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

"Presentation Short Course: Beyond Beta and Applications"

***November 20th, 2018**, La Sapienza*

Donatella Vicari[1], **Johan Rene van Dorp**[2]

[1] Department of Statistics, Probability and Applied Statistics, University of Rome "La Sapienza", Rome, Italy

[2] Corresponding Author, Department of Engineering Management and Systems Engineering, The George Washington University, Washington D.C., USA

OUTLINE

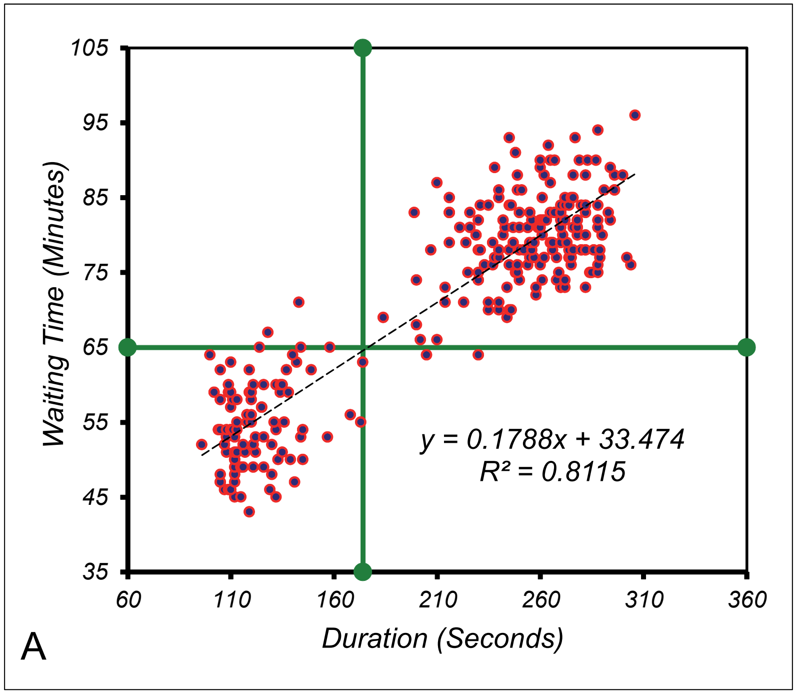
1. **Introduction**
2. A two-sided framework of univariate distributions
3. PDF and CDF of TS-EP distributions
4. Maximum Likelihood Estimation
5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit
6. Concluding Remarks
7. Some References

1. INTRODUCTION... Old-Faithfull Geyser Data

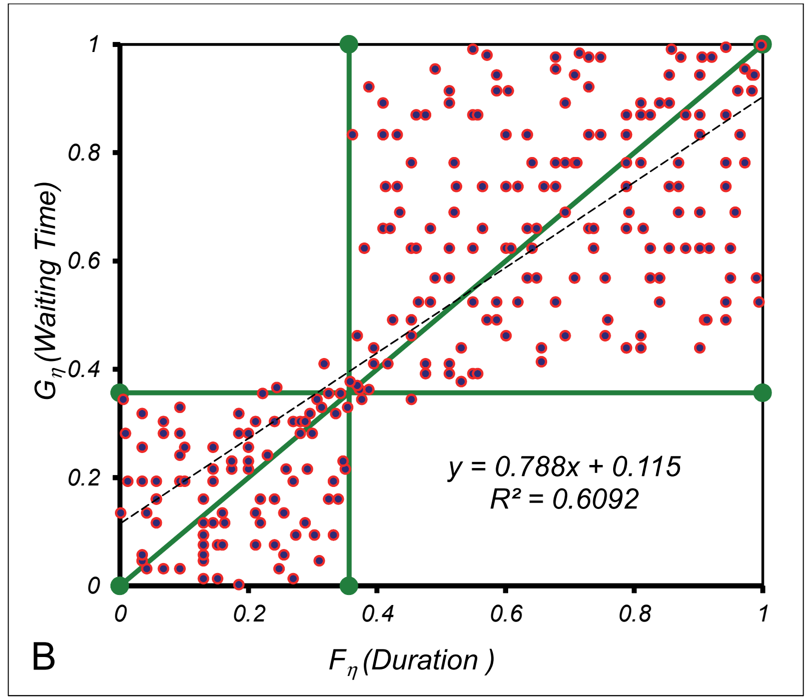
- **The Old Faithful geyser at Yellowstone National Park**, Wyoming, USA, was observed from August 1st to August 15th, 1985. During that time, data were collected on eruptions. **There were 272 durations (D) of and waiting time (W) between eruptions observed**, of which 15 data points are listed below.

	Duration (s)	Waiting Time (Min)
1	119	43
2	112	45
3	115	45
4	132	45
5	107	46
6	109	46
7	110	46
8	110	46
9	129	46
10	105	47
11	105	47
12	112	47
13	141	47
14	105	48
15	112	48

- Is has been a popular data set to demonstrate a variety of statistical techniques e.g: **kernel density estimation** (e.g. Silverman, 1986), **time-series analysis** (e.g. Azzalini and Bowman, 1990), **clustering** (e.g. Atkinson and Riani, 2006), and **distribution theory** (e.g. Eilers and Borgdorff, 2007), to name a few. **This presentation/paper falls in the latter category**.



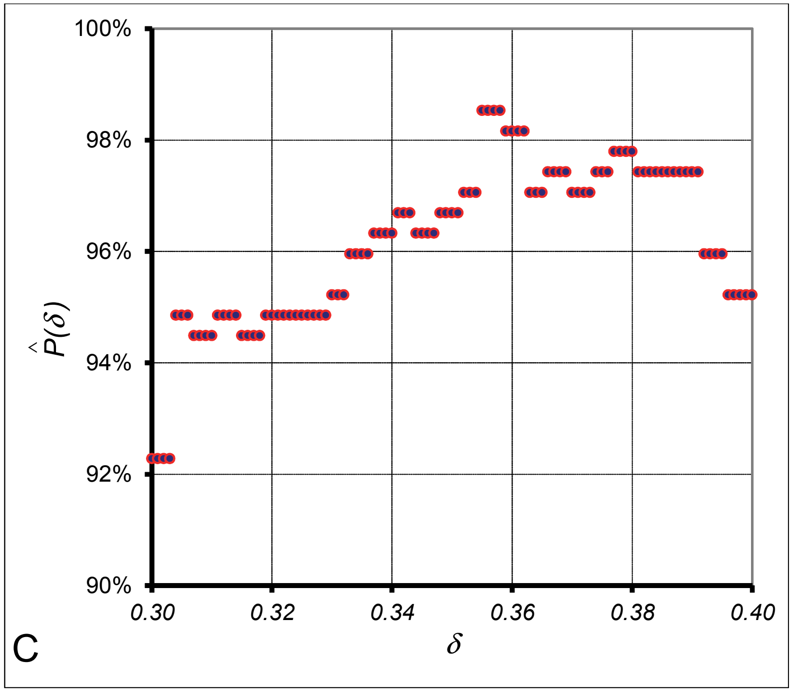
Scatter plot of (d_i, w_i).



Scatter plot of $(F_\eta(d_i), G_\eta(w_i))$.

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

- One observes a **clear clustering** and **a strong statistical dependence** in (D, W).

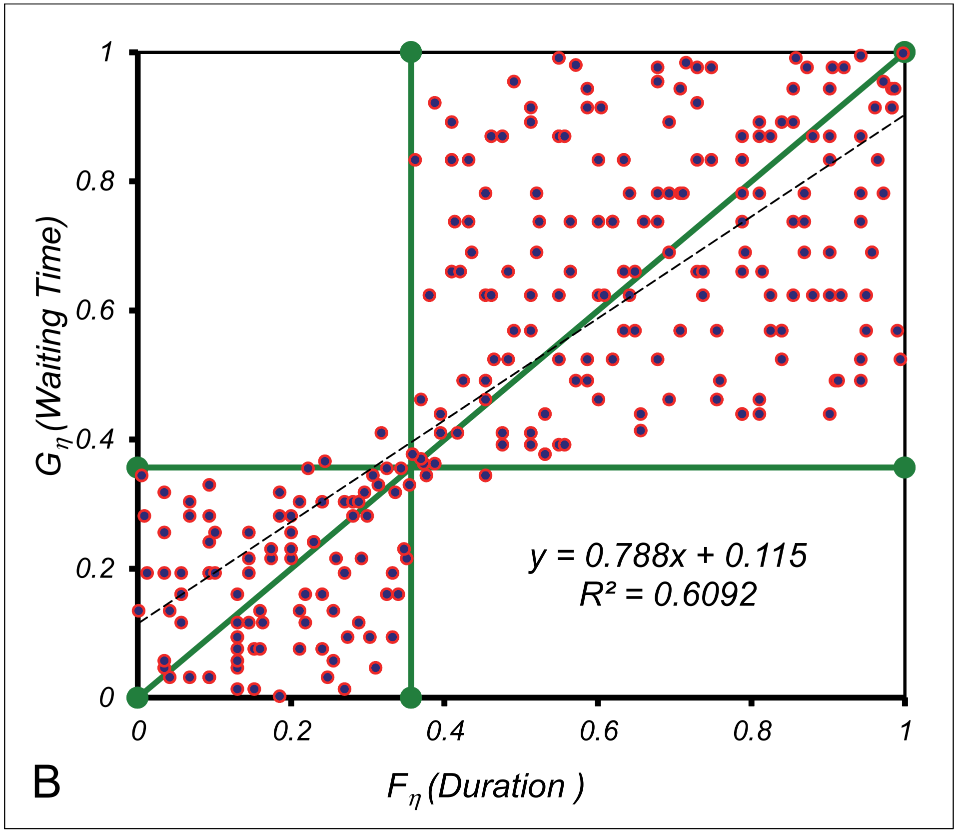


$$\widehat{P}(\delta) = \frac{\#\text{ data points in } \in [0,\delta]^2 \cup (\delta,1]^2}{\eta}$$

- **98.5%** of data $(F_\eta(d_i), G_\eta(w_i)) \in [0,\delta]^2 \cup (\delta,1]^2$ for $\delta \in (0.355, 0.358)$.
- In fact, **only 4 out of the 272 data points** do not fall within this area.
- A challenge in modeling **a bivariate distribution** for this data is the lack of observations in $[0,\delta] \times [\delta,1]$ and $[\delta,1] \times [0,\delta]$ areas.
- A copula is **a joint distribution with $U[0,1]$ marginals.**
- Use **a two-dimensional mixture technique** with $S(\cdot,\cdot)$ and $T(\cdot,\cdot)$ **copulas.**

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

It is not difficult to show that **$C(u,v|\delta)$ then is too a copula** on $[0,1]^2$.



- **Set δ equal to mid-point of interval:**

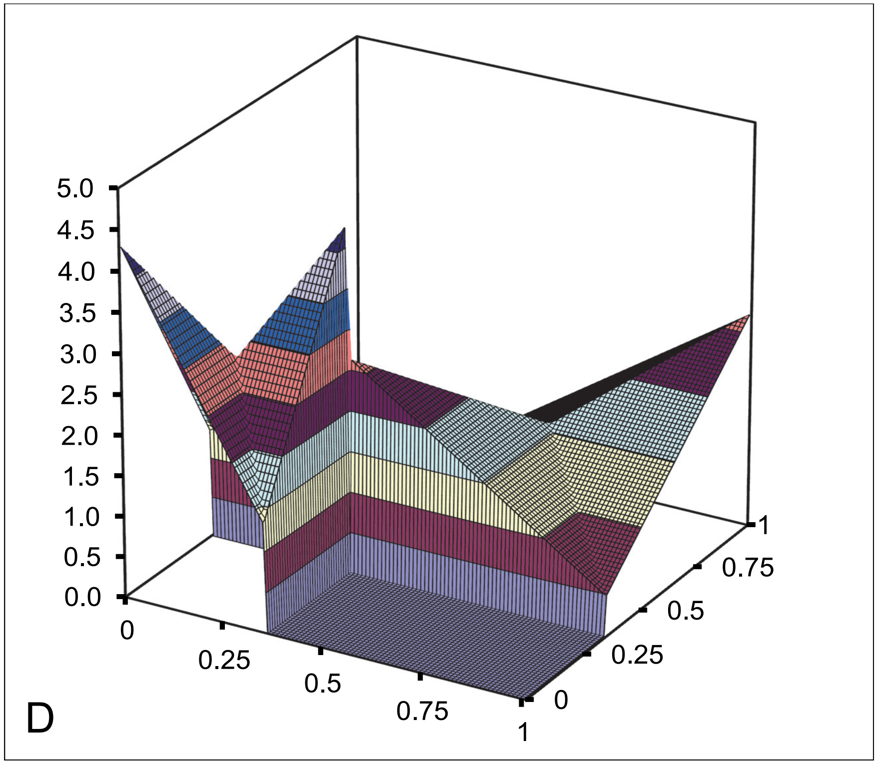
$$\delta = (0.355 + 0.358)/2 = 0.3565.$$

- **$S(\cdot,\cdot)$ a copula with $\rho = 0.214$, $T(\cdot,\cdot)$ a copula with $\rho = 0.278$** $\Rightarrow$ bivariate copula $C(u, v|\delta)$ captures these characteristics of figure to the left.

- For $S(\cdot,\cdot|\alpha_s)$ and $T(\cdot,\cdot|\alpha_t)$ select **generalized diagonal band copulas with Two-Sided slope generating densities** (see, Kotz and van Dorp, 2010) with

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

$$\rho(\alpha_s) = 0.214,\ \rho(\alpha_t) = 0.278 \Rightarrow \alpha_s = 1.535,\ \alpha_t = 1.696.$$



- We have for **the density $S(u,v\,|\,\alpha_s)$** :

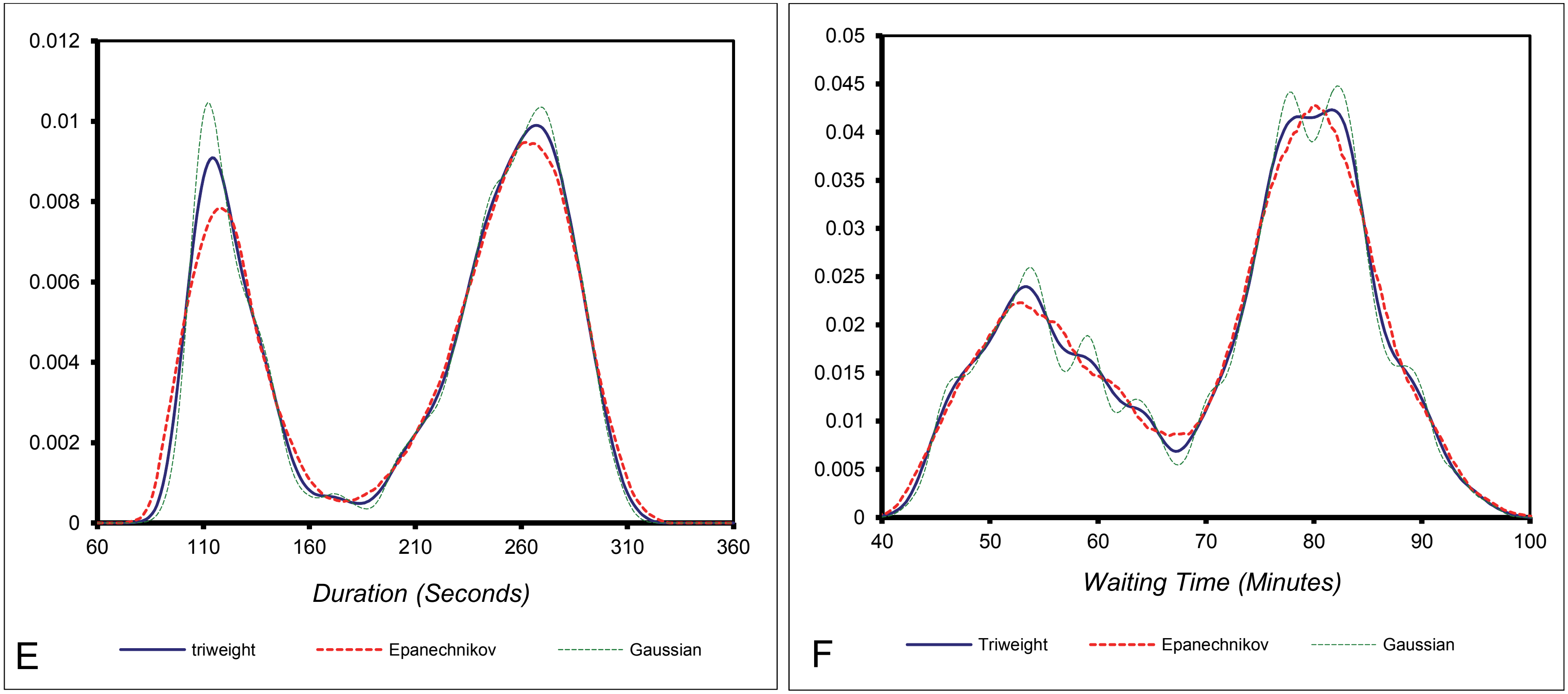
$$\begin{cases} \alpha_s - 2(\alpha_s - 1)v, & (x,y) \in A_1, \\ \alpha_s - 2(\alpha_s - 1)u, & (x,y) \in A_2, \\ \alpha_s - 2(\alpha_s - 1)(1-u), & (x,y) \in A_3, \\ \alpha_s - 2(\alpha_s - 1)(1-v), & (x,y) \in A_4, \end{cases}$$

where $0 \leq \alpha_s \leq 2$ and **areas:**

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

- **Same for the density $T(u,v\,|\,\alpha_t)$**

- **Advantages of copulas $S(u,v\,|\,\alpha_s), T(u,v|\alpha_T)$ and $C(u,v|\delta)$** are they have **closed form pdfs and cdfs**. Most importantly, they were s**hown to be approximately least informative in the entropy sense** given their correlation constraints (see, Kotz and van Dorp, 2010).



- **All that remains** is the **modeling of the marginal distributions** for the old faithfull geyser data. Naturally, one could follow the traditional approach of non-parametric density estimation through **kernel density estimation.**

- Our goal is to **develop a parametric model for these marginals.**

OUTLINE

1. Introduction
2. **A two-sided framework of univariate distributions**
3. PDF and CDF of TS-EP distributions
4. Maximum Likelihood Estimation
5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit
6. Concluding Remarks
7. Some References

- **Vicari et al.'s (2008) Two-Sided (TS) framework** using cdfs $G(\cdot\,|\,\Phi)$ and $H(\cdot\,|\,\Psi)$:

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

where $\Theta = (\Omega, \Phi, \Psi)$, $\Omega = (\theta, m, n)$ are the TS power parameters and

$$p(\Omega) = \frac{\theta n}{(1-\theta)m + \theta n} \text{ (not a function of } G(\cdot\,|\,\Phi) \text{ and } H(\cdot\,|\,\Psi)).$$

- One obtains for the corresponding **probability density function (pdf)** :

$$f_Y(y|\Theta) = \frac{mn}{(1-\theta)m + \theta n} \begin{cases} g\big(\tfrac{y}{\theta}\big)\Big\{G\big(\tfrac{y}{\theta}\,|\,\Phi\big)\Big\}^{m-1}, & \text{for } 0 < y < \theta, \\ h\big(\tfrac{y-\theta}{1-\theta}\big)\Big\{1 - H\big(\tfrac{y-\theta}{1-\theta}\,|\,\Psi\big)\Big\}^{n-1}, & \text{for } \theta \leq y < 1, \end{cases}$$

- $f_Y(\theta^+|\Theta) - f_Y(\theta^-|\Theta) = h\big(0\,|\,\Psi\big) - g\big(1\,|\,\Phi\big) \Rightarrow$ **discontinuous when $\neq 0$**.

OUTLINE

1. Introduction
2. A two-sided framework of univariate distributions
3. **PDF and CDF of TS-EP distributions**
4. Maximum Likelihood Estimation
5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit
6. Concluding Remarks
7. Some References

- Use a [**reflected**] **elevated power distribution on $[0,1]$** García et al (2011) with pdfs:

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

where ϕ, $\psi \in [0,1]$, $\alpha, \beta \geq 1$. Observe **elevated power distributions** are themselves **mixtures of a uniform and a power distribution**.

- One obtains the **probability density function (pdf):**

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

- $f_Y(\theta^+|\Theta) - f_Y(\theta^-|\Theta) = \psi - \phi \Rightarrow$ **continuous on $[0,1]$ provided $\psi = \phi$**.

OUTLINE

1. Introduction
2. A two-sided framework of univariate distributions
3. PDF and CDF of TS-EP distributions
4. **Maximum Likelihood Estimation**
5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit
6. Concluding Remarks
7. Some References

- For a random ordered sample $\underline{X} = (X_{(1)}, \ldots, X_{(s)})$ the **log likelihood function** is :

$$Log\{L(\underline{X}, \Theta)\} = sLog\{\frac{mn}{(1-\theta)m+\theta n}\} +$$
$$\sum_{i=1}^{r} Log\{g(\frac{X_{(i)}}{\theta}|\alpha,\phi)\} + (m-1)\sum_{i=1}^{r} Log\{G(\frac{X_{(i)}}{\theta}|\alpha,\phi)\} +$$
$$\sum_{i=r+1}^{s} Log\{h(\frac{X_{(i)}-\theta}{1-\theta}|\beta,\psi)\} + (n-1)\sum_{i=r+1}^{s} Log\Big\{1 - H\Big(\frac{X_{(i)}-\theta}{1-\theta}|\beta,\psi\Big)\Big\},$$

and ***r* is a positive integer** such that

$$X_{(r)} \leq \theta < X_{(r+1)}.$$

- The log likelihood function **does not have to be concave**.
- There is **no guarantee an MLE 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.**

- **We propose the following algorithm** to maximize the **log likelihood** $Log\{L(\underline{X},\Theta)\}$ 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 $k = 1$, $m_1 = m^*, n_1 = n^*, \theta_1 = \theta^*$, $\Omega_1 = (m_1, n_1, \theta_1)$
$\alpha_1 = \alpha^*, \phi_1 = \phi^*, \Phi_1 = (\alpha_1, \phi_1), \beta_1 = \beta^*, \psi_1 = \psi^*, \Psi_1 = (\beta_1, \psi_1)$

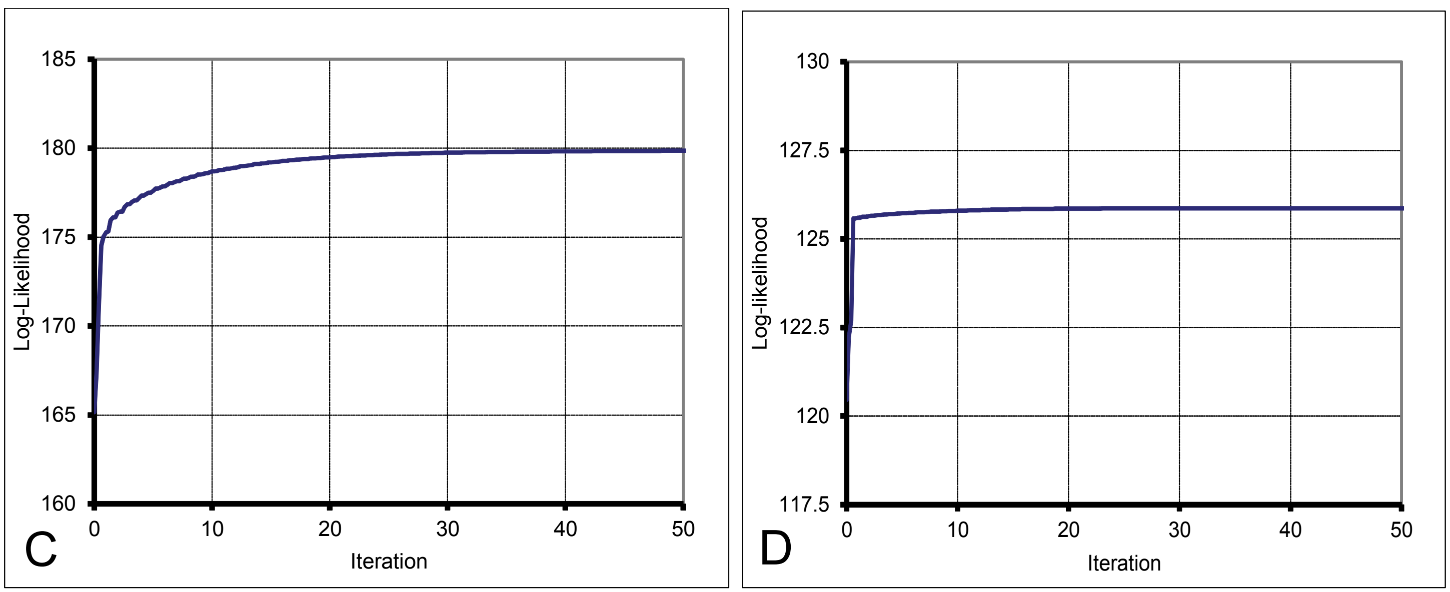
Step 1: Determine $\boldsymbol{\Omega}_{k+1}$ by maximizing $Log\{L(\underline{X}|\Omega, \boldsymbol{\Phi}_k, \boldsymbol{\Psi}_k)\}$ over $\Omega = (m, n, \theta)$.

Step 2: Determine $\boldsymbol{\Phi}_{k+1}$ by maximizing $Log\{L(\underline{X}|\boldsymbol{\Omega}_{k+1}, \Phi, \boldsymbol{\Psi}_k)\}$ over $\Phi = (\alpha, \phi)$.

Step 3: Determine $\boldsymbol{\Psi}_{k+1}$ by maximizing $Log\{L(\underline{X}|\boldsymbol{\Omega}_{k+1}, \boldsymbol{\Phi}_{k+1}, \Psi)\}$ over $\Psi = (\beta, \psi)$.

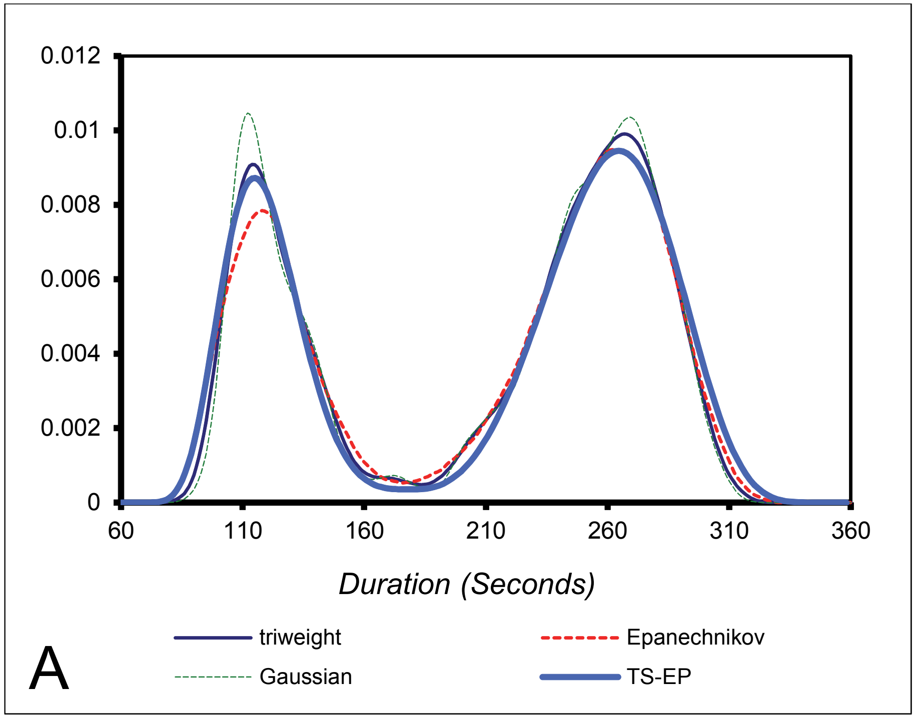
Step 4: If $|Log\{L(\underline{X}|\boldsymbol{\Omega}_{k+1}, \boldsymbol{\Phi}_{k+1}, \boldsymbol{\Psi}_{k+1})\} - Log\{L(\underline{X}|\boldsymbol{\Omega}_k, \boldsymbol{\Phi}_k, \boldsymbol{\Psi}_k)\}| < \epsilon$
$STOP$
Else $k = k + 1$ and Goto Step 1.

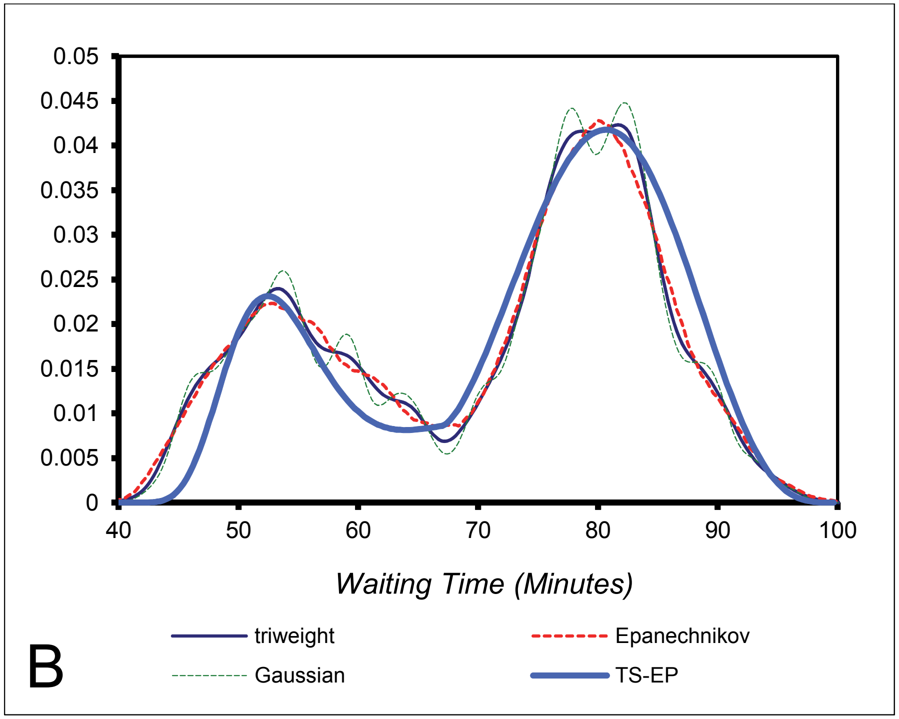
	Duration (seconds)	Waiting Time (minutes)
Starting Point		
TS framework	$m = 12,\ n = 10,\ \theta = 0.4$	$m = 20,\ n = 12,\ \theta = 0.45$
Left branch	$\phi = 0.01,\ \alpha = 4.5$	$\phi = 0.05,\ \alpha = 4.5$
Right branch	$\psi = \phi = 0.01,\ \beta = 3.5$	$\psi = \phi = 0.05,\ \beta = 3$
MLE estimates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



Log-Likelihood progression by iteration

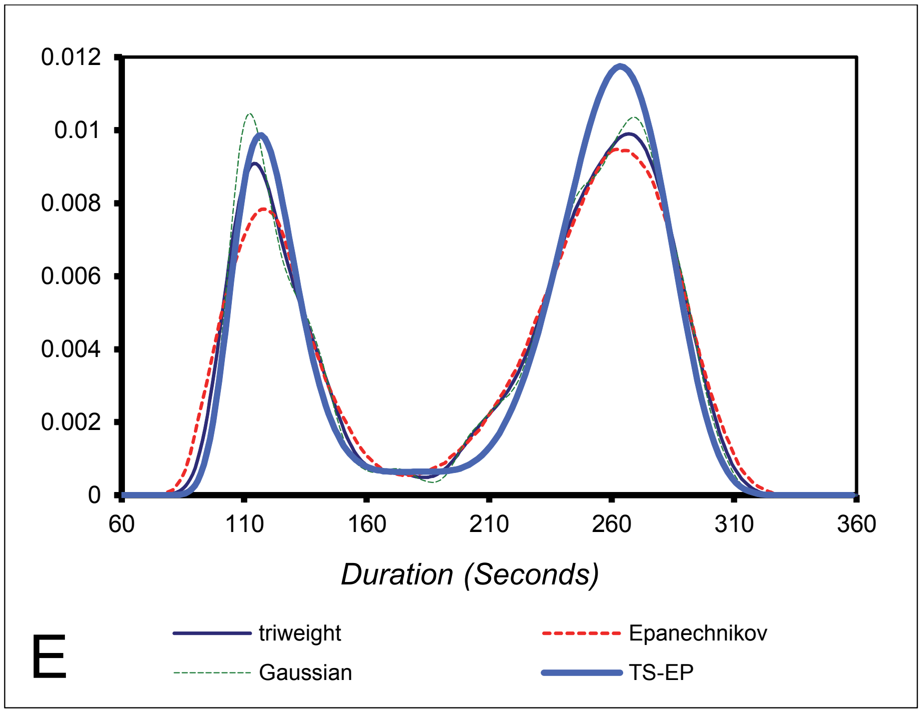
STARTING POINT MLE ALGORITHM



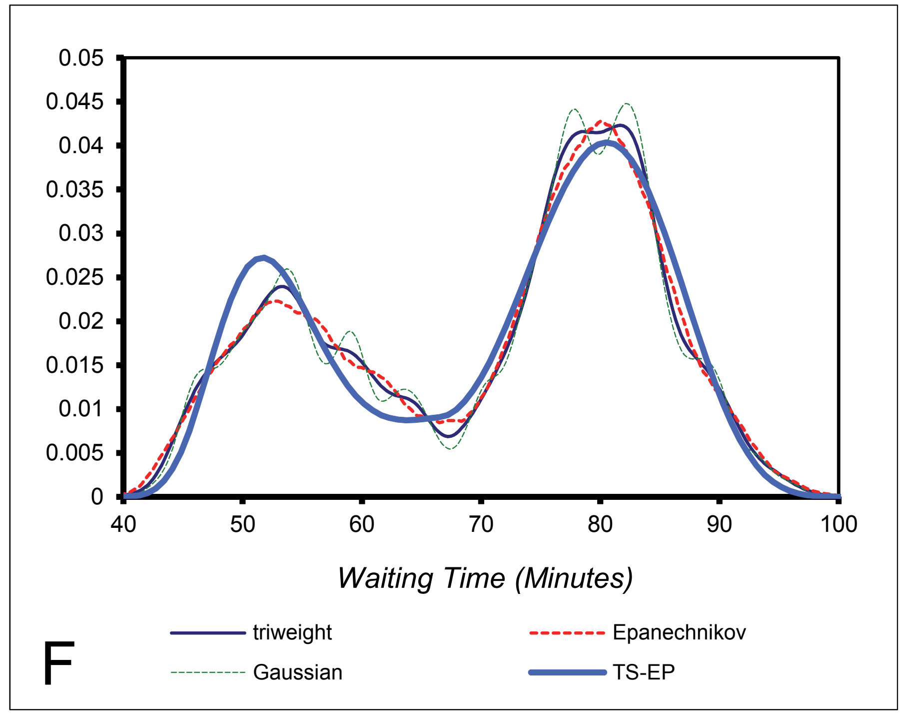


Duration **Waiting Time**

SOLUTION MLE ALGORITHM



Duration



Waiting Time

OUTLINE

1. Introduction
2. A two-sided framework of univariate distributions
3. PDF and CDF of TS-EP distributions
4. Maximum Likelihood Estimation
5. **A comparison of the Joint TS-EP fit to a bivariate mixture model fit**
6. Concluding Remarks
7. Some References

- **Compare fit** to the Old-Faithful data using construction method **herein totaling 15 parameters** with modeling **bivariate multi modal distribution**

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

where **its 11 parameters** are estimated at:

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

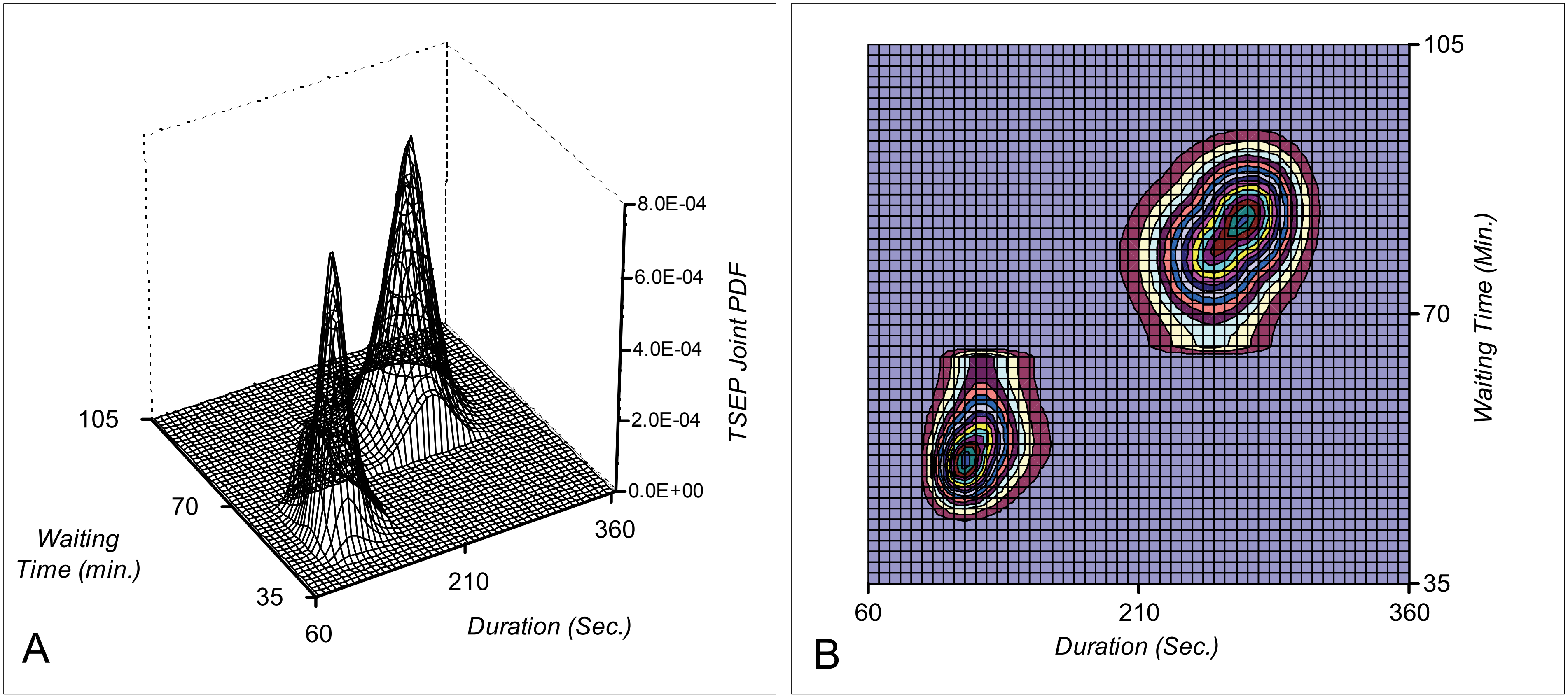
$$\Sigma_1 = \begin{pmatrix} 6.90e-3 & 1.26e-3 \\ 1.26e-3 & 2.77e-3 \end{pmatrix},\ \Sigma_2 = \begin{pmatrix} 7.27e-3 & 2.67e-3 \\ 2.67e-3 & 6.78e-3 \end{pmatrix}. \tag{22}$$

- From Σ_1 we have a **correlation of 0.28** (**0.38**) for **the first** (**second**) component.

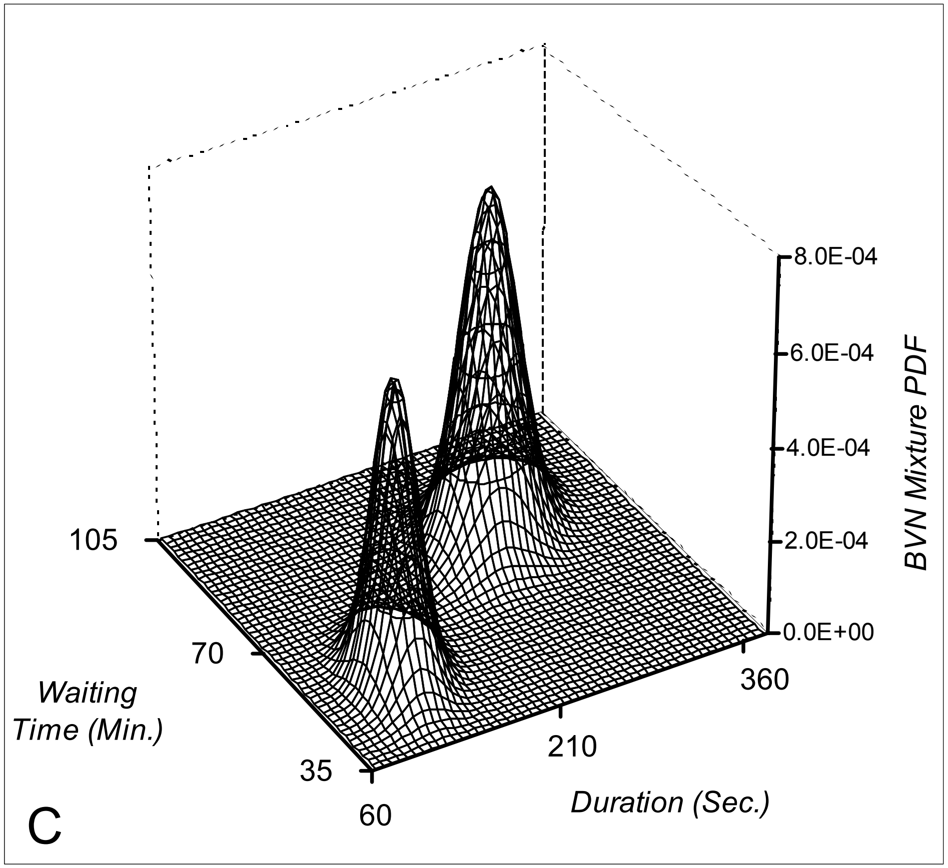
- **A two component BVN mixture with 11 parameters** was favored over **a three component BVN one with 17 parameters** using **the least squares criterion.**

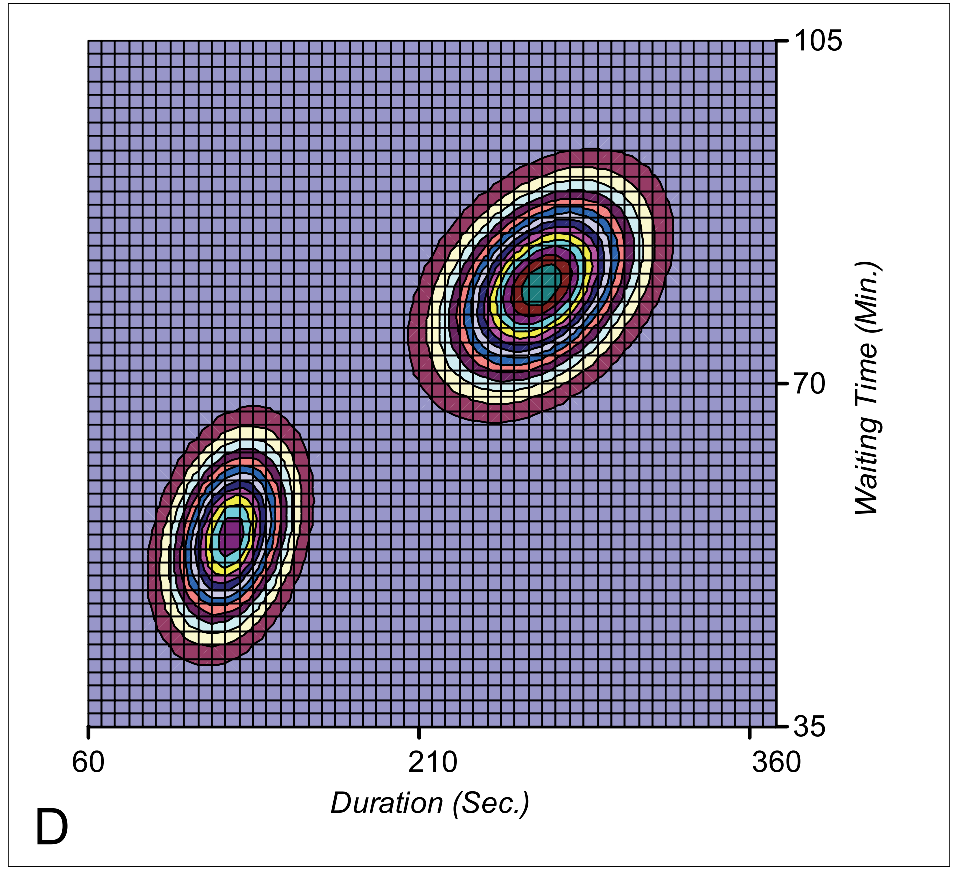
- **Parameters** of two component mixture of bivariate normal distributions **were estimated via the EM algorithm** (see, e.g., Meng and Rubin, 1993).

Joint Probability Density Function Two-Sided Elevated Power Model



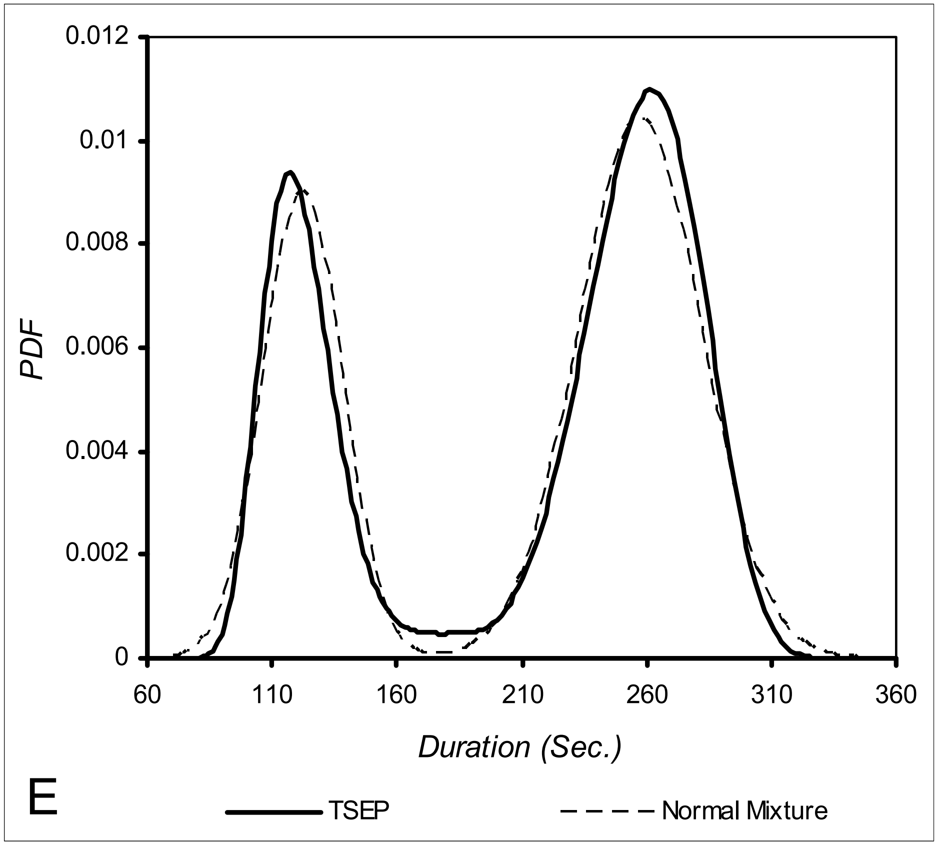
Joint Probability Density Function Bivariate Normal Mixture

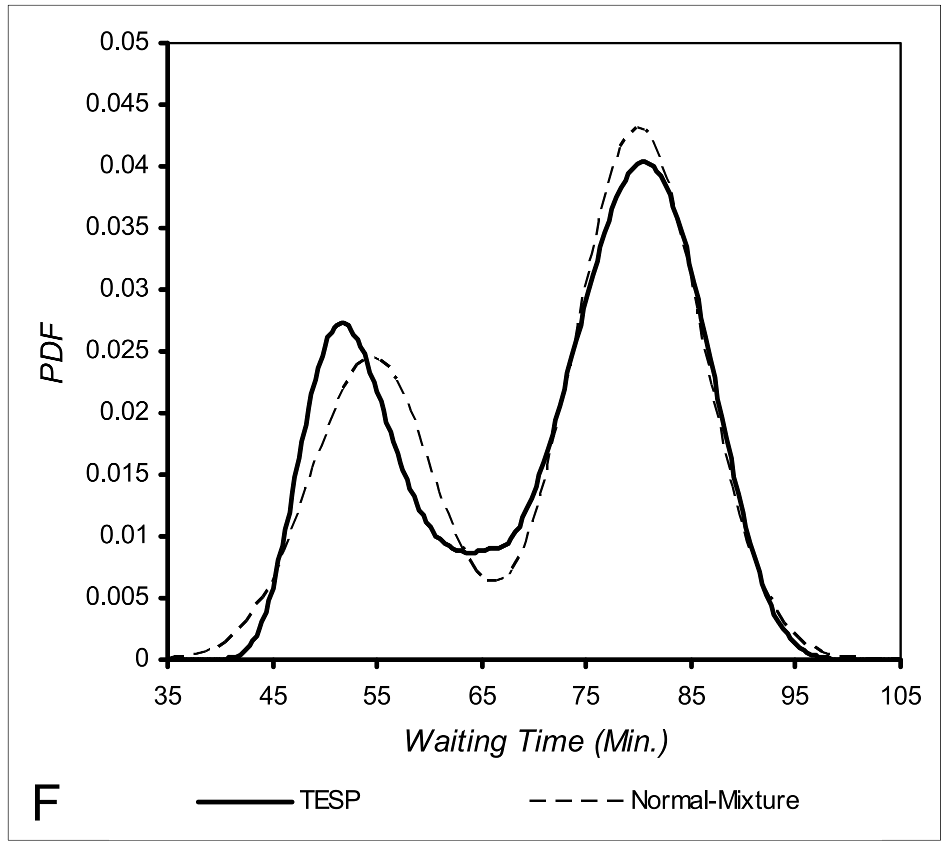




Durations Old Faithfull Data

Waiting Times Old Faithfull Data





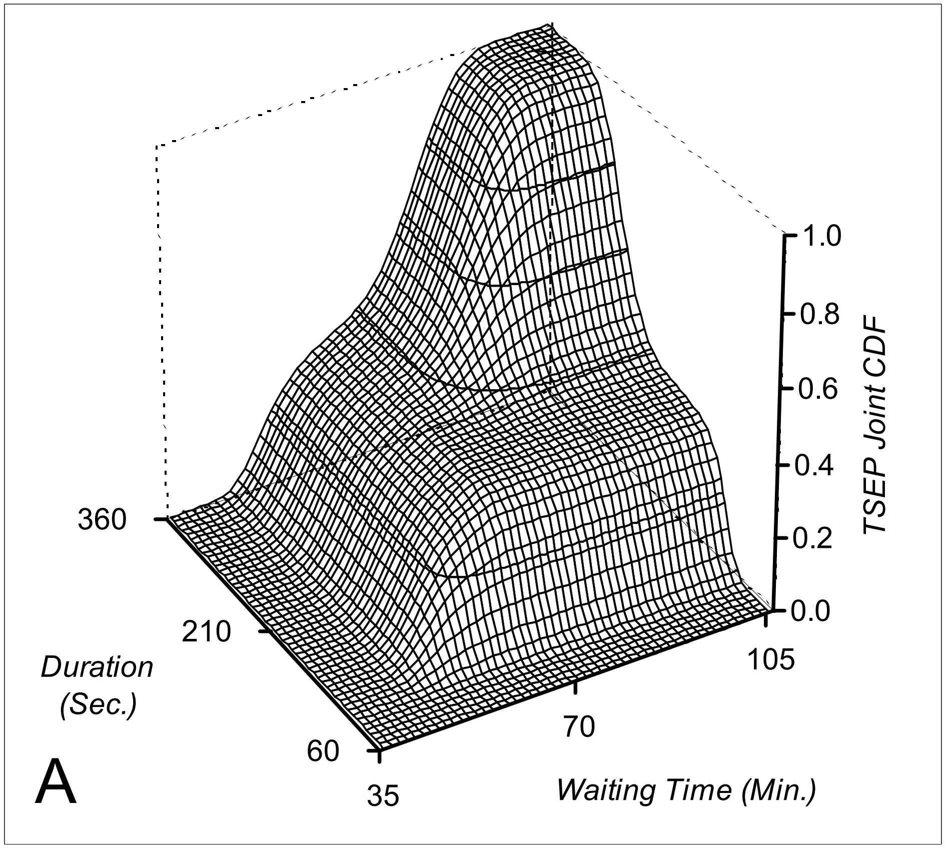
Fit Comparison Durations Old-Faithful data

	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

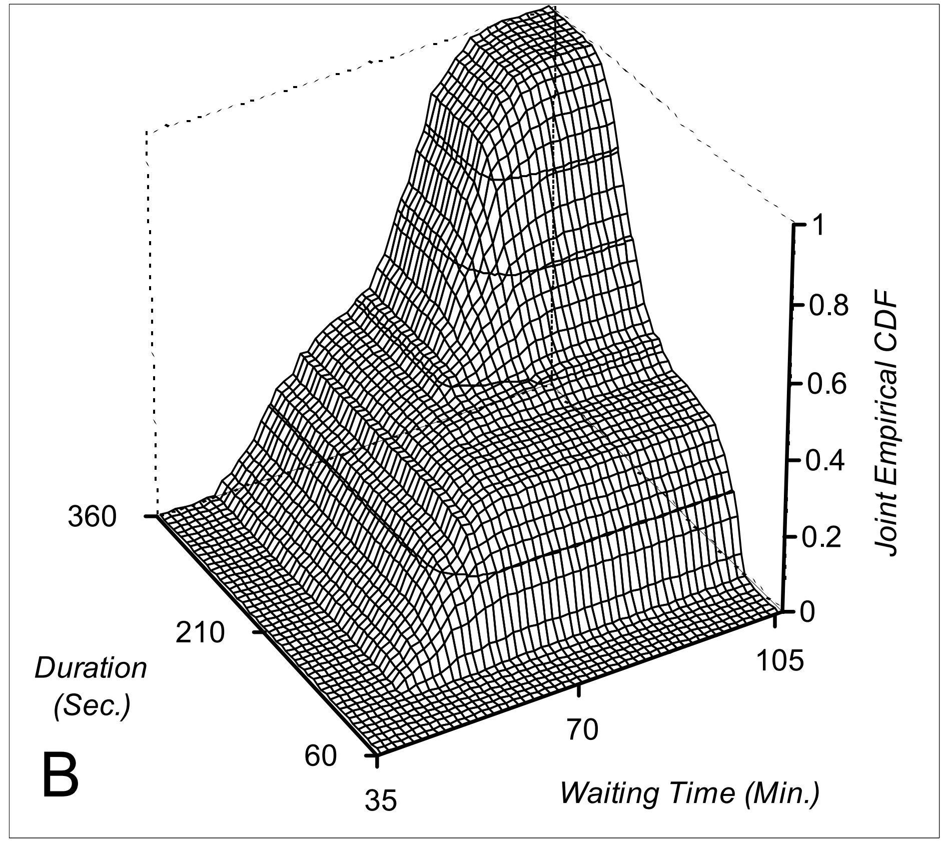
Fit Comparison Waiting Time Data Old-Faithful data

	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**

Joint Cumulative Distribution Function

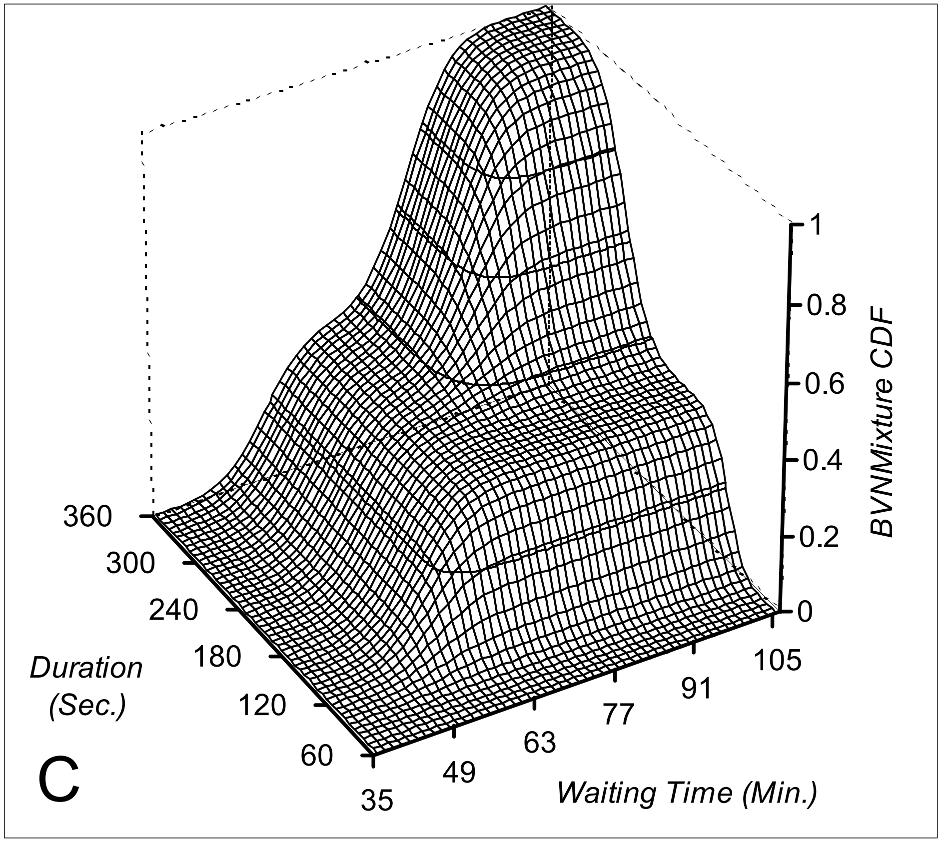


Two-Sided Elevated Power Model

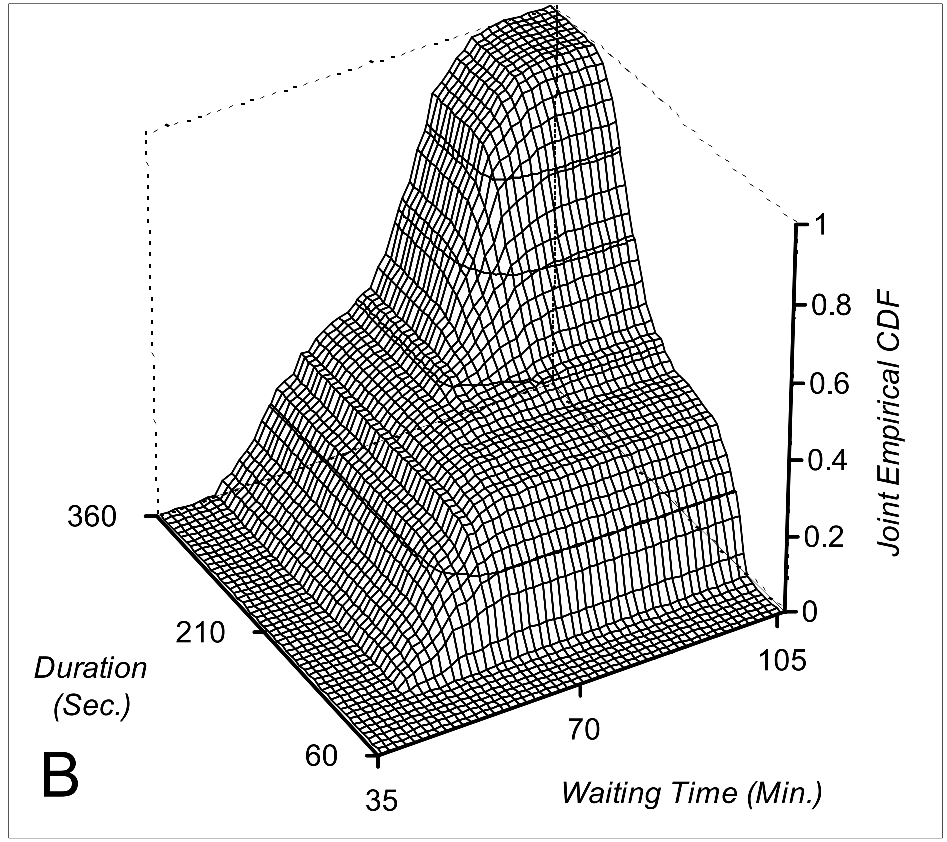


Empirical CDF

Joint Cumulative Distribution Function

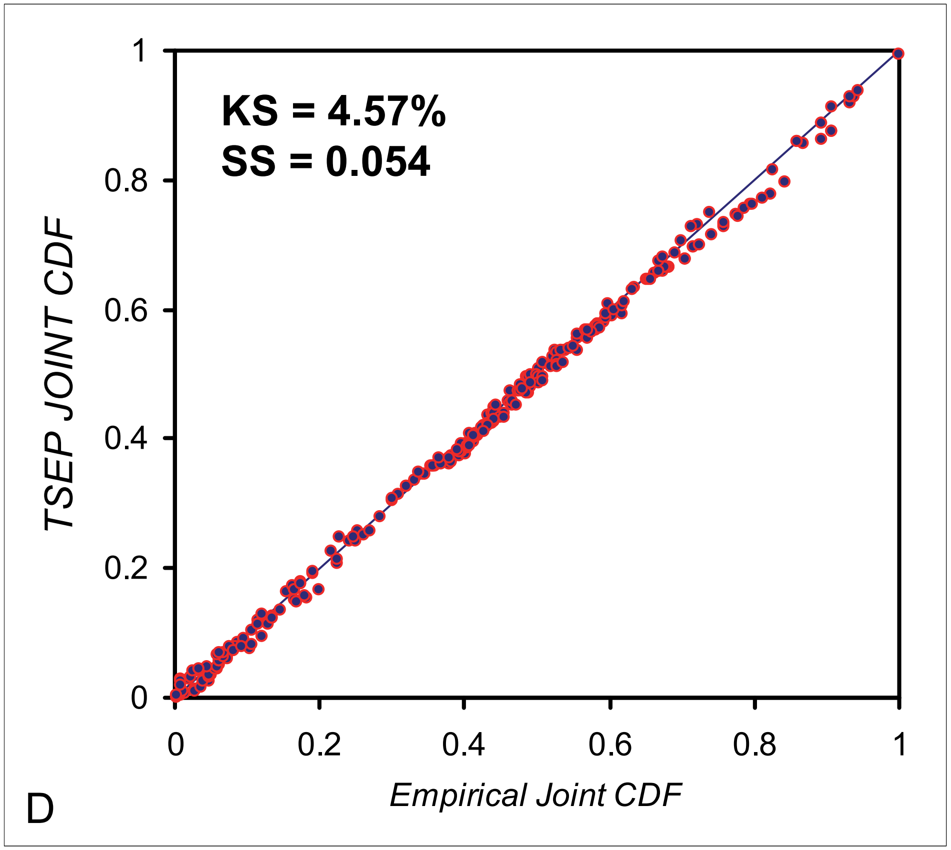


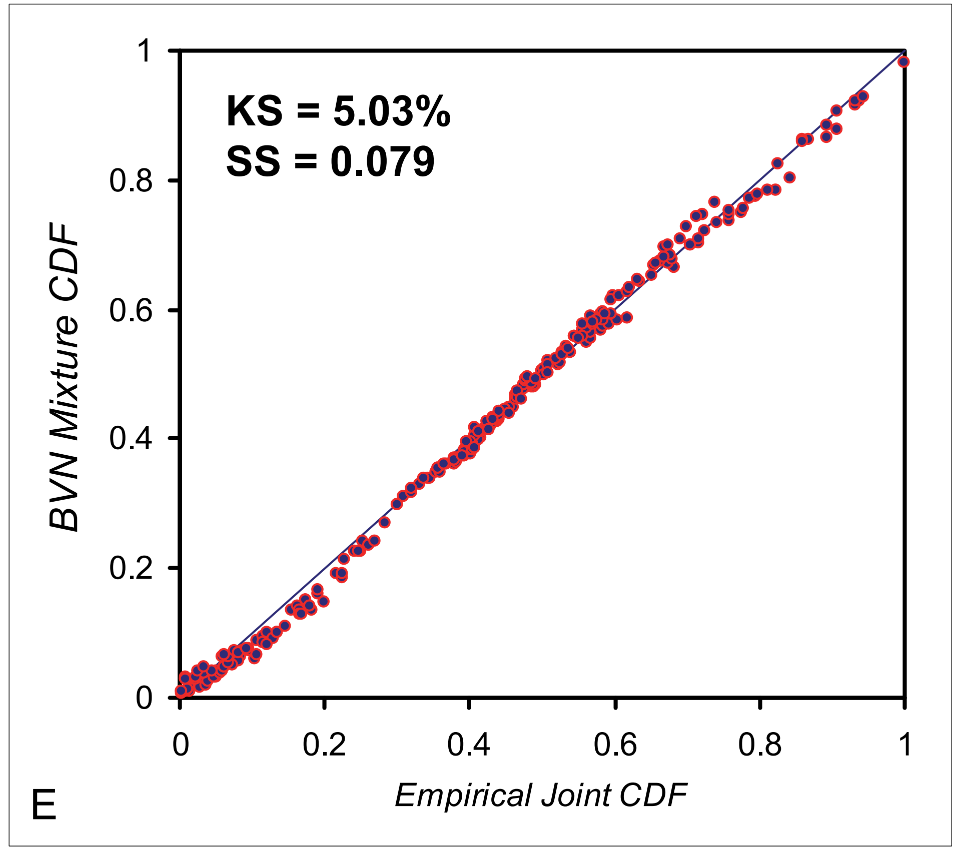
Bivariate Normal Mixture Model



Empirical CDF

Probability-Probabity Plots of Joint CDF Functions





Two-Sided Elevated Power Model

Bivariate Normal Mixture Model

- **Not much can be concluded by visually** comparing **the empirical cdf graphs**.
- **From the joint P-P plots** we visually observe that **the TS-EP joint cdf outperforms** the joint Gaussian mixture cdf **in the lower quantile ranges** translating in a better Kolmogorov-Smirnov (KS) and Sum-of-Squares (SS) criterion.
- **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**.
- From **a marginal distribution 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.**

OUTLINE

1. Introduction
2. A two-sided framework of univariate distributions
3. PDF and CDF of TS-EP distributions
4. Maximum Likelihood Estimation
5. A comparison of the Joint TS-EP fit to a bivariate mixture model fit
6. **Concluding Remarks**
7. Some References

6. CONCLUDING REMARKS…

- **A novel procedure for modeling the classical Old Faithful data set** was presented.

- In particular **two aspects** deserve attention:

  (a) The **introduction of a two-sided bivariate mixture technique utilizing two copulas** as its components and the resulting mixture again being a copula.

  (b) The second one is the introduction of **a novel univariate distribution for modeling bimodal distributions** with **a closed form cdf.**

- **Both aspects** are **integrated in the bivariate 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**.

7. 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.