Doping distribution functions for improving data fits

Meitner Cadena, Alejandro Yerovi

Doping distribution functions for improving data fits

Meitner Cadena* & Alejandro Yerovi

April 28, 2020

Abstract

Families of parametric functions $g(x)$ for improving data fits of given distribution functions $F(x)$ via distribution functions $F(g(x))$ are provided. These new distribution functions are called doped distribution functions. They are tailored in function of a given data set by varying $g(x)$. Conditions for guaranteeing that $F(g(x))$ be a distribution function are discussed. The design $F(g(x))$ increases nonlinearities in the likelihood function due to the need of supplementary parameters related to $g(x)$. This makes parameter estimation harder and in some cases, with non-unique solutions. This drawback is managed through the use of optimization procedures without calculation of derivatives when parameters are estimated. In spite of this increase in model complexity, examinations on a number of real data sets and known distribution functions put in evidence advantages on the use of these functions $g(x)$, showing in many cases improvements on data fits. Some of these improvements are remarkable.

Keywords: Improvement data fit, Doped distribution function, Inverse parabolic interpolation, Powell's method

1 Introduction

Data generated in fields present diverse behaviors, with trends related to the field nature as in information technology, health, engineering, finance, insurance and economics among others. Considering statistical analysis, the need for describing such behaviors has continuously motivated the development of a number of distribution functions (dfs) in order to capture particular data behaviors. A df which stands out among them is the normal df. This df is frequently used mainly due to data asymptotic properties under mild conditions when data size increases. Data features have also influenced df designs, for instance the data heaviness when dealing with data behaviors at tails.

*Corresponding author: mncadena2@espe.edu.ec

We review in this paper a number of typical dfs $F(x)$ by doping them through functions $g(x)$ suitably chosen in order to reach new dfs $F(g(x))$ that improve fits to data. The term dope is borrowed from physics where a doped material means a mixture of materials producing a new one with remarkable physical properties. Furthermore, it is noted that the use of families of functions $g(x)$ is an alternative method for generating dfs. In effect, functions $g(x)$ can be found between couples of given dfs $F(x)$ and $G(x)$ such that $G(x) - F(x)$ be a positive increasing function. If $h(x) = G(x) - F(x)$, then we have

$$G(x) = F(x) + h(x) = F(g(x)),$$

where $g(x) = F^{-1}(G(x)) = F^{-1}(F(x) + h(x))$. Note that $g(x)$ is increasing. An example of these functions is when considering the exponential df given by $G(x) = 1 - e^{-\lambda x}$, for $x > 0$ and for some $\lambda > 0$, and the Pareto df given by $F(x) = 1 - \min\{1, x^{-\alpha}\}$, for $x > 0$ and for some $\alpha > 0$. Choosing α such that $\frac{\lambda}{\alpha} > 1$, we have that, for $0 < x < 1$, $0 < 1 - e^{-\lambda x}$ implies $F(x) < G(x)$; whereas for $x \geq 1$, $\frac{\lambda}{\alpha} x > \ln x$ implies $e^{-\lambda x} < x^{-\alpha}$, and then $G(x) > F(x)$ too.

When the support of $F(x)$ is right-unbounded, for guaranteeing that $F(g(x))$ be a df, it is required that the support of $g(x)$ must contain the support of $F(x)$, $g'(x) \geq 0$ and $g(x) \to \infty$ as $x \to \infty$. Hence, assuming $F(x)$ and $g(x)$ are differentiable, the probability density function (pdf) of $F(g(x))$ is $f(g(x))g'(x)$ where $f(x)$ is the pdf of $F(x)$.

On the other hand, when fitting a sample of independent and identically distributed (iid) random variables (rvs) $X_1, \ldots, X_n$ by using a df F, we are interested in assessing if an alternative model $F(g(x))$ may improve that fit to such a sample. This means that the transformed rvs $g(X_1), \ldots, g(X_n)$, which are also iid, may follow the df $F(x)$. In other words, a transformed rv may present better characteristics for being represented it by using F than modeling such a rv without any transformation.

Focusing on dfs with right-unbounded supports, in this paper we present two families of functions $g(x)$, say $g_1(x)$ and $g_2(x)$, and others by combining this couple of functions, i.e. $g_3(x) = cg_1(x) + (1 - c)g_2(x)$ for some positive value $0 < c < 1$. Note that $g_3(x)$ also contains the support of $F(x)$, satisfies the condition $g_3'(x) \geq 0$ and presents the asymptotic behavior $g_3(x) \to \infty$ as $x \to \infty$.

A drawback to our doped dfs is the introduction of more parameters through $g(x)$, which makes the estimation of all parameters harder. In order to circumvent this issue, we propose a combination of optimization methods. Among these methods we take into account ones without calculation of derivatives as that called *inverse parabolic interpolation*, see e.g. [Press et al. (2007)], incorporating it in Powell's method [Powell (1964)].

The families of functions $g(x)$ introduced are applied to a number of real data sets and known dfs. Improvements of data fits fixing data sets or dfs are analyzed. The functions $g(x)$ involved in the doped models that fit best are also examined.

The remainder of this paper is organized as follows. In the next section, proposals for functions $g(x)$ are discussed. Characteristics and properties of such functions are shown. In Section 3 a procedure for estimating the parameters of $F(g(x))$ is described. In Section 4 a number of numerical illustrations are performed. These illustrations involve a variety of distributions frequently presented in literature and a number of real data sets publicly available, coming from different fields. Concluding remarks are presented in the last section.

2 Families of functions $g(x)$

Let $F(x)$ be a df and $g(x)$ be a function. We begin presenting the following result on the conditions for guaranteeing that $F(g(x))$ is a df.

Proposition 2.1. *Let $F(x)$ be a differentiable df and $f(x)$ be its pdf. Let $g(x)$ be a differentiable function satisfying $g'(x) \geq 0$, with support containing the one of F. We have:*

(i) Assume $F(x)$ has right-unbounded support and $x_0 = \inf\{x : F(x) > 0\}$. Assume $g(x) \to \infty$ as $x \to \infty$. Then the function defined by $H(x) = F(g(x))$ for $x > x_0$, and $H(x) = F(g(x_0))$ (≥ 0) for $x = x_0$, is a df with pdf $f(g(x))g'(x)$ for $x > x_0$.

(ii) Assume $supp(F) = \mathbb{R}$. Assume $g(x) \to \infty$ $(-\infty)$ as $x \to \infty$ $(-\infty)$. Then $F(g(x))$ is a df with pdf $f(g(x))g'(x)$.

Proof. (i) $H(x)$ is non-negative, $H'(x) = [F(g(x))]' = f(g(x))g'(x) \geq 0$, for $x > x_0$, and, by using the change of variable $y = g(x)$,

$$\lim_{x\to\infty} F(g(x)) = \lim_{y\to\infty} F(y) = 1.$$

Hence, the claim (i) follows.

(ii) $F(g(x))$ is well-defined, $[F(g(x))]' = f(g(x))g'(x) \geq 0$ and, by using the change of variable $y = g(x)$,

$$\lim_{x\to\infty} F(g(x)) = \lim_{y\to\infty} F(y) = 1.$$

and

$$\lim_{x\to-\infty} F(g(x)) = \lim_{y\to-\infty} F(y) = 0.$$

Hence, the claim (ii) follows. □

Let X be a rv with df $F(x)$.

In what follows we present two families of functions g which satisfy the conditions enunciated in Proposition 2.1. We distinguish these families according to the presence or not of a parameter of $F(x)$ for representing the mean of X. All these functions and other derived from them are used throughout the paper.

2.1 Without a parameter for the mean

In this subsection we assume that $F(x)$ has right-unbounded support and $x_0 = \inf\{x : F(x) > 0\} = 0$

If there is no parameter for representing the mean of X, the following functions $g(x)$ are then considered. For $x > 0$,

$$g_1(x) = x^k \big[\ln(x^2+1)\big]^d \tag{1}$$

$$g_2(x) = \frac{x^k \big[\ln(x^2+1)\big]^d}{\big[x^2+q\big]^p}, \tag{2}$$

where $k \in \mathbb{R} \setminus \{0\}$, $d, p \in \mathbb{R}$ and $q > 0$.

In order to guarantee that $g_1(x)$ and $g_2(x)$ be well behaved when $x \to 0^+$, we need that the condition indicated in the following result be satisfied.

Proposition 2.2. *Let the functions $g_1(x)$ and $g_2(x)$ be defined by (1) and (2), respectively. If $k + 2d > 0$, then $g_1(x) \to 0^+$ and $g_2(x) \to 0^+$ as $x \to 0^+$.*

Proof. Consider the function $g_1(x)$. We have, by applying the L'Hopital rule,

$$\begin{aligned} \lim_{x\to 0^+} g_1(x) &= \lim_{x\to 0^+} x^k \big[\ln(x^2+1)\big]^d \\ &= \lim_{x\to 0^+} x^{k+2d} \\ &= 0, \end{aligned}$$

because of the hypothesis. The claim for $g_2(x)$ also follows because the previous result and $g_2(x)/g_1(x) \to 1/q^p$ as $x \to 0^+$. □

As mentioned above, another function $g(x)$ to be considered is $g_3(x) = cg_1(x) + (1-c)g_2(x)$ for some positive value $0 < c < 1$. Thus, the condition provided in Proposition 2.2 also applies to this mixture of functions.

Note that if $k = 1$ and $d = 0$, then $F(g_1(x)) = F(x)$, and when moreover $q = 1$ and $p = 0$, then $F(g_2(x)) = F(x)$. Thus, $F(g_1(x))$ and $F(g_2(x))$ are generalizations of $F(x)$. Further, $g_2(x)$ is a generalization of $g_1(x)$, i.e. it would be enough to consider $g_2(x)$. However, in practical terms, given $g_2(x)$, the additional conditions $q = 1$ and $p = 0$ that $g_1(x)$ has, may allow it the provision of different behaviors to the ones given by $g_2(x)$. This feature of $g_1(x)$ will be exploited when considering $g_3(x)$ because $g_1(x)$ and $g_2(x)$ may intervene in a complementary way.

In order to have that $F(g_1(x))$ and $F(g_2(x))$ be dfs, from Proposition 2.1 we see that the derivatives of $g_1(x)$ and $g_2(x)$ need to be non-negative. These derivatives are

$$
\begin{aligned}
g_1'(x) &= \frac{x^{k-1}\big[\ln(x^2+1)\big]^{d-1}}{x^2+1} \times \big(k(x^2+1)\ln(x^2+1)+2dx^2\big) \\
g_2'(x) &= \frac{x^{k-1}\big[\ln(x^2+1)\big]^{d-1}}{(x^2+1)\big[x^2+q\big]^{p+1}} \\
&\quad \times \big((x^2+1)\ln(x^2+1)\left[k(x^2+q)-2px^2\right]+2dx^2(x^2+q)\big).
\end{aligned}
$$

Note that the first factors on the right side of both previous equalities are positive for $x > 0$. Thus it remains to guarantee that the corresponding second factors are positive for $x > 0$. We denote these second factors as follows. Taking $y = x^2$, for $y > 0$,

$$
\begin{aligned}
r_1(y) &= k(y+1)\ln(y+1)+2dy \\
r_2(y) &= (y+1)\ln(y+1)\left[k(y+q)-2py\right]+2dy(y+q).
\end{aligned}
$$

In order to find conditions for guaranteeing that $r_1(y)$ and $r_1(y)$ are positive for $y > 0$, we go to use the following well-known result obtained by application of Lagrange's Mean Value Theorem.

Lemma 2.3. *Let $u(x)$ and $v(x)$ be differentiable functions such that $u(a) \leq v(a)$ and $u'(x) < v'(x)$ for $x \geq a$. Then $u(x) < v(x)$ for $x > a$.*

Note that $r_1(0) = 0$ and $r_2(0) = 0$. In order to apply Lemma 2.3, we then need to prove that the derivatives of $r_1(y)$ and $r_2(y)$ are positive for $y > 0$. These derivatives are:

$$
\begin{aligned}
r_1'(y) &= k\ln(y+1)+k+2d \\
r_2'(y) &= \ln(y+1)\left[kq+k-2p+2y(k-2p)\right]+(k-2p+4d)y+q(k+2d).
\end{aligned}
$$

For having $r_1'(y) > 0$ for $y > 0$, it is thus required either $k = 0$ and $d > 0$, or $k > 0$ and $d \geq -k/2$. With respect to $r_2'(y)$, we go to again apply Lemma 2.3. Note that $r_2'(0) = q(k+2d)$, implying that it is needed to have $k > -2d$. We then need to prove that the derivative of $r_2'(y)$ is positive for $y > 0$. This derivative is:

$$
r_2^{(2)}(y) = 3(k-2p)+4d+\frac{kq-k+2p}{y+1}+2(k-2p)\ln(y+1).
$$

For having $r_2^{(2)}(y) > 0$ for $y > 0$, it is thus required $k \geq 2p$ and $k > 4(p-d)/(2+q)$.

We have obtained the following result.

Proposition 2.4. *Let $g_1(x)$ and $g_2(x)$ be the functions defined by (1) and (2), respectively. We have:*

(i) If either $k = 0$ and $d > 0$, or $k > 0$ and $d \geq -1/2$, then $g_1'(x) > 0$ for $x > 0$.

(ii) If $k > \max\{-2d, 2p, 4(p-d)/(q+2)\}$, then $g_2'(x) > 0$ for $x > 0$.

Remarks 2.5.

(i) According to Proposition 2.4(i), k is at least 0. However, the condition $k > \max\{-2d, 2p, 4(p-d)/(q+2)\}$ given by Proposition 2.4(ii) allows the parameter k may take negative values, for instance if $p < 0$ or $p < d$.

(ii) According to Proposition 2.4:

(a) For $g_1'(x)$, if either $0 \leq k < 1$, or $k = 1$ and $d < 0$, $g_1'(x) \to \infty$ as $x \to 0^+$; if $k = 1$ and $d = 0$, $g_1'(x) \to 1$ as $x \to 0^+$; and, if either $k = 0$ and $d > 0$, or $k > 1$, $g_1'(x) \to 0$ as $x \to 0^+$.

(b) For $g_2'(x)$, if either $k < 1$, or $k = 1$ and $d < 0$, $g_1'(x) \to \infty$ as $x \to 0^+$; if $k = 1$ and $d = 0$, $g_1'(x) \to q$ as $x \to 0^+$; and, if either $k = 0$ and $d > 0$, or $k > 1$, $g_1'(x) \to 0$ as $x \to 0^+$.

For having that $g_1(x)$ and $g_2(x)$ satisfy all hypothesis in Proposition 2.1, we also need the following result.

Proposition 2.6. *Let $g_1(x)$ and $g_2(x)$ be the functions defined by (1) and (2), respectively. We have:*

(i) If $k > 0$, then $g_1(x) \to \infty$ as $x \to \infty$.

(ii) If $k > 2p$, then $g_2(x) \to \infty$ as $x \to \infty$.

Proof. Let us begin proving (ii). Assume $k > 2p$.

If $d = 0$, then

$$\begin{aligned} \lim_{x\to\infty} g_2(x) &= \lim_{x\to\infty} \frac{x^k}{[x^2+q]^p} \\ &= \lim_{x\to\infty} \frac{x^{k-2p}}{[1+qx^{-2}]^p} \\ &= \infty. \end{aligned}$$

Now assume $d \neq 0$. Let $\epsilon = (k-2p)/2$ (> 0). Writing

$$\begin{aligned} g_2(x) &= \frac{x^k \big[\ln(x^2+1)\big]^d}{\big[x^2+q\big]^p} \\ &= \big[x^{\epsilon/d} \ln(x^2+1)\big]^d \times \frac{x^{k-\epsilon}}{\big[x^2+q\big]^p}, \end{aligned}$$

we have, on the one hand, that, if $d > 0$, $\big[x^{\epsilon/d}\ln(x^2+1)\big]^d \to \infty$ as $x \to \infty$, and if $d < 0$, by taking $y = x^2$ and then applying the L'Hopital rule,

$$\begin{aligned}\lim_{x\to\infty} x^{\epsilon/d}\ln(x^2+1) &= \lim_{y\to\infty} \frac{\ln(y+1)}{y^{-\epsilon/(2d)}} \\ &= \lim_{y\to\infty} \frac{1/(y+1)}{-\frac{\epsilon}{2d}y^{-\epsilon/(2d)-1}} \\ &= -\frac{2d}{\epsilon}\lim_{y\to\infty}\frac{y^{\epsilon/(2d)+1}}{y+1} \\ &= 0,\end{aligned}$$

implying $\big[x^{\epsilon/d}\ln(x^2+1)\big]^d = 1\big/\big[x^{\epsilon/d}\ln(x^2+1)\big]^{-d} \to \infty$ as $x \to \infty$. On the other hand, noting that $k - \epsilon - 2p = (k-2p)/2 > 0$, arguments as above allow the deduction that

$$\lim_{x\to\infty}\frac{x^{k-\epsilon}}{\big[x^2+q\big]^p} = \infty.$$

Thus we have $g_2(x) \to \infty$ as $x \to \infty$, and the proposition follows.

Since $g_1(x)$ is obtained from $g_2(x)$ by taking $p = 0$, then (i) follows from (ii). □

By Propositions 2.1(i), 2.2, 2.4 and 2.6, we obtain the following result.

Proposition 2.7. *Let $g_1(x)$ and $g_2(x)$ be the functions defined by (1) and (2), respectively. We have:*

(i) If $k > \max\{0, -2d\}$ and $d \geq -1/2$, then $F(g_1(x))$ is a df.

(ii) If $k > \max\{-2d, 2p, 4(p-d)/(q+2)\}$, then $F(g_2(x))$ is a df.

(iii) Let $g_3(x) = cg_1(x)+(1-c)g_2(x)$ for some $0 < c < 1$. If $k_1 > \max\{0, -2d_1\}$, $d_1 \geq -1/2$ and $k_2 > \max\{-2d_2, 2p, 4(p-d_2)/(q+2)\}$, where k_i and d_i are the parameters k and d, respectively, related to $g_i(x)$, $i \in \{1,2\}$, then $F(g_3(x))$ is a df.

2.2 With a parameter for the mean

In this subsection we assume that the support of $F(x)$ is $\mathbb{R}$.

If there is a parameter μ for the mean of the df analyzed $F(x)$, then x in (1) and (2) is replaced by $x - \mu$ and $k = 1$ is taken. Under these changes, $g_1(x)$ and $g_2(x)$ are expressed as follows. For $x \in \mathbb{R}$,

$$g_{\mu,1}(x) = (x-\mu)\big[\ln((x-\mu)^2+1)\big]^d \tag{3}$$

$$g_{\mu,2}(x) = \frac{(x-\mu)\big[\ln((x-\mu)^2+1)\big]^d}{\big[(x-\mu)^2+q\big]^p}. \tag{4}$$

In order to have that $F(g_{\mu,1}(x))$ and $F(g_{\mu,2}(x))$ be dfs, according to Proposition 2.1 we need that the derivatives of $g_{\mu,1}(x)$ and $g_{\mu,2}(x)$ be positive. These derivatives are, respectively, using the change of variable $y = (x-\mu)^2$ (≥ 0),

$$\begin{aligned} h_1'(y) &= \frac{\big[\ln(y+1)\big]^{d-1}}{y+1} \times \{(y+1)\ln(y+1) + 2dy\} \\ h_2'(y) &= \frac{\big[\ln(y+1)\big]^{d-1}}{\big[y+1\big]\big[y+q\big]^{p+1}} \\ &\quad \times \big\{(y+1)\ln(y+1)\big[(1-2p)y+q\big] + 2dy(y+q)\big\}. \end{aligned}$$

Note that the first factors on the right side of both previous equalities are positive for $y > 0$. Hence it then remains to guarantee that the corresponding second factors are positive for $y > 0$. We denote these second factors as follows. For $y \geq 0$,

$$\begin{aligned} r_1(y) &= (y+1)\ln(y+1) + 2dy \\ r_2(y) &= (y+1)\ln(y+1)\big[(1-2p)y+q\big] + 2dy(y+q). \end{aligned}$$

Note that $r_1(0) = 0$ and $r_2(0) = 0$. In order to apply Lemma 2.3, we then need to prove that the derivatives of $r_1(y)$ and $r_2(y)$ are positive for $y > 0$. These derivatives are:

$$\begin{aligned} r_1'(y) &= \ln(y+1) + 1 + 2d \\ r_2'(y) &= \big[2(1-2p)y + 1 - 2p + q\big]\ln(y+1) + (1-2p+4d)y + q(1+2d). \end{aligned}$$

It is thus enough to take $d \geq -1/2$ for having $r_1'(y) > 0$. Moreover to this condition, taking $p \leq 1/2$ and $d \geq (1-2p)/4$, we guarantee $r_2'(y) > 0$ for $y > 0$. We have then proved the following result.

Proposition 2.8. *Let $g_{\mu,1}(x)$ and $g_{\mu,2}(x)$ be the functions defined by (3) and (4), respectively. We have:*

(i) If $d > -1/2$, then $g_{\mu,1}'(x) > 0$ for $x \in \mathbb{R} \setminus \{\mu\}$.

(ii) If $p \leq 1/2$ and $d \geq (1-2p)/4$, then $g_{\mu,2}'(x) > 0$ for $x \in \mathbb{R} \setminus \{\mu\}$.

Remarks 2.9. *For $i \in \{1,2\}$: $g_{\mu,i}(\mu) = 0$ and $g_{\mu,i}(x)$ is continuous at $x = \mu$, but $g_{\mu,i}'(x)$ is not continuous at $x = \mu$ when $d < 0$.*

For having that $g_{\mu,1}(x)$ and $g_{\mu,2}(x)$ satisfy all hypothesis in Proposition 2.1, we also need the following result.

Proposition 2.10. *Let $g_{\mu,1}(x)$ and $g_{\mu,2}(x)$ be the functions defined by (3) and (4), respectively. We have:*

(i) $g_{\mu,1}(x) \to \infty$ $(-\infty)$ *as* $x \to \infty$ $(-\infty)$.

(ii) $g_{\mu,2}(x) \to \infty$ $(-\infty)$ *as* $x \to \infty$ $(-\infty)$, *if* $p = 1/2$ *and* $d > 0$, *or* $p < 1/2$; $g_{\mu,2}(x) \to 1$ (-1) *as* $x \to \infty$ $(-\infty)$, *if* $p = 1/2$ *and* $d = 0$.

Proof. (i) If $d = 0$, the claim is immediate. Assume $d < 0$. Taking $y = x - \mu$ and then applying the L'Hopital rule, give

$$\begin{aligned}
\lim_{x\to\infty} g_{\mu,1}(x) &= \lim_{y\to\infty} \left[\frac{y^{-1/d}}{\ln(y^2+1)}\right]^{-d} \\
&= \lim_{y\to\infty} \left[-\frac{1}{2d} y^{-1/d-2}(y^2+1)\right]^{-d} \\
&= \lim_{y\to\infty} \left[-\frac{1}{2d} y^{-1/d}\left(1+\frac{1}{y^2}\right)\right]^{-d} \\
&= \infty.
\end{aligned}$$

and we then have, by taking $y = -(x-\mu)$ and then applying the previous result,

$$\begin{aligned}
\lim_{x\to-\infty} g_{\mu,1}(x) &= -\lim_{y\to\infty} y\big[\ln(y^2+1)\big]^d \\
&= -\infty.
\end{aligned}$$

(ii) Assume $p = 1/2$ and $d > 0$. Writing

$$\frac{(x-\mu)\big[\ln((x-\mu)^2+1)\big]^d}{\big[(x-\mu)^2+q\big]^{1/2}} = \frac{(x-\mu)}{|x-\mu|\left[1+\frac{q}{(x-\mu)^2}\right]^{1/2}}\big[\ln((x-\mu)^2+1)\big]^d,$$

the claim follows immediately.

Assume $p < 1/2$. Writing, for $x > \mu$,

$$\frac{(x-\mu)\big[\ln((x-\mu)^2+1)\big]^d}{\big[(x-\mu)^2+q\big]^{p}} = (x-\mu)^{1-2p}\big[\ln((x-\mu)^2+1)\big]^d \times \frac{1}{\left[1+\frac{q}{(x-\mu)^2}\right]^{p}},$$

applying arguments as tose for proving (i) the claim follows, and, for $x < \mu$, taking the change of variable $y = -(x-\mu)$,

$$\frac{(x-\mu)\big[\ln((x-\mu)^2+1)\big]^d}{\big[(x-\mu)^2+q\big]^{p}} = -\frac{y\big[\ln(y^2+1)\big]^d}{\big[y^2+q\big]^{p}},$$

the claim follows by the previous result.

Now assume $p = 1/2$ and $d = 0$. Writing

$$\frac{(x-\mu)}{\big[(x-\mu)^2+q\big]^{1/2}} = \frac{(x-\mu)}{|x-\mu|\left[1+\frac{q}{(x-\mu)^2}\right]^{1/2}},$$

the claim follows immediately.

□

By Propositions 2.1(ii), 2.8 and 2.10, the following result is obtained.

Proposition 2.11. *Let $g_{\mu,1}(x)$ and $g_{\mu,2}(x)$ be the functions defined by (3) and (4), respectively. We have:*

(i) If $d > -1/2$, then $F(g_{\mu,1}(x))$ is a df.

(ii) If $p = 1/2$ and $d > 0$, or $p < 1/2$ and $d \geq (1-2p)/4$, then $F(g_{\mu,2}(x))$ is a df.

(iii) Let $g_{\mu,3}(x) = cg_{\mu,1}(x) + (1-c)g_{\mu,2}(x)$ for some $0 < c < 1$. If $d_1 > -1/2$, $p \leq 1/2$ and $d_2 > (1-2p)/4$, where d_i is the parameter d related to $g_{\mu,i}(x)$, $i \in \{1, 2\}$, then $F(g_{\mu,3}(x))$ is a df.

Remark 2.12. *Note that μ as parameter of $F(x)$ may vary when considering $F(g_{\mu,1}(x))$ and $F(g_{\mu,2}(x))$, i.e. we would have $F(g_{\mu_1,1}(x))$ and $F(g_{\mu_2,2}(x))$ where μ_1 and μ_2 may be different. Thus, in what follows, we consider the following two models for $g_{\mu,3}(x)$:*

(i) $g_{\mu,3a}(x) = cg_{\mu,1}(x) + (1-c)g_{\mu,2}(x)$ for some $0 < c < 1$, when assumed that $\mu_1 = \mu_2 = \mu$.

(ii) $g_{\mu,3b}(x) = cg_{\mu_1,1}(x) + (1-c)g_{\mu_2,2}(x)$ for some $0 < c < 1$, when assumed that $\mu_1 \neq \mu_2$.

3 Method for estimating parameters

For estimating the parameters of $F(g(x))$ we propose the maximum likelihood method. Such estimators are thus obtained by maximizing the log-likelihood function, defined by

$$\ell = \ell(\theta_1, \ldots, \theta_r | x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \left[F(g(x_i))\right]',$$

where $x_1, \ldots, x_n$ are outcomes of a sample of iid rvs $X_1, \ldots, X_n$ each following the df $F(g(x))$.

The proposed doped dfs $F(g(x))$ require 1, 2 or 3 parameters more according to the functions $g(x)$ defined in (1), (2), (3) or (4) respectively. If one focuses on $g_3(x)$, i.e. a combination of $g_1(x)$ and $g_2(x)$, then such numbers of parameters increase in 7. In the case of the functions $g_{\mu,3a}(x)$ and $g_{\mu,3b}(x)$, its parameters increase in 7 and 6 respectively. These supplementary parameters may complicate computations of all their estimates due to new complex non-linearities introduced from the new df designs. This issue limits the application of typical methods for estimating parameters, as the well-known Newton method or any of its variants.

In order to circumvent this limitation, we propose first the application of typical methods for estimating $F(x)$ since such a df is well-known. Next, using the previous parameter estimates as initial values for the model $F(g(x))$, we

apply methods without calculation of derivatives as Nelder and Mead's method [Nelder and Mead (1965)] or the one called inverse parabolic interpolation, see e.g. [Press et al. (2007)], incorporating it in Powell's method [Powell (1964)]. In this way, data fits of the new models start from the ones given by data fits using well-known models. This procedure thus guarantees that the new models fit data as well as at least the well-known models.

Next, considering the following methods without calculation of derivatives. On the one hand, Nelder and Mead's method (NMM), which is a heuristic optimization technique for searching local optima in multidimensional non-convex functions. Briefly, this method evaluates the function in several points in order to decide another one for addressing a better estimate. On the other hand, inverse parabolic interpolations (IPI) which are unidimensional approximations to the function in a given direction, from which a better estimate can be determined. This unidimensional procedure implemented in the direction set built from Powell's method, provides a technique for also finding multidimensional local optima. In practice, we find that NMM may eventually produce infeasible solutions since it works for unconstrained optimization. This issue does not occur with the second technique proposed since it always allows the control on undesirable solutions in order to avoid them. However, this second technique may be very time-consuming due to very slow convergence.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_F, \boldsymbol{\theta}_g)$ where $\boldsymbol{\theta}_F = (\theta_{F,1}, \ldots, \theta_{F,r_F})$ with $\theta_{F,i}$, $i = 1, \ldots, r_F$, are the parameters to be estimated related to $F(x)$, and $\boldsymbol{\theta}_g = (\theta_{g,1}, \ldots, \theta_{g,r_g})$ with $\theta_{g,i}$, $i = 1, \ldots, r_g$, are the parameters to be estimated related to $g(x)$. As stopping criteria of Algorithm A we consider in step $k+1$ some of the following two conditions: (1) $\sum_{i=1}^{r} \left|\theta_{k+1,i} - \theta_{k,i}\right| < 10^{-5}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$ are the concerned parameters with $r = r_F + r_g$, or (2) $k + 1 > 10^4$. Note that this algorithm may be trapped at local maxima and consequently would fail to reach global maxima.

The previous description of the proposed iterative method for estimating $\boldsymbol{\theta}$, is consolidated in the instructions given in Algorithm 1.

It follows an immediate property of Algorithm 1.

Proposition 3.1. *Let $F(x)$ be a df. Let $g(x)$ be a function as defined in Propositions 2.7 or 2.11. Let $X_1, \ldots, X_n$ be a sample of iid rvs following the distribution function $F(g(x))$. Assume Algorithm 1. We have that the sequence $\big(\ell(\theta_m)\big)_{m\in\{0,1,\ldots\}}$ is non-decreasing.*

The following result provides conditions for proving the existence of estimators for the parameters of $F(g(x))$.

Proposition 3.2. *Let $F(x)$ be a df. Let $g(x)$ be a function as defined in Propositions 2.7 or 2.11. Let $X_1, \ldots, X_n$ be a sample of iid rvs following the df $F(g(x))$. We have that if there exists maximum likelihood estimators for the parameters of $F(x)$, then there exists maximum likelihood estimators for the parameters of $F(g(x))$.*

Proof. Let h be any strictly monotonic function defined on the range of X_1. The rvs $h(X_1), \ldots, h(X_n)$ are thus iid and follow the df $F(x)$. This fact

Algorithm 1 Calculation of parameter estimates

Set a df F
Set a function g as defined in Propositions 2.7 or 2.11
Set $x_1, \ldots, x_m$ outcomes of iid rvs $X_1, \ldots, X_m$ following $F \circ g$
Compute $\widehat{\boldsymbol{\theta}}_{F,0}$ by fitting F to data x
Initialize $\widehat{\boldsymbol{\theta}}_{g,0}$ such that $g(x) = x$
Define $\widehat{\boldsymbol{\theta}}_0 = (\widehat{\boldsymbol{\theta}}_{F,0}, \widehat{\boldsymbol{\theta}}_{g,0})$
Set $m = 0$
while $\sum_{i=1}^{r} \left|\widehat{\theta}_{m,i} - \widehat{\theta}_{m-1,i}\right| \geq 10^{-5}$ if $m > 0$, and $m \leq 10^4$ **do**
    $m \leftarrow m + 1$
    Compute $\widetilde{\boldsymbol{\theta}}_m$ from $\widehat{\boldsymbol{\theta}}_{m-1}$ by applying NMM or IPI to ℓ
    if $\ell(\widetilde{\boldsymbol{\theta}}) > \ell(\widehat{\boldsymbol{\theta}}_{m-1})$ **then**
        $\widehat{\boldsymbol{\theta}}_m \leftarrow \widetilde{\boldsymbol{\theta}}_m$
    else
        $\widehat{\boldsymbol{\theta}}_m \leftarrow \widehat{\boldsymbol{\theta}}_{m-1}$
    end if
end while
return $\widehat{\boldsymbol{\theta}}_m$

and by hypothesis, imply that there exists a maximum likelihood estimator for $\boldsymbol{\theta}_F \big| h(X_1), \ldots, h(X_n)$. Then, the equality

$$\ell(\boldsymbol{\theta}_{F\circ h} \big| X_1, \ldots, X_n) = \ell(\boldsymbol{\theta}_F \big| h(X_1), \ldots, h(X_n)),$$

implies that $F \circ h$ has a maximum likelihood estimator for $\boldsymbol{\theta}_{F\circ h} \big| X_1, \ldots, X_n$. This result applies in particular to $F \circ g$. □

4 Numerical applications

In this section, we present a number of illustrations in order to assess the performance of the introduced doped models with respect to well-known models. To this aim, we consider the following right-unbounded models frequently used in practice (acronyms for their representations in this paper are indicated between brackets): normal (N), gamma (G), lognormal (LN), inverse-normal (IN), Weibull (W), chi-squared (CS), t-Student (T), Birnbaum-Saunders (BS) and Laplace (L) distributions. Other dfs of interest are also considered, they are: generalized lambda (GL), skew-normal (SN) and skew-t (ST) distributions. For each of these models, their fits to several real data are performed. Such fits are made using undoped and doped models. In this way, we can examine the impact of doping on the models indicated above. Note that the performance of the doped models always is at least as good as that of the models without doping because the method for estimating parameters that is used takes as initial parameters those of the undoped model.

Table 1: Data sets to be analyzed

Data (acronym)	Description	Source	Number of observations
DS_1	Strengths of 1.cm fibre	[Smith and Naylor (1987)]	63
DS_2	Study on self-awareness	[Dana (1990)]	19
DS_3	Human heart rate (beats per minute)	[Mackowiak et al. (1992)]	130
DS_4 (‡)	Monthly estimates of Carbon Dioxide emissions from the U.S.A.	[Blasing et al. (2004)]	276
DS_5	Alkaline Phosphatase International Units/Liter	[Boyd et al. (1998)]	276
DS_6	Calcium mmol/L	[Boyd et al. (1998)]	$275^{\dagger}$
DS_7	Inorganic Phosphorus mmol/L	[Boyd et al. (1998)]	$275^{\dagger}$
DS_8	Weight of the euro coin in grams	[Shkedy et al. (2006)]	2000
DS_9	Father heights in inches (with random noise added)	[Pearson and Lee (1903)]	1078
DS_{10}	Son heights in inches (with random noise added)	[Pearson and Lee (1903)]	1078
DS_{11}	Length of thorax of male fruitflies, in mm	See e.g. [Hanley and Shapiro (1994)]	125
DS_{12}	Transparency value	[Su (2007)]	176
DS_{13}	Lifetimes in hours of lamps	[Davis (1952)]	417
DS_{14}	Diastolic blood pressure in mmHg (diabetic patients)	[Lee and Wang (2003)]	149
DS_{15}	Systolic blood pressure in mmHg (diabetic patients)	[Lee and Wang (2003)]	149
DS_{16}	Age at diagnosis (diabetic patients)	[Lee and Wang (2003)]	149
DS_{17}	Body mass index (diabetic patients)	[Lee and Wang (2003)]	149
DS_{18}	Lifetimes of Strips of Aluminum Coupon	[Birnbaum and Saunders (1958)]	101
DS_{19}	Concentration of PCB in the yolk lipids of pelican eggs	[Risebrough (1972)]	65
DS_{20}	Survival time in days of infected guinea pigs	[Bjerkedal (1960)]	72
DS_{21}	Times between successive failures of air condition equipment	[Bhaumik et al. (2009)]	30
DS_{22}	Maximum flood level for a four year period	[Dumonceaux and Antle (1973)]	20
DS_{23}	Cadmium concentration (μg/g) in horse kidneys	[Elinder et al. (1981)]	69
DS_{24}	Over-all heights of fragmentation bomb bases	[Duncan (1986)]	145
DS_{25}	Life of fatigue fracture of Kevlar 373/epoxy	[Alizadeh et al. (2017)]	76
DS_{26}	Breaking stress of carbon fibers of 50 mm length (GPa)	[Nichols and Padgett (2006)]	100
DS_{27}	Maximum stress per cycle 31,000 psi	[Birnbaum and Saunders (1969)]	101
DS_{28}	Survival times (weeks) of patients with acute myelogenous leukemia	[Feigl and Zelen (1965)]	17
DS_{29}	Annual maxima of river flows (m^3/s)	[Shao et al. (2004)]	98
DS_{30}	Floyd river flood rates (ft^3/s)	[Mudholkar and Huston (1996)]	39

† One existing missing value is excluded
‡ All original data have been multiplied by -1

All data sets to be fitted are real and publicly available. They are described in Table 1, indicating their source and their number of observations. Acronyms for representing these data sets in this paper are assigned.

As measures for assessing performance we use the parsimony criteria AIC (acronym of Akaike's information criterion) and BIC (acronym of Bayesian information criterion). Minimum of these values are usually used as model selection criteria because lower values for AIC and BIC indicate better model fits. These values are expressed by, see e.g. [Hastie et al. (2008)],

$$\begin{aligned} \text{AIC} &: 2 \times NLL + 2 \times n_p \\ \text{BIC} &: 2 \times NLL + n_p \times \ln n, \end{aligned}$$

where NLL is the negative log-likelihood of the model fitted to a given data set, n is the data size and n_p is the parameter number of that model. Since the BIC value involves the data size, it penalizes the model more for its complexity than the AIC value. This means that under BIC values, more complex models may be discarded due to larger scores, except that such complexities provide remarkable improvements in model fits. On the other hand, when considering AIC values, the data size does not influence the selection of models.

Table 2 presents for each analyzed data set the models selected among the models examined, according to the lowest AIC or BIC. These outputs refer only to the best fits among all considered models, distinguished between undoped and doped. Doped models are written as UM(g) where UM is an undoped model and $g \in \{g_1, g_2, g_3, g_{3a}, g_{3b}\}$ with g_3 as defined in Proposition 2.7(iii) and

g_{3a} and g_{3b} as defined in Remark 2.12. Corresponding parsimony values are indicated between brackets and the lowest ones are highlighted. These results show that in most cases the doped models perform better than the undoped ones. As noted through almost all cases, this improvement is promoted not only by the use of doped models, but also by the use of other dfs $G(x)$ different to the ones used as undoped models $F(x)$. In other words, given a data set and a df $F(x)$ for fitting those data, a model fitting better than $F(x)$ may be $G(g(x))$ where the model $G(x)$ may be different from $F(x)$. On the other hand, it is notorious that many doped models are based on the most complex functions of $g(x)$, i.e. $g_3(x)$, $g_{3a}(x)$ $g_{3b}(x)$. This feature is remarkable with respect to the following two aspects. First, it shows that better fits may be obtained in spite of the additional complexity introduced in the undoped models. Second, the functions $g_1(x)$ and $g_2(x)$ may become to be related to each other in such a way that this integration provide models with better performances than the ones based on $g_1(x)$ and $g_2(x)$ separately.

On the other hand, the results presented in Table 2 put in evidence the effect of the use of AIC or BIC. The number of doped models with better performance is low for BIC than for AIC since the former criterion is much more penalizing.

The results obtained show that for 25 from 30 data sets (83.3 %), their fits are improved by using doped models, when considering AIC values; whereas for 24 from 30 data sets (80.0 %) when considering BIC values.

In this paper, functions $g(x)$ are the key ingredients for building doped models of interest. Figure 1 presents plots of these functions, which correspond to the best doped models described in Table 2. For each data set analyzed, plots of the graphs of the functions $g(x)$ associated to values AIC (colored black) and BIC (colored brown) are shown. In many cases this couple of functions are coincident, and only a few ones differences between them are found, as for DS_1, DS_4, DS_5, DS_{13}, DS_{16}, DS_{17} and DS_{18}. On the other hand, these functions may take particular behaviors as those for DS_{13} and DS_{18}, all of them being tailored for allowing improvements on data fits.

Next, Figure 2 shows plots of data fits using the models $F(x)$ (solid lines) and $F(g(x))$ (dashed lines) indicated in Table 2. The cases AIC (colored black) and BIC (colored brown) both are included in each plot. These plots illustrate how well the alternative models may improve data fits. In the cases where improvements on data fits are obtained, differences between doped and undoped models may be not so remarkable as for DS_{27}, and notorious as for DS_4, DS_{18} and DS_{24}. Among all plots, it is noted the skill of doped models for producing multimodal dfs, which may be eventually required.

Each analyzed model $F(x)$ can also be examined through all data sets considered, by taking or not $g(x)$ into account. In this way, one has an empirical assessment on eventual improvements of $F(x)$ when considering one of the alternative functions $g(x)$ examined. Plots for each model and distinguished by the AIC and BIC values are represented in Figures 3 and 4, respectively. In these plots, each point is associated with a data set. A diagonal line is incorporated in each plot in order to identify the cases under such a line, which indicates that a doped model fit better than the undoped models for a given data set and a

Table 2: The best AIC and BIC values associated to data and models

Data	AIC				BIC			
set	Undoped model		Doped model		Undoped model		Doped model	
DS_1	ST	(31.6)	ST(g_1)	(**29.7**)	L	(36.2)	L(g_1)	(**36.2**)
DS_2	GL	(265.9)	T(g_1)	(**261.3**)	IG	(268.6)	T(g_1)	(**265.0**)
DS_3	N	(880.2)	L(g_{3b})	(**767.9**)	N	(885.9)	L(g_{3b})	(**790.9**)
DS_4	GL	(545.5)	L(g_{3b})	(**516.5**)	GL	(560.0)	SN(g_1)	(**541.5**)
DS_5	IG	(1684.2)	L(g_{3b})	(**1617.3**)	IG	(1690.6)	L(g_1)	(**1636.0**)
DS_6	N	(−200.7)	L(g_{3a})	(**−336.2**)	N	(−194.4)	L(g_{3a})	(**−314.0**)
DS_7	N	(−107.3)	L(g_{3a})	(**−232.6**)	N	(−101.0)	L(g_{3a})	(**−210.4**)
DS_8	ST	(−7869.9)	ST(g_1)	(**−7881.6**)	T	(−7849.9)	T(g_1)	(**−7855.0**)
DS_9	N	(5239.2)	N(g_2)	(**5239.2**)	N	(5249.2)	N(g_2)	(**5249.2**)
DS_{10}	T	(5285.3)	T(g_2)	(**5285.3**)	T	(5300.3)	T(g_2)	(**5300.3**)
DS_{11}	GL	(−306.6)	L(g_{3a})	(**−772.8**)	SN	(−296.9)	L(g_{3a})	(**−753.0**)
DS_{12}	GL	(−582.7)	G(g_3)	(**−3938.6**)	GL	(−570.0)	G(g_3)	(**−3910.1**)
DS_{13}	T	(5558.0)	L(g_{3b})	(**5529.4**)	T	(5570.1)	L(g_{3a})	(**5559.2**)
DS_{14}	N	(1136.0)	L(g_{3a})	(**887.0**)	N	(1142.0)	L(g_{3a})	(**908.0**)
DS_{15}	LN	(1311.8)	T(g_{3b})	(**1254.4**)	LN	(1317.8)	T(g_{3b})	(**1278.5**)
DS_{16}	IG	(1129.9)	SN(g_{3a})	(**1031.5**)	IG	(1135.9)	L(g_{3a})	(**1052.3**)
DS_{17}	LN	(981.5)	SN(g_1)	(**976.2**)	LN	(**987.5**)	LN(g_1)	(**987.5**)
DS_{18}	N	(1495.5)	L(g_{3b})	(**1489.3**)	N	(1500.7)	N(g_2)	(**1500.7**)
DS_{19}	L	(740.1)	L(g_1)	(**726.8**)	L	(744.4)	L(g_1)	(**733.4**)
DS_{20}	G	(**855.6**)	G(g_1)	(**855.6**)	G	(**860.2**)	G(g_1)	(**860.2**)
DS_{21}	LN	(**307.2**)	LN(g_1)	(**307.2**)	LN	(**310.0**)	LN(g_1)	(**310.0**)
DS_{22}	SN	(−29.4)	SN(g_1)	(**−29.4**)	SN	(−26.4)	SN(g_1)	(**−26.4**)
DS_{23}	G	(673.3)	SN(g_{3a})	(**664.9**)	G	(677.8)	SN(g_1)	(**675.3**)
DS_{24}	L	(−1094.5)	L(g_{3b})	(**−1189.6**)	L	(−1088.5)	L(g_{3b})	(**−1165.8**)
DS_{25}	G	(248.5)	ST(g_1)	(**241.0**)	G	(253.2)	ST(g_1)	(**252.6**)
DS_{26}	W	(**287.1**)	W(g_1)	(**287.1**)	W	(**292.3**)	W(g_1)	(**292.3**)
DS_{27}	G	(916.8)	L(g_{3a})	(**855.5**)	G	(922.0)	L(g_{3a})	(**873.8**)
DS_{28}	G	(177.7)	L(g_{3a})	(**171.0**)	G	(179.4)	L(g_{3a})	(**176.8**)
DS_{29}	LN	(**25.0**)	LN(g_1)	(**25.0**)	LN	(**30.2**)	LN(g_1)	(**30.2**)
DS_{30}	LN	(**218.6**)	LN(g_1)	(**218.6**)	LN	(**221.9**)	LN(g_1)	(**221.9**)

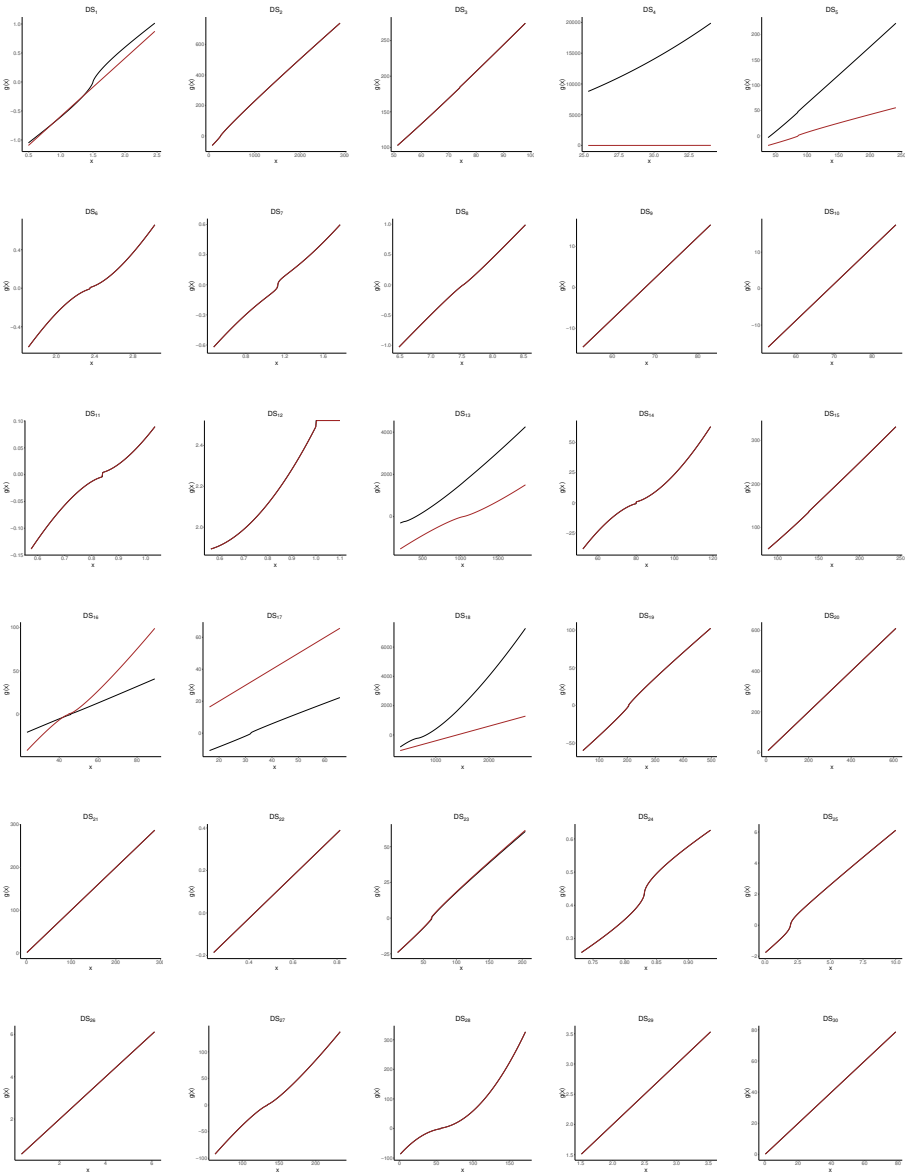

Figure 1: Original values x and modified values $g(x)$. AIC curves in black, BIC curves in brown

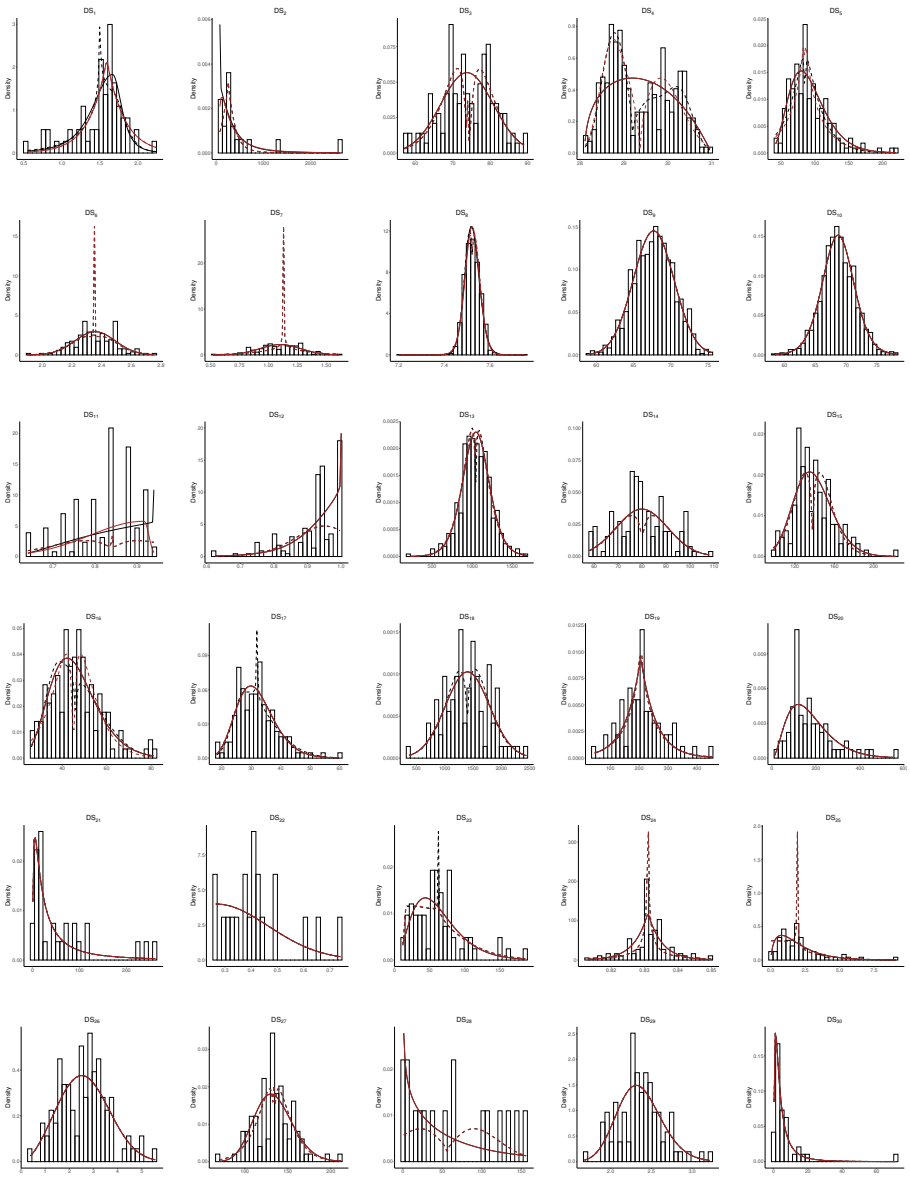

Figure 2: Data fits using the models $F(x)$ and the best models $F(g(x))$. Undoped models in solid lines, doped models in dotted lines. AIC curves in black, BIC curves in brown

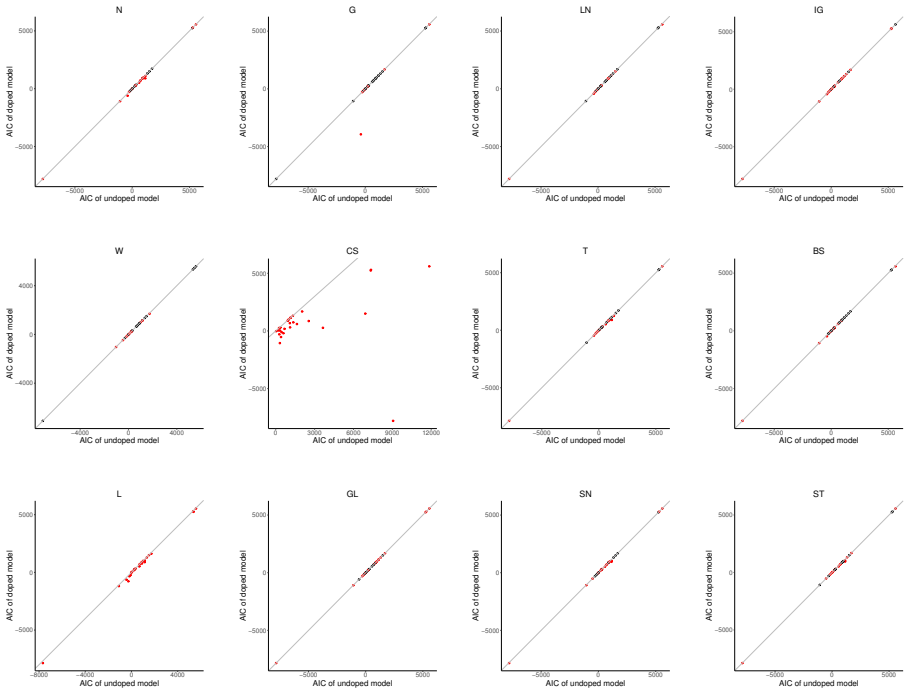

Figure 3: Distributions of AIC values when fitting the data sets using $F(x)$ (undoped model) and $F(g(x))$ (doped model). Fit improvement cases in red

given model. These favorite models are identified in red in order to help their visualization. Graphical examinations of these plots show that the corresponding plots between Figures 3 and 4 are very similar, which puts in evidence that for a given model the AIC and BIC values in general present coincidence on to select or not a model. Next, the plots for the IN, CS and L models show that doped models may frequently improve data fits. Among these models, the CS model under AIC is exceptional since the doped models always improve the undoped ones. Other models show a few number of cases where doped models may be favorite, for instance the G, W and BS models.

Information presented in Figures 3 and 4 can be synthesized in percentages of the number of data sets where some doped model $F(g(x))$ fits better than the undoped one $F(x)$, given the model $F(x)$. Table 3 presents these percentages, distinguished by parsimony criteria. These results put in evidence that these percentages may increase more if AIC values are used rather than BIC values because higher penalizations under BIC values. There the percentages related to the models IN, CS and L are higher as observed in Figures 3 and 4. These results also show that in general doped models improve data fits.

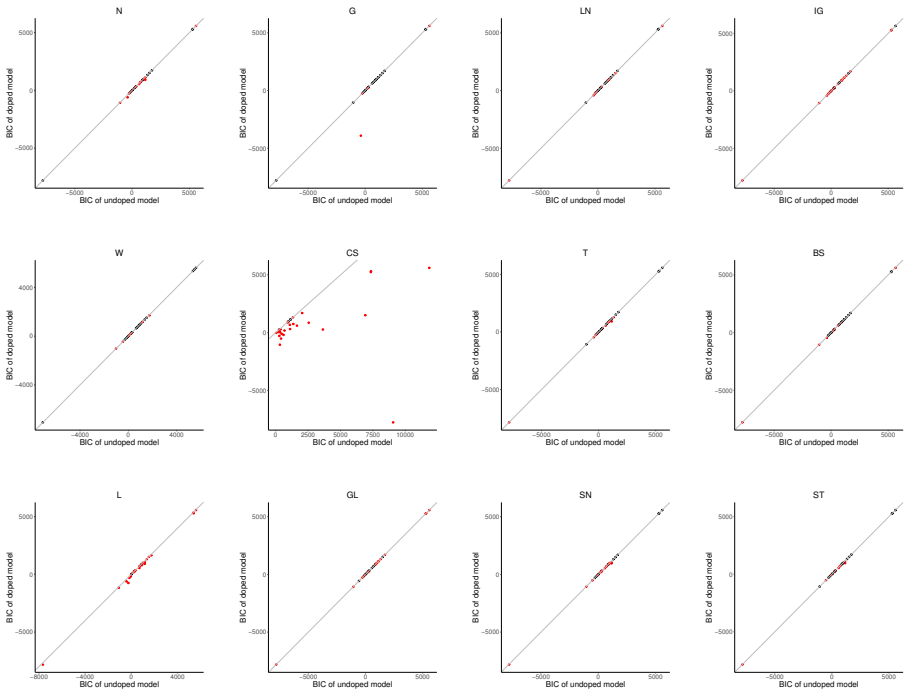

Figure 4: Distributions of BIC values when fitting the data sets using $F(x)$ (undoped model) and $F(g(x))$ (doped model). Fit improvement cases in red

Table 3: Percentages of doped models $F(g(x))$ that fit better than undoped models $F(x)$, when $F(x)$ is given

	N	G	LN	IN	W	CS	T	BS	L	GL	SN	ST
AIC	63.3	26.7	43.3	76.7	36.7	100.0	53.3	36.7	96.7	60.0	63.3	63.3
BIC	53.3	20.0	40.0	73.3	23.3	86.7	36.7	33.3	90.0	60.0	56.7	40.0

5 Concluding remarks

In this paper we have introduced a couple of families of functions $g(x)$ in order to improve data fits of given distributions $F(x)$ through the new distribution $F(g(x))$ which has been called a doped distribution. Examinations of several well-known distributions through a number of real data sets have shown that the involved data fits may be often improved by using doped models. A finding has been to identify that such improvements have been frequently associated with combinations of this couple of functions $g(x)$. These results are astonishing since these combinations increase model complexity because the increase of the number of parameters dut to $g(x)$, but at the same time these new doped models produce the best data fits with respect to parsimony criteria as AIC or BIC. These results therefore motivate the research for establishing more families of functions $g(x)$ improving data fits. In a paper in developing we will present new families of functions of this type.

From a different point of view, the introduction of the family of functions $g(x)$ can be used as a method for generating new continuous distributions, i.e. by applying strictly increasing functions to the argument of the distributions. In effect, given a random variable X satisfying $P(X < x) = F(x)$, the new distribution $F(g(x))$ is related to the random variable $Y = g^{-1}(X)$ because of $P(X < g(x)) = F(g(x))$. This method is thus an alternative way to other presented in literature as the one developed in [Alzaatreh et al. (2013)]. In a forthcoming paper we exploit this idea in order to show its advantages.

Acknowledgments

The authors thank Dr. Steve Su for kindly facilitating us data used in his paper [Su (2007)].

References

[Alizadeh et al. (2017)] Alizadeh, M., Merovciz, F. and Hamedani, G.G. (2017). Generalized Transmuted Family of Distributions: Properties and Applications. *Hacet J. Math. Stat.* **46** 645–667.

[Alzaatreh et al. (2013)] Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron* **71** 63–79.

[Bhaumik et al. (2009)] Bhaumik, D.K., Kapur, K. and Gibbons, R.D. (2009). Testing Parameters of a Gamma Distribution for Small Samples. *Technometrics* **51** 326–334.

[Birnbaum and Saunders (1958)] Birnbaum, Z.W. and Saunders, S.C. (1958). A Statistical Model for Life-Length of Materials. *J. Am. Stat. Assoc.* **53** 151–160.

[Birnbaum and Saunders (1969)] Birnbaum, Z.W. and Saunders, S.C. (1969). Estimation for a Family of Life Distributions with Applications to Fatigue, *J. Appl. Probab.* **6** 328–347.

[Bjerkedal (1960)] Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Am. J. Epidemiol.* **72** 130–148.

[Blasing et al. (2004)] Blasing, T.J., Broniak, C. and Marland, G. (2004). Monthly Estimates of C-13/C-12 (per mil) in Fossil-Fuel Carbon Dioxide Emissions from the U.S.A.. Data file accessed at cdiac.esd.ornl.gov/ftp/trends/emis_mon/emis_mon_c13.dat

[Boyd et al. (1998)] Boyd, J., Delost, M. and Holcomb, J. (1998). Calcium, phosphorus, and alkaline phosphatase laboratory values of elderly subjects. *Clinical Laboratory Science* **11** 223–227.

[Dana (1990)] Dana, E. (1990). *Salience of the self and salience of standards: Attempt to match self to a standard*. Unpublished PhD dissertation, Dept of Psychology, University of Southern California.

[Davis (1952)] Davis, D.J. (1952). An Analysis of Some Failure Data. *J. Am. Stat. Assoc.* **47** 113–150.

[Dumonceaux and Antle (1973)] Dumonceaux, R. and Antle, C.E. (1973). Discrimination between the Log-Normal and the Weibull Distributions, *Technometrics* **15** 923–926.

[Duncan (1986)] Duncan, A.J. (1986). *Quality Control and Industrial Statistics*. 5th ed. Irwin, Homewood, Illinois.

[Elinder et al. (1981)] Elinder, C-G, Jönsson, L., Piscator, M. and Rahnster, B. (1981). Histopathological changes in relation to cadmium concentration in horse kidneys. *Environ. Res.* **26** 1–21.

[Feigl and Zelen (1965)] Feigl, P. and Zelen, M. (1965). Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics* **21** 826–838.

[Hanley and Shapiro (1994)] Hanley, J.A. and Shapiro, S.H. (1994). Sexual Activity and the Lifespan of Male Fruitflies: A Dataset That Gets Attention. *J. Stat. Educ.* **2** 1–4.

[Hastie et al. (2008)] Hastie, T., Tibshirani, R. and Friedman, J. (2008). *The Elements of Statistical Learning*. 2nd ed. Springer Series in Statistics, New York.

[Lee and Wang (2003)] Lee, E.T. and Wang, J.L. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, Inc., New Jersey.

[Mackowiak et al. (1992)] Mackowiak, P.A., Wasserman, S.S. and Levine, M.M. (1992). A Critical Appraisal of 98.6 F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich. *JAMA* **268** 1578–1580.

[Mudholkar and Huston (1996)] Mudholkar, G.S. and Huston, A.D. (1996). The exponentiated Weibull family: Some properties and a flood data application. *Commun. Stat.-Theor. M.* **23** 1149–1171.

[Nelder and Mead (1965)] Nelder, J.A. and Mead, R. (1965). A Simplex Method for Function Minimization. *Comput. J.* **7** 308–313.

[Nichols and Padgett (2006)] Nichols, M.D. and Padgett, W.J. (2006). A Bootstrap Control Chart for Weibull Percentiles, *Qual. Reliab. Eng. Int.* **22** 141–151.

[Pearson and Lee (1903)] Pearson, K. and Lee, A. (1903). On the Laws of Inheritance in Man: I. Inheritance of Physical Characters. *Biometrika* **2** 357–462.

[Powell (1964)] Powell, M.J.D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **7** 155–162.

[Press et al. (2007)] Press, W.H., Teukolsky, S.A., Bethe, H.A., Vetterling, W.T. and Flannery, B.P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge University Press, Cambridge.

[Risebrough (1972)] Risebrough, R.W. (1972). Effects of environmental pollutants upon animals other than man. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.* **6** 443–463.

[Shao et al. (2004)] Shao, S., Wong, H., Xia, J. and Ip, W-C (2004). Models for extremes using the extended three parameter Burr XII system with application to flood frequency analysis. *Hydrolog. Sci. J.* **49** 685–702.

[Shkedy et al. (2006)] Shkedy, Z., Aerts, M. and Callaert, H. (2006). The Weight of Euro Coins: Its Distribution Might Not Be As Normal As You Would Expect. *J. Stat. Educ.* **14** 1–14.

[Smith and Naylor (1987)] Smith, R.L. and Naylor, J.C. (1987). A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution. *J. R. Stat. Soc. C-Appl.* **36** 358–369.

[Su (2007)] Su, S. (2007). Numerical maximum log likelihood estimation for generalized lambda distributions. *Comput. Stat. Data An.* **51** 3983–3998.