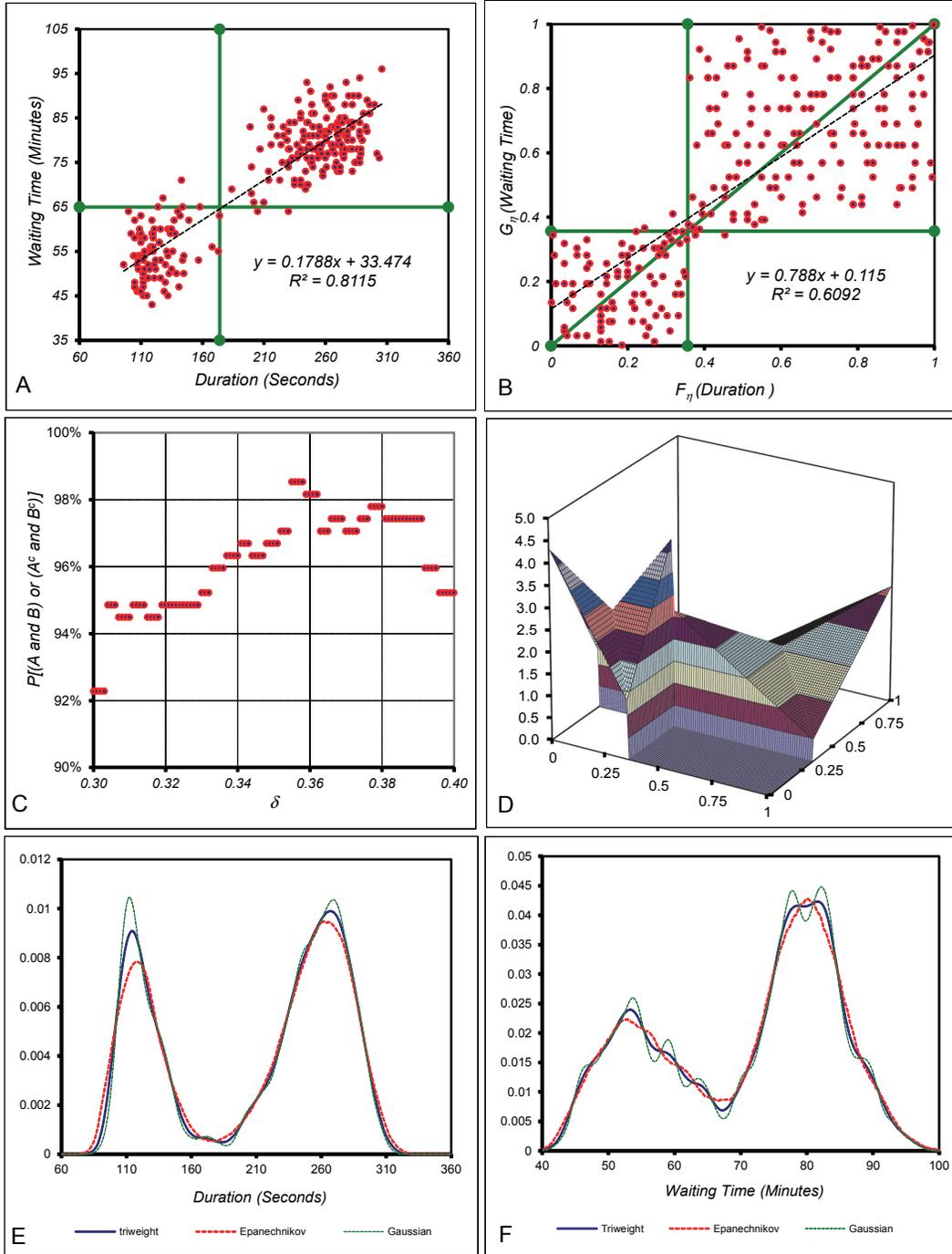
Fig. 1. An exploratory analysis of the bivariate Old-Faithful data set (d_i, w_i) in Dekking et. al (2005), $i = 1, \ldots, n$, $n = 272$. A : Scatter plot of (d_i, w_i); B : Scatter plot of $(F_\eta(d_i), G_\eta(w_i))$, C : Graph of $\widehat{P}\{F_\eta(d_i), \ G_\eta(w_i) \in [0, \delta]^2 \cup (\delta, 1]^2\}$ as a function of δ; D : A copula model for $(F_\eta(D), G_\eta(W))$; E : Kernel density estimates for duration D; F: Kernel density estimates for waiting time W.

3

Moreover, $C(u, v|\delta)$ has no probability mass in neither $[0, \delta] \times [\delta, 1]$ nor $[\delta, 1] \times [0, \delta]$. If we now select for $S(\,\cdot\,, \,\cdot\,)$ a copula with correlation 0.214 (i.e. for the lower quadrant in Figure 1B) and for $T(\,\cdot\,, \,\cdot\,)$ a copula with correlation 0.278, we have defined a bivariate copula that captures these characteristics of Figure 1B.

Given the analysis in Figure 1C, we shall set

$$\delta = (0.355 + 0.358)/2 = 0.3565.$$

For $S(\,\cdot\,, \,\cdot\,|\alpha_s)$ and $T(\,\cdot\,, \,\cdot\,|\alpha_T)$ in (1) we select generalized diagonal band copulas with Two-Sided slope generating densities (see, Kotz and van Dorp, 2010). We have for the density $S(u, v|\alpha_s)$:

$$S(u, v|\alpha_s) = \begin{cases} \alpha_s - 2(\alpha_s - 1)v, & (x, y) \in A_1, \\ \alpha_s - 2(\alpha_s - 1)u, & (x, y) \in A_2, \\ \alpha_s - 2(\alpha_s - 1)(1 - u), & (x, y) \in A_3, \\ \alpha_s - 2(\alpha_s - 1)(1 - v), & (x, y) \in A_4, \end{cases} \tag{2}$$

where $0 \leq \alpha_s \leq 2$ and

$$\begin{aligned} A_1 &= \{(u, v) \in [0, 1]^2 \,|\, 0 < u + v \leq 1, \, -1 < u - v \leq 0\}, \\ A_2 &= \{(u, v) \in [0, 1]^2 \,|\, 0 < u + v \leq 1, \quad 0 < u - v < 1\}, \\ A_3 &= \{(u, v) \in [0, 1]^2 \,|\, 1 < u + v \leq 2, \, -1 < u - v \leq 0\}, \\ A_4 &= \{(u, v) \in [0, 1]^2 \,|\, 1 < u + v \leq 2, \quad 0 < u - v < 1\}. \end{aligned}$$

The correlation for $S(u, v|\alpha_s)$ follows as a linear relationship

$$\rho(\alpha_s) = -\frac{2}{5} + \frac{2}{5}\alpha_s \in [-0.4, 0.4]. \tag{3}$$

Thus setting $\rho(\alpha_s) = 0.214$ yields directly $\alpha_s = 1.535$. Similarly, we obtain for $T(\,\cdot\,, \,\cdot\,|\alpha_T)$ that $\alpha_T = 1.696$. Advantages of copulas $S(u, v|\alpha_s)$, $T(u, v|\alpha_T)$ and $C(u, v|\theta)$ given by (1) are: (i) they have closed form probability density functions (pdf's) and cumulative distribution functions (cdf's); (ii) the functional form of their pdf and cdf are defined by linear and quadratic relationships, respectively; and (iii) the correlation (3) too follows a straightforward linear form. Most importantly, however, $S(u, v|\alpha_s)$ and $T(u, v|\alpha_T)$ were shown to be approximately least informative in the entropy sense given their correlation constraints (see, Kotz and van Dorp, 2010). Figure 1D plots the copula density (1) with parameters

$$\delta = 0.3565, \, \alpha_s = 1.535, \, \alpha_T = 1.696 \tag{4}$$

to model the data set in Figure 1B.

Having modeled the statistical dependence between random variable D and W using copula (1) all that remains is the modeling of the marginal distributions for the data in Figure 1A. Naturally, one could follow the traditional approach of non-parametric density estimation. Figures 1E and 1F demonstrate such fits utilizing kernel density estimation (Silverman, 1986) using three separate kernels: the triweight, Epanechnikov and Gaussian kernels. Observe that different kernels result in different density estimates. More importantly, however, by using a non-parametric approach we lose the specific advantage of the closed form copula expression for its pdf (1) utilizing expressions the like of (2).

In the remainder of this paper we set as our goal to develop a parametric model that captures the univariate bimodal behavior of Figures 1E and 1F. To reach this goal we utilize the flexible two-sided framework of univariate distribution introduced by Vicari et al. (2008). This framework is summarized in Section 2. In Section 3 we develop an instance of this framework allowing for bimodal shapes using elevated power distributions as their generating densities (García et al., 2011). In Section 4, an approximate maximum likelihood procedure is presented to estimate the parameters of the distributional model in Section 3. Finally, in Section 5 we integrate the copula density model (1) with parameter settings (4) with MLE fitted using the data in Figures 1E and 1F and show an improvement of overall fit as compared to a bivariate Gaussian mixture model (see, e.g., Titterington et. al, 1985) with parameters estimated using the EM algorithm (see, e.g., Meng and Rubin, 1993).

2. A two-sided framework of univariate distributions

Vicari et al. (2008) introduced the following two-sided framework of distributions using continuous cumulative distribution functions $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi)$ with the support $[0,1]$:

$$Pr(Y \leq y|\Theta) = \begin{cases} 0, & \text{for } y \leq 0, \\ p(\Omega)\Big\{G\big(\frac{y}{\theta}|\Phi\big)\Big\}^m, & \text{for } 0 < y < \theta, \\ 1 - \{1 - p(\Omega)\}\Big\{1 - H\big(\frac{y-\theta}{1-\theta}|\Psi\big)\Big\}^n, & \text{for } \theta \leq y < 1, \\ 1, & \text{for } y \geq 1. \end{cases} \tag{5}$$

where

$$p(\Omega) = \frac{\theta n}{(1-\theta)m + \theta n}. \tag{6}$$

The parameter vector $\Theta = (\Omega, \Phi, \Psi)$, where $\Omega = (\theta, m, n)$ are the two-sided power parameters, and Φ and Ψ are the parameters of the two respective branch generating distributions $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi)$, both potentially vector valued as well. The cdf (5) is

continuous at θ with value $Pr(Y \le \theta|\Theta) = p(\Omega)$ which does not depend on the structure of either $G(\,\cdot\,|\Phi)$ or $H(\,\cdot\,|\Psi)$.

By taking the derivative with respect to y in (5) one obtains for the corresponding probability density function (pdf) :

$$f_Y(y|\Theta) = \frac{mn}{(1-\theta)m + \theta n} \begin{cases} g\left(\frac{y}{\theta}\right)\left\{G\left(\frac{y}{\theta}|\Phi\right)\right\}^{m-1}, & \text{for } 0 < y < \theta, \\ h\left(\frac{y-\theta}{1-\theta}\right)\left\{1 - H\left(\frac{y-\theta}{1-\theta}|\Psi\right)\right\}^{n-1}, & \text{for } \theta \le y < 1, \\ 0, & \text{elsewhere,} \end{cases} \tag{7}$$

(where $g(\,\cdot\,|\Phi)$ and $h(\,\cdot\,|\Psi)$ are the pdf's of the cdf's $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi)$, respectively). It follows from (7) that

$$f_Y(\theta^+|\Theta) - f_Y(\theta^-|\Theta) = h\big(0|\Psi\big) - g\big(1|\Phi\big), \tag{8}$$

and hence the pdf (7) is not necessarily continuous at θ. Moreover, $f_Y(\theta^-|\Theta) > 0$ $[f_Y(\theta^+|\Theta) > 0]$ if and only if $g(1|\Phi) > 0$ $[h(0|\Psi) > 0]$. When $h\big(0|\Psi\big) = g\big(1|\Phi\big)$ the density $f_Y(y|\Theta)$ (7) becomes continuous at θ and strictly positive at θ if and only if $g\big(1|\Phi\big), h(0|\Psi) > 0$.

The quantile function $F^{-1}(z|\Theta)$ of the cdf $F(y||\Theta) = Pr(Y \le y|\Theta)$ defined by (5) follows as

$$F^{-1}(z|\Theta) = \begin{cases} 0, & \text{for } z \le 0, \\ \theta G^{-1}\left(\sqrt[m]{\frac{z}{p(\Omega)}}|\Phi\right), & \text{for } 0 < z < p(\Omega), \\ (1-\theta)H^{-1}\left(1 - \sqrt[n]{\frac{1-z}{1-p(\Omega)}}|\Psi\right) + \theta, & \text{for } p(\Omega) \le z < 1, \\ 1, & \text{for } z \ge 1. \end{cases} \tag{9}$$

where $G^{-1}(\,\cdot\,|\Phi)$ and $H^{-1}(\,\cdot\,|\Psi)$ are the quantile functions of the branch generating distributions $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi)$ in (5). Thus sampling from (5) is straightforward and direct when the branch quantile functions $G^{-1}(\,\cdot\,|\Phi)$ and $H^{-1}(\,\cdot\,|\Psi)$ are available in a closed form.

Setting $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi))$ to be the uniform cdf's on $[0, 1]$, expressions (5) and (7) reduce to those of a GTSP random variables presented in Herrerías-Velasco et. al (2009). Vicari et al. (2008) originally considered slope distributions on $[0, 1]$ as candidates for $G(\,\cdot\,|\Phi)$ and $H(\,\cdot\,|\Psi)$ leading to their two-sided generalized Topp and Leone distributions.

3. PDF and CDF of TS-EP distributions

We shall now provide an example of the density construction (7) above by letting $H(\,\cdot\,)$ $[G(\,\cdot\,)]$ to be a [reflected] elevated power distribution on $[0,1]$ given by:

$$\begin{cases} G(x|\alpha,\phi) = 1 - \phi(1-x) - (1-\phi)(1-x)^{\alpha}, \\ H(x|\beta,\psi) = \psi x + (1-\psi)x^{\beta}, \end{cases} \tag{10}$$

where $\phi, \psi \in [0,1], \alpha, \beta \geq 1$ with the corresponding pdf's:

$$\begin{cases} g(x|\alpha,\phi) = \phi + (1-\phi)\alpha(1-x)^{\alpha-1}, \\ h(x|\beta,\psi) = \psi + (1-\psi)\beta x^{\beta-1}, \end{cases} \tag{11}$$

Substituting (10) and (11) into (7) one obtains the density:

$$f_Y(y|\Theta) = \frac{mn}{(1-\theta)m + \theta n} \times \tag{12}$$

$$\begin{cases} \left\{\phi + (1-\phi)\alpha\left(\frac{\theta-y}{\theta}\right)^{\alpha-1}\right\}\left\{1 - \phi\left(\frac{\theta-y}{\theta}\right) - (1-\phi)\left(\frac{\theta-y}{\theta}\right)^{\alpha}\right\}^{m-1}, & \text{for } 0 < y < \theta, \\ \left\{\psi + (1-\psi)\beta\left(\frac{y-\theta}{1-\theta}\right)^{\beta-1}\right\}\left\{\psi\left(\frac{y-\theta}{1-\theta}\right) + (1-\psi)\left(\frac{y-\theta}{1-\theta}\right)^{\beta}\right\}^{n-1}, & \text{for } \theta \leq y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where as above $\Theta = (\Omega, \Phi, \Psi), \Omega = (\theta, m, n), \Phi = (\alpha, \phi)$ and $\Psi = (\beta, \psi)$ with the cdf

$$F_Y(y|\Theta) = \begin{cases} 0, & \text{for } y \leq 0, \\ p(\Omega)\left\{1 - \phi\left(\frac{\theta-y}{\theta}\right) - (1-\phi)\left(\frac{\theta-y}{\theta}\right)^{\alpha}\right\}^{m}, & \text{for } 0 < y < \theta, \\ 1 - \{1 - p(\Omega)\}\left\{\psi\left(\frac{y-\theta}{1-\theta}\right) + (1-\psi)\left(\frac{y-\theta}{1-\theta}\right)^{\beta}\right\}^{n}, & \text{for } \theta \leq y < 1, \\ 1, & \text{for } y \geq 1. \end{cases} \tag{13}$$

We have from (11) and (8) that

$$f_Y(\theta^+|\Theta) - f_Y(\theta^-|\Theta) = \psi - \phi. \tag{14}$$

Hence, the density (12) is continuous on $[0,1]$ provided $\psi = \phi$. Moreover, $f_Y(\theta|\Theta) > 0$, when $\psi = \phi > 0$.

4. Maximum Likelihood Estimation

For a random sample $\underline{X} = (X_1, \ldots, X_s)$ of size s from the distribution (12), the log likelihood function is, by definition,

$$Log\{L(\underline{X}, \Theta)\} = sLog\{\frac{mn}{(1-\theta)m + \theta n}\} + \tag{15}$$

$$\sum_{i=1}^{r}Log\{g(\frac{X_{(i)}}{\theta}|\alpha,\phi)\} + (m-1)\sum_{i=1}^{r}Log\{G(\frac{X_{(i)}}{\theta}|\alpha,\phi)\} +$$

$$\sum_{i=r+1}^{s}Log\{h(\frac{X_{(i)}-\theta}{1-\theta}|\beta,\psi)\} + (n-1)\sum_{i=r+1}^{s}Log\{1 - H(\frac{X_{(i)}-\theta}{1-\theta}|\beta,\psi)\},$$

where $g(\,\cdot\,|\alpha,\phi), G(\,\cdot\,|\alpha,\phi), h(\,\cdot\,|\beta,\psi), H(\,\cdot\,|\beta,\psi)$ are defined by (10) and (11),

$$X_{(1)} < X_{(2)} < \ldots < X_{(s)} \tag{16}$$

are the order statistics of $\underline{X}$, and r is a positive integer such that

$$X_{(r)} \le \theta < X_{(r+1)}. \tag{17}$$

By convention $X_{(0)} = -\infty, X_{(s+1)} = +\infty$.

We propose the following algorithm to maximize the log likelihood $Log\{L(\underline{X}, \Theta)\}$ (15) to determine the ML estimates of the parameters Θ using a feasible starting point

$$\Theta^* = (\Omega^*, \Phi^*, \Psi^*),\ \Omega^* = (m^*, n^*, \theta^*),\ \Phi^* = (\alpha^*, \phi^*),\ \Psi^* = (\beta^*, \psi^*)$$

and as its k-th iteration:

Step 0: Set $k = 1$, $m_1 = m^*, n_1 = n^*, \theta_1 = \theta^*$, $\Omega_1 = (m_1, n_1, \theta_1)$
$\alpha_1 = \alpha^*, \phi_1 = \phi^*, \Phi_1 = (\alpha_1, \phi_1), \beta_1 = \beta^*, \psi_1 = \psi^*, \Psi_1 = (\beta_1, \psi_1)$

Step 1: Determine Ω_{k+1} by maximizing $Log\{L(\underline{X}|\Omega, \Phi_k, \Psi_k)\}$ over $\Omega = (m, n, \theta)$.

Step 2: Determine Φ_{k+1} by maximizing $Log\{L(\underline{X}|\Omega_{k+1}, \Phi, \Psi_k)\}$ over $\Phi = (\alpha, \phi)$.

Step 3: Determine Ψ_{k+1} by maximizing $Log\{L(\underline{X}|\Omega_{k+1}, \Phi_{k+1}, \Psi)\}$ over $\Psi = (\beta, \psi)$.

Step 4: If $|Log\{L(\underline{X}|\Omega_{k+1}, \Phi_{k+1}, \Psi_{k+1})\} - Log\{L(\underline{X}|\Omega_k, \Phi_k, \Psi_k)\}| < \epsilon$
$STOP$
Else $k = k + 1$ and Goto Step 1.

Since, the log likelihood does not have to be concave and may posses local minima, there is no guarantee the algorithm above converges to a global maximum. This stresses the importance of specifying a reasonable starting Θ^* which can be obtained through some exploratory analysis of log-likelihood profile functions. We shall demonstrate this procedure in an illustrative example.

Step 1 in the algorithm above determines the optimal parameters of the general framework (5) given branch parameters Φ_k, Ψ_k and uses (15) as its objective function. Step 2 (Step 3) determines the optimal left (right) branch parameters given Ω_{k+1}, Ψ_k

$(\Omega_{k+1}, \Phi_{k+1})$ and only requires the second (third) line of log likelihood (15) in its objective function. As usual, ϵ in Step 4 may be chosen arbitrarily small.

To obtain an initial starting solution Θ^* in Step 0 one could, for example, select Θ to visually match a plot of pdf (12) to that of an empirical pdf. To further aid in the selection of θ we have from (6)

$$Pr(Y \le \theta|\Theta) = p(\Omega) = \frac{\theta n}{(1-\theta)m + \theta n} \tag{18}$$

and with (18) it follows that

$$\frac{r}{s} \le \frac{\theta n}{(1-\theta)m + \theta n} \le \frac{r+1}{s}. \tag{19}$$

Hence, given sample size s and values for m and n candidate search intervals $(X_{(r)}, X_{(r+1)}]$ for the threshold parameter θ may be determined via (19).

Figures 2A and 2B provide starting point pdf's for the Old-Faithful data obtained in this manner and compares them to the kernel density estimates in Figures 1E and 1F. We select $\phi = \psi$ to ensure continuity of the pdf (12). Figures 2C and 2D depict the log-likelihood progression through 50 iterations of the algorithm. Finally, Figures 2E and 2F plots the MLE fitted TS-EP densities fitted to the Old-Faithful data utilizing the MLE algorithm above. Starting point parameters and MLE parameter estimates are provided in Table 1 below.

Table 1: Starting point and end results parameters from MLE Algorithm for Old-Faithful duration (seconds) and waiting time (minutes) data.

	Duration (seconds)	Waiting Time (minutes)
Starting Point		
TS framework	$m = 12, n = 10, \theta = 0.4$	$m = 20, n = 12, \theta = 0.45$
Left branch	$\phi = 0.01, \alpha = 4.5$	$\phi = 0.05, \alpha = 4.5$
Right branch	$\psi = \phi = 0.01, \beta = 3.5$	$\psi = \phi = 0.05, \beta = 3$
MLE estimates		
TS framework	$m = 27.23, n = 20.51, \theta = 0.4142$	$m = 14.89, n = 10.14, \theta = 0.4468$
Left branch	$\phi = 0.0085, \alpha = 5.808$	$\phi = 0.0526, \alpha = 4.168$
Right branch	$\psi = \phi = 0.0085, \beta = 4.184$	$\psi = \phi = 0.0526, \beta = 2.888$

5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit

In this section we shall compare the fit to the Old-Faithful data depicted in Figure 1 using the parameter settings in Table 1 and copula construct (1) totaling 15 parameters to a more
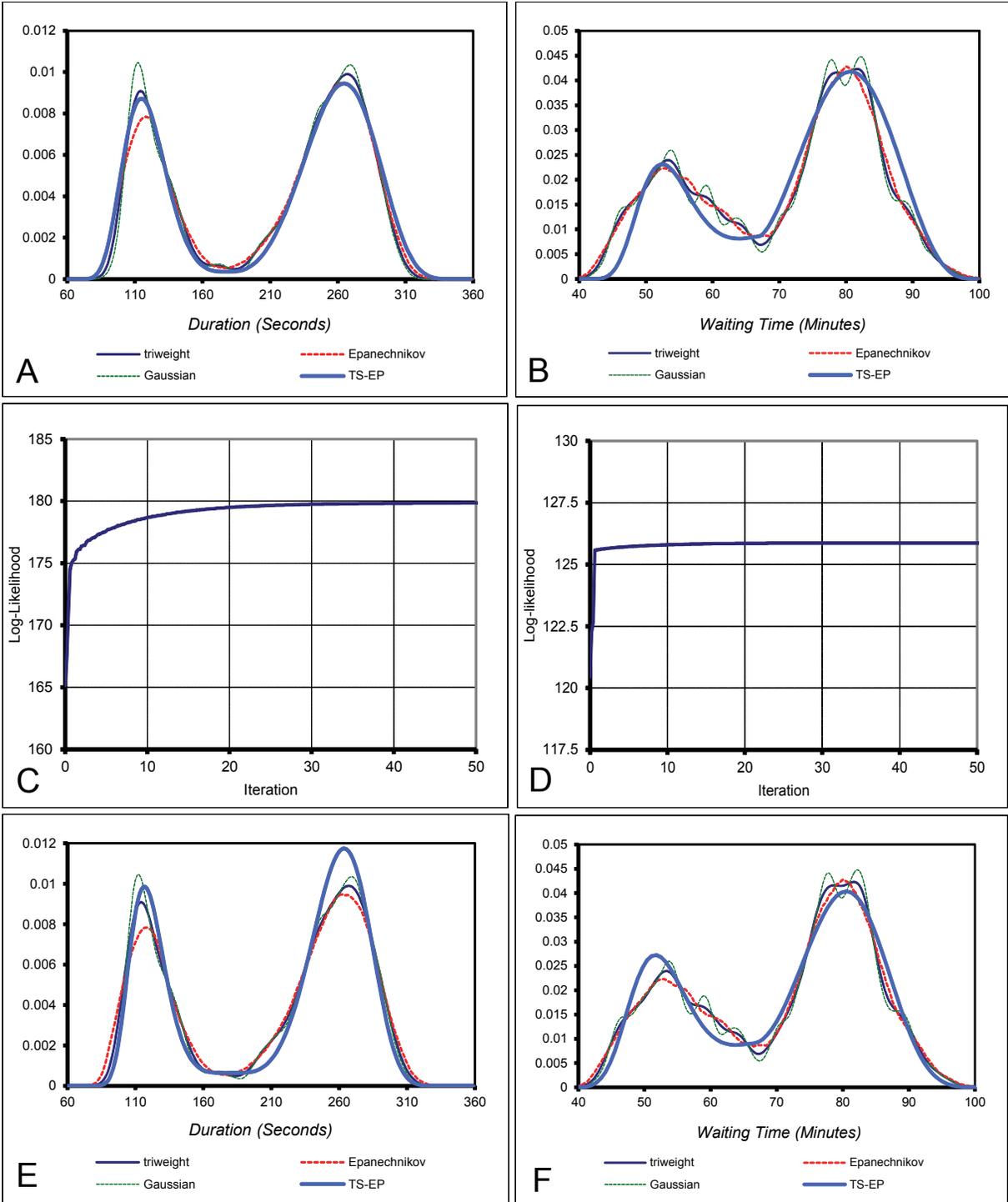
Fig. 2: A-B: Starting points for MLE Algorithm A - Duration, B - Waiting Time.

C-D: Likelihood progression MLE Algorithm C - Duration, D - Waiting Time.

E-F: ML solutions MLE Algorithm E - Duration, F - Waiting Time.

traditional technique in modeling bivariate multi modal distributions. Specifically, we shall compare the fit to a two component mixture of bivariate normal distributions with parameters estimated via the EM algorithm (see, e.g., Meng and Rubin, 1993), i.e. with pdf

$$f(x, y) = \lambda BVN_1(x, y | \underline{\mu}_1, \Sigma_1) + (1 - \lambda)BVN_2(x, y | \underline{\mu}_2, \Sigma_2), \tag{20}$$

where its 11 parameter$s$ are estimated at:

$$\lambda = 0.356, \ \underline{\mu}_1 = \begin{pmatrix} 0.2785 \\ 0.2076 \end{pmatrix}, \ \underline{\mu}_2 = \begin{pmatrix} 0.6427 \\ 0.65802 \end{pmatrix}, \tag{21}$$

$$\Sigma_1 = \begin{pmatrix} 6.90e - 3 & 1.26e - 3 \\ 1.26e - 3 & 2.77e - 3 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 7.27e - 3 & 2.67e - 3 \\ 2.67e - 3 & 6.78e - 3 \end{pmatrix}. \tag{22}$$

From Σ_1 we have a correlation of 0.28 (0.38) for the first (second) component. A two component mixture with 11 parameters was favored over a three component one with 17 using the least squares criterion, hence we selected the two member one (20) with parameter settings (21), (22) for comparison.

Firstly, Figures 3A-3D presents a visual comparison of resulting joint pdfs and contour plots. Whereas in case of Figure 3D one recognizes the typical ellipsoid contours associated with multivariate Gaussian distributions, the copula mixture technique (1) with separately estimated TS-EP marginals does not posses ellipsoid contours. Figures 3E-3F provides a visual comparison of the marginal distribution estimated utilizing both modeling techniques. One observe larger differences between the two in case of the waiting time plot 3F than the duration plot 3E.

In Tables 2 and 3 we provide a more formal comparison of the fitted marginal distributions using the following criteria: χ^2 p-value (evaluated using the equal-probability method with 17 bins), Log-Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov (KS) and Sum of Squares (SS). Of these, only the first four appropriately discount for the number of estimated parameters. Observe from Table 2 that for the duration Old-Faithful data in all six criteria the TS-EP marginal outperforms the Gaussian Mixture. Observe from Table 3 that the TS-EP marginal is preferred in only three out of the six criteria. That being said, the p-value of 0.15 in Table 3 for the TS-EP model is certainly respectable, whereas the $4.13e - 3$ value in Table 2 for the Gaussian mixture model is not.

Finally, in Figure 4 we provide a visual and QQ-plot comparison of the joint TS-EP cdf and joint Gaussian mixture. Not much can be concluded by visually comparing the cdf graphs 4A-4C. From Figures 4D and 4E we visually observe that the TS-EP joint cdf
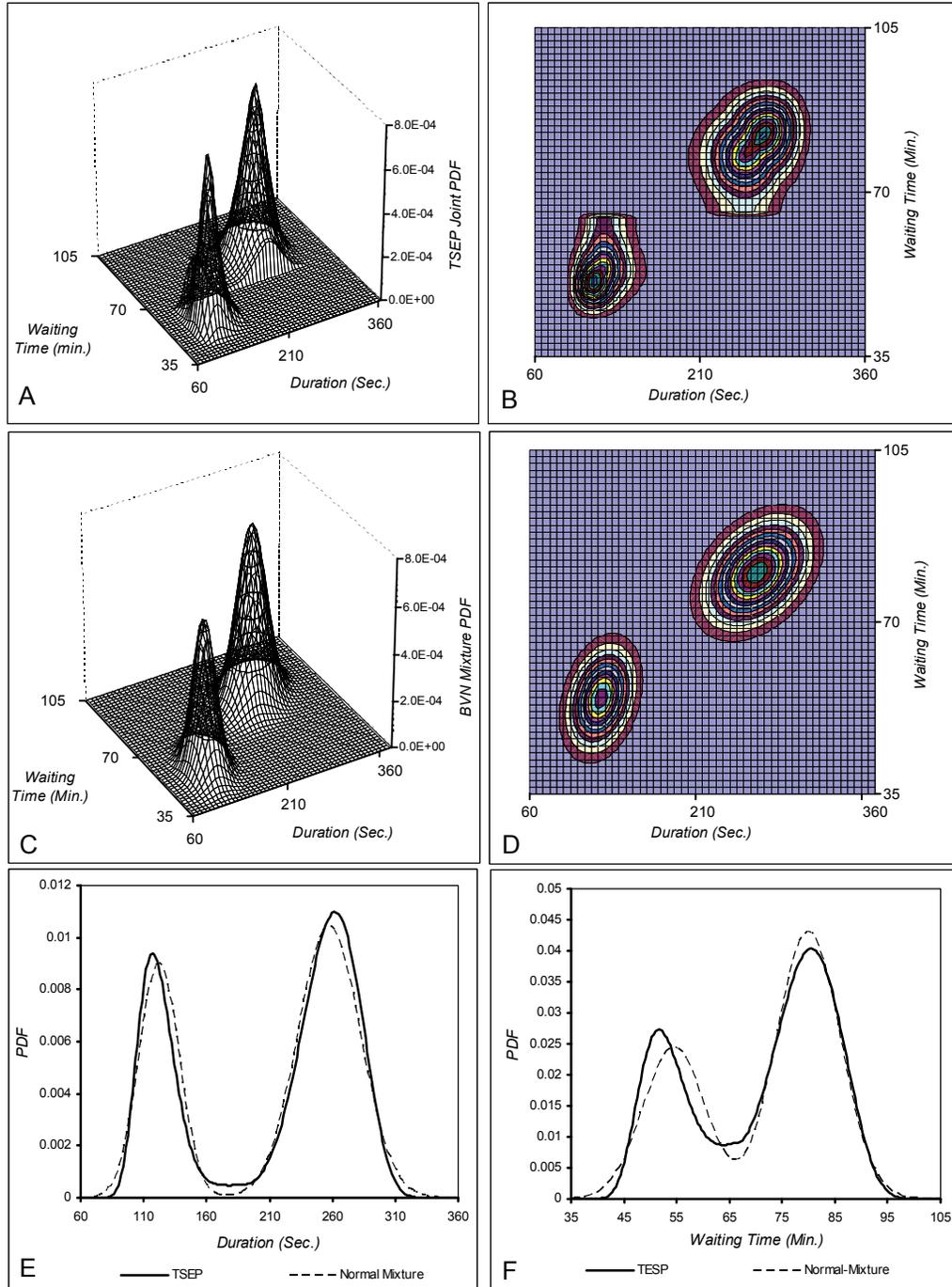
Fig. 3: Density plot and contour plot comparison of bivariate fitted TS-EP and bivariate Gaussian mixture models to Old-Faithful data set. A-B: joint pdf with TS-EP marginals (A) and its contour plot (B); C-D: joint bivariate normal mixture pdf (C) and its contour plot (D); E: Plot of TS-EP and normal mixture marginals for duration; E: Plot of TS-EP and normal mixture marginals for waiting time.

Table 2: Fit analysis comparison for TS-EP and Gaussian mixture fitted marginal distributions of the duration Old-Faithful data in Figure 1A. A bold font indicates the preferred pdf using a particular criterion.

	TS-EP pdf	Normal mixture pdf
χ^2 p-value	**0.260**	$4.13e-3$
Log-Likelihood	**178.59**	160.94
AIC	$-$**345.19**	-311.88
BIC	$-$**323.55**	-293.85
KS - criterion	**0.037**	0.045
SS - criterion	**0.043**	0.11

Table 3: Fit analysis comparison for TS-EP and Gaussian mixture fitted marginal distributions of the waiting time Old-Faithful data in Figure 1A.

	TS-EP pdf	Normal mixture pdf
χ^2 p-value	0.150	**0.353**
Log-Likelihood	**125.81**	122.39
AIC	$-$**239.63**	-234.78
BIC	$-$**217.99**	-216.75
KS - criterion	0.037	**0.044**
SS - criterion	0.043	**0.058**

outperforms the joint Gaussian mixture cdf in the lower quantile ranges. This translates in a better Kolmogorov-Smirnov (KS) and Sum-of-Squares (SS) criterion for the TS-EP joint cdf in figure 4D than the joint Gaussian mixture cdf in Figure 4E. Truth be told however, the joint Gaussian mixture cdf only uses 11 parameters, whereas the TS-EP joint cdf estimates a total of 15 parameters.

Summarizing, whereas from a marginal perspective one could argue the TS-EP models perform at least as well as the Gaussian Mixture model, such a conclusion cannot be made in a joint sense due to the difference in number of estimated parameters. Preference of either model may depend on the application context. For example, the closed form expressions of the copula mixture model using TS-EP marginals allow for a straightforward bivariate sampling algorithm. This is certainly more challenging in case of a mixture of bivariate Gaussian distributions.
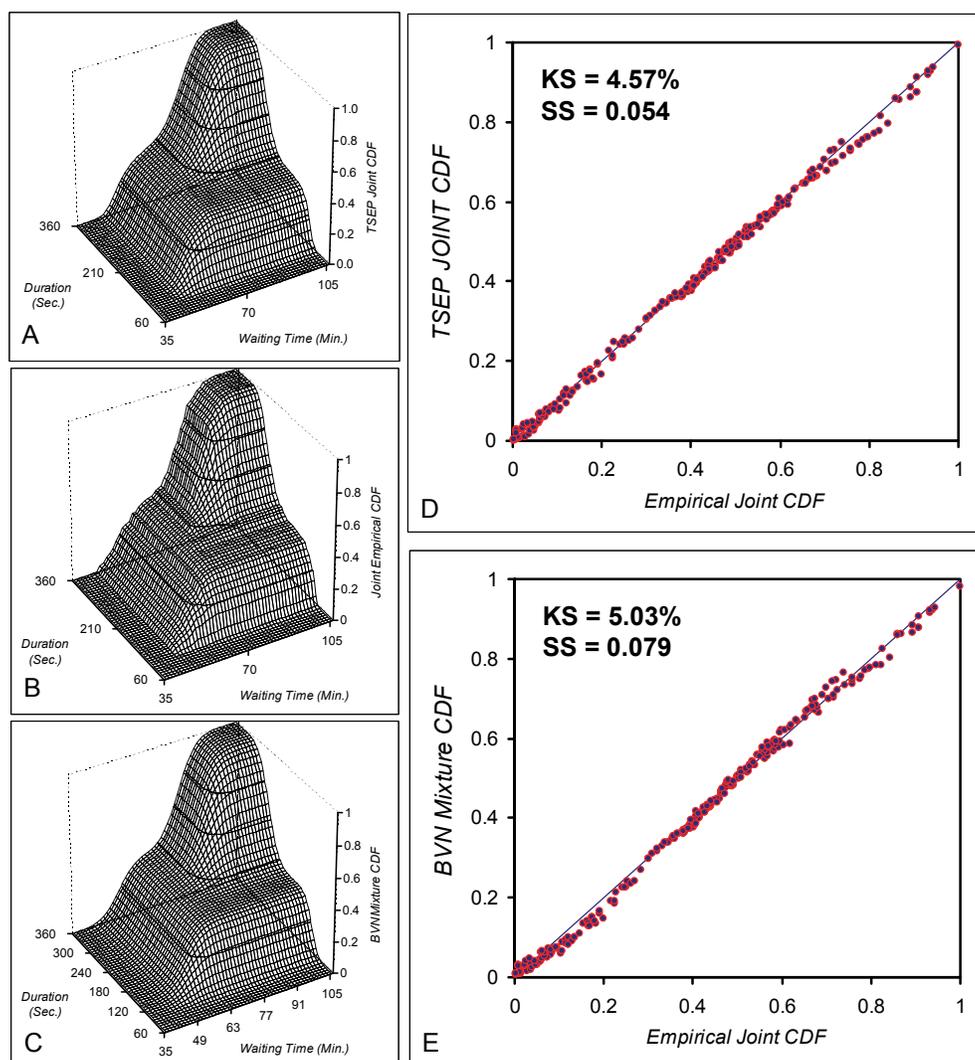
Fig. 4: A bivariate fit comparison of TS-EP joint cdf and bivariate Gaussian mixture model. A: TS-EP joint cdf; B: Empirical CDF; C: BVN mixture cdf; D: QQ Plot TS-EP joint cdf; E: QQ plot Gaussian Mixture CDF.

Concluding Remarks

In this paper we have presented a novel procedure for modeling the classical Old Faithful data set. In particular two aspects deserve attention. The first one is the introduction of a two-sided bivariate mixture technique given by (1) utilizing two copulas as its components and the resulting mixture again being a copula. The second one is the introduction of a novel univariate distribution for modeling bimodal distributions. Both aspects are integrated in the distribution model for the Old Faithful Geyser data set outperforming the tradition bivariate normal mixture approach. While tempting to compare the correlations

0.287 and 0.380 of the components of the bivariate normal mixture with the copula component correlations 0.214 and 0.278, one needs to recognize that the former are Pearson moment correlations, whereas the latter are Spearman rank correlations.

References

Atkinson, A.C, Riani, M. (2007). Exploratory tools for clustering multivariate data, *Computational Statistics* & *Data Analysis* 52 (1) 272 – 285

Azzalini, A. and Bowman, A.W. (1990). A look at some data on the Old Faithful geyser. *Applied Statistics*, 39, pp. 357–365.

Dekking, F.M., Kraaikamp, C., Lopuhaä, H.P. and Meester, L.E. (2005). *A Modern Introduction to Probability and Statistics, Understanding Why and How*, Springer-Verlag.

Eilers, P.H.C. and Borgdorff, M.W.(2007). Non-parametric log-concave mixtures, *Computational Statistics & Data Analysis*, 51(11), 5444-5451.

García, C.B., García Pérez, J. , and van Dorp, J.R. (2011). Modeling Heavy-Tailed, Skewed and Peaked Uncertainty Phenomena with Bounded Support, *Statistical Methods and Applications,* Vol. 20, No. 4., pp. 463–-482.

Kotz, S. and van Dorp, J.R. (2010), Generalized Diagonal Band Copulas with Two-Sided Generating Densities, *Decision Analysis*, 7 (2), 196-214.

Nelsen, R.B. (1999). *An Introduction to Copulas.* Springer, New York.

Meng, X.L., and Rubin, D.B, (1993). Maximum likelihood estimation via the ECM algorithm: a general framework, *Biometrika* 80, 267-278.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall (London).

Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons, Chichester.

Vicari, D., van Dorp, J.R. and Kotz, S. (2008), Two-sided generalized Topp and Leone (TS-GTL) distributions, *Journal of Applied Statistics*, Vol. 35 (10), pp. 1115 - 1129.