Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/120605

Exponentially-Modified Logistic Distribution with Application to Mining and Nutrition Data

Jimmy Reyes[1], *Osvaldo Venegas*[2,*] *and Héctor W. Gómez*[1]

[1] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile.
[2] Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Chile.

Received: 4 May 2018, Revised: 1 Sep. 2018, Accepted: 12 Sep. 2018
Published online: 1 Nov. 2018

Abstract: In this work we introduce a modification of the exponentially-modified Gaussian distribution. This new distribution is obtained by combining a logistic distribution with an exponential distribution, and is more flexible than other similar distributions. We provide a closed expression for the density function and obtain some important properties useful for making inferences, such as moment estimators and maximum likelihood estimators. By way of illustration, and using real data to show the effectiveness of the new model, we compare it with known related models, showing that the new model achieves a better fit.

Keywords: Exponentially modified Gaussian distribution, Logistic distribution, moments, maximum likelihood estimates

1 Introduction

Logistic (LOG) distribution has been used in various areas of scientific research, for example bioassay problems see [4], income distribution see [5], survival analysis see [8]. The properties and applications of LOG distribution are available in Balakrishnan [2]. LOG distribution is symmetrical, and it is considered an alternative to normal distribution in various practical uses. Below we discuss some basic properties of LOG distribution which are very useful in the new distribution, allowing better and clearer notation compared to other models.

1.1 Logistic distribution

We say that a random variable X follows a logistic distribution if its density function is of the form

$$f_X(x) = \frac{e^x}{(1+e^x)^2}, \quad x \in \mathbb{R}, \tag{1}$$

which we denote as $X \sim LOG(0,1)$.

1.1.1 Properties

Let $X \sim LOG(0,1)$

1. If f_X is the density function of X then $f_X(-x) = f_X(x)$
2. The distribution function of the random variable X is given by $F_X(x) = \frac{e^x}{1+e^x}$
3. The quantile function of the random variable X is given by $F_X^{-1}(x) = \ln\left(\frac{p}{1-p}\right)$
4. The moment of order r of X is $\mu_r = \mathbb{E}(X^r) = (2^r - 2)\pi^r |B_r|$ if r is even and $\mu_r = 0$ if r is odd. Where B_r is the r-th Bernoulli number given by

$$|B_r| = \begin{cases} 1 & r=0 \\ 1/2 & r=1 \\ 1/6 & r=2 \\ 0 & r=3 \\ 1/30 & r=4 \end{cases}.$$

1.2 Incorporating parameters of location and scale

Let $X \sim LOG(0,1)$ and $Y = \alpha + \beta X$ then we can say that Y has a logistic distribution with location parameter α and scale parameter β, which we denote by $Y \sim LOG(\alpha, \beta)$ if

* Corresponding author e-mail: ovenegas@uct.cl

its density function of Y is

$$f_Y(y;\alpha,\beta)=\frac{e^{\left(\frac{y-\alpha}{\beta}\right)}}{\beta\left(1+e^{\left(\frac{y-\alpha}{\beta}\right)}\right)^2},\quad y,\alpha\in\mathbb{R},\beta>0 \quad (2)$$

1.2.1 Properties

Let $X\sim LOG(0,1)$ and $Y\sim LOG(\alpha,\beta)$ then

1. The distribution function of Y is $F_Y(y)=\frac{e^{\left(\frac{y-\alpha}{\beta}\right)}}{1+e^{\left(\frac{y-\alpha}{\beta}\right)}}$.
2. The quantile function of Y is $F_Y^{-1}(p)=\alpha+\beta\ln\left(\frac{p}{1-p}\right),\quad 0<p<1$
3. The moment of order r of Y for $r=1,2,3,\ldots$. is $\mu_r'=\mathbb{E}(Y^r)=\sum_{j=0}^{r}\binom{r}{j}\beta^j\alpha^{r-j}\mu_j$
where $\mu_j=\mathbb{E}\left(X^j\right)=(2^j-2)\pi^j|B_j|$ if j is even then
$\mu_1'=\mathbb{E}(Y)=\alpha$
$\mu_2'=\mathbb{E}\left(Y^2\right)=\alpha^2+\frac{\pi^2}{3}\beta^2$
$\mu_3'=\mathbb{E}\left(Y^3\right)=\alpha^3+\pi^2\alpha\beta^2$
$\mu_4'=\mathbb{E}\left(Y^4\right)=\alpha^4+2\pi^2\alpha^2\beta^2+\frac{7}{15}\pi^4\beta^4$
4. $Var(Y)=\frac{\pi^2}{3}\beta^2$
5. The coefficients of asymmetry and kurtosis of Y are $\sqrt{\beta_1}=0$ and $\beta_2=\frac{21}{5}$ respectively.

Grushka [6] introduced the exponentially-modified Gaussian (EMG) distribution, defined as a convolution of an exponential (Exp) distribution with parameter λ and a normal distribution with parameters α and β, which are independent of each other. The EMG model has been used in various areas of science such as chromatography and some classes of phenomena in biology. We say that a random variable Y follows an EMG distribution if its density function is of the form

$$f_Y(y;\alpha,\beta,\lambda)=\lambda e^{-\frac{\lambda}{2}(2y-2\alpha-\lambda\beta^2)}\Phi\left(-\lambda\beta-\frac{\alpha-y}{\beta}\right), \quad (3)$$

where $\Phi(\cdot)$ is the distribution function of the standard normal, denoted by $Y\sim EMG(\alpha,\beta,\lambda)$. For a more complete review of the EMG model, see [7].

Using the methodology of Grushka, [6], to construct the EMG model, the principal object of this article is to change the normal model for the LOG model in the representation of the EMG model; thus we obtain a model with greater kurtosis, since the kurtosis of the normal model is 3 while that of the logistic model is $\frac{21}{5}$. This gives us a new, more flexible distribution that we call the exponentially-modified logistic (EMLOG) distribution.

The article is organized as follows: In Section 2 we show the representation, density function, distribution function, hazard function, moments, and coefficients of asymmetry and kurtosis of this new distribution. In Section 3 we make inferences using methods for moment and maximum likelihood estimation and present a simulation study. Section 4 contains two applications with real data. In Section 5 we offer some conclusions.

2 EMLOG distribution

In this section we introduce the representation, density function and basic properties of the EMLOG distribution. It should be noted that this distribution has applications in different areas, by way of example we enunciate some of them such as in biology: to describe how species behave in competitive environments, in psychology: to describe the learning process, in energy: to study the diffusion and substitution of some primary energy sources by others, in technology: to describe how technologies become popular and compete with each other, in marketing: to study the diffusion of new products, among others.

2.1 Stochastic representation

The following expression is the representation of the EMLOG distribution.

$$Y=Z+T, \quad (4)$$

where $Z\sim LOG(\alpha,\beta)$ and $T\sim Exp(\beta)$ are independent random variables. It is denoted by $Y\sim EMLOG(\alpha,\beta)$, making this a very flexible distribution with support in all real numbers. The same scale parameter was considered in both distributions (LOG and Exp) to obtain a parsimonious model. The EMLOG density is the result of a convolution of an Exp distribution with parameter β and a LOG distribution with parameters α and β, which are independent of each other.

2.2 Density function

Proposition 1. *Let $Y\sim EMLOG(\alpha,\beta)$ then the density function of Y is*

$$f_Y(y;\alpha,\beta)=\frac{1}{\beta(e^x+1)}\left[(e^{-x}+1)\log(e^x+1)-1\right], \quad (5)$$

where $-\infty<y<\infty$, $x=\frac{y-\alpha}{\beta}$, $\alpha\in\mathbb{R}$ and $\beta>0$.

Proof. Using the stochastic representation in (4), we have

$$Z\sim LOG(\alpha,\beta)\Rightarrow f_Z(z)=\frac{e^{\left(\frac{z-\alpha}{\beta}\right)}}{\beta\left(1+e^{\left(\frac{z-\alpha}{\beta}\right)}\right)^2},$$

where $-\infty<z<\infty$,

$$T\sim Exp(\beta)\Rightarrow f_T(t)=\frac{1}{\beta}e^{-\frac{1}{\beta}t},\ t>0$$

by convolution, it follows that:

$$f_{Y,W}(y,w) = f_{Z,T}(y-w,w)$$
$$f_{Y,W}(y,w) = f_Z(y-w)f_T(w),\ -\infty<y<\infty,\ w>0.$$
$$= \frac{e^{\left(\frac{y-w-\alpha}{\beta}\right)}}{\beta\left(1+e^{\left(\frac{y-w-\alpha}{\beta}\right)}\right)^2}\frac{1}{\beta}e^{-\frac{1}{\beta}w}$$

Marginalizing with respect to the variable W, we have

$$f_Y(y;\alpha,\beta) = \frac{1}{\beta^2}e^{\frac{y-\alpha}{\beta}}\int_0^\infty e^{-\frac{2w}{\beta}}\left[1+e^{\frac{y-w-\alpha}{\beta}}\right]^{-2}\mathrm{d}w,$$

where $-\infty<y<\infty$. Making the change of variable $u = 1+e^{\frac{y-\alpha}{\beta}}e^{\frac{-w}{\beta}}$ we obtain the result. □

We note that the parameters α and β, have control on the model the location and the scale respectively.

Figure 1, shows the form of the density function for some values of the parameter β.

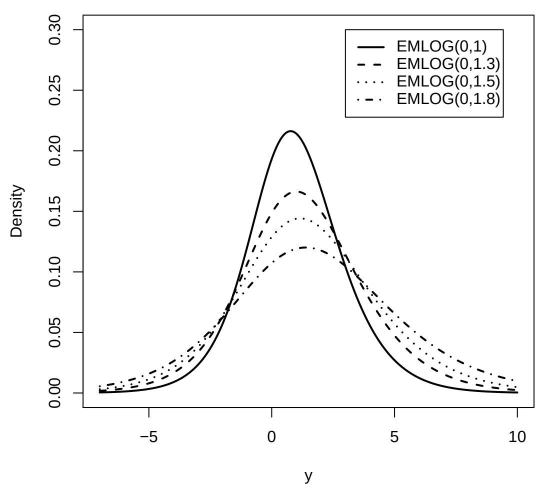


Fig. 1: EMLOG pdf for different values of β

Proposition 2. *Let $Y \sim EMLOG(\alpha,\beta)$, where $\alpha\in\mathbb{R}$ and $\beta>0$. Then, the distribution function of Y is given by*

$$F_Y(t) = 1-e^{-\frac{t-\alpha}{\beta}}\log(e^{\frac{t-\alpha}{\beta}}+1). \tag{6}$$

Proof. Let $x=\frac{y-\alpha}{\beta}$, then we have that

$$\begin{aligned}F_Y(t) &= \int_{-\infty}^t f_Y(y)dy\\ &= \frac{1}{\beta}\int_{-\infty}^t\frac{1}{e^x+1}[(e^{-x}+1)\log(e^x+1)-1]dy\\ &= \frac{1}{\beta}\int_{-\infty}^t\frac{1}{e^x+1}[e^{-x}(e^x+1)\log(e^x+1)-1]dy\\ &= \frac{1}{\beta}\int_{-\infty}^t\left[e^{-x}\log(e^x+1)-\frac{1}{e^x+1}\right]dy.\end{aligned}$$

Making the change of variable $u=e^x+1$ we obtain

$$F_Y(t) = \int_1^{e^t+1}\left[\frac{\log u}{(u-1)^2}-\frac{1}{u(u-1)}\right]du.$$

By calculating this improper integral we obtain the result. □

Corollary 1. *The hazard rate function for the random variable $Y\sim EMLOG(\alpha,\beta)$ is given by*

$$h(t) = \frac{f_Y(t)}{1-F_Y(t)} = \frac{1}{\beta}\left[1-\frac{e^{\frac{t-\alpha}{\beta}}}{(e^{\frac{t-\alpha}{\beta}}+1)\log(e^{\frac{t-\alpha}{\beta}}+1)}\right]. \tag{7}$$

Figure 2, shows the form of the hazard function for some values of the parameter β.

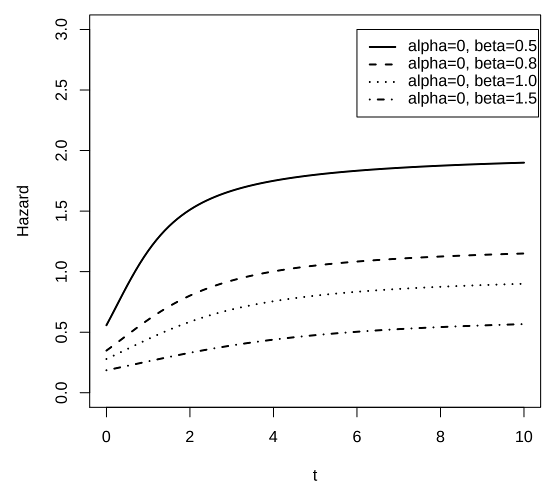


Fig. 2: Hazard rate function of $EMLOG(\alpha=0,\beta)$ model and different values for β.

2.3 Moments

The following proposition shows the moments for the EMLOG distribution. Essentially, these moments depend on the moments of the logistic distribution and the exponential distribution.

Proposition 3. *Let $Y \sim EMLOG(\alpha,\beta)$. Hence, for $r = 1,2,3,\ldots$ we have*

$$\mu_r = \mathbb{E}(Y^r) = \sum_{j=0}^{r} \binom{r}{j} (2^j - 2)\pi^j |B_k| \beta^{r-j} \Gamma(r-j+1).$$

Proof. From (3), since $Z \sim LOG(\alpha,\beta)$ and $T \sim Exp(\beta)$ are independent, we have

$$\begin{aligned}\mu_r &= \mathbb{E}(Y^r) = \mathbb{E}(Z+T)^r = \mathbb{E}\left(\sum_{j=0}^{r}\binom{r}{j} Z^j T^{r-j}\right)\\ &= \sum_{j=0}^{r}\binom{r}{j}\mathbb{E}\left(Z^j\right)\mathbb{E}\left(T^{r-j}\right),\end{aligned}$$

where $\mathbb{E}\left(Z^j\right)$ are the jth moments of the logistic distribution and $\mathbb{E}\left(T^{r-j}\right) = \beta^{r-j}\Gamma(r-j+1)$ are the $(r-j+1)$th moments of the exponential distribution. □

Corollary 2. *Let $Y \sim EMLOG(\alpha,\beta)$ then*

$$\begin{aligned}\mu_1 &= \mathbb{E}(Y) = \alpha+\beta\\ \mu_2 &= \mathbb{E}\left(Y^2\right) = (\alpha+\beta)^2 + \left(1+\frac{\pi^2}{3}\right)\beta^2\\ \mu_3 &= \mathbb{E}\left(Y^3\right) = (\alpha+\beta)^3 + (\pi^2+5)\beta^3 + (3+\pi^2)\alpha\beta^2\\ \mu_4 &= \mathbb{E}\left(Y^4\right) = (\alpha+\beta)^4 + \left(\frac{7\pi^4}{15} + 4\pi^2 + 23\right)\beta^4\\ &\quad + 4(\pi^2+5)\alpha\beta^3 + 2(\pi^2+3)\alpha^2\beta^2\\ Var(Y) &= \left(1+\frac{\pi^2}{3}\right)\beta^2.\end{aligned}$$

Corollary 3. *Let $Y \sim EMLOG(\alpha,\beta)$, then the asymmetry coefficient and the kurtosis coefficient of Y are given by*

$$\gamma_1 = \frac{2}{(1+\frac{\pi^2}{3})^{3/2}} = 0.2251$$

and

$$\gamma_2 = \frac{\frac{7\pi^4}{15} + 2\pi^2 + 9}{(1+\frac{\pi^2}{3})^2} = 4.0318$$

respectively.

Proof. By definition of the asymmetry coefficient we have

$$\gamma_1 = \frac{\mathbb{E}\left[Y-\mathbb{E}(Y)\right]^3}{[\mathrm{var}(Y)]^{3/2}} = \frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{\left[(\mu_2-\mu_1^2)\right]^{3/2}},$$

for the kurtosis coefficient we have

$$\gamma_2 = \frac{\mathbb{E}\left[Y-\mathbb{E}(Y)\right]^4}{[\mathrm{var}(Y)]^2} = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2}. \quad \square$$

3 Inference

In this section, we study the parameter estimation of the new model using the maximum likelihood and moments approach.

3.1 Method of moments estimation

The following proposition shows explicitly the moments estimators α and β.

Proposition 4. *Let $Y_1,\ldots,Y_n$ be a random sample of a distribution of the random variable $Y \sim EMLOG(\alpha,\beta)$, then the moments estimators of $\boldsymbol{\theta} = (\alpha,\beta)$ are*

$$\widehat{\alpha}_M = \overline{Y} - \sqrt{\frac{3\left(\overline{Y^2} - \overline{Y}^2\right)}{3+\pi^2}} \quad \text{and} \quad \widehat{\beta}_M = \sqrt{\frac{3\left(\overline{Y^2} - \overline{Y}^2\right)}{3+\pi^2}},$$

where $\overline{Y}$ is the mean of the sample, and $\overline{Y^2}$ is the mean of the sample of the observations squared.

Proof. Using Corollary 2 we have that

$$\mathbb{E}(Y) = \alpha+\beta \;\; \text{and} \;\; \mathbb{E}(Y^2) = (\alpha+\beta)^2 + \beta^2\left(1+\frac{\pi^2}{3}\right), \tag{8}$$

and then substituting $\mathbb{E}(Y)$ for $\overline{Y}$ and $\mathbb{E}(Y^2)$ for $\overline{Y^2}$ in (8) we obtain a two-by-two system of equations. Solving this system, we obtain the moments estimators $(\widehat{\alpha}_M, \widehat{\beta}_M)$ of (α,β). □

3.2 Maximum likelihood estimators

Given an observed sample $Y_1,\ldots,Y_n$ from the EMLOG(α,β) distribution, the log-likelihood function for the parameters α and β, given $\mathbf{y} = (y_1,\ldots,y_n)^\top$, can be written as

$$\begin{aligned} l(\alpha,\beta) &= -n\log(\beta) - \sum_{i=1}^{n}\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right) \qquad (9)\\ &\quad + \sum_{i=1}^{n}\log\left[\left(e^{-\frac{y_i-\alpha}{\beta}}+1\right)\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right)-1\right].\end{aligned}$$

The maximum likelihood equations are given by

$$\sum_{i=1}^{n}\frac{e^{\frac{y_i-\alpha}{\beta}}}{e^{\frac{y_i-\alpha}{\beta}}+1} + \sum_{i=1}^{n}\frac{e^{-\frac{y_i-\alpha}{\beta}}\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right)-1}{\left(e^{-\frac{y_i-\alpha}{\beta}}+1\right)\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right)-1} = 0, \tag{10}$$

$$\sum_{i=1}^{n}\frac{(y_i-\alpha)e^{\frac{y_i-\alpha}{\beta}}}{e^{\frac{y_i-\alpha}{\beta}}+1} + \sum_{i=1}^{n}\frac{(y_i-\alpha)\left(e^{-\frac{y_i-\alpha}{\beta}}\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right)-1\right)}{\left(e^{-\frac{y_i-\alpha}{\beta}}+1\right)\log\left(e^{\frac{y_i-\alpha}{\beta}}+1\right)-1} = n\beta. \tag{11}$$

The solution for the system of equations given in equations (10) - (11) can be obtained using the Newton-Raphson numerical method. The maximum likelihood estimator (MLE) can also be obtained by maximizing directly the log-likelihood function given in (9), and various existing software programmes can be used, such as the R software [9].

3.2.1 Simulation study

By using the quantile function to generate random numbers of a random logistic variable and the representation given in (4), it is possible to generate random numbers for the EMLOG(α,β) distribution, which leads to the following algorithm

1. Generate $V_i \sim Uniform(0,1)$, $i = 1,2,\ldots,n$.
2. Compute $S_i = \alpha + \beta \ln\left(\frac{V}{1-V}\right)$, $i = 1,2,\ldots,n$.
3. Generate $T_i \sim Exp(\beta)$, $i = 1,2,\ldots,n$.
4. Compute $Y_i = S_i + T_i$, $i = 1,2,\ldots,n$.

It then follows that $Y_i \sim EMLOG(\alpha,\beta)$, $i = 1,2,\ldots,n$.

Table 1: Empirical means and SD for the MLE estimators of α and β.

		$n=50$		$n=100$	
α	β	$\widehat{\alpha}$ (SD)	$\widehat{\beta}$ (SD)	$\widehat{\alpha}$ (SD)	$\widehat{\beta}$ (SD)
1	1	1.0128(0.2931)	0.9821(0.1147)	1.0080(0.1982)	0.9901(0.0827)
	2	1.0199(0.5420)	1.9700(0.2207)	1.0019(0.4256)	1.9918(0.1706)
	3	1.1523(0.7735)	2.9546(0.3272)	1.0583(0.5735)	2.9706(0.2309)
2	1	1.9996(0.2998)	0.9884(0.1152)	1.0583(0.5735)	2.9706(0.2309)
	2	2.0601(0.5956)	1.9639(0.2305)	2.0110(0.2010)	0.9899(0.0793)
	3	2.0634(0.6592)	2.9594(0.3485)	2.0081(0.6056)	2.9832(0.2585)
3	1	3.0059(0.2829)	0.9901(0.1150)	3.0038(0.2024)	0.993904(0.0828)
	2	3.0342(0.5755)	1.9849(0.2299)	2.9885(0.4082)	1.9944(0.1659)
	3	3.0217(0.8668)	2.9714(0.3426)	3.0373(0.6177)	2.9724(0.2475)
		$n=150$		$n=200$	
α	β	$\widehat{\alpha}$ (SD)	$\widehat{\beta}$ (SD)	$\widehat{\alpha}$ (SD)	$\widehat{\beta}$ (SD)
1	1	1.0031(0.1655)	0.9950(0.0672)	1.0090(0.1530)	0.9966(0.0591)
	2	1.0087(0.3430)	1.9901(0.1487)	1.0025(0.2952)	1.9977(0.1269)
	3	1.0200(0.4716)	2.9964(0.1986)	1.0151(0.4311)	2.9797(0.1776)
2	1	2.0090(0.1490)	0.9956(0.0591)	2.0090(0.1490)	0.9956(0.0591)
	2	1.9961(0.3374)	1.9869(0.1355)	1.9993(0.2934)	1.9839(0.1175)
	3	2.0453(0.4869)	2.9761(0.1937)	2.0040(0.4279)	2.9981(0.1806)
3	1	3.0010(0.1710)	0.9953(0.050)	3.0020(0.1471)	0.9963(0.0583)
	2	3.0267(0.3489)	1.9959(0.1273)	3.0144(0.2860)	1.9889(0.1127)
	3	3.0043(0.5204)	2.9935(0.1964)	3.0136(0.4629)	2.9942(0.1768)

Table 1 shows the results of simulations studies illustrating the behavior of the MLEs for $1,000$ generated samples of sizes 50, 100, 150 and 200 from a population EMLOG(α,β) distribution. For each sample generated, MLEs are computed numerically using a Newton-Raphson procedure. Means and standard deviations (SD) are reported. It is observed that the bias becomes smaller as the sample size n increases, as one would expect.

4 Applications

In this Section we present two applications to real data sets. In the first we compare the new model with the EMG model which serves as the inspiration for the construction of the EMLOG model, and in the second we compare it to various models widely used in statistical literature.

4.1 Application 1

We consider data for the concentration of Zirconium in 86 soil samples obtained by the Mines Department of Universidad de Atacama, Chile. Table 2 shows the descriptive statistics, where we use the notation b_1 and b_2 to represent sample asymmetry and kurtosis coefficients respectively. From the results in Section 3.1, the moments estimates for the parameters of the EMLOG model are $\widehat{\alpha}_M$ =147.288 and $\widehat{\beta}_M$=26.573. These estimates are useful as starting values required to implement estimation via maximum likelihood using a numerical method. Table 3 shows the MLEs for the parameters of the EMLOG and EMG models. The standard errors of the maximum likelihood estimates are calculated using the Hessian matrix corresponding to each model.

Table 2: Summary statistics for a data set of Zirconium concentrations.

n	$\overline{Y}$	S_Y	b_1	b_2
86	173.860	55.361	1.2925	8.6525

We compare the EMLOG distribution with the EMG model. We calculate the Akaike information criterion AIC (see Akaike, 1974) and the Bayesian information criterion BIC (see Schwarz, 1978). For these data, the values in the table indicate that the EMLOG distribution leads to a better fit than the EMG distribution.

Table 3: : Summary statistics for Zirconium concentrations in a data set of 86 samples: Model, ML estimates, AIC and BIC values.

Model	MLEs $\widehat{\alpha}(SD)$, $\widehat{\beta}(SD)$, $\widehat{\lambda}(SD)$	AIC	BIC
EMLOG	145.780 , 25.004 , - (5.400) , (2.134) , -	**927.166**	**932.015**
EMG	107.032 , 10.502 , 1.529 (4.921) , (2.083) , (0.090)	936.684	944.047

Figure 3 presents the histogram for the data with the fitted densities and Figure 4 shows the qq-plots of the two fitted densities.

4.2 Application 2

This data set consists of several variables recorded in 202 Australian athletes, reported in [3]. In particular we analyse measurements of body mass index (BMI). Table 4 presents basic descriptive statistics for the data set. From the results in Section 3.1, the moments estimates for the parameters of the EMLOG model are $\widehat{\alpha}_M$ =21.573 and

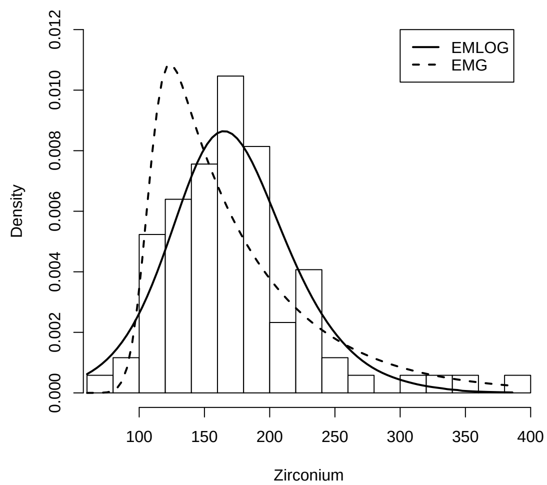


Fig. 3: Models fitted by maximum likelihood method for data set

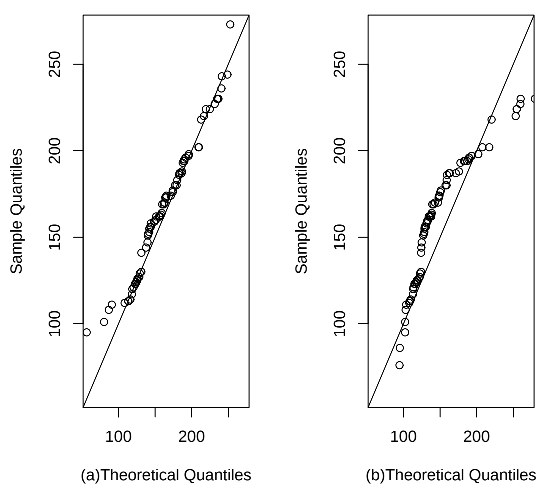


Fig. 4: SThis represents the qq-plots for the Zirconium concentration data set: EMLOG model (a) and EMG model (b)

.

$\widehat{\beta}_M$=1.383. These estimates are useful as the starting values required to implement estimation via maximum likelihood using a numerical method. Table 5 shows the MLEs for the parameters of the EMLOG, LOG, WEIBULL, GAMMA and GUMBEL models. The standard errors of the maximum likelihood estimates are calculated using the Hessian matrix corresponding to each model. Table 5 shows the corresponding AIC and BIC for each model. For these data, the values in the table indicate that the EMLOG distribution leads to a better fit than the LOG, WEIBULL, GAMMA and GUMBEL distributions.

Table 4: Summary statistics for data set of the body mass index of 202 Australian athletes.

n	$\overline{Y}$	S_Y	b_1	b_2
202	22.926	2.866	0.940	5.132

Table 5: Summary statistics for data set of the body mass index of 202 Australian athletes: Model, ML estimates, AIC and BIC values.

Model	MLEs $\widehat{\alpha}(SD)$, $\widehat{\beta}(SD)$	AIC	BIC
EMLOG(α,β)	21.523 (0.073), 1.332 (0.071)	**983.475**	**990.091**
LOG(α,β)	22.787 (0.186), 1.529 (0.090)	986.924	993.540
WEIBULL(α,β)	7.281 (0.340), 24.259 (0.249)	1053.042	1059.658
GAMMA(α,β)	67.729 (0.6.723), 2.950 (0.294)	989.453	996.069
GUMBEL(α,β)	21.636 (0.182), 2.442 (0.127)	987.132	993.748

Figure 5 shows the histogram for the data with the fitted densities and the qq-plots of the EMLOG distribution. Figures 6 and 7 show the qq-plots for the LOG, WEIBULL, GAMMA and Gumbel models, which are calculated with the estimates of the parameters in each model.

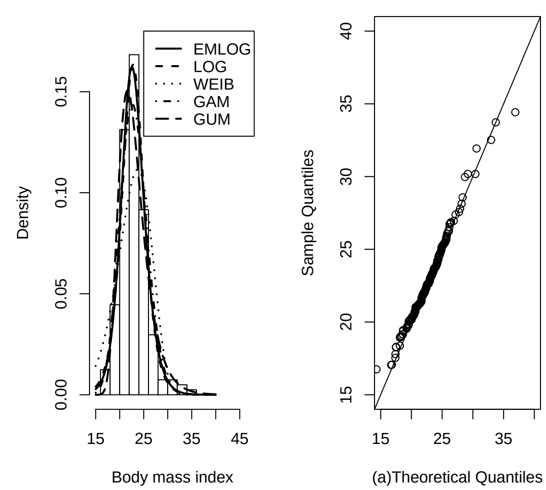


Fig. 5: Histogram for the body mass index data set with fitted densities and qq-plot for the EMLOG model (a)

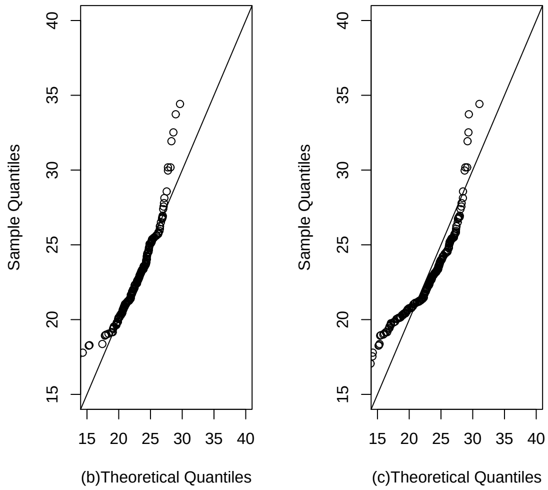


Fig. 6: Shows the qq-plots: LOG model (b), WEIBULL model (c)

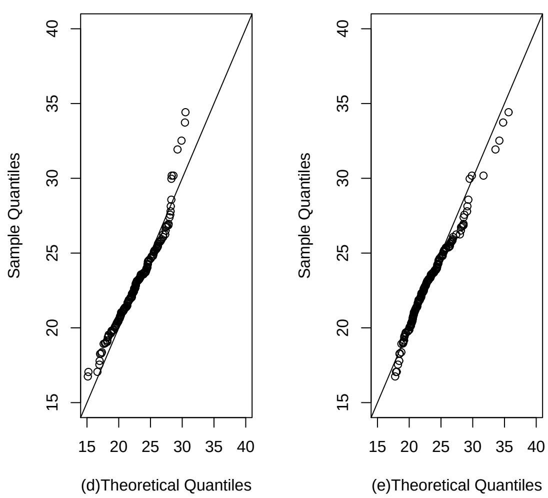


Fig. 7: Shows the qq-plots: GAMMA model (d), GUMBEL model (e)

5 Conclusions

The main focus of this work is to study a modification of the EMG model, both to change normal distribution to logistic distribution and to reduce the number of parameters. This model involves two parameters and is an alternative to other models with the same number of parameters. Maximum likelihood estimation methods are used to estimate the parameters; the results of a simulation study indicate that this model has good properties for small and moderate sample sizes. We compare the models using the AIC and BIC criteria; two applications using real data indicate that this model can produce a better fit than other distributions.

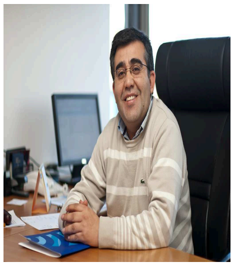
Acknowledgments

The research of J. Reyes and H.W. Gómez was supported by Grant SEMILLERO UA-2018.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control, Vol. 19, No. 6, pp. 716-723 (1974).
[2] N. Balakrishnan, Handbook of Logistic Distribution, New York: Dekker (1992).
[3] R.D. Cook and S. Weisberg, An Introduction to Regression Graphics, New York: Wiley (1994).
[4] D.J. Finney, Statistical Methods in Biological Assay, New York: Hafner Publications (1952).
[5] P.R. Fisk, The graduation of income distributions, Econometrica, Vol. 29, pp. 171-185 (1961).
[6] E. Grushka, Characterization of Exponentially Modified Gaussian Peaks in Chromatography, Analytical Chemistry, Vol. 44, pp. 1733-1738 (1972).
[7] M.S. Jeansonne and J.P. Foley, Review of the Exponentially Modified Gaussian (EMG) Function Since 1983, Journal of Chromatographic Science, Vol. 29, No. 6, pp. 258-266 (1991).
[8] R.L. Plackett, The analysis of life test data, Technometrics, Vol. 1, pp. 9-19 (1959).
[9] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2016).
[10] G. Schwarz, Estimating the dimension of a model, Annals of Statistics, Vol. 6, pp. 461-464 (1978).

Jimmy Reyes received the Magister degree in mathematical statistic at CIENES-Universidad de Chile (1978). He is an Associate Professor at the Universidad de Antofagasta. His research is focused in distribution theory.



Osvaldo Venegas is Associate Professor at Universidad Católica de Temuco, Temuco, Chile. He received his PhD in Mathematics at the Pontificia Universidad Católica of Chile (2005). His research interests are mathematical statistics, reaction-diffusion equations and complex analysis.

Héctor W. Gómez received the PhD degree in Statistics at the Pontificia Universidad Católica of Chile (2004). He is currently a Titular Professor at the Universidad de Antofagasta - Chile. His research is focused in distribution theory.