ON THE GENERALIZED EXPONENTIATED EXPONENTIAL LINDLEY DISTRIBUTION

Hok Shing Kwong[*] , Saralees Nadarajah[2]

University of Manchester

Abstract

The generalized exponentiated exponential Lindley distribution is a novel three parameter distribution due to Hussain et al. (2017). They studied its properties including estimation issues and illustrated applications to four datasets. Here, we show that several known distributions including those having two parameters can provide better fits. We also correct errors in the derivatives of the likelihood function.

Keywords: Consumption relations, Brand cluster, City siting

[*] Corresponding author:
Email: hokshing.kwong@postgrad.manchester.ac.uk

1   Introduction

Hussain et al. (2017) introduced a three parameter distribution given by the probability density function

$$f(x) = \frac{\theta^2}{\theta + \beta}\{\alpha[1 - \exp(-\theta x)]^{\alpha-1} + \beta x\}\exp(-\theta x) \qquad (1)$$

for > 0, $\theta > 0$, $\alpha > 0$ and $\beta \geq -10^{-3}$. It was referred to as the generalized exponentiated exponential Lindley distribution.Other recent generalizations of the Lindley distribution include the beta Lindley distribution (MirMostafaee et al., 2015), the log generalized Lindley-Weibull distribution (Oluyede et al., 2015) and an extended inverse Lindley distribution (Sharma and Khandelwal, 2017).

Hussain et al. (2017) studied various properties of the GEEL distribution, including shape of density function, shape of hazard rate function, moments, conditional moments, L moments, mean deviations, information generating function, Shannon entropy (Shannon, 1948), order statistics, and characterizations in terms of hazard rate function. Hussain et al. (2017) also wrote down the log-likelihood function, derived its derivatives, discussed inference issues and provided four real data applications.

We would like to point out that the derivatives of the log-likelihood function do not appear correct.This incorrect log-likelihood function however does not appear to have affected the es- timates presented in Hussain et al.   (2017).It appears that Hussain et al. (2017) have coded   the log-likelihood function correctly (in spite of the given form for log-likelihood function being incorrect).In Section 2 of this note, we state the correct log-likelihood function and its derivatives.

The three parameter GEEL distribution may provide excellent fits for many datasets. For datasets considered by Hussain et al. (2017), however, several known distributions including those having two parameters can provide better fits. This is illustrated in Section 3. A discussion is given in Section 4.

2 Maximum likelihood estimation

Suppose $x_1, x_2, \ldots, x_n$ is a random sample from (1). The log-likelihood function is

$$\log L(\theta, \alpha, \beta) = 2n \log \theta - n \log(\theta + \beta) + \sum_{i=1}^{n} \log\{\alpha[1 - \exp(-\theta x_i)]^{\alpha-1} + \beta x_i\} - \theta \sum_{i=1}^{n} x_i \qquad (2)$$

The first order partial derivatives of (2) with respect to θ, α and β given in Hussain et al. (2017) do not appear correct. According to our calculations,

$$\frac{\partial \log L(\theta,\alpha,\beta)}{\partial \theta} = \frac{2n}{\theta} - \frac{n}{\theta+\beta} + \alpha(\alpha-1) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-2} x_i \exp(-\theta x_i)}{\alpha[1-\exp(-\theta x_i)^{\alpha-1}+\beta x_i]} - \sum_{i=1}^{n} x_i$$

$$\frac{\partial \log L(\theta, \alpha, \beta)}{\partial \theta} = \sum_{i=1}^{n} \frac{[1 - \exp(-\theta x_i)]^{\alpha-1}\{1 + \alpha \log[1 - \exp(-\theta x_i)]\}}{\alpha[1 - \exp(-\theta x_i)^{\alpha-1} + \beta x_i]}$$

and

$$\frac{\partial \log L(\theta, \alpha, \beta)}{\partial \beta} = \frac{n}{\theta + \beta} + \sum_{i=1}^{n} \frac{x_i}{\alpha[1 - \exp(-\theta x_i)]^{\alpha-1} + \beta x_i}$$

The second order partial derivatives of (2) given in Hussain et al. (2017) do not appear correct either. Furthermore, only three second order partial derivatives are given. According to our calculations, the six second order partial derivatives are

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial \theta^2} = -\frac{2n}{\theta^2} + \frac{n}{(\theta+\beta)^2} + \alpha(\alpha-1)(\alpha-2) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-3} x_i^2 \exp(-2\theta x_i)}{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i} -$$

$$\alpha(\alpha-1) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-2} x_i^2 \exp(-\theta x_i)}{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i} - \alpha^2(\alpha-1)^2 \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{2\alpha-4} x_i^2 \exp(-2\theta x_i)}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2}$$

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial \theta^2} = -\frac{2n}{\theta^2} + \frac{n}{(\theta+\beta)^2} + \alpha(\alpha-1)(\alpha-2) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-3} x_i^2 \exp(-2\theta x_i)}{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i} -$$

$$\alpha(\alpha-1) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-2} x_i^2 \exp(-\theta x_i)}{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i} - \alpha^2(\alpha-1)^2 \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{2\alpha-4} x_i^2 \exp(-2\theta x_i)}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2}$$

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial \theta \partial \alpha} = \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{\alpha-2} x_i \exp(-\theta x_i)\{2\alpha-1+\alpha(\alpha-1)\log[1-\exp(-\theta x_i)]\}}{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i} - \alpha(\alpha-$$

$$1) \sum_{i=1}^{n} \frac{[1-\exp(-\theta x_i)]^{2\alpha-3} x_i \exp(-\theta x_i)\{1+\alpha \log[1-\exp(-\theta x_i)]\}}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2} ,$$

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial\theta\partial\beta} = \frac{n}{(\theta+\beta)^2} - \alpha(\alpha-1)\sum_{i=1}^{n}\frac{[1-\exp(-\theta x_i)]^{\alpha-2}\log[\exp(-\theta x_i)]\{2+\alpha\log[\exp(-\theta x_i)]\}}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2} \, ,$$

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial\alpha^2} = \sum_{i=1}^{n}\frac{[1-\exp(-\theta x_i)]^{\alpha-1}x_i^2\exp(-\theta x_i)}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2} -$$

$$\sum_{i=1}^{n}\frac{[1-\exp(-\theta x_i)]^{2\alpha-2}\{1+\alpha\log[1-\exp(-\theta x_i)]\}^2}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2} \, ,$$

$$\frac{\partial^2 \log L(\theta,\alpha,\beta)}{\partial\alpha\partial\beta} = -\sum_{i=1}^{n}\frac{[1-\exp(-\theta x_i)]^{\alpha-1}\{1+\alpha\log[1-\exp(-\theta x_i)]\}}{\{\alpha[1-\exp(-\theta x_i)]^{\alpha-1}+\beta x_i\}^2}$$

and

$$\frac{\partial^2 \log L(\theta, \alpha, \beta)}{\partial\beta^2} = \frac{n}{(\theta + \beta)^2} - \sum_{i=1}^{n}\frac{x_i^2}{\{\alpha[1 - \exp(-\theta x_i)]^{\alpha-1} + \beta x_i\}^2}$$

These second order derivatives can be used for confidence intervals and tests of hypotheses.

3   Data analysis

Here, we compare the fit of the three parameter GEEL distribution (Hussain et al., 2017) to known distributions including those having two parameters. We show evidence that known distributions provide better fits. This is not a criticism of Hussain et al.'s novel contribution. The GEEL distribution of Hussain et al. (2017) may give adequate fits to many datasets.

We consider fitting the following seven distributions to the four datasets in Hussain et al.(2017):

1.   the gamma distribution (Jambunathan, 1954) with pdf

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)$$

for > 0 , $\alpha > 0$, $\beta > 0$,where

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$$

denotes the gamma function;

2. the Weibull distribution (Weibull, 1951) with pdf

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp[-(x/\beta)^{\alpha}]$$

for > 0 , $\alpha > 0$, $\beta > 0$;

3. the exponentiated exponential distribution (Gupta and Kundu, 1999) with pdf

$$f(x) = \alpha\beta[1 - \exp(-\beta x)]^{\alpha-1} \exp(-\beta x)$$

for > 0 , $\alpha > 0$, $\beta > 0$;

4. the Gompertz distribution (Gompertz, 1825) with pdf

$$f(x) = \alpha\beta\exp(\alpha + \beta x) \exp[-\alpha \exp(\beta x)]$$

for > 0 , $\alpha > 0$, $\beta > 0$;

5. the generalized gamma distribution (Stacy, 1962) with pdf

$$f(x) = \frac{\theta/\alpha^{\beta}}{\Gamma(\beta/\theta)} x^{\beta-1}\exp\left[-(x/\alpha)^{\theta}\right]$$

for > 0 , $\alpha > 0$, $\beta > 0$ and $\theta > 0$;

6. the generalized Gompertz distribution (El-Gohary et al., 2013) with pdf

$$f(x) = \beta\alpha \exp(\theta x)\exp\left\{-\frac{\alpha}{\theta}[\exp(\theta x) - 1]\right\}\left[1 - \exp\left\{-\frac{\alpha}{\theta}[\exp(\theta x) - 1]\right\}\right]^{\beta-1}$$

for > 0 , $\alpha > 0$, $\beta > 0$ and $\theta > 0$;

7. the gamma Gompertz distribution (Manton et al., 1986) with pdf

$$f(x) = \frac{\alpha\beta\theta}{[\theta - 1 + \exp(\alpha x)]^{\beta+1}}$$

for > 0 , $\alpha > 0$, $\beta > 0$ and $\theta > 0$.

Each distribution was fitted by the method of maximum likelihood. Tables 1, 4, 7 and 10 show the log-likelihood value, value of Akaike Information Criterion (AIC) due to Akaike (1974) and value of Bayesian Information Criterion (BIC) due to Schwarz (1978) for the fitted distributions. n denotes the sample size of the datasets and k the number of estimated parameters. Tables 3, 6, 9 and 12 give the parameter estimates and standard errors. Tables 2, 5, 8 and 11 perform goodness of fit tests based on Kolmogorov Smirnov, Anderson Darling and Cramer von Mises tests. Figures 1, 2 and 3 show the density plot, quantile-quantile plot and probability-probability plot of the GEEL distribution as well as three distributions (generalized Gompertz, exponentiated exponential and gamma Gompertz distributions) that fitted best for all four datasets.
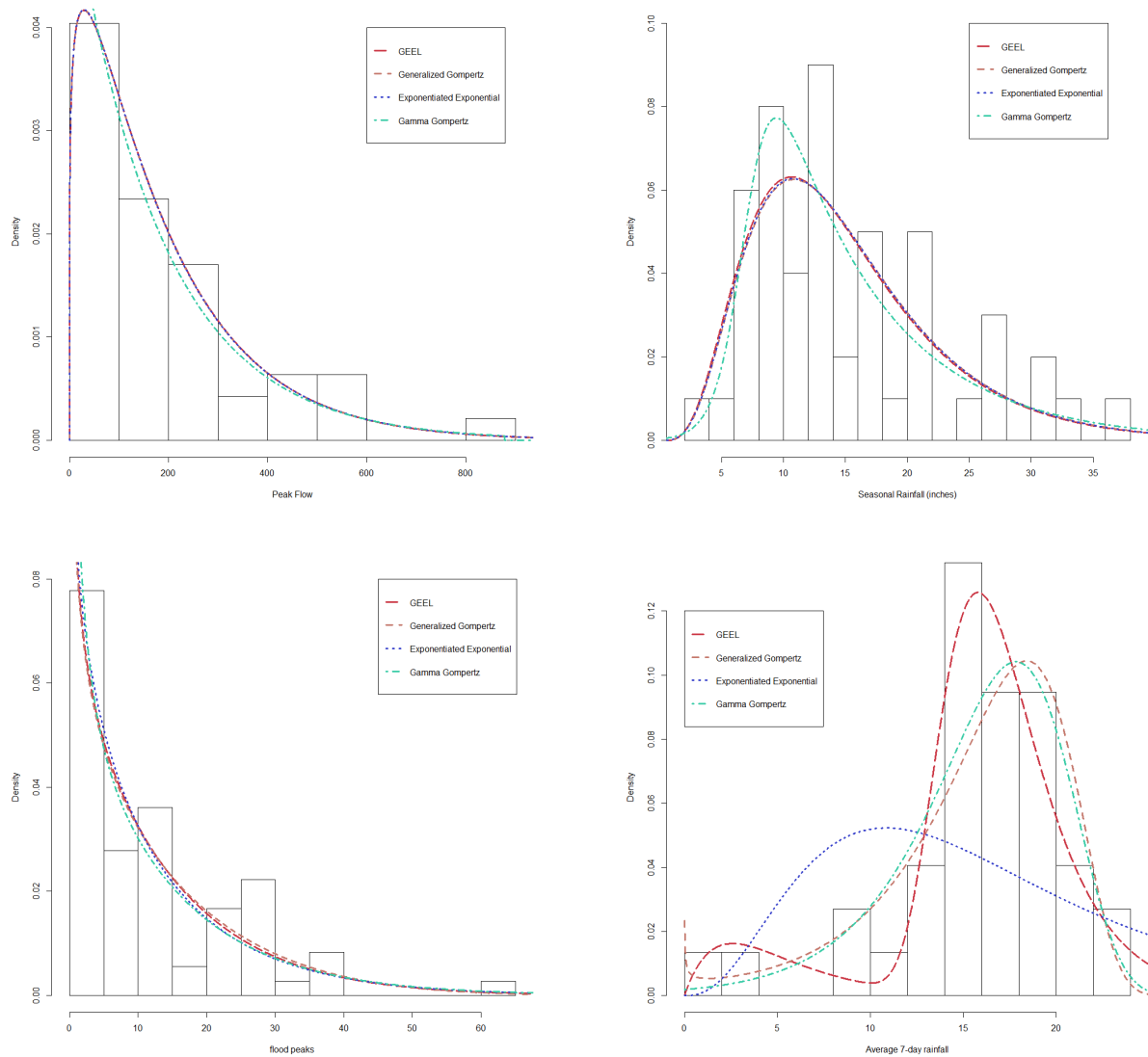

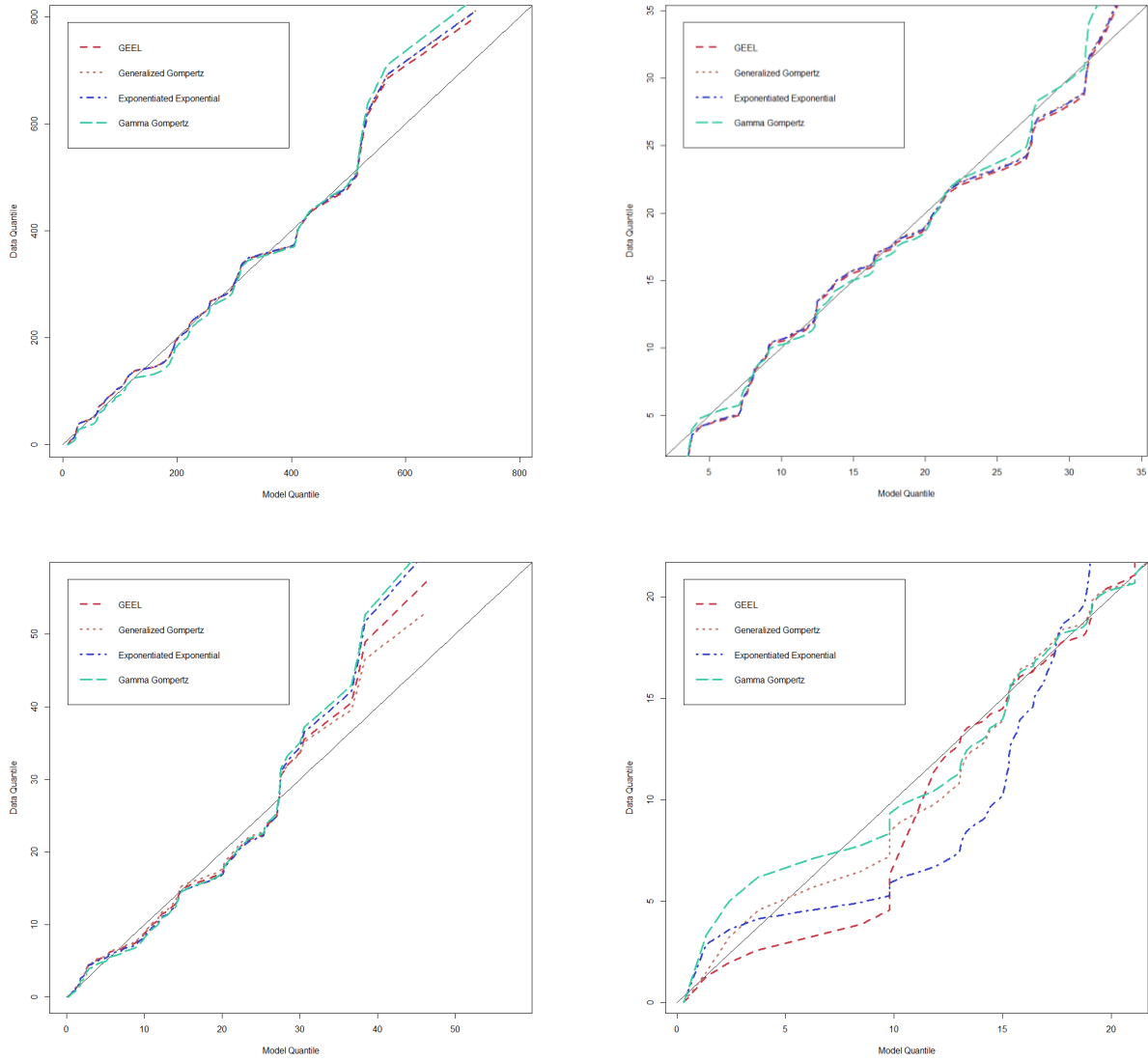Figure 1: Density plots; Left to right, top to bottom, datasets 1 to 4

Figure 2: Quantile - Quantile plots; Left to right, top to bottom, datasets 1 to 4.
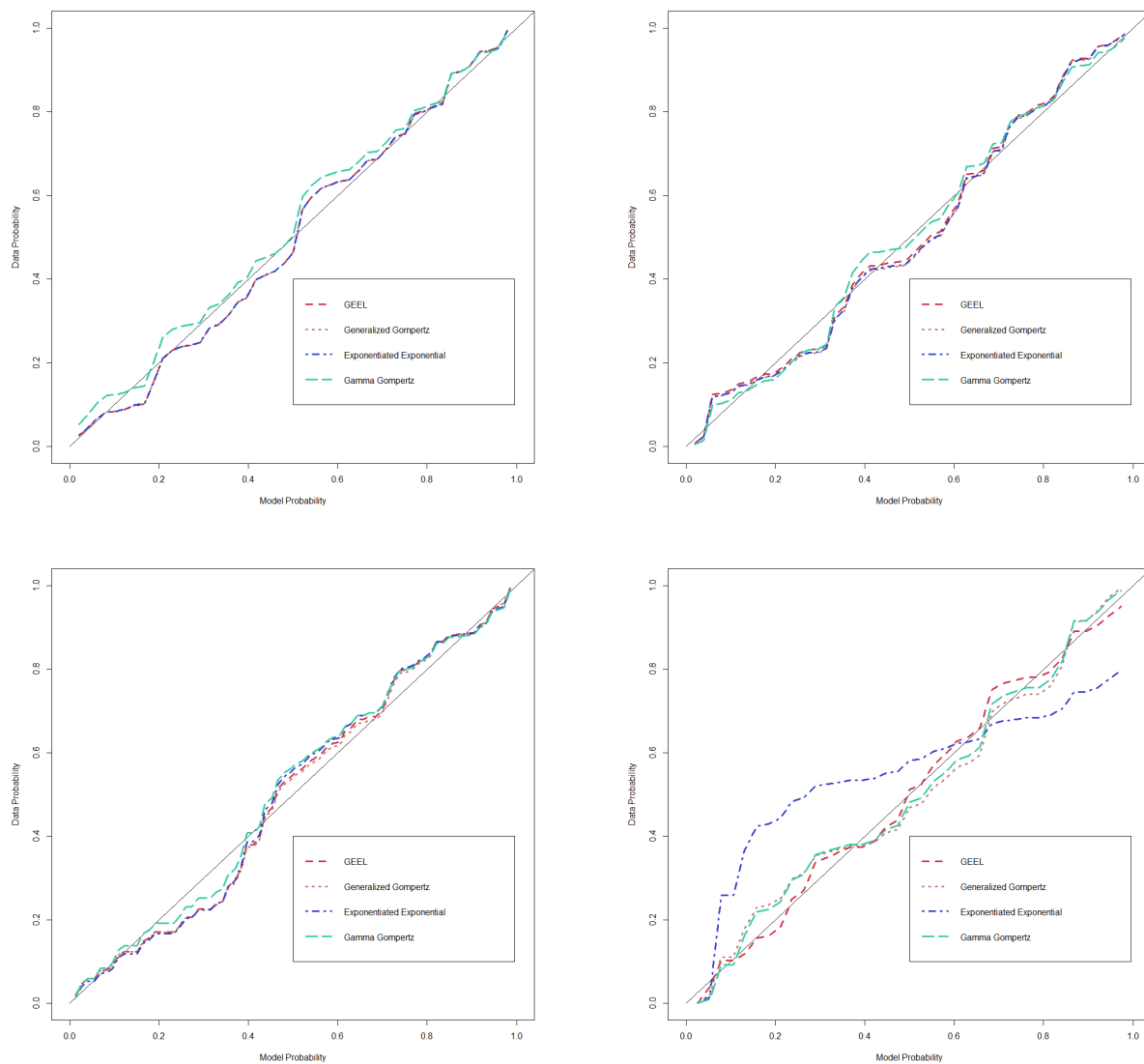
Figure 3: Probability - Probability plots; Left to right, top to bottom, datasets 1 to 4

Table 1 Information criteria for dataset 1.

Distribution	n	k	Log-likelihood	AIC	BIC
GEEL	47	3	-293.05	592.11	591.13
Gamma	47	2	-293.07	590.13	589.48
Weibull	47	2	-293.15	590.29	589.64
Exponentiated exponential	47	2	-293.06	590.11	589.46
Gompertz	47	2	-293.29	590.59	589.93
Generalized gamma	47	3	-292.97	591.93	590.95
Generalized Gompertz	47	3	-293.06	592.11	591.13
Gamma Gompertz	47	3	-291.99	589.98	588.99

Table 2 Goodness of fit tests for dataset 1.

Dataset 1	k	KS	*p-value*	AD	*p-value*	CVM	*p-value*
GEEL	3	0.069	0.969	0.246	0.972	0.036	0.953
Gamma	2	0.076	0.927	0.345	0.900	0.055	0.850
Weibull	2	0.062	0.988	0.230	0.980	0.033	0.967
Exponentiated exponential	2	0.068	0.970	0.243	0.974	0.036	0.955
Gompertz	2	0.067	0.973	0.302	0.936	0.044	0.915
Generalized gamma	3	0.079	0.909	0.252	0.969	0.038	0.943
Generalized Gompertz	3	0.068	0.970	0.243	0.974	0.036	0.954
Gamma Gompertz	3	0.093	0.773	0.402	0.846	0.057	0.836

Table 3:Parameter estimates and standard errors in parentheses for dataset 1.

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$
Gamma	1.1940	0.0060	-
	(0.22)	(0.00137)	-
Weibull	1.0900	195.700	-
	(0.0395)	(1.47)	-
Exponentiated exponential	1.1860	0.0059	-
	(0.234)	(0.00108)	-
Gompertz	11.4380	0.0004	-
	(4.66)	(0.000152)	-
Generalized gamma	72.5150	1.4220	0.7490
	(165)	(0.709)	(0.491)
Generalized Gompertz	0.0059	1.1863	0.0000
	(0.00205)	(0.307)	(0.00109)
Gamma Gompertz	0.80808	0.006	0.2686
	(0.000413)	(0.000995)	(0.477)

Table 4:Information criteria for dataset 2.

Distribution	n	k	Log-likelihood	AIC	BIC
GEEL	50	3	-168.03	342.06	341.16
Gamma	50	2	-168.37	340.74	340.14
Weibull	50	2	-170.16	344.33	343.72
Exponentiated exponential	50	2	-167.99	339.99	339.38
Gompertz	50	2	-175.13	354.25	353.65
Generalized gamma	50	3	-167.98	341.96	341.06
Generalized Gompertz	50	3	-167.99	341.99	341.08
Gamma Gompertz	50	3	-167.38	340.76	339.85

Table 5:Goodness of fit tests for dataset 2.

Dataset 2	k	KS	p-value	AD	p-value	CVM	p-value
GEEL	3	0.085	0.836	0.411	0.837	0.056	0.841
Gamma	2	0.094	0.728	0.527	0.719	0.084	0.669
Weibull	2	0.118	0.458	0.825	0.463	0.135	0.438
Exponentiated exponential	2	0.085	0.830	0.415	0.833	0.061	0.810
Gompertz	2	0.182	0.063	1.644	0.146	0.253	0.184
Generalized gamma	3	0.089	0.787	0.426	0.822	0.068	0.765
Generalized Gompertz	3	0.087	0.812	0.419	0.829	0.063	0.798
Gamma Gompertz	3	0.075	0.918	0.293	0.943	0.041	0.930

Table 6:Parameter estimates and standard errors in parentheses for dataset 2.

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$
Gamma	3.9250	0.2580	-
	(0.754)	(0.0529)	-
Weibull	2.0490	17.2690	-
	(0.0395)	(1.47)	-
Exponentiated exponential	5.2650	0.1540	-
	(1.39)	(0.0209)	-
Gompertz	0.3592	0.074	-
	(0.0509)	(0.00104)	-
Generalized gamma	0.0720	7.3880	0.5060
	(NaN)	(0.897)	(NaN)
Generalized Gompertz	0.1535	5.2648	0.0000
	(0.0593)	(2.54)	(0.0217)
Gamma Gompertz	0.8297	0.1433	354.1567
	(0.324)	(0.0661)	(699)

Table 7 Information criteria for dataset 3.

Distribution	n	k	Log-likelihood	AIC	BIC
GEEL	72	3	-251.01	508.02	507.59
Gamma	72	2	-251.34	506.69	506.40
Weibull	72	2	-251.50	507.00	506.71
Exponentiated exponential	72	2	-251.29	506.59	506.30
Gompertz	72	2	-252.13	508.26	507.97
Generalized gamma	72	3	-251.07	508.13	507.70
Generalized Gompertz	72	3	-250.85	507.70	507.27
Gamma Gompertz	72	3	-250.23	506.46	506.03

Table 8 Goodness of fit tests for dataset 3.

Dataset 2	k	KS	p-value	AD	p-value	CVM	p-value
GEEL	3	0.103	0.433	0.652	0.600	0.109	0.542
Gamma	2	0.102	0.446	0.770	0.503	0.135	0.438
Weibull	2	0.105	0.404	0.844	0.450	0.149	0.394
Exponentiated exponential	2	0.102	0.447	0.744	0.523	0.130	0.458
Gompertz	2	0.141	0.115	1.473	0.183	0.234	0.210
Generalized gamma	3	0.108	0.371	0.647	0.604	0.106	0.559
Generalized Gompertz	3	0.105	0.406	0.623	0.626	0.103	0.573
Gamma Gompertz	3	0.079	0.765	0.550	0.696	0.094	0.615

Table 9: Parameter estimates and standard errors in parentheses for dataset 3.

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$
Gamma	0.8380 (0.121)	0.2580 (0.0529)	- -
Weibull	0.9010 (0.0855)	17.2690 (1.47)	- -
Exponentiated exponential	0.8280 (0.123)	0.1540 (0.0209)	- -
Gompertz	117.7995 (NaN)	0.0007 (NaN)	- -
Generalized gamma	24.3830 (11.4)	0.7110 (0.156)	1.4920 (0.727)
Generalized Gompertz	0.0575 (0.0177)	0.7503 (0.137)	0.0125 (0.013)
Gamma Gompertz	0.3620 (0.241)	0.1966 (0.121)	0.4337 (0.194)

Table 10: Information criteria for dataset 4

Distribution	n	k	Log-likelihood	AIC	BIC
GEEL	37	3	-105.48	216.95	215.66
Gamma	37	2	-125.06	254.11	253.25
Weibull	37	2	-114.20	232.40	231.54
Exponentiated exponential	37	2	-127.80	259.60	258.74
Gompertz	37	2	-104.38	212.76	211.89
Generalized gamma	37	3	-108.72	223.44	222.14
Generalized Gompertz	37	3	-103.78	213.56	212.27
Gamma Gompertz	37	3	-104.39	214.77	213.48

Table 11: Goodness of fit tests for dataset 4.

Dataset 2	k	KS	p-value	AD	p-value	CVM	p-value
GEEL	3	0.075	0.986	0.263	0.963	0.027	0.987
Gamma	2	0.272	0.008	4.141	0.008	0.748	0.009
Weibull	2	0.182	0.171	1.658	0.143	0.246	0.193
Exponentiated exponential	2	0.289	0.004	4.783	0.004	0.887	0.004
Gompertz	2	0.086	0.948	0.411	0.837	0.052	0.864
Generalized gamma	3	0.176	0.201	1.293	0.234	0.219	0.233
Generalized Gompertz	3	0.103	0.829	0.461	0.786	0.069	0.758
Gamma Gompertz	3	0.086	0.948	0.408	0.839	0.052	0.868

Table 12: Parameter estimates and standard errors in parentheses for dataset 4.

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$
Gamma	4.2410	0.2670	-
	(0.95)	(0.0635)	-
Weibull	3.6640	17.2250	-
	(0.518)	(0.792)	-
Exponentiated exponential	3.8030	0.1230	-
	(0.941)	(0.0175)	-
Gompertz	0.0066	0.2810	-
	(0.00507)	(0.0385)	-
Generalized gamma	22.1270	2.3290	14.8010
	(0.972)	(0.453)	(8.7)
Generalized Gompertz	0.0003	0.6167	0.3737
	(NaN)	(0.0432)	(NaN)
Gamma Gompertz	0.2826	102.1808	15824.0980
	(0.0428)	(1000)	(152000)

For dataset 1, we see that all of the fitted distributions (including the two parameter gamma, Weibull and exponentiated exponential distributions) fit better than the GEEL distribution in terms of AIC and BIC. The best fitted distribution in terms of AIC and BIC is the gamma Gompertz distribution, AIC and BIC are 589.98 and 588.99, respectively. The best fitted distribution in terms of goodness of fit tests is the Weibull distribution.

For dataset 2, gamma, exponentiated exponential, generalized Gompertz, generalized gamma and gamma Gompertz distributions fit better than the GEEL distribution in terms of AIC and BIC. The best fitted distribution in terms of AIC and BIC is the exponentiated exponential distribution, AIC and BIC are 339.99 and 339.38, respectively. The second best fitted distribution in terms of AIC and BIC is the gamma Gompertz distribution. The gamma Gompertz distribution appearsto give a slightly better fit than the GEEL, exponentiated exponential and generalized Gompertz distributions in terms of density, Q-Q and P-P plots. The best fitted distribution in terms of goodness of fit tests is also the gamma Gompertz distribution.

For dataset 3, all fitted distributions except for the Gompertz and generalized gamma distributions fit better than the GEEL distribution in terms of AIC and BIC. The best fitted distribution in terms of AIC and BIC is the gamma Gompertz distribution. The best fitted distribution in terms of goodness of fit tests is also the gamma Gompertz distribution.

For dataset 4, Gompertz, generalized Gompertz and gamma Gompertz distributions fit better than the GEEL distribution. Most of the other fitted distributions have much higher AIC and BIC than the GEEL distribution. They should be considered unsuitable for this dataset. We can see from the density plot that, even though the GEEL, generalized Gompertz and gamma Gompertz distributions give similar fits in terms of AIC and BIC, their shapes are quite different. The GEEL pdf has a bimodal shape, the gamma Gompertz pdf has a unimodal shape with negative skewness and the generalized Gompertz pdf has negative skewness and approaches infinity near zero. We also see that both the generalized Gompertz and gamma Gompertz distributions fit considerably better than the GEEL distribution, see the P-P and Q-Q plots. The GEEL distribution provides the best p-values of the goodness of fit tests, but this is due to the bimodal shape. We shall show later that the bimodal shape is unrealistic.

Both the gamma Gompertz and generalized Gompertz distributions perform better than the GEEL distribution in all four datasets and the gamma Gompertz distribution performs slightly better than the generalized Gompertz distribution in terms of density, Q-Q and P-P plots. For the first three datasets, apart from the differences in AIC and BIC, the best fitted

distributions give similar shapes and they should all be considered as suitable for these datasets. For the fourth dataset, the shapes of the pdfs of the top three best fitted distributions are quite different. We are skeptical of the bimodality of the fitted GEEL distribution, as the dataset describe the annual lowest seven-day average flows at a gauging station. We also consider the behavior of the generalized Gompertz pdf near zero unsuitable. The most natural looking pdf exhibiting a unimodal shape with negative skewness is given by the gamma Gompertz distribution. The gamma Gompertz distribution also provides the second best p-values of the goodness of fit tests. The Gompertz distribution (the limiting case of the gamma Gompertz distribution for $\theta \to \infty$) could also be considered most natural as it has smaller AIC and BIC values. But its p-values are slightly smaller.

Hussain et al. (2017) appear to have proposed a "one-for-all" distribution to describe many kinds (maximum, average, minimum type) of hydrological data. We question this approach. We have shown, with empirical evidence, that some existing distributions (for example, the gamma Gompertz distribution and the generalized Gompertz distribution) are more suitable for the datasets. The generalized Gompertz distribution is a flexible distribution. It contains the Gompertz distribution as a particular case for $\beta = 1$. It accommodates both maximum stability and minimum stability, properties desirable to model datasets consisting of maximum order statistics and minimum order statistics. We believe that it can be a more suitable distribution than the GEEL distribution for many kinds of hydrological datasets.

4   Discussion

There is no doubt that the GEEL distribution is a very flexible distribution, which can attain desirable shapes in many situations. However, the GEEL distribution is a mixture of exponentiated exponential and gamma distributions. Similar fits can be achieved by many other mixture distributions. The GEEL distribution is highly prone to overfit datasets when the sample size is small. Taking dataset 4 as an example, we can see from Figure 1 that the bimodality of the data is very insignificant since the sample size is only 37. Therefore, the GEEL distribution should be considered as overfitting this dataset.

There are various tests in the literature to test for unimodality. We performed the following tests for unimodality for each of the four datasets: Silverman (1981)'s test, Hall and York (2001)'s test, Hartigan and Hartigan (1985)'s test, Cheng and Hall (1998)'s test and Ameijeiras-Alonso et al. (2016)'s test. The p-values are given in Table 13. All of the p-values are well above 0.05. Hence, there is no evidence that any of the four datasets are multimodal. Hence, fitting a distribution to these datasets that accommodates bimodality is

inappropriate.

Table 13: *p*-values of the tests for unimodality.

Test	Dataset 1	Dataset 2	Dataset 3	Dataset 4
SI	0.264	0.936	0.196	0.290
HY	0.184	0.830	0.118	0.164
HH	0.868	0.486	0.920	0.640
CH	0.556	0.100	0.540	0.268
NP	0.504	0.098	0.594	0.308

In statistical modelling of extreme values, it is often the case that there exists a small cluster of outliers. Sometimes likelihood can be improved a lot by adding another distribution to model that cluster of outliers, but will the fitted distribution be always reasonable? We fitted a mixture of two Weibull distributions, a mixture of two gamma distributions and a mixture of two exponentiated exponential distributions to all four datasets. Many of these distributions fitted better than all other fitted distributions for all four datasets in terms of likelihood, AIC and BIC. But their fits did not appear reasonable.

We believe that unimodality or bimodality of a dataset should not be solely determined in terms of likelihood based values. Unimodality or bimodality should be determined based more on the physical mechanism underlying the data. We do not want the GEEL distribution giving a bimodal fit to a dataset which is truly unimodal. A better practice is to choose mixture models on a case by case basis.

Acknowledgments

References

[1]   Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

[2]   Ameijeiras-Alonso, J., Crujeiras, R. and Alberto Rodrguez-Casal, M. (2016). Mode testing, critical bandwidth and excess mass. arXiv preprint: 1609.05188.

[3]   Cheng, M. Y. and Hall, P. (1998). Calibrating the excess mass and dip tests of modality.

[4]   Journal of the Royal Statistical Society, B, 60, 579-589.

[5]   El-Gohary, A., Alshamrani, A. and Al-Otaibi, A. N. (2013). The generalized Gompertz distribution. Applied Mathematical Modelling, 37, 13-24.

[6]   Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London, 115, 513-583.

[7]   Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions. Australian and New Zealand Journal of Statistics, 41, 173-188.

[8]   Hall, P. and York, M. (2001). On the calibration of Silverman's test for multimodality. Statistica Sinica, 11, 515-536.

[9]   Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. Annals of Statistics, 13, 70-84.

[10]   Hussain, T., Bakouch, H. S. and Iqbal, Z. (2017). A new probability model for hydrologic events: Properties and applications. Journal of Agricultural, Biological, and Environmental Statistics, doi: 10.1007/s13253-017-0313-6

[11]   Jambunathan, M. V. (1954). Some properties of beta and gamma distributions. Annals of Mathematical Statistics, 25, 401-405.

[12]   Manton, K. G., Stallard, E. and Vaupel, J. W. (1986). Alternative models for the heterogene- ity of mortality risks among the aged. Journal of the American Statistical Association, 81, 635-644.

[13]   MirMostafaee, S. M. T. K., Mahdizadeh, M. and Nadarajah, S. (2015). The beta Lindley distribution. Journal of Data Science, 13, 603-626.

[14] Oluyede, B. O., Mutiso, F. and Huang, S. (2015). The log generalized Lindley-Weibull distribution with application. Journal of Data Science, 13, 281-310.

[15] Schwarz, G. E. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464. [15]Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-432.

[16] Sharma, V. K. and Khandelwal, P. (2017). On the extension of inverse Lindley distribution. Journal of Data Science, 15, 205-220.

[17] Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. Journal of the Royal Statistical Society, B, 43, 97-99.

[18] Stacy, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187-1192.

[19] Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanics, 18, 293-297.