

## Research Article

# Comparing Estimation Methods for the FPLD

**Agostino Tarsitano**

*Dipartimento di Economia e Statistica, Università della Calabria, Via Pietro Bucci, Cubo 1C,  
87136 Rende (CS), Italy*

Correspondence should be addressed to Agostino Tarsitano, agotar@unical.it

Received 7 July 2009; Revised 22 December 2009; Accepted 31 March 2010

Academic Editor: Tomasz J. Kozubowski

Copyright © 2010 Agostino Tarsitano. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We use the quantile function to define statistical models. In particular, we present a five-parameter version of the generalized lambda distribution (FPLD). Three alternative methods for estimating its parameters are proposed and their properties are investigated and compared by making use of real and simulated datasets. It will be shown that the proposed model realistically approximates a number of families of probability distributions, has feasible methods for its parameter estimation, and offers an easier way to generate random numbers.

## 1. Introduction

Statistical distributions can be used to summarize, in a small number of parameters, the patterns observed in empirical work and to uncover existing features of data, which are not immediately apparent. They may be used to characterize variability or uncertainty in a dataset in a compact way. Furthermore, statistical distributions may both facilitate the mathematical analysis of the basic structure of empirical observations and interrelate information from two or more sources. The aim of this paper is to make a contribution to the use of quantile statistical methods in fitting a probability distribution to data, following the line of thought indicated by Parzen [1] and Gilchrist [2, Section 1.10]. In particular, we adopt a five-parameter version of the generalized lambda distribution

$$X(p, \lambda) = \lambda_1 + \frac{\lambda_2}{2} \left\{ (1 - \lambda_3) \left[ \frac{p^{\lambda_4} - 1}{\lambda_4} \right] - (1 + \lambda_3) \left[ \frac{q^{\lambda_5} - 1}{\lambda_5} \right] \right\}, \quad (1.1)$$

where  $0 \leq p \leq 1$ ,  $q = 1 - p$ .  $X(p, \lambda)$  is the quantile function and  $p \in [0, 1]$ . If  $\lambda_2 \geq 0$  and  $\lambda_3 \in [-1, 1]$ , then (1.1) is a continuous and increasing function of  $p$ . Here,  $\lambda_1$  controls,

albeit not exclusively, the location of an FPLD; the parameter  $\lambda_2$  acts for a multiplier to the translated quantile function  $[X(p, \boldsymbol{\lambda}) - \lambda_1]$  and is, thus, a scale parameter;  $\lambda_3, \lambda_4, \lambda_5$  influence the shape of  $[X(p, \boldsymbol{\lambda})]$ . The limiting cases  $\lambda_4 \rightarrow 0$  and/or  $\lambda_5 \rightarrow 0$  must be interpreted according to L'Hopital's rule. The references [2–7] are illustrative regarding the history of various parametrizations of the distribution (see also [8]). The (1.1) version was proposed by Gilchrist [2, page 163] as an extension of the generalized lambda distribution (GLD) in the parametrization given by Freimer et al. [5]

$$X(p, \boldsymbol{\lambda}) = \lambda_1 + \frac{1}{\lambda_2} \left\{ \left[ \frac{p^{\lambda_4} - 1}{\lambda_4} \right] - \left[ \frac{q^{\lambda_5} - 1}{\lambda_5} \right] \right\}, \quad (1.2)$$

where the scale parameter  $\lambda_2$  appears as a reciprocal and  $\lambda_3 = 0$ .

The FPLD can be useful in problem solving when there are difficulties with specifying the precise form of an error distribution, because it can assume a wide variety of curve shapes and, more importantly, it uses only one general formula over the entire range of data. Its versatility, however, is obtained at the cost of an unusually large number of parameters if compared with the array of studies referred to in the literature on fitting distributions to data. Gilchrist [9] notes the rather strange fact that, throughout the history of statistics, researchers seem to have been quite happy to use many parameters in, for example, regression, yet they have kept to two or three parameters in most distributional models. Clearly, a high number of parameters are a less challenging task in our age of readily available computer power. Moreover, the number of individual data in a statistical sample has considerably increased from what was customary a few decades ago, so there should be no special reason for not using five- (or more) parameter distributional models.

Regrettably, the richness of shape in (1.1) is not accompanied by a comparable development in parameter estimation techniques. For instance, one of the classical methods for computing  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_5)$  can be based on matching the first five moments of the sample observations. In this case, though, the problem of modelling the data would be restricted to distributions possessing fairly light tails because the fifth moment must be finite. In addition, the percentiles estimates (e.g., Karian and Dudewicz [10]) and methods using location and scale-free shape functionals (e.g., King and MacGillivray [11]) depend markedly on the particular choice of percentage points. On the other hand, maximum likelihood estimation is computationally demanding because of the complex and unwieldy objective function (see Tarsitano [12]). Asquith [13] as well as Karvanen and Nuutinen [14] equated sample L-moments to those of the fitted distribution. However, higher L-moments are inelastic to changes in the parameters, leading to difficulties in estimation. These questions have been the focus of intense academic research in the past few years, for example, Ramberg et al. [4], King and MacGillivray [15], Karian and Dudewicz [7], Lakhany and Mausser [16], Rayner and MacGillivray [17], Fournier et al. [18], and Su [19].

In the present paper, we start from the fact that the vector of parameters  $\boldsymbol{\lambda}$  can be divided

into linear and nonlinear parameters, so that we can apply the nonlinear least squares (NLS) method, proposed by Öztürk and Dale [20], which is driven by the minimization of the sum of squared differences between the observed order statistics and their expected

values. Furthermore, we will develop a criterion driven by the minimization of the sum of the least absolute deviations between the observed order statistics and their medians. The two estimation procedures are based on a two-stage inner/outer optimization process. The first stage is the inner step in which the linear parameters are estimated for fixed values of the nonlinear parameters. The second stage is the outer step in which a controlled random search (CRS) technique is applied in order to determine the best companion value for the nonlinear parameters. The two stages are carried out consecutively, and the process is stopped when the difference between the values of the minima of two successive steps becomes smaller than a prefixed tolerance threshold.

The rest of the paper is organized as follows. The Section 2 describes the density of the FPLD random variables and underlines the advantage of its use in practical work. The Section 3 derives the nonlinear least squares for the estimation of  $\lambda$ . In Section 4, we outline a procedure for applying the criterion of least absolute deviations to the FPLD. The CRS, an optimization algorithm based on a purely heuristic, derivative-free technique is presented in the fifth section. In Section 6, the effectiveness of the proposed methods is demonstrated and compared using real and simulated data. Finally, we conclude and outline future research directions in Section 7.

## 2. Overview of the Five-Parameter Lambda Distribution (FPLD)

The FPLD is a highly flexible and adaptable tool for modelling empirical and theoretical distributions and, hence, might have many potential applications. In this section, we summarize its general features. The density-quantile function of a FPLD random variable is

$$f[X(p, \lambda)] = \frac{\lambda_2^{-1}}{\gamma_1 p^{\lambda_4-1} + \gamma_2 q^{\lambda_5-1}}, \quad \gamma_1 = \frac{(1 - \lambda_3)}{2}, \quad \gamma_2 = \frac{(1 + \lambda_3)}{2}. \quad (2.1)$$

The range of a FPLD is given by  $[X(0, \lambda), X(1, \lambda)]$  and is different for different values of  $\lambda$ ; in particular, the extremes are  $[\lambda_1 - \lambda_2\gamma_1, \lambda_1 + \lambda_2\gamma_2]$  if  $\lambda_4, \lambda_5 > 0$ ;  $[\lambda_1 - \lambda_2\gamma_1, \infty)$  if  $\lambda_4 > 0, \lambda_5 \leq 0$ ; and  $(-\infty, \lambda_1 + \lambda_2\gamma_2]$  if  $\lambda_4 \leq 0, \lambda_5 > 0$ . Hence, the FPLD can model distributions whose support extends to infinity in either or both directions as well as those with finite support.

The  $r$ th moment of the linear transformation  $Z = 2(X - \lambda_1)/\lambda_2 + k_1 - k_2$  where  $k_1 = (1 - \lambda_3)/\lambda_4, k_2 = (1 + \lambda_3)/\lambda_5$  is

$$E(Z^r) = \sum_{j=0}^r (-1)^j \binom{r}{j} k_1^{(r-j)} k_2^j B[1 + (r-j)\lambda_4, 1 + j\lambda_5]. \quad (2.2)$$

The symbol  $B(\alpha, \beta)$  with  $\alpha, \beta > 0$  denotes the complete beta function. As a consequence, the  $r$ th moment of  $Z$  exists and is finite if  $\lambda_4$  and  $\lambda_5$  are greater than  $r^{-1}$ .

The central moments of  $X$  which can be derived from (2.2) are

$$\begin{aligned}
 \sigma^2 &= \alpha_1 - 2k_1k_2B(\lambda_4 + 1, \lambda_5 + 1), \\
 \gamma_1 &= \frac{\alpha_2}{\sigma^3} - \frac{3(k_1k_2)^2}{\sigma^3} \left[ \frac{B(2\lambda_4 + 1, \lambda_5 + 1)}{k_2} - \frac{B(\lambda_4 + 1, 2\lambda_5 + 1)}{k_1} \right], \\
 \gamma_2 &= \frac{\alpha_3}{\sigma^4} - \frac{4(k_1k_2)^3}{\sigma^4} \left[ \frac{B(3\lambda_4 + 1, \lambda_5 + 1)}{k_2^2} + \frac{B(\lambda_4 + 1, 3\lambda_5 + 1)}{k_1^2} + \right. \\
 &\quad \left. - \frac{3B(2\lambda_4 + 1, 2\lambda_5 + 1)}{2k_1k_2} \right], \\
 \alpha_i &= \frac{k_1^{i+1}}{[(i+1)\lambda_4 + 1]} + \frac{(-1)^{i+1}k_2^{i+1}}{[(i+1)\lambda_5 + 1]}, \quad i = 1, 2, 3.
 \end{aligned} \tag{2.3}$$

The shape of the FPLD is sensitive to its parameter values; in fact, the moments of  $X(p, \boldsymbol{\lambda})$  are determined by a combination of all of the vector  $\boldsymbol{\lambda}$  elements. If  $\lambda_4 = \lambda_5$  and  $\lambda_3 = 0$ , then the FPLD is symmetric about  $\lambda_1$  because, in this case, (1.1) satisfies the condition  $X(p, \boldsymbol{\lambda}) = -X(q, \boldsymbol{\lambda})$ . If  $\lambda_4 = \lambda_5$ , then (2.1) is skewed to the left (right) if  $\lambda_3 < 0$  ( $\lambda_3 > 0$ ), which suggests a natural interpretation of  $\lambda_3$  as a parameter which is prevalently related to the asymmetry of  $X(p, \boldsymbol{\lambda})$ . For a FPLD, interchanging  $\lambda_4$  and  $\lambda_5$  and simultaneously changing the sign of  $\lambda_3$  reverses the type of skewness. If  $\lambda_3 = -1$  ( $\lambda_3 = 1$ ), then  $\lambda_4$  ( $\lambda_5$ ) indicates the left (right) tail flatness in the sense that the smaller  $\lambda_4$  ( $\lambda_5$ ) the flatter the left (right) tail of the FPLD density. In practice,  $\lambda_4$  and  $\lambda_5$  capture the tail order on the two sides of the support of  $X(p, \boldsymbol{\lambda})$ .

The densities (2.1) are zeromodal if  $\{\max(\lambda_4, \lambda_5) > 1\} \wedge \{\min(\lambda_4, \lambda_5) < 1\}$ , unimodal with continuous tails if  $\{\max(\lambda_4, \lambda_5) < 1\}$ , unimodal with truncated tails if  $\{\min(\lambda_4, \lambda_5) > 2\}$ , U-shaped if  $1 < \lambda_4, \lambda_5 < 2$ , and S-shaped if  $\{\max(\lambda_4, \lambda_5) > 2\} \wedge \{\min(\lambda_4, \lambda_5) > 1\}$ . Curves corresponding to large positive values of  $\lambda_4$  and  $\lambda_5$  have extreme peakedness and short high tails. For  $\lambda_2 = 0$  the FPLD degenerates  $X(p, \boldsymbol{\lambda}) = \lambda_1$  to a one-point distribution.

The FPLD family is a good model for use in Monte Carlo simulation and robustness studies because it contains densities which cover a large spectrum of arbitrarily shaped densities. For example, if  $\lambda_4 \rightarrow 0, \lambda_5 \rightarrow 0$ , then  $X(p, \boldsymbol{\lambda})$  converges on a potentially asymmetric form of the logistic distribution; if  $\lambda_4 \rightarrow \infty$  and  $\lambda_5 \rightarrow 0$ , then  $X(p, \boldsymbol{\lambda})$  is an exponential distribution whereas for  $\lambda_4 \rightarrow 0, \lambda_5 \rightarrow \infty$ ,  $X(p, \boldsymbol{\lambda})$  becomes a reflected exponential distribution. Furthermore, FPLD fits data containing extreme values well; in fact, for  $\lambda_4 \rightarrow \infty, |\lambda_5| < \infty$ , the FPLD corresponds to the generalized Pareto distribution. For  $\lambda_5 \rightarrow \infty, |\lambda_4| < \infty$ , the FPLD generates the power-function distribution. The rectangular random variable is present in four versions:  $(\lambda_3 = -1, \lambda_4 = 1)$ ,  $(\lambda_3 = 1, \lambda_5 = 1)$ ,  $(\lambda_4 = 1, \lambda_5 = 1)$ , and  $(\lambda_3 = 0, \lambda_4 = 2, \lambda_5 = 2)$ .

The FPLD could have an important role as a general (not necessarily Gaussian) distribution if it is able to produce good approximations to many of the densities commonly encountered in applications. There are several measures of closeness and overlapping of two statistical distributions based on the distance between density functions or between distributions functions (see, e.g., [21] and [7, pages 196–201]). On the other hand, the FPLD is expressed in the quantile domain, so, for a reliable assessment of the agreement between the exact distribution and that provided by the FPLD, we should use a measure based

**Table 1:** Comparison of FPLD approximations for symmetric distributions.

Model	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	Maxd
Gaussian (0, 1)	0.00000	1.35921	0.00000	0.13312	0.13312	0.0065
Laplace: (0, 1)	0.00000	1.32397	0.00000	-0.09556	-0.09556	0.0957
Student's t: $\nu = 2$	0.00000	0.91014	0.00000	-0.45155	-0.45155	0.0973
Cauchy (0, 1)	0.00000	0.66267	0.00000	-0.99250	-0.99250	0.1673
Beta (0.5, 0.5)	0.00000	1.63202	0.00000	1.45456	1.45456	0.0593

on quantiles. Statistical literature is not very helpful in providing criteria to quantify the proximity between distributions naturally arising from a quantile framework. In the present paper we have used the Tchebycheff metric, that is, the closeness between the theoretical quantile function  $X(p)$  and the FPLD is quantified by the maximum absolute difference (Maxd) between observed and fitted quantiles:

$$\max_{1 \leq i \leq 500} |X(p_i) - X(p_i, \lambda)|, \quad \text{where } p_i = \frac{i}{501}, \quad i = 1, 2, \dots, 500. \quad (2.4)$$

The Tchebycheff metric is attractive because it leads to the minimization of the worst case and does not imply a heavy computational burden.

The five parameters of  $X(p, \lambda)$  have been estimated using the controlled random search method described in Section 5. Table 1 shows the accuracy of the FPLD fit to some standard symmetric distributions.

Our findings suggest that the FPLD fit is reasonably good for all of the models included in the table, with the only possible exception being the Cauchy random variable.

One of the major motives for generalizing the four-parameter GLD to the five-parameter FPLD is to have a more flexible, and, hence, more suitable, distribution for fitting purposes. How much better, though, is the FPLD when compared to the GLD? Are the fitting improvements that the FPLD provides significant enough to offset the complexity that is introduced through a fifth parameter? Table 2 shows the results of the fitting procedure for some asymmetric distributions. The first row presents the parameter estimates and the Tchebycheff metric for the FPLD approximation; the second row presents the same quantities for the GLD approximation using the FMKL parametrization (see Freimer et al. [5]). The values in the table indicate that the upper bound on the discrepancy between the true and the approximated distribution is systematically lower for the FPLD than it is for the GLD. The amounts appear to be large enough to justify the inclusion of an additional linear parameter in the GLD so as to form the FPLD.

The most important point to note in Tables 1 and 2 is that the FPLD is a flexible class of very rich distributional models which cover the Gaussian and other common distributions. In this sense, the FPLD is a valid candidate for being fitted to data when the experimenter does not want to be committed to the use of a particular distribution.

### 3. Nonlinear Least Squares Estimation

Öztürk and Dale [20] proposed a nonlinear least squares (NLS) framework to estimate vector  $\lambda$  for a four-parameter generalized distribution. The method can easily be generalized to handle a FPLD.

**Table 2:** Comparison of FPLD and GLD approximations for asymmetric distributions.

Model	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	Maxd
Chi-square $\nu = 3$	1.10010	2.84313	0.61505	0.48579	0.01338	0.0046
	2.01822	1.07570		0.97596	-0.03394	0.0827
Lognormal (4, 2)	3.44560	1.07305	0.16147	0.21008	0.04776	0.0048
	3.56591	3.64067		0.26715	0.01550	0.0193
Weibull (1, 3)	0.92565	0.50134	-0.15389	0.30875	0.12579	0.0013
	0.87453	8.10780		0.26347	0.16478	0.0052
Gumbel (0, 1)	-0.00050	1.50239	0.32984	0.19211	-0.00046	0.0020
	0.34406	2.44744		0.33889	-0.04845	0.0549
Inv. Gaussian (0.5, 6)	0.45521	0.18606	0.17494	0.25786	0.02961	0.0014
	0.47729	20.64410		0.32817	0.00039	0.0044

For a given random sample  $S = \{x_1, x_2, \dots, x_n\}$  of  $n$ -independent observations from  $X(p, \lambda)$ , the  $i$ th order statistic may be expressed as

$$x_{i:n} = E(x_{i:n}) + e_i, \quad i = 1, 2, \dots, n, \quad (3.1)$$

in which the ordered observations are compared with their expected values under the hypothesized distribution  $X(p, \lambda)$  and  $e_i$  is a measure of the discrepancy between the observed and modelled  $i$ th value. The error terms  $\{e_i, i = 1, 2, \dots, n\}$  are such that  $E(e_i) = 0$  and  $\sigma^2(e_i) = \sigma_i^2$ . The heteroskedasticity in (3.1) is not the only violation of standard assumptions of the classical regression model. In fact, the error terms are not independent and do not come from a symmetrical distribution (except for the median). However, since the purpose was to obtain an approximate solution, Öztürk and Dale [20] ignored these inadequacies.

The expected value of the  $i$ th order statistic from a FPLD is available in closed form

$$E(x_{i:n}) = \lambda_1 + \frac{(\beta_1/\lambda_4) \int_0^1 p^{\lambda_4+i-1} q^{n-i} dp + (\beta_2/\lambda_5) \left[1 - \int_0^1 p^{i-1} q^{\lambda_5+n-i} dp\right]}{B(i, n+1-i)}, \quad (3.2)$$

where  $\beta_1 = 0.5(1 - \lambda_3)\lambda_2$ ,  $\beta_2 = 0.5(1 + \lambda_3)\lambda_2$ . It follows that the deterministic component of (3.1) can be written as

$$\bar{x}_{i:n}(\lambda) = \beta_0 U_{0,i} + \beta_1 U_{1,i} + \beta_2 U_{2,i}, \quad i = 1, 2, \dots, n, \quad (3.3)$$

with  $\beta_0 = \lambda_1$  and

$$U_{0,i} = 1, \quad U_{1,i} = \lambda_4^{-1} \left[ \frac{\Gamma(n+1)\Gamma(i+\lambda_4)}{\Gamma(i)\Gamma(n+1+\lambda_4)} - 1 \right], \quad (3.4)$$

$$U_{2,i} = \lambda_5^{-1} \left[ 1 - \frac{\Gamma(n+1)\Gamma(n+1-i+\lambda_5)}{\Gamma(n+1-i)\Gamma(n+1+\lambda_5)} \right].$$

The least squares approach calls for  $\boldsymbol{\beta}$  in (3.3) to be chosen so as to minimize

$$S_2(\boldsymbol{\lambda}) = \sum_{i=1}^n [x_{i:n} - \bar{x}_{i:n}(\boldsymbol{\lambda})]^2. \quad (3.5)$$

The  $S_2(\boldsymbol{\lambda})$  is linear in  $\beta_1, \beta_2, \beta_3$  but not in  $(\lambda_4, \lambda_5)$ . According to Lawton and Sylvestre [22] the solution to (3.5) can be obtained by fixing the value of  $(\lambda_4, \lambda_5)$  and, then, applying the ordinary least squares to solve the linear problem

$$\hat{\boldsymbol{\beta}}(\lambda_4, \lambda_5) = (\mathbf{U}_\lambda^t \mathbf{U}_\lambda)^{-1} \mathbf{U}_\lambda^t \mathbf{x} \quad (3.6)$$

for  $\boldsymbol{\beta}$ , where  $\mathbf{U}_\lambda$  denotes the  $(n \times 3)$  matrix with elements given by (3.4) and  $\mathbf{x}$  is the  $(n \times 1)$  vector of the ordered observations. The conversion of  $\beta_0, \beta_1, \beta_2$  into  $\lambda_1, \lambda_2, \lambda_3$  is straightforward:

$$\hat{\lambda}_1 = \hat{\beta}_0, \quad \hat{\lambda}_2 = \frac{\hat{\beta}_1}{0.5(1 - \hat{\lambda}_3)} = \frac{\hat{\beta}_2}{0.5(1 + \hat{\lambda}_3)}, \quad \hat{\lambda}_3 = \frac{1 - (\hat{\beta}_1/\hat{\beta}_2)}{1 + (\hat{\beta}_1/\hat{\beta}_2)}. \quad (3.7)$$

Öztürk and Dale [20] used an approximation of  $E(x_{i:n})$  to overcome the problems created by repeated use of the gamma function in (3.4). More specifically, they used  $E(x_{i:n}) = X(p_i^+, \boldsymbol{\lambda})$ , with the plotting positions  $p_i^+ = i(n+1)^{-1}$ ,  $i = 1, 2, \dots, n$ . In this case, the pseudoregressors are

$$V_{0,i} = 1, \quad V_{1,i} = \lambda_4^{-1} [(p_i^+)^{\lambda_4} - 1], \quad V_{2,i} = \lambda_5^{-1} [(p_{n+1-i}^+)^{\lambda_5} - 1]. \quad (3.8)$$

Parameter estimates are then obtained as in (3.5) through (3.7) with  $V_{j,i}$  in place of  $U_{j,i}$  for  $j = 1, 2, 3$ . We can now define the reduced form of the quantile function (3.5) in which  $\hat{\boldsymbol{\beta}}$  is substituted into (3.3). Clearly, the reduced quantile function only depends on the pair  $(\lambda_4, \lambda_5)$ . We might try to find a better estimate of  $(\lambda_4, \lambda_5)$  by using one of the function minimization techniques to solve (3.5) and then go through exactly the same procedure as the one described above. The two-stage process is repeated until the correction for  $S_2(\boldsymbol{\lambda})$  becomes sufficiently small.

#### 4. Least Absolute Deviations Estimation

Filliben [23] noted that order statistics means have three undesirable properties: (1) no uniform technique exists for generating the  $E(x_{i:n})$  for all distributions; (2) for almost all distributions  $E(x_{i:n})$  is difficult or time consuming to compute and so must be stored or approximated; (3) for other distributions (e.g., Cauchy), the expected value of the order statistics  $E(x_{i:n})$  may not always be defined. All three of these drawbacks are avoided in general by choosing to measure the location of the  $i$ th order statistic by its median rather than by its mean.

The use of the median in preference to the mean leads naturally to the absolute error being the loss function instead of the squared error. Gilchrist [2] called distributional

residuals the deviations of the observed data from a given quantile function  $e_i = x_{i:n} - X(p_i, \boldsymbol{\lambda})$ ,  $i = 1, 2, \dots, n$ . In this context, the method of nonlinear least squares discussed in the previous section is defined as distributional least squares since it involves using the ordered data as a dependent variable, whereas the deterministic component of the model is defined by a quantile function.

The least squares criterion is reasonable when data come from a normal distribution or when linear estimates of the parameters are required. While data may contain outliers or highly influential values, the method of least squares has a clear disadvantage in that it may be pulled by extremely large errors. The least absolute deviation estimation of parameters is more suitable in such cases.

In this section we consider the problem of finding a vector of constants  $\mathbf{y} \in \mathbf{R}^n$  which minimizes the objective function

$$\min_{\mathbf{y} \in \mathbf{R}^n} \sum_{i=1}^n |x_{i:n} - y_i|, \quad (4.1)$$

where  $x_{i:n}$  is the  $i$ th order statistic of the sample  $S = \{x_1, x_2, \dots, x_n\}$  from the quantile function  $X(p, \boldsymbol{\lambda})$ . Rao [24] observed that the solution to this problem is the vector of marginal medians, that is, the  $i$ th element of  $\mathbf{y}$  is the median of the order statistics  $x_{i:n}$ .

The quantile function (see, e.g., Gilchrist [2, page 86]) of the  $i$ th order statistic for a sample of size  $n$  from an FPLD is  $X[p_i^*(p), \boldsymbol{\lambda}]$  with

$$p_i^*(p) = B^{-1}(p, i, n + 1 - i), \quad (4.2)$$

where  $B^{-1}(\cdot)$  is the inverse of the incomplete beta function. Consequently, the median of the  $i$ th order statistic can be computed using

$$p_i^* = X[B^{-1}(0.5, i, n + 1 - i), \boldsymbol{\lambda}], \quad i = 1, 2, \dots, n. \quad (4.3)$$

The quantities in (4.3) are a solution of (4.1) in the sense that, for a fixed value of  $\boldsymbol{\lambda}$ , the following expression

$$\sum_{i=1}^n |x_{i:n} - X[p_i^*, \boldsymbol{\lambda}]| \quad (4.4)$$

attains its minimum value. It follows that the estimate of  $\boldsymbol{\lambda}$  that minimizes (4.4) can be obtained by the use of distributional absolutes (DLAs).

There is nothing in the previous section that cannot be extended to the criterion (4.4). In the light of the decomposition of  $\boldsymbol{\lambda}$  into two groups, linear parameters  $(\lambda_1, \lambda_2, \lambda_3)$ , and nonlinear parameters  $(\lambda_4, \lambda_5)$ , another two-stage procedure can be devised. In the inner stage, we determine the vector  $\tilde{\boldsymbol{\beta}}$  which minimizes

$$S_1(\boldsymbol{\lambda}) = \sum_{i=1}^n |x_{i:n} - \tilde{x}_{i:n}(\boldsymbol{\lambda})|, \quad \text{with } \tilde{x}_{i:n}(\boldsymbol{\lambda}) = \sum_{j=0}^2 \beta_j W_{i,j}, \quad (4.5)$$

with respect of the linear parameters  $\beta$  and for a given value of the nonlinear parameters  $(\lambda_4, \lambda_5)$ . In this case, the pseudo regressors are

$$W_{0,i} = 1, \quad W_{1,i} = \lambda_4^{-1} \{ [p_i^*]^{\lambda_4} - 1 \}, \quad W_{2,i} = \lambda_5^{-1} \{ [1 - p_i^*]^{\lambda_5} - 1 \}. \quad (4.6)$$

The computation of the LAD estimates for  $\tilde{\beta}$  in (4.5) can be formulated as a linear programming optimization and the standard simplex method can be employed to solve this problem. For a detailed treatment of the method of least absolutes see, for example, Koenker and D'Orey [25]. The conversion of  $\beta_0, \beta_1, \beta_2$  into  $\lambda_1, \lambda_2, \lambda_3$  takes place as shown in (3.7). Once  $(\lambda_1, \lambda_2, \lambda_3)$  have been estimated, we can obtain a new value for  $(\lambda_4, \lambda_5)$  with the same outer step as that used in the previous section. The inner and outer steps can be alternated until a satisfactory result is found.

Least absolute deviations have, undoubtedly, greater sophistication than least squares, although two important points should be made regarding this issue. First, the annoying computation of  $p_i^*$  has to be executed just once. In addition, the fact that  $B(p, i, n + 1 - i) = 1 - B(1 - p, n + 1 - i, i)$  implies that  $p_{n+1-i}^* = 1 - p_i^*$  and, consequently, the number of medians to be computed is halved.

## 5. Controlled Random Search (CRS)

The inner/outer scheme outlined in previous sections has the merit of reducing the dimension of the nonlinear problem. To apply it, however, we must minimize two relatively complex objective functions  $S_1(\lambda)$  and  $S_2(\lambda)$  of two unknowns,  $\lambda_4, \lambda_5$ , which generally have more than one minimum. The solutions to these equations are difficult to obtain through traditional optimization approaches. In order to perform this special task, we resorted to a direct search optimization technique (CRS) proposed by Price [26], which is suitable for searching the global minima for continuous functions that are not differentiable, or whose derivatives are difficult to compute or to approximate (see [27, 28]).

Some well-known direct search procedures such as the downhill simplex method of Nelder and Mead [29] and the pattern-search method of Hooke and Jeeves [30] are familiar to statisticians. However, these techniques are really local optimization methods; in practice, they are designed so as to converge on a single optimum point and, therefore, they, unavoidably, discard information related to all other possible optima. In contrast, a wise application of the CRS overcomes this drawback since this method is especially suited to multimodal functions. In effect, CRS procedures have more randomness in their decision process, thus increasing the chances of finding a global optimum. Of course, global optimality cannot be guaranteed in finite time.

A CRS procedure employs a storage  $A$  of points, the number of which is determined by the user for the particular problem to be solved. Repeated evaluation of the function to be optimized,  $f(\mathbf{x})$ , is performed at points randomly chosen from the storage of points and the search region is progressively contracted by substituting the worst point with a better one. The search continues until an iteration limit is reached, or until a desired tolerance between minimum and maximum values in the stored  $f(\mathbf{x})$  values is attained. Apparently, these methods are less known and less frequently used in statistics (two works that have appeared in literature are those by Shelton et al. [31] as well as Křivý and Tvrdík [32]).

For practical purposes, it is necessary to confine the search within a prescribed bounded domain. Let  $C \subset \mathbf{R}^k$  be the set of admissible values of the  $k = 2$  nonlinear parameters of the FPLD model. The most promising region that we have found in our applications is the rectangle delimited by

$$C = \{(\lambda_4, \lambda_5) \mid -0.999 \leq \lambda_4 \leq 3; -0.999 \leq \lambda_5 \leq 3\}. \quad (5.1)$$

Furthermore,  $C$  ensures that the mean of the order statistics from a FPLD always exists and the ranges of  $\lambda_4, \lambda_5$  cover most of the distributional shapes commonly encountered.

The theory and practice of global optimization have progressed rapidly over the last few years, and a wide variety of modifications of the basic CRS algorithm are now available. In particular, we refer to the scheme discussed in Brachetti et al. [33].

Any CRS implies many subjective choices of constants and weights. The version used in our paper is as follows.

- (1) Generate a set  $A$  of  $m_0 > (k + 1)$  random admissible points in  $C$ . Find the minimum and the maximum in  $A$  and their function values:  $(\mathbf{x}_L, f_L), (\mathbf{x}_H, f_H)$ .
- (2) If  $|f_H - f_L| / (10^{-7} + |f_H| + |f_L|) < 10^{-8}$ , then stop. Also stop if the limit to the number of function evaluations has been reached. The point with the lowest objective function value found is returned. If required, return all the points in  $A$ .
- (3) Choose at random, without duplication, a set  $B$  of  $k$  points in  $(A - \mathbf{x}_L)$ , include  $\mathbf{x}_L$  in  $B$ , and compute the centroid  $\bar{\mathbf{x}} = (k + 1)^{-1} \sum_{\mathbf{x} \in B} \mathbf{x}$ . Define  $A_1 = A - B$ .
- (4) Select randomly a point  $\mathbf{x}_r \in A_1$ .
- (5) Generate uniform random numbers  $d_1, d_2, \dots, d_k$  in  $[0, 2]$  and compute the trial point  $\mathbf{x}_{i,T} = (1 + d_i)\bar{\mathbf{x}}_i - d_i\mathbf{x}_{r,i}$ , for  $i = 1, 2, \dots, k$ ; thus, the exploration of the space around the centroid is based on a randomized reflection in the simplex originally used by Price [26].
- (6) If  $\mathbf{x}_T \in \mathbf{M}$ , then compute  $f_T = f(\mathbf{x}_T)$  and go to step (7). If  $\mathbf{x}_T \notin \mathbf{M}$ , then repeat step (5) a fixed number  $m_1$  of times. If  $\mathbf{x}_T$  is still unfeasible, then exclude  $\mathbf{x}_r$  from  $A_1$ , redefine the set  $A_1$  and, if it is not empty, repeat step (4) for at most  $m_2$  times; otherwise go to step (9).
- (7) If  $f_L < f_T < f_H$ , then replace  $f_H$  with  $f_T$  and  $\mathbf{x}_H$  with  $\mathbf{x}_T$  and then redetermine the maximum  $\mathbf{x}_H$  and  $f_H$ .
- (8) If  $f_T < f_L$ , then replace  $f_H$  with  $f_L$ ,  $\mathbf{x}_H$  with  $\mathbf{x}_L$ ,  $f_L$  with  $f_T$ ,  $\mathbf{x}_L$  with  $\mathbf{x}_T$ , and  $\mathbf{x}_H$  and  $f_H$ .
- (9) Randomly choose two distinct points  $\mathbf{x}_2, \mathbf{x}_3$  in  $(A - \mathbf{x}_L)$  and compute

$$d_{Q,i} = (x_{2,i} - x_{3,i})f_L + (x_{3,i} - x_{L,i})f_2 + (x_{L,i} - x_{2,i})f_3, \quad (5.2)$$

for  $i = 1, 2, \dots, k$ .

- (10) If  $|d_{Q,i}| > 10^{-7}$  for each  $i$ , then build the following quadratic interpolation of the objective function:

$$\mathbf{x}_{Q,i} = \left[ \frac{(x_{2,i} - x_{3,i})^2 f_L + (x_{3,i} - x_{L,i})^2 f_2 + (x_{L,i} - x_{2,i})^2 f_3}{2d_{Q,i}} \right] \quad (5.3)$$

for  $i = 1, 2, \dots, k$ . If some of the  $|d_{Q,i}| \leq 10^{-7}$ , then generate  $k$  uniform random numbers  $d_i, i = 1, 2, \dots, k$ , in  $[0, 1]$  and compute

$$x_{i,Q} = d_i x_{i,L} + (1 - d_i) x_{i,H}, \quad \text{for } i = 1, 2, \dots, k. \quad (5.4)$$

- (11) If  $\mathbf{x}_Q \notin \mathbf{M}$ , then repeat steps (9)-(10) no more than  $m_3$  times. If  $\mathbf{x}_Q$  is still unfeasible, then return to step (2).
- (12) Execute steps (7)-(9) with  $\mathbf{x}_Q$  in place of  $\mathbf{x}_T$  and  $f_Q$  in place of  $f_T$  and then return to step (2).

We experimented by using the CRS algorithm, with  $\mathbf{x} = \boldsymbol{\lambda}$ ,  $f(\mathbf{x}) = S_2(\boldsymbol{\lambda})$ ,  $m_0 = 30k$ ,  $m_1 = m_2 = m_3 = 4(k + 1)$ , and  $k = 2$ . For the random component of our implementation we used quasirandom Faure sequences (Faure [34]) rather than pseudorandom numbers. This choice was made in order to cover the search region  $\mathbf{C}$  more evenly (see, e.g., Morokoff and Caflisch [35]).

The CRS procedure has several desirable features. Boundaries and objective function can be arbitrarily complex because the procedure does not require the calculation of derivatives. Shelton et al. [31] observed that, if several points within the optimization region are equally good, the procedure will make this clear to the user.

The general outline presented here should give some appreciation of the philosophy underlying the design of CRS. It is important not to forget that, although this method proved very robust in many applications, typical nonlinear problems which have a complicated bound-constrained, possibly multiextremal objective function are challenging tasks and the success of an automated algorithm is still dependent upon the researcher's guidance. In particular, intervention may be needed to contract the box constraints on the parameters in order to perform a finer search in a final stage.

## 6. Comparison of the Procedures

In this section we present results of some numerical experiments carried out in order to study the behavior of the FPLD model. Moreover, we perform a comparison between estimated and theoretical values to test the accuracy of the estimators proposed in the previous sections and investigate their properties through a Monte Carlo sampling process.

### 6.1. Fitting to Data

To illustrate the efficacy of the FPLD as a modeling device we applied it to four sets of real data. To judge the overall agreement between estimated and observed quantiles we used the correlation between the empirical values  $x_{i:n}$  centered on the arithmetic mean  $\bar{x}$  and the hypothesized values

$$r(\boldsymbol{\lambda}) = \frac{\sum_{i=1}^n [x_{i:n} - \bar{x}] [X(p_i, \boldsymbol{\lambda}) - \mu(\boldsymbol{\lambda})]}{\left\{ \sum_{i=1}^n [x_{i:n} - \bar{x}]^2 \sum_{i=1}^n [X(p_i, \boldsymbol{\lambda}) - \mu(\boldsymbol{\lambda})]^2 \right\}^{0.5}}, \quad (6.1)$$

where  $p_i = i/(n + 1)$ , and

$$\mu(\lambda) = \int_0^1 X(p, \lambda) dp = \lambda_1 + \frac{\lambda_2}{2} \left[ \frac{1 + \lambda_3}{1 + \lambda_5} - \frac{1 - \lambda_3}{1 + \lambda_4} \right]. \quad (6.2)$$

From another point of view, (6.1) can be considered a measurement of the linearity of the Q-Q plot for empirical versus theoretical standardized quantiles. The coefficient (6.1) will always be positive since  $x_{i:n}$  and  $X(p_i, \lambda)$  are both arranged in ascending order of magnitude. More specifically,  $r(\lambda)$  takes values in the  $[0, 1]$  interval. The case  $r(\lambda) = 0$  corresponds to a noninformative model  $X(p_i, \lambda) = \lambda_1$ ,  $i = 1, 2, \dots, n$ ; a value of the coefficient  $r(\lambda)$  close to one will suggest a good fit of the FPLD. The model that causes data to be most like a straight line on its Q-Q plot is the FPLD that most closely resembles the underlying distribution of the data. Correlations that are too small indicate a lack of fit. Finally,  $r(\lambda) = 1$  means that  $x_{i:n} = X(p_i, \lambda)$ ,  $i = 1, 2, \dots, n$ , that is, the observed data correspond exactly to their expected values.

Table 3 reports our findings on the following datasets:

- (i) recorded annual drought of Iroquois River recorded near Chebanse (IL),  $n = 32$ , Ghosh [36],
- (ii) times in seconds to serve light vans of the Severn Bridge River crossing in Britain,  $n = 47$ , Gupta and Parzen [37],
- (iii) peak concentrations of accidental releases of toxic gases,  $n = 100$ , Hankin and Lee [38],
- (iv) observations on overall length of bolts,  $n = 200$ , Pal [39].

The symbol  $\nu$  denotes the number of function evaluations required to find the global optimum of the criterion. The acronym OD refers to the plotting positions  $p_i^+$  proposed by Öztürk and Dale [20].

The best overall performance is achieved by NLS; comparably good results are also obtained for DLA and they are only slightly inferior for OD. Based on these few values, the computational difficulties encountered in NLS may be worth the effort. The estimates provided by DLA can be very different from the others because they are based on a different distance metric. The complexity of DLA is higher than that of NLS or OD, due to its more intricate search. The limited impact on the modelling adequacy does not compensate for the extra energy expended for the DLA procedure.

## 6.2. Simulations

In this section we carry out a simulation study to evaluate the performance of the proposed methods mainly with respect to their biases and squared errors for different sample sizes. In particular, we consider the parameter combination that allows the FPLD to approximate to the standard Gaussian distribution:

$$\lambda_0 = (0.00000, 1.35921, 0.00000, 0.13312, 0.13312). \quad (6.3)$$

**Table 3:** Comparison of FPLD interpolations.

Method	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\nu$	$r(\lambda)$
Ghosh [36]							
NLS	13.5553	48.5361	0.9978	-0.5401	0.1851	1283	0.7812
OD	14.7086	46.5057	0.9917	-0.6384	0.1140	1158	0.3821
DLA	15.2217	51.2328	0.9160	0.9585	0.2036	9287	0.3771
Gupta and Parzen [37]							
NLS	5.4370	2.9152	-0.1544	0.6544	-0.1097	1097	0.8468
OD	6.3223	3.4372	-0.6018	0.8189	-0.4031	1139	0.9100
DLA	4.9864	2.6533	0.1058	0.5644	-0.0895	2673	0.7579
Hankin and Lee [38]							
NLS	5.4882	4.6259	0.6731	-0.0152	0.2025	1095	0.6982
OD	5.5427	4.4453	0.6759	-0.0896	0.1620	1049	0.5605
DLA	5.9392	4.0234	0.5977	-0.0764	0.0935	5395	0.6147
Pal [39]							
NLS	6.3981	0.1982	0.7108	-0.0518	0.2184	961	0.6771
OD	6.3971	0.1933	0.7330	-0.1266	0.1983	1082	0.5337
DLA	6.4183	0.1803	0.5849	-0.0087	0.1329	2272	0.6142

The vector  $\lambda_0$  will act as our reference point. The reason we use the Gaussian model is that the normality assumption is one of the most extensively used in statistics and a new method should at least work in this default case.

Two simple coefficients of performance have been considered for comparing

$$\text{MRB}(\lambda_j) = N^{-1} \sum_{i=1}^N \left| \frac{\lambda_{i,j} - \lambda_{0,j}}{\alpha_j} \right|, \quad \text{where } \alpha_j = \begin{cases} \lambda_{0,j} & \text{if } \lambda_{0,j} \neq 0 \\ 1 & \text{if } \lambda_{0,j} = 0 \end{cases} \quad (6.4)$$

$$\text{RMSE}(\lambda_j) = \left[ N^{-1} \sum_{i=1}^N (\lambda_{i,j} - \lambda_{0,j})^2 \right]^{0.5}.$$

The mean relative bias (MRB) quantifies the average relative magnitude of the accuracy for each parameter and the root mean squared error (RMSE) represents the mean Euclidean distance between simulation and measurement. Methods that yield estimates which are closer to the true parameter value have lower bias, higher precision, and lower RMSEs.

The statistics in (6.4) were calculated by generating  $N = 10,000$  different random samples of size  $n \in (30, 60, 120, 250, 500)$  from  $X(p, \lambda_0)$ . The results are summarized in Table 4

For all methods, MRB and RMSE decrease as the sample size increases, an indication that all of the estimators are consistent. The sign of the bias for the exponential parameters is almost always negative showing a systematic underestimation of  $(\lambda_4, \lambda_5)$  probably due to the fact that the Gaussian has an infinite range (which would imply a FPLD with negative values of the exponential parameters), but it is well approximated by a FPLD with a limited support.

As might have been expected, the bias and the standard error of the exponential parameters  $(\lambda_4, \lambda_5)$  yielded little or no discernible differences, particularly for the largest samples. The bias of the NLS technique is generally smaller than those of the other methods.

Table 4: Simulation results.

Method	Index	$n$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$
NLS	MRB	30	0.0049	0.5616	-0.0042	0.3627	0.3382
		60	-0.0061	0.1994	0.0037	-0.1199	-0.0756
		120	0.0004	0.0789	-0.0015	-0.0565	-0.0687
		250	0.0014	0.0342	-0.0017	-0.0205	-0.0265
		500	-0.0021	0.0176	0.0027	-0.0110	0.0035
OD		30	0.0033	0.5080	-0.0055	-0.7632	-0.7697
		60	-0.0149	0.1782	0.0140	-1.0219	-0.8967
		120	0.0004	0.0636	-0.0013	-0.5128	-0.5173
		250	-0.0047	0.0190	0.0055	-0.2752	-0.2395
		500	0.0021	0.0035	-0.0016	-0.1402	-0.1441
DLA		30	0.0083	0.5739	-0.0047	-0.0693	-0.0791
		60	-0.0046	0.2045	0.0020	-0.4196	-0.4449
		120	0.0072	0.0702	-0.0082	-0.2173	-0.2610
		250	0.0058	0.0280	-0.0058	-0.0898	-0.1174
		500	0.0008	0.0111	-0.0014	-0.0495	-0.0535
NLS	RMSE	30	0.8814	2.1470	0.7079	0.7572	0.7578
		60	0.6729	0.9853	0.5984	0.4033	0.4026
		120	0.4728	0.3686	0.4602	0.2140	0.2128
		250	0.3225	0.1567	0.3340	0.1300	0.1310
		500	0.2208	0.1004	0.2369	0.0879	0.0875
OD		30	0.8610	2.1256	0.7446	0.8617	0.8664
		60	0.7257	0.9355	0.6628	0.5121	0.5104
		120	0.5290	0.3262	0.5226	0.2832	0.2823
		250	0.3567	0.1630	0.3765	0.1636	0.1630
		500	0.2410	0.1013	0.2630	0.1027	0.1029
DLA		30	0.8733	2.1730	0.7405	0.8594	0.8580
		60	0.7040	1.0139	0.6358	0.4775	0.4784
		120	0.5058	0.3063	0.4951	0.2412	0.2426
		250	0.3378	0.1606	0.3537	0.1447	0.1455
		500	0.2314	0.1011	0.2497	0.0944	0.0946

This effect is particularly noticeable in the exponential parameters. The standard error of OD and DLA are generally larger than those for the NLS so that this method can be considered as the best of those looked at in this paper. However, for moderate sample sizes ( $n > 120$ ), OD is nearly as good as NLS, but takes less time to compute, so the choice of method depends on the particular dataset. The DLA solution does not seem to be a real improvement over methods based on the least squares criterion.

## 7. Conclusions

Statistics professionals are often faced with the problem of fitting a set of data by means of a stochastic approximation. For this purpose, it is essential to select a single parametric family of distributions that offers simple but still realistic descriptions of most of the potential models of the underlying phenomenon. A very useful and tractable parametric model is the five-parameter generalized lambda distribution. By means of the FPLD model we have

obtained some satisfactory fitting of empirical data and at least one of the many subordinate models can provide a good approximation to several common distributions.

More specifically, we have analyzed three methods for estimating the parameters of the FPLD (two of which are completely new). The common denominator of all of the methods is the replacement of linear parameters by their least squares or least absolutes estimates, given the value of the nonlinear parameters that, in turn, are determined by using a controlled random search technique. The proposed procedures are simple to apply as they only require a derivative-free optimization over a bounded region. Thus, many obstacles to the use of the FPLD are resolved. The efficiency of the new algorithms has been verified by applications to real and simulated data.

Our findings indicate that NLS estimators outperform least absolute deviation estimators with respect to bias and mean square error. For large sample, however, the difference between NLS and the approximated, but more manageable, scheme OD proposed by Öztürk and Dale [20] becomes imperceptible.

As for future research, we plan to compare different diagnostic tools used to analyze the fit of a quantile model. It would also be interesting to incorporate in the methods described in this paper a weighting scheme (e.g., weights proportional to the local density) so that the tails and/or the middle portion of the samples become more detectable.

## References

- [1] E. Parzen, "Nonparametric statistical data modeling," *Journal of the American Statistical Association*, vol. 74, no. 365, pp. 105–131, 1979.
- [2] W. G. Gilchrist, *Statistical Modelling with Quantile Functions*, Chapman & Hall, Boca Raton, Fla, USA, 2000.
- [3] J. S. Ramberg and B. W. Schmeiser, "An approximate method for generating asymmetric random variables," *Communications of the Association for Computing Machinery*, vol. 17, pp. 78–82, 1974.
- [4] J. S. Ramberg, E. J. Dudewicz, P. R. Tadikamalla, and E. F. Mykytka, "A probability distribution and its uses in fitting data," *Technometrics*, vol. 21, no. 2, pp. 201–214, 1979.
- [5] M. Freimer, G. S. Mudholkar, G. Kollia, and C. T. Lin, "A study of the generalized Tukey lambda family," *Communications in Statistics. Theory and Methods*, vol. 17, no. 10, pp. 3547–3567, 1988.
- [6] Z. A. Karian, E. J. Dudewicz, and P. McDonald, "The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the "final word" on moment fits," *Communications in Statistics. Simulation and Computation*, vol. 25, no. 3, pp. 611–642, 1996.
- [7] Z. A. Karian and E. J. Dudewicz, *Fitting Statistical Distributions*, CRC Press, Boca Raton, Fla, USA, 2000, The generalized lambda distribution and generalized bootstrap method.
- [8] R. A. R. King, "gld: Estimation and use of the generalised (Tukey) lambda distribution," R package version 1.8.4, 2009, <http://tolstoy.newcastle.edu.au/~rking/publ/rprogs/information.html>.
- [9] W. G. Gilchrist, "Modeling and fitting quantile distributions and regressions," *American Journal of Mathematical and Management Sciences*, vol. 27, no. 3-4, pp. 401–439, 2007.
- [10] Z. A. Karian and E. J. Dudewicz, "Comparison of GLD fitting methods: superiority of percentile fits to moments in  $L_2$  norm," *Journal of the Iranian Statistical Society*, vol. 2, pp. 171–187, 2003.
- [11] R. A. R. King and H. L. MacGillivray, "Fitting the generalized lambda distribution with location and scale-free shape functionals," *American Journal of Mathematical and Management Sciences*, vol. 27, no. 3-4, pp. 441–460, 2007.
- [12] A. Tarsitano, "Estimation of the generalized lambda distribution parameters for grouped data," *Communications in Statistics. Theory and Methods*, vol. 34, no. 8, pp. 1689–1709, 2005.
- [13] W. H. Asquith, "L-moments and TL-moments of the generalized lambda distribution," *Computational Statistics & Data Analysis*, vol. 51, no. 9, pp. 4484–4496, 2007.
- [14] J. Karvanen and A. Nuutinen, "Characterizing the generalized lambda distribution by L-moments," *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 1971–1983, 2008.

- [15] R. A. R. King and H. L. MacGillivray, "A starship estimation method for the generalized  $\lambda$  distributions," *Australian & New Zealand Journal of Statistics*, vol. 41, no. 3, pp. 353–374, 1999.
- [16] A. Lakhany and H. Mausser, "Estimating the parameters of the generalized lambda distribution," *Algo Research Quarterly*, vol. 3, pp. 47–58, 2000.
- [17] G. D. Rayner and H. L. MacGillivray, "Numerical maximum likelihood estimation for the  $g$ -and- $k$  and generalized  $g$ -and- $h$  distributions," *Statistics and Computing*, vol. 12, no. 1, pp. 57–75, 2002.
- [18] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, and R. Wilcox, "Estimating the parameters of a generalized lambda distribution," *Computational Statistics & Data Analysis*, vol. 51, no. 6, pp. 2813–2835, 2007.
- [19] S. Su, "Numerical maximum log likelihood estimation for generalized lambda distributions," *Computational Statistics & Data Analysis*, vol. 51, no. 8, pp. 3983–3998, 2007.
- [20] A. Öztürk and R. F. Dale, "Least squares estimation of the parameters of the generalized lambda distributions," *Technometrics*, vol. 27, no. 1, pp. 81–84, 1985.
- [21] S. N. Mishra, A. K. Shah, and J. J. Lefante Jr., "Overlapping coefficient: the generalized  $t$  approach," *Communications in Statistics: Theory and Methods*, vol. 15, no. 1, pp. 123–128, 1986.
- [22] W. H. Lawton and E. A. Sylvestre, "Elimination of linear parameters in nonlinear regression," *Technometrics*, vol. 13, pp. 461–467, 1971.
- [23] J. J. Filliben, "The probability plot correlation coefficient test for normality," *Technometrics*, vol. 17, no. 1, pp. 111–117, 1975.
- [24] C. R. Rao, "Methodology based on the  $L_1$ -norm, in statistical inference," *Sankhya*, vol. 50, no. 3, pp. 289–313, 1988.
- [25] R. Koenker and V. D'Orey, "Computing regression quantiles," *Applied Statistics*, vol. 36, pp. 383–393, 1987.
- [26] W. L. Price, "Global optimization by controlled random search," *Computer Journal*, vol. 20, pp. 367–370, 1977.
- [27] M. M. Ali, A. Törn, and S. Viitanen, "A numerical comparison of some modified controlled random search algorithms," *Journal of Global Optimization*, vol. 11, no. 4, pp. 377–385, 1997.
- [28] M. M. Ali, C. Storey, and A. Törn, "Application of stochastic global optimization algorithms to practical problems," *Journal of Optimization Theory and Applications*, vol. 95, no. 3, pp. 545–563, 1997.
- [29] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.
- [30] R. Hooke and T. A. Jeeves, "Direct search solution of numerical and statistical problems," *Journal of the Association for Computing Machinery*, vol. 8, pp. 212–229, 1961.
- [31] J. T. Shelton, A. I. Khuri, and J. A. Cornell, "Selecting check points for testing lack of fit in response surface models," *Technometrics*, vol. 25, no. 4, pp. 357–365, 1983.
- [32] I. Křivý and J. Tvrdík, "The controlled random search algorithm in optimizing regression models," *Computational Statistics and Data Analysis*, vol. 20, no. 2, pp. 229–234, 1995.
- [33] P. Brachetti, M. D. F. Ciccoli, G. D. Pillo, and S. Lucidi, "A new version of the Price's algorithm for global optimization," *Journal of Global Optimization*, vol. 10, no. 2, pp. 165–184, 1997.
- [34] H. Faure, "Discrèpançe de suites associées à un système de numération (en dimension  $s$ )," *Acta Arithmetica*, vol. 41, no. 4, pp. 337–351, 1982.
- [35] W. J. Morokoff and R. E. Caflisch, "Quasi-Monte Carlo integration," *Journal of Computational Physics*, vol. 122, no. 2, pp. 218–230, 1995.
- [36] A. Ghosh, "A FORTRAN program for fitting Weibull distribution and generating samples," *Computers & Geosciences*, vol. 25, pp. 729–738, 1999.
- [37] A. Gupta and E. Parzen, "Input modelling using quantile statistical methods," in *Proceedings of the Winter Simulation Conference*, vol. 1, pp. 716–724, 2004.
- [38] R. K. S. Hankin and A. Lee, "A new family of non-negative distributions," *Australian & New Zealand Journal of Statistics*, vol. 48, no. 1, pp. 67–78, 2006.
- [39] S. S. Pal, "Evaluation of nonnormal process capability indices using generalized lambda distribution," *Quality Engineering*, vol. 17, no. 1, pp. 77–85, 2005.