Revista Colombiana de Estadística
Junio 2013, volumen 36, no. 1, pp. 95 a 114

Properties and Inference for Proportional Hazard Models

Propiedades e inferencia para modelos de Hazard proporcional

Guillermo Martínez-Flórez[1,a], Germán Moreno-Arenas[2,b], Sandra Vergara-Cardozo[3,c]

[1]Departamento de Matemáticas y Estadística, Facultad de Ciencias, Universidad de Córdoba, Montería, Colombia

[2]Escuela de Matemáticas, Facultad de Ciencias, Universidad Industrial de Santander, Bucaramanga, Colombia

[3]Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract

We consider an arbitrary continuous cumulative distribution function $F(x)$ with a probability density function $f(x) = dF(x)/dx$ and hazard function $h_f(x) = f(x)/[1 - F(x)]$. We propose a new family of distributions, the so-called proportional hazard distribution-function, whose hazard function is proportional to $h_f(x)$. The new model can fit data with high asymmetry or kurtosis outside the range covered by the normal, t-student and logistic distributions, among others. We estimate the parameters by maximum likelihood, profile likelihood and the elemental percentile method. The observed and expected information matrices are determined and likelihood tests for some hypotheses of interest are also considered in the proportional hazard normal distribution. We show an application to real data, which illustrates the adequacy of the proposed model.

***Key words*:** Hazard function, Kurtosis, Method of moments, Profile likelihood, Proportional hazard model, Skewness, Skew-normal distribution.

Resumen

Consideramos una función de distribución continua arbitraria $F(x)$ con función de densidad de probabilidad $f(x) = dF(x)/dx$ y función de riesgo $h_f(x) = f(x)/[1 - F(x)]$. En este artículo proponemos una nueva familia de distribuciones cuya función de riesgo es proporcional a la función de riesgo $h_f(x)$. El modelo propuesto puede ajustar datos con alta asimetría o curtosis

[a]Professor. E-mail: gmartinez@correo.unicordoba.edu.co
[b]Associate professor. E-mail: gmorenoa@uis.edu.co
[c]Assistant Professor. E-mail: svergarac@unal.edu.co

fuera del rango de cobertura permitido por la distribución normal, t-Student, logística, entre otras. Estimamos los parámetros del modelo usando máxima verosimilitud, verosimilitud perfilada y el método elemental de percentiles. Calculamos las matrices de información esperada y observada. Consideramos test de verosimilitudes para algunas hipótesis de interés en el modelo con función de riesgo proporcional a la distribución normal. Presentamos una aplicación con datos reales que ilustra que el modelo propuesto es adecuado.

***Palabras clave*:** asimetría, curtosis, distribución skew-normal, función de riesgo, método de los momentos, modelo de riesgo proporcional, verosimilitud perfilada.

1. Introduction

When data originates from heavy-tailed or asymmetrical distributions, the normality-based inferential processes are inadequate. In these situations many authors choose to transform the variables in order to attain symmetry or normality. These transformations produce unsatisfactory results because the interpretation of the results becomes cumbersome. Although the class of elliptic distributions is a good alternative for situations with heavy-tailed behavior, this is not appropriate when the distribution is asymmetric. These circumstances prompted the search for new distributions, better suited to fit data with high asymmetry or kurtosis. The literature on families of flexible distributions has experienced great increase in the last three or four decades. Some early results include Lehmann (1953), Roberts (1966) and O'Hagan & Leonard (1976), among others. Azzalini (1985), Durrans (1992), Fernández & Steel (1998), Mudholkar & Hutson (2000), Gupta, Chang & Huang (2002), Arellano-Valle, Gómez & Quintana (2004, 2005), Gómez, Venegas & Bolfarine (2007), Arnold, Gómez & Salinas (2009), Pewsey, Gómez & Bolfarine (2012) represent some of the important contributions.

Azzalini (1985) defines a probability density function of a random variable Z with skew-normal distribution and parameter λ, given by

$$f_{SN}(z;\lambda) = 2\phi(z)\Phi(\lambda z),\ z \in \mathbb{R} \tag{1}$$

where ϕ and Φ denote the standard normal density and the cumulative distribution functions, respectively. The skewness is controlled by the parameter λ. We denote this by $Z \sim SN(\lambda)$. The asymmetry and kurtosis coefficients for this distribution are in the intervals $(-0.9953, 0.9953)$ and $[3, 3.8692)$, respectively. The skew-normal distribution was first introduced by O'Hagan & Leonard (1976) as a prior distribution for estimating a normal location parameter. The density (1) has also been studied widely by Henze (1986), Chiogna (1998), Pewsey (2000) and Gómez et al. (2007).

Durrans (1992), in a hydrological context, introduced the fractional order statistics distribution with density function given by

$$g_F(z;\alpha) = \alpha f(z)\{F(z)\}^{\alpha-1},\ z \in \mathbb{R},\ \alpha \in \mathbb{R}^+ \tag{2}$$

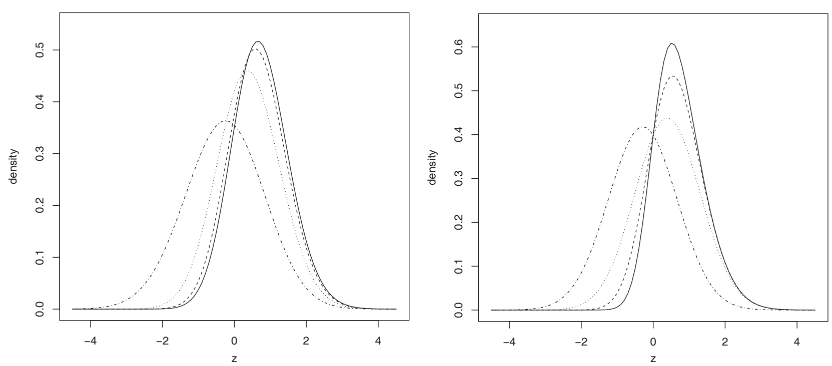
where F is an absolutely continuous distribution function, f is a corresponding density function and α is a shape parameter that controls the amount of asymmetry in the distribution. We refer to this model as the power distribution. We use the notation $Z \sim AP(\alpha)$.

Following the idea of Durrans, Gupta & Gupta (2008) we define the power-normal distribution whose distribution function is given by

$$g_{\Phi}(z;\alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \ z \in \mathbb{R}, \ \alpha \in \mathbb{R}^{+} \tag{3}$$

We use the notation $Z \sim PN(\alpha)$. Pewsey, Gómez and Bolfarine (2012) showed that the expected information matrix is nonsingular for the neighborhood of $\alpha = 1$, contrary to the skew-normal distribution where the information matrix is singular under the symmetry hypothesis ($\lambda = 0$). They also found that the asymmetry and kurtosis coefficients for this distribution are in the intervals $[-0.6115, 0.9007]$ and $[1.7170, 4.3556]$, respectively.

Figure 1 shows how the parameters α and λ control the asymmetry and kurtosis of the (1) and (3) models.



(a) (b)

FIGURE 1: Probability density function (a) $PN(\alpha)$ for $\alpha = 0.746$ (dashed-dotted line), 1.626 (dotted line), 2.254 (dashed line) and 2.516 (solid line). (b) $SN(\lambda)$ for λ =-0.40 (dashed-dotted line), 0.60 (dotted line), 1.40 (dashed line) and 2.20 (solid line).

In this paper we present a new family of distributions so-called proportional hazard distribution-functions. The paper is presented as follows. In Section 2 we define the proportional hazard distribution-function, study some of its properties and discuss maximum likelihood estimation. The location-scale extension for proportional hazard distribution-function is presented in Section 3. In Section 4, we define the *location-scale proportional hazard normal* model and different methods for parameter estimation; we derive the information matrix and discuss likelihood ratio tests for some hypotheses of interest. Further, the asymptotic distribution of maximum likelihood estimators is obtained. The usefulness of the proposed model is illustrated in an application to real data in Section 5. Finally, some concluding remarks are found in Section 6.

2. Proportional Hazard Distribution-Function

Let $F(x)$ be a continuous cumulative distribution function with probability density function $f(x)$ and hazard function $h_f(x) = f(x)/(1 - F(x))$. We will say that Z has proportional hazard distribution-function associated with F and f and parameter $\alpha > 0$ if its probability density function is

$$\varphi_F(z; \alpha) = \alpha f(z)\{1 - F(z)\}^{\alpha-1}, \quad z \in \mathbb{R} \tag{4}$$

where α is a positive real number and F is a continuous distribution function with continuous density function f. We use the notation $Z \sim PHF(\alpha)$. The distribution function of the PHF model is given by

$$\mathbb{F}(z) = 1 - \{1 - F(z)\}^{\alpha}, \ z \in \mathbb{R} \tag{5}$$

We observe that the name "proportional hazard distribution-function" is appropriate because its hazard function with respect to the density φ_F is

$$h_{\varphi_F}(x, \alpha) = \alpha h_f(x)$$

The inversion method can be used to generate a $PHF(\alpha)$ distribution. Thus, if U is a uniform random variable on $(0, 1)$,

$$Z = F^{-1}(1 - (1 - u)^{1/\alpha})$$

obeys a $PHF(\alpha)$ distribution, whose median, $Z_{0.5}$, can be found from the inverse of F through

$$Z_{0.5} = F^{-1}\left(\frac{2^{1/\alpha} - 1}{2^{1/\alpha}}\right)$$

where F^{-1} is the inverse of the distribution F. In general, the *p-th* percentile can be computed by

$$Z_p = F^{-1}\left(1 - (1 - p)^{1/\alpha}\right)$$

The distribution mode is the solution to the non-linear equation

$$[1 - F(z)]\, f'(z) - (\alpha - 1)f^2(z) = 0$$

where f' is the derivative of F.

In the next section we present some particular cases of the PHF distribution.

2.1. Proportional Hazard Normal Distribution

When $F = \Phi$, the standard normal distribution function, we obtain the *proportional hazard normal* distribution, which we denote by $PHN(\alpha)$. Its density function is given by

$$\varphi_\Phi(z; \alpha) = \alpha\phi(z)\{1 - \Phi(z)\}^{\alpha-1}, \ z \in \mathbb{R} \tag{6}$$

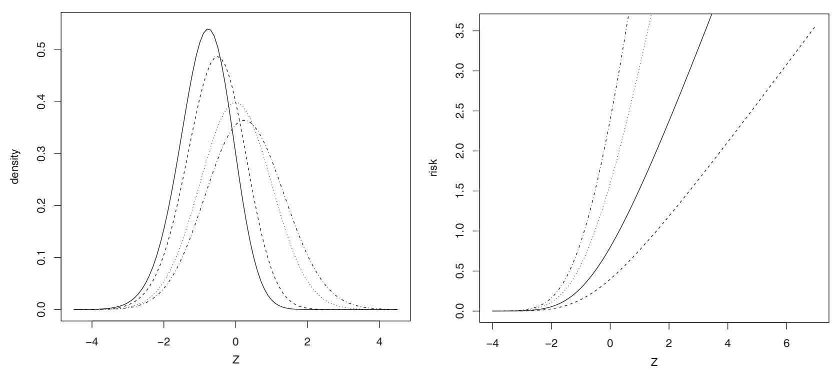
This model is also an alternative to accommodate data with asymmetry and kurtosis that are outside the ranges allowed by the normal distribution. The PHN is a special case of Eugene, Lee & Famoye (2002)'s *beta-normal* distribution. A simple comparison makes clear that $PHN(1) = SN(0) = PN(1) = N(0, 1)$.

The survival function and the hazard function are given, respectively, by

$$S(t) = \{1 - \Phi(t)\}^{\alpha} \quad \text{and} \quad h_{\varphi_\Phi}(t) = \alpha h_{\phi}(t)$$

That is to say, the PHN model's hazard function is directly proportional to the normal model's hazard function. It can then be said that the hazard function is a non decreasing (and unimodal) function of T, but an increasing function of parameter α. It can also be said that for $\alpha > 1$, the PHN's model hazard is greater than the normal's model, while for $\alpha > 1$ the opposite occurs.

In Figure 2-(a) we can see the behavior of the $PHN(\alpha)$ density and Figure 2-(b) shows the model's hazard function for a few values of the parameter α.



(a) (b)

FIGURE 2: (a) $PHN(\alpha)$ for α =0.75 (solid line), 1.0 (dashed line), 2.0 (dotted line), and 3.0 (dashed-dotted line) (b) $h_{\varphi_\Phi}(z)$ for α =0.25 (dashed line), 1.0 (solid line), 2.0 (dotted line), 3.0 (dashed-dotted line)

2.2. Proportional Hazard Logistic Distribution

The *proportional hazard logistic* distribution is defined by the probability density function

$$\varphi_L(z; \alpha) = \alpha \exp(x) \left\{ \frac{1}{1 + \exp(x)} \right\}^{\alpha+1} \tag{7}$$

We denote it by $PHL(\alpha)$. Figure 3 shows the behaviour of the this distribution for diferents values of the α.

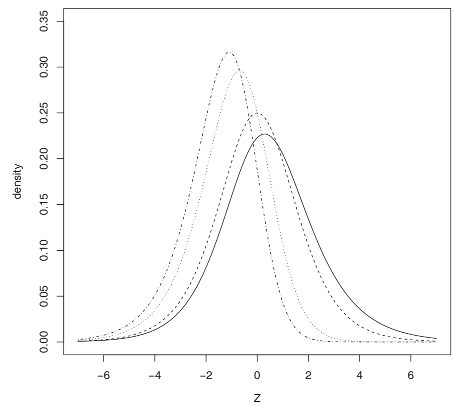


FIGURE 3: $PHL(\alpha)$ distribution for α =0.75 (solid line), 1.0 (dashed line), 2.0 (dotted line) and 3.0 (dashed-dotted line)

2.3. Proportional Hazard t-Student Distribution

The *proportional hazard t-student* distribution is defined by the probability density function

$$\varphi_T(z;\alpha,v) \quad = \frac{\alpha\Gamma(\frac{v+1}{2})}{(v\pi)^{1/2}\Gamma(\frac{v}{2})}\left[1+\frac{z^2}{v}\right]^{-(v+1)/2}\{1-T(z)\}^{\alpha-1} \tag{8}$$

where T is the cumulative distribution function of the t-student distribution and v is the number of degrees of freedom. The notation we use is $PHt(v,\alpha)$. Figure 4 shows the behavior of this distribution.

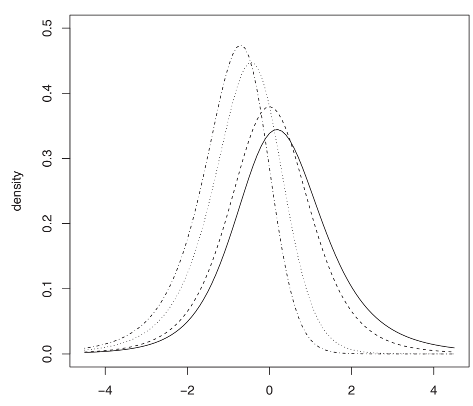


FIGURE 4: $PHt(v,\alpha)$ distribution for α =0.75 (solid line), 1.0 (dashed line), 2.0 (dotted line) and 3.0 (dashed-dotted line)

2.4. Proportional Hazard Cauchy Distribution

When $v = 1$ in $PHt(v, \alpha)$ gives the *proportional hazard Cauchy* distribution, whose probability density function is

$$\varphi_C(z; \alpha) \quad = \tfrac{\alpha}{\pi[1+z^2]} \left\{ \tfrac{1}{2} - \tfrac{1}{\pi} \arctan(z) \right\}^{\alpha - 1} \tag{9}$$

We denote it by $PHC(\alpha)$. Figure 5 shows the behavior of this distribution for different values of the α parameter.

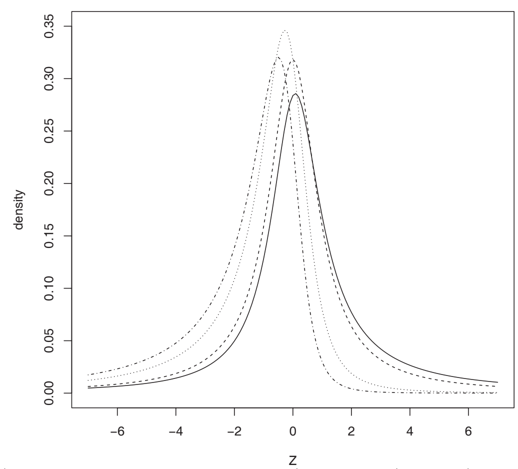


Figure 5: $PHC(\alpha)$ distribution for α =0.75 (solid line), 1.0 (dashed line), 2.0 (dotted line) and 3.0 (dashed-dotted line)

2.5. Moments of the PHF

The moment generating function for the PHF distribution is given by

$$M(t) = \alpha \int_0^1 \exp\left\{tF^{-1}(y)\right\} (1 - y)^{\alpha - 1} dy \tag{10}$$

There is no closed form for the moments of a random variable Z with distribution $PHF(\alpha)$; these are computed numerically.

The r-th Z moment for the random variable $Y \sim PHF$ can be obtained with the expression

$$\mu_r = \alpha \int_0^1 \left\{F^{-1}(y)\right\}^r (1 - y)^{\alpha - 1} dy, \; r = 0, 1, 2, \ldots \tag{11}$$

This expectation agrees with the expected value of the function $\left\{F^{-1}(y)\right\}^r$ where Y is a random variable with a Beta distribution with parameters α and 1. The central moments $\acute{\mu_r} = E(Z - E(Z))^r$ for $r = 2, 3, 4$ can be found from the expressions

$\acute{\mu_2} = \mu_2 - \mu_1^2$, $\acute{\mu_3} = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$ and $\acute{\mu_4} = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4$. Consequently, the variance asymmetry and kurtosis coefficients are $\sigma^2 = Var(Z) = \acute{\mu_2}$, $\sqrt{\beta_1} = \acute{\mu_3}/[\acute{\mu_2}]^{3/2}$ and $\beta_2 = \acute{\mu_4}/[\acute{\mu_2}]^2$, respectively.

For $F = \Phi$, that is, the case of the $PHN(\alpha)$ distribution, the r-th Z moment is given by

$$\mu_r = \alpha \int_0^1 \left\{\Phi^{-1}(y)\right\}^r (1-y)^{\alpha-1} dy, \ r = 0, 1, 2, \dots \tag{12}$$

Thus, for α values between 0.0005 and 9,000 the asymmetry and kurtosis coefficients, $\sqrt{\beta_1}$ and β_2 of the variable $Z \sim PHN(\alpha)$ belong to the intervals (-1.1578, 0.9918) and (1.1513,4.3023), respectively. Therefore the PHN distribution clearly fits data with less negative asymmetry and more platykurtic than the SN and PN distributions do. It also fits distributions with a higher positive asymmetry than PN and more leptokurtic than SN. It is evident that the PHN distribution fits data with as much positive asymmetry as SN distribution does and as much kurtosis as PN distribution does.

3. Location-Scale PHF

Let $Z \sim PHF(\alpha)$ with $\alpha \in \mathbb{R}^+$. The family of PHF distributions with location-scale parameters is defined as the distribution of $X = \xi + \eta Z$ for $\xi \in \mathbb{R}$ and $\eta > 0$. The corresponding density function is given by

$$\varphi_F(x; \xi, \eta, \alpha) = \frac{\alpha}{\eta} f\left(\frac{x-\xi}{\eta}\right) \left\{1 - F\left(\frac{x-\xi}{\eta}\right)\right\}^{\alpha-1}, \ x \in \mathbb{R} \tag{13}$$

where ξ is the location parameter and η is the scale parameter. We use the notation $PHF(\xi, \eta, \alpha)$.

3.1. Estimation and Inference for the Location-Scale PHF

We now deduce the maximum likelihood estimators (MLE) for the parameters of the $PHF(\xi, \eta, \alpha)$ distribution and the respective observed and expected information matrices.

For n observations, $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ from the $PHF(\xi, \eta, \alpha)$ distribution, the log-likelihood function of $\boldsymbol{\theta} = (\xi, \eta, \alpha)'$, given $\mathbf{x}$, is

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = n\log(\alpha) - n\log(\eta) + \sum_{i=1}^n \log(f(z_i)) + (\alpha - 1)\sum_{i=1}^n \log(1 - F(z_i))$$

where $z_i = \frac{x_i-\xi}{\eta}$. Thus, under the assumption that the derivative of f exists, the score function is given by:

$$\begin{aligned}
U(\xi) &= -\frac{1}{\eta}\sum_{i=1}^{n}\frac{f'(z_i)}{f(z_i)}+\frac{\alpha-1}{\eta}\sum_{i=1}^{n}\frac{f(z_i)}{1-F(z_i)},\\
U(\eta) &= -\frac{n}{\eta}-\frac{1}{\eta}\sum_{i=1}^{n}z_i\frac{f'(z_i)}{f(z_i)}+\frac{\alpha-1}{\eta}\sum_{i=1}^{n}z_i\frac{f(z_i)}{1-F(z_i)},\\
U(\alpha) &= \frac{n}{\alpha}+\sum_{i=1}^{n}\log[1-F(z_i)]
\end{aligned}$$

MLE estimators are the solutions to this system of equations usually solved by iterative numerical methods. It is usual to use a software algorithm implemented in R (R Development Core Team 2013).

3.1.1. Observed Information Matrix for the Location-Scale PHF

Assuming the existence of the second derivative of f and putting $w_i = \frac{f(z_i)}{1-F(z_i)}$, $s_i = \frac{f'(z_i)}{1-F(z_i)}$, $t_i = \frac{f''(z_i)}{f(z_i)}$ and $v_i = \frac{f'(z_i)}{f(z_i)}$, the observed information matrix entries, $j_{\xi\xi}, j_{\eta\xi}, \dots, j_{\alpha\alpha}$, are obtained:

$$j_{\xi\xi} = -\frac{n}{\eta^2}\left\{(\overline{v^2}-\overline{t})+(\alpha-1)\left[\overline{w^2}+\overline{s}\right]\right\}$$

$$j_{\eta\xi} = -\frac{n}{\eta^2}(\overline{v}+\overline{t}-\overline{v^2})+n\frac{\alpha-1}{\eta^2}\left[\overline{zw^2}+\overline{zs}+\overline{w}\right]$$

$$j_{\eta\eta} = -\frac{n}{\eta^2}+\frac{n}{\eta^2}\left[-2\overline{zv}-\overline{z^2t}+\overline{z^2v^2}\right]+n\frac{\alpha-1}{\eta^2}\left[2\overline{zw}+\overline{z^2s}+\overline{z^2w^2}\right]$$

$$j_{\alpha\xi} = -\frac{n}{\eta}\overline{w} \qquad j_{\alpha\eta} = -\frac{n}{\eta}\overline{zw} \qquad j_{\alpha\alpha} = \frac{n}{\alpha^2}$$

where $\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$, $\overline{v^2} = \frac{1}{n}\sum_{i=1}^{n} v_i^2, \dots, \overline{z^2w^2} = \frac{1}{n}\sum_{i=1}^{n} z_i^2 w_i^2$.

3.1.2. Expected Information Matrix for the Location-Scale PHF

The expected information matrix entries are n^{-1} times the expected value of the observed information matrix elements, that is,

$$I_{\theta_r\theta_p} = n^{-1}E\left\{-\frac{\partial^2\ell(\boldsymbol{\theta};\mathbf{x})}{\partial\theta_r\partial\theta_p}\right\},\ r,p=1,2,3,\ \text{with } \theta_1=\xi,\ \theta_2=\eta \text{ and } \theta_3=\alpha.$$

Considering the notation below (Pewsey et al. 2012):

$$\begin{aligned}
a_{kj} &= E\{z^k\,(f(z)/[1-F(z)])^j\}\\
b_k &= E\{z^k f'(z)/[1-F(z)]\}\\
c_{kj} &= E\{z^k(f'(z)/f(z))^j\}\\
d_k &= E\{z^k f''(z)/f(z)\} \text{ for } k=0,1,2 \text{ and } j=1,2
\end{aligned}$$

the observed information matrix elements these are given by

$$
\begin{aligned}
I_{\xi\xi} &= \{(c_{02} - d_0) + (\alpha - 1)(a_{02} + b_0)\}/\eta^2, \\
I_{\xi\eta} &= \{(c_{12} - c_{01} - d_1) + (\alpha - 1)(a_{12} + b_1 + a_{01})\}/\eta^2 \\
I_{\xi\alpha} &= E(w)/\eta = a_{01}/\eta \\
I_{\eta\eta} &= \{(c_{22} - d_2 - 2c_{11} - 1) + (\alpha - 1)(a_{22} + b_2 + 2a_{11})\}/\eta^2 \\
I_{\eta\alpha} &= E(zw)/\eta = a_{11}/\eta, \text{ and } I_{\alpha\alpha} = 1/\alpha^2
\end{aligned}
$$

In general, these expected values are computed using numerical integration. When $\alpha = 1$, we have $\varphi(x; \xi, \eta, 1) = \frac{1}{\eta} f\left(\frac{x-\xi}{\eta}\right)$, the location-scale f function model, thus the matrix information is reduced to

$$
\begin{pmatrix}
(c_{02} - d_0)/\eta^2 & (c_{12} - c_{01} - d_1)/\eta^2 & a_{01}/\eta \\
(c_{12} - c_{01} - d_1)/\eta^2 & (c_{22} - d_2 - 2c_{11} - 1)/\eta^2 & a_{11}/\eta \\
a_{01}/\eta & a_{11}/\eta & 1
\end{pmatrix}
$$

The properties of this matrix depend on the function f.

4. Location-Scale Proportional Hazard Normal

A very special particular case of the $PHF(\xi, \eta, \alpha)$ model occurs when $F = \Phi$, the standard normal distribution function. In this case the probability density function is

$$
\varphi_\Phi(x; \xi, \eta, \alpha) = \frac{\alpha}{\eta}\phi\left(\frac{x - \xi}{\eta}\right)\left\{1 - \Phi\left(\frac{x - \xi}{\eta}\right)\right\}^{\alpha - 1}, \; x \in \mathbb{R} \tag{14}
$$

which we call *location-scale proportional hazard normal.* Note that when $\alpha = 1$ we are in the case of the location-scale normal distribution.

In what follows we discuss estimation by moments, maximum likelihood, profiled likelihood and elemental percentile method for the $PHN(\xi, \eta, \alpha)$ model and show the respective observed and information matrices for a PHN random variable.

4.1. Estimation by the Method of Moments for the Location-Scale PHN

The mean (μ), variance (σ^2) and asymmetry coefficient ($\sqrt{\beta_1}$) in the location-scale case are:

$$
\mu = \xi + \eta\Phi_1(\alpha), \quad \sigma^2 = \eta^2\Phi_2(\alpha) \;\text{ and }\; \sqrt{\beta_1} = \frac{\mu_3'}{\sigma^3} = \Phi_3(\alpha)
$$

Thus, the estimators for α, ξ and η can be obtained by substituting, in the above expressions, μ, σ^2 and $\sqrt{\beta_1}$ for their respective sample moments $\bar{y}$, s^2 and

$\sqrt{b_1}$. First the α estimator is obtained as in the standard case and its value can be used to estimate $\Phi_1(\alpha)$ and $\Phi_2(\alpha)$, leaving a simple 2×2 system of linear equations to solve, whose solution gives the ξ and η estimators. The asymptotic distribution of moment estimators is widely studied in Sen & Singer (1993) and Sen, Singer & Pedroso de Lima (2010).

4.2. Maximum Likelihood Estimation for the Location-Scale PHN

For n observations, $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ from the $PHN(\xi, \eta, \alpha)$ distribution, the log-likelihood function of $\theta = (\xi, \eta, \alpha)^\top$ given $\mathbf{x}$ is

$$\ell(\theta; \mathbf{x}) = n\log(\alpha) - n\log(\eta) + \sum_{i=1}^{n} \log(\phi(z_i)) + (\alpha - 1)\sum_{i=1}^{n} \log(1 - \Phi(z_i))$$

where $z_i = \frac{x_i - \xi}{\eta}$. Thus, the score function, defined as the derivative of the log-likelihood function with respect to each of the parameters, is:

$$U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log[1 - \Phi(z_i)]$$

$$U(\xi) = \frac{1}{\eta}\sum_{i=1}^{n} z_i + \frac{\alpha - 1}{\eta}\sum_{i=1}^{n} \frac{\phi(z_i)}{1 - \Phi(z_i)}$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 + \frac{\alpha - 1}{\eta}\sum_{i=1}^{n} z_i \frac{\phi(z_i)}{1 - \Phi(z_i)}$$

Setting these expressions to zero, we get the corresponding score equations whose numerical solution leads to the MLE estimators.

4.2.1. Observed information matrix for location-scale PHN

The observed information matrix follows from minus the second derivatives of the log-likelihood function, which are denoted by $j_{\xi\xi}, j_{\xi\eta}, \dots, j_{\alpha\alpha}$, and are given by

$$j_{\xi\xi} = \frac{n}{\eta^2} + n\frac{\alpha - 1}{\eta^2}\left[\overline{w^2} - \overline{zw}\right]$$

$$j_{\xi\eta} = \frac{2n}{\eta^2}\overline{z} + n\frac{\alpha - 1}{\eta^2}\left[-\overline{zw^2} - \overline{z^2w} + \overline{w}\right]$$

$$j_{\eta\eta} = -\frac{n}{\eta^2} + \frac{3n}{\eta^2}\overline{z^2} + n\frac{\alpha - 1}{\eta^2}\left[2\overline{zw} + \overline{z^2w^2} - \overline{z^3w}\right]$$

$$j_{\xi\alpha} = -\frac{n}{\eta}\overline{w} \qquad j_{\eta\alpha} = -\frac{n}{\eta}\overline{zw} \qquad j_{\alpha\alpha} = \frac{n}{\alpha}, \text{ where } w_i = \frac{\phi(z_i)}{1 - \Phi(z_i)}$$

$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$, $\overline{w^2} = \frac{1}{n}\sum_{i=1}^{n} w_i^2$, $\overline{zw} = \frac{1}{n}\sum_{i=1}^{n} z_i w_i$ $\dots$, $\overline{z^2w^2} = \frac{1}{n}\sum_{i=1}^{n} z_i^3 w_i^2$

4.2.2. Expected Information Matrix for the Location-Scale PHN

Considering $a_{kj} = \mathbb{E}\{z^k w^j\}$, the expected information matrix entries are:

$$I_{\xi\xi} = \frac{1}{\eta^2}\left[1 + (\alpha - 1)(a_{02} - a_{11})\right] \qquad I_{\eta\xi} = \frac{2}{\eta^2}a_{10} + \frac{\alpha - 1}{\eta^2}\left[a_{01} - a_{02} + a_{12}\right]$$

$$I_{\eta\eta} = -\frac{1}{\eta^2} + \frac{3}{\eta^2}a_{20} + \frac{\alpha - 1}{\eta^2}\left[a_{22} + 2a_{11} - a_{31}\right]$$

$$I_{\alpha\xi} = -\frac{1}{\eta}a_{01} \qquad I_{\alpha\eta} = -\frac{1}{\eta}a_{11} \qquad I_{\alpha\alpha} = \frac{1}{\alpha^2}$$

The expected values of the above variables are generally calculated using numerical integration. When $\alpha = 1$, $\varphi(x; \xi, \eta, 1) = \frac{1}{\eta}\phi\left(\frac{x-\xi}{\eta}\right)$, the location-scale normal density function. Thus, the information matrix becomes

$$I(\theta) = \begin{pmatrix} 1/\eta^2 & 0 & -a_{01}/\eta \\ 0 & 2/\eta^2 & -a_{11}/\eta \\ -a_{01}/\eta & -a_{11}/\eta & 1 \end{pmatrix}$$

Numerical integration shows that the determinant is

$$|I(\theta)| = \frac{1}{\eta^4}[2 - a_{11}^2 - 2a_{01}^2] = \frac{0.013687}{\eta^4} \neq 0$$

so in the case of a normal distribution the information matrix of the model is non-singular. The upper left 2×2 submatrix is the normal distribution's information matrix for the normal distribution.

For large n and under regularity conditions we have

$$\hat{\theta} \xrightarrow{A} N_3(\theta, I(\theta)^{-1})$$

and the conclusion follows that $\hat{\theta}$ is consistent and asymptotically approaches the normal distribution with $I(\theta)^{-1}$ as the covariance matrix, for large samples.

4.3. Profile Likelihood Estimation for the Location-Scale PHN

Maximum likelihood estimators of the $PHN(\xi, \eta, \alpha)$ distribution parameters usually display high levels of bias in the estimation of the shape parameter α when the sample size is small. Other estimation techniques can be used that result in a more consistent estimation of α. Among these are the profile likelihood and the modified profile likelihood (see Barndorff-Nielsen 1983, Severini 1998). Thus, calling $\boldsymbol{\tau} = (\xi, \eta)^\top$ the vector of parameters of interest and $\phi = \alpha$ the nuisance parameter, the profile likelihood is

$$L_p(\boldsymbol{\tau}) = L(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}})$$

where $\hat{\phi}_{\boldsymbol{T}} = \hat{\alpha}(\xi,\eta) = -n\left\{\sum_{i=1}^{n}\log\left[1-\Phi\left(\frac{x_i-\xi}{\eta}\right)\right]\right\}^{-1}$. Substituting $\hat{\alpha}(\xi,\eta)$ in the original likelihood we obtain the profile log-likelihood, defined as the logarithm of the profile likelihood:

$$\ell_p(\xi,\eta) = n\left[\log(n) - \log\left(-\sum_{i=1}^{n}\log\left[1-\Phi\left(z_i\right)\right]\right) - \log(\eta) - \frac{1}{2}\log(2\pi) - 1\right] - \frac{1}{2}\sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n}\log\left[1-\Phi\left(z_i\right)\right] \tag{15}$$

where $z_i = (x_i - \xi)/\eta$.

Consequently, the profiled maximum likelihood estimators for ξ and η, that is, $\hat{\xi}_p$ and $\hat{\eta}_p$, are the solutions to the nonlinear equations $u_p(\xi) = \frac{\partial \ell_p(\xi,\eta)}{\partial \xi} = 0$ and $u_p(\eta) = \frac{\partial \ell_p(\xi,\eta)}{\partial \eta} = 0$, which are obtained with iterative numerical methods.

Since sometimes the estimation of parameters by maximum likelihood can be inconsistent or inefficient, Barndorff-Nielsen (1983) proposes a modified profiled likelihood. Severini (2000) presents an alternative that is easier to apply in certain models like $PHN(\xi,\eta,\alpha)$. The profiled likelihood is not an actual likelihood, because some of the likelihood properties are not verified. for instance, the score function may have a nonzero mean and the observed information can have a bias. Nevertheless this function has go some interesting properties that make it look like an actual likelihood. For more examples, properties and uses of estimation by modified or unmodified profiled likelihood see Farias, Moreno & Patriota (2009).

4.4. Estimation by the Elemental Percentile Method for the Location-Scale PHN

The elemental percentile method can also be used in the estimation of the $PHN(\xi,\eta,\alpha)$ parameters applying the theory developed in Castillo & Hadi (1995).

Estimation of ξ and η when α is known. If the shape parameter α is known, the elemental percentile method for two order statistics $x_{(i)}$ and $x_{(j)}$, with $i < j$, leads to the equations

$$\hat{\eta}(i,j) = \frac{x_{(j)} - x_{(i)}}{\Phi^{-1}\left(1-\left(\frac{(n-j)+1}{n+1}\right)^{1/\alpha}\right) - \Phi^{-1}\left(1-\left(\frac{(n-i)+1}{n+1}\right)^{1/\alpha}\right)}$$

and

$$\hat{\xi}(i,j) = x_{(j)} - \hat{\eta}(i,j)\Phi^{-1}\left(1-\left(\frac{(n-j)+1}{n+1}\right)^{1/\alpha}\right)$$

Then, proceeding like in the previous case (for α), we select m samples of two order statistics and estimate the parameters ξ and η and again using robust statistics we finally get the estimators for these parameters.

A second estimation method, in two steps, using percentiles is illustrated next. It is motivated in the fact that the maximum likelihood method gives fairly good estimations of the location and scale (ξ and η) parameters.

Initially it is assumed that the location and scale parameters are known and their actual values are the MLE estimators, and we estimate α like in the standard case. Once the α estimator is known, the second step is to suppose that this is the actual value of the parameter and then we estimate the ξ and η values under the assumption that α is known. The standard errors for the parameter estimations can be computed using resampling techniques such as Jackknife or Bootstrap (see Efron (1982, 1979)). In both cases above we took $p_i = i/(n+1)$, given that we know $\mathbb{E}(\mathbb{F}) = i/(n+1)$.

4.5. Simulation Study

To study the MLE estimator properties of the $PHN(\xi, \eta, \alpha)$ distribution, a simulation was carried out for $\alpha = 0.75$, 1.5 and 3.0. Without loss of generality the location and scale parameters were set at $\xi = 0$ and $\eta = 1$.

The sample sizes in the simulation were $n = 50$, 100, 200 and 500 with 2,000 replications in each case. The random variable X with distribution $PHN(\xi, \eta, \alpha)$ was obtained with the algorithm

$$X = \xi + \eta\Phi^{-1}(1 - (1-u)^{1/\alpha}),$$

where u is a uniform random variable $U(0, 1)$. In all cases, the bias and root mean square errors of the MLE estimators were calculated.

The results shown in Tables (1) and (2) demonstrate that when the sample size increases, the bias and root mean square error decrease, that is, the estimators are asymptotically consistent. Still, a high bias in the shape parameter α for small sample sizes is evident. In conclusion, this estimation process would be recommended for very large sample sizes. Using the profiled likelihood estimation method for α we found biases 0.2511 and 0.7241 for values $\alpha = 0.75$ and 1.5 respectively with a sample size 100.

TABLE 1: Bias of the MLE from PHN model parameters.

	$\alpha = 0.75$			$\alpha = 1.5$			$\alpha = 3.0$		
n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$
50	0.1546	-0.0635	1.5529	0.0947	-0.0700	2.0300	-0.1128	-0.0915	1.9252
100	0.1523	-0.0061	0.4897	0.0899	-0.0163	1.4511	-0.0722	-0.0547	1.8106
200	0.0725	-0.0020	0.1831	0.0636	-0.0054	0.5113	0.0665	-0.0148	1.2823
500	0.0307	0.0001	0.0600	0.0321	-0.0005	0.1519	0.0325	0.0005	0.4562

Tables (3) and (4) show the behavior of estimators by the elemental percentile method for the $PHN(\xi, \eta, \alpha)$ model. As can been seen, these also are asymptotically consistent and their biases are less than the biases of the maximum likelihood estimators for a small sample. However, the bias of the α estimator is still too large. For small sample sizes, Jackknife or Bootstrap estimators can be applied to correct the bias of the MLE estimators (see, Efron 1982, Efron & Tibshirani 1993).

TABLE 2: $\sqrt{MSE}$ of the MLE from PHN model parameters.

	$\alpha = 0.75$			$\alpha = 1.5$			$\alpha = 3.0$		
n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$
50	1.5939	0.5583	3.6623	1.3763	0.4745	4.3718	1.1312	0.3808	4.5231
100	1.2367	0.4147	1.5684	1.0863	0.3440	3.4739	0.9468	0.2917	3.9510
200	0.8756	0.2945	0.7383	0.8353	0.2585	1.6110	0.7430	0.2169	3.4404
500	0.5102	0.1756	0.3607	0.5313	0.1633	0.7819	0.5374	0.1517	1.9056

TABLE 3: Bias of the PHN model percentile estimators.

	$\alpha = 0.75$			$\alpha = 1.5$			$\alpha = 3.0$		
n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$
50	0.1448	-0.0628	1.3740	0.0875	-0.0392	1.8700	-0.1013	-0.0740	1.6296
200	0.0902	0.0099	0.3897	0.0829	0.0107	0.7995	0.0533	0.0042	1.2500
500	0.0157	-0.0027	0.1018	0.0253	0.0038	0.2850	0.0511	0.0081	0.6931
1,500	0.0134	0.0020	0.0371	0.0118	0.0017	0.0811	0.0210	0.0039	0.2578
5,000	0.0040	0.0009	0.0104	-0.0003	0.0009	0.0178	0.0019	0.0004	0.0549

TABLE 4: $\sqrt{MSE}$ of the PHN model percentile estimators.

	$\alpha = 0.75$			$\alpha = 1.5$			$\alpha = 3.0$		
n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$
50	1.6300	0.6170	3.5602	1.4147	0.4979	4.3722	1.1907	0.4184	4.5468
200	0.8751	0.2966	1.4306	0.8616	0.2674	2.4950	0.7802	0.2269	3.5839
500	0.5135	0.1781	0.5298	0.5466	0.1699	1.2777	0.5313	0.1515	2.4152
1,500	0.2810	0.0983	0.2516	0.2896	0.0904	0.5403	0.3256	0.0919	1.3219
5,000	0.1553	0.0539	0.1304	0.1580	0.0497	0.2708	0.1728	0.0494	0.6281

5. Illustration

In this illustration we use a dataset related to 1,150 heights measured at 1 micron intervals along a roller drum (i.e. parallel to the roller's axis). This was part of an extensive study of the roller's surface roughness. It is available for download at `http://lib.stat.cmu.edu/jasadata/laslett`.

The data set to illustrate the PHN model has the following summary statistics: mean $\bar{x} = 3.535$ and variance $s^2 = 0.422$. Teh quantities $\sqrt{b_1} = -0.986$ and $b_2 = 4.855$ correspond to sample asymmetry and kurtosis coefficients. According to the asymmetry $(\sqrt{b_1})$ and kurtosis (b_2) values there is a strong evidence that an asymmetric model may provide a better fit for these data. We see that the skewness and kurtosis values are outside the range allowed by the SN and PN models, and even though the kurtosis value is greater than the one found in this paper for the PHN model, the latter may provide a better fit than the SN and PN models.

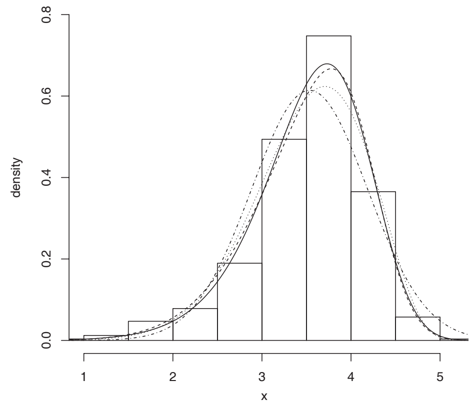
We proceed then to fit the models PN, SN and PHN to the data set. Maximum likelihood estimators for each model are presented in Table (5), with standard errors in parenthesis, obtained by inverting the observed information matrix. The Kolmogorov-Smirnov test rejects the normality assumption (p-value $= 0$); while the equality hypothesis of the roller variables' mean is not rejected (p-value$= 0.1308$), which justifies the fitness of the PHN model.

Table 5: Parameter estimators (standard error) for N, PN, SN and PHN models.

Estimates	**N**	**PN**	**SN**	**PHN**
$log(lik)$	-1135.866	-1085.241	-1071.362	-1066.994
AIC	2275.488	2176.482	2148.724	2139.988
$\hat{\xi}$	3.5347(0.0191)	4.5495(0.0572)	4.2503(0.0284)	7.0723(0.3194)
$\hat{\eta}$	0.6497(0.0135)	0.1982(0.0279)	0.9694(0.0304)	1.4380(0.0648)
$\hat{\alpha}$	–	0.0479(0.0156	-2.7864(0.2529)	86.8309(28.6166)

To implement model comparison between the models considered above, we use the AIC (Akaike Information Criterion), which penalizes the maximized likelihood function by the excess of model parameters ($AIC = -2\widehat{\ell}(\cdot) + 2k$, where k is the number of parameters in the model), see Akaike (1974).

According to this criterion the model that best fits the data is the one with the lowest AIC. By this criterion the PHN model gives the best fit to the roller data set. Graphs for the fitted models are shown in Figure 6. Figure 7-(a) shows the qqplot calculated with the roller's variable percentiles and the percentile of the PHN variable calculated with the estimates of the parameters, while Figure 7-(b) shows the empirical cumulative distribution functions and the estimated PHN model.



(a)

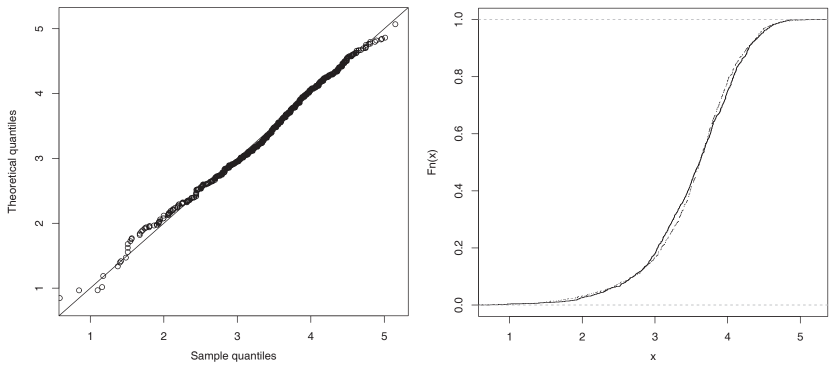
Figure 6: Graphs for distributions: $N(3.5347, 0.6497)$ (dashed dotted line), $PN(4.5495, 0.1982, 0.0479)$ (dashed line), $SN(4.2503, 0.9694, -2.7864)$ (dotted line) and $PHN(7.0723, 1.4380, 86.8309)$ (solid line).

We also conducted a hypothesis test to compare the fitness of the normal (N) model versus that of the PHN model. Formally we have the hypotheses

$$H_0 : \alpha = 1 \text{ versus } H_1 : \alpha \neq 1$$

then, using the statistic likelihood of ratio,

$$\Lambda = \frac{\ell_N(\widehat{\xi}, \widehat{\eta})}{\ell_{PHN}(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha})}$$



(a) (b)

Figure 7: (a) Q-Qplot roller variable (b) CDF, roller variable (dotted line), PHN variable (solid line).

Substituting the estimated values, we obtain

$$-2\log(\Lambda) = -2(1135.866 - 1066.994) = 137.744$$

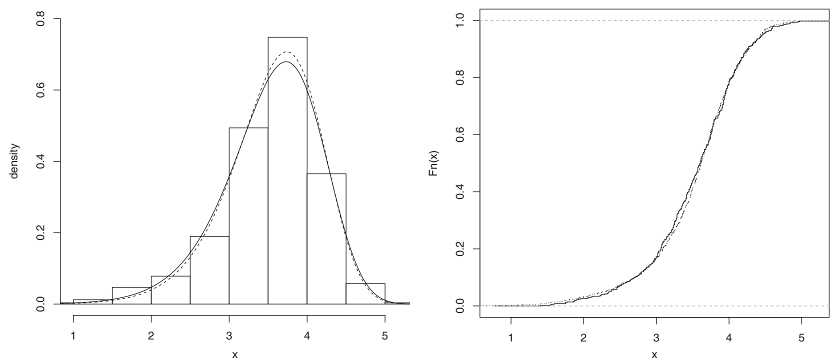
which when compared with the 95% critical value of the $\chi_1^2 = 3.84$ indicate that the null hypotheses is clearly rejected. The PHN model is a good alternative for modelling data.

According to the AIC criterion the PHN model fits the roller data better than the Normal, SN and PN models. So the PHN model captures the asymmetry and kurtosis that the other models fail to capture. A reason for this situation is in the fact that the asymmetry and kurtosis of these particular data are outside the range allowed in the SN and PN models.

We also estimated the model parameters using the two-step percentile method, obtaining: $\widehat{\xi}_p = 6.8219(0.0133)$, $\widehat{\eta}_p = 1.3574(0.0028)$ and $\widehat{\alpha}_p = 75.3801(1.0902)$ (where the estimation errors, in parentheses, were calculated with the Jackknife method). Figure 8-(a) shows the PHN densities from MLE estimation (solid line) and elemental percentile estimation (dash-dot line); Figure 8-(b) shows the corresponding cumulative density functions. Note that this method provides estimates that give a fairly good fit to the PHN model in comparison with the one fitted by maximum likelihood, but the graphs of cumulative distributions give a better fit to the distribution function estimated by maximum likelihood.

6. Concluding Remarks

We have defined a new family of distributions whose hazard function is proportional to hazard function concerning to original distribution function. We discussed several of its properties and provided and estimation of parameters via maximum likelihood, profile likelihood and elemental percentile methods. This is supported



(a) (b)

Figure 8: (a) $PHN(\xi, \eta, \alpha)$ density: MLE (solid line), estimation by elemental percentile method (dash-dot line) (b) CDF, roller variable (dotted line), PHN variable from estimation by elemental percentile method (solid line).

with an application to real data in which we show that the PHN model provides consistently better fits than the SN and PN models. The outcome of this practical demonstration shows that the new family is very general, quite flexible and widely applicable.

Acknowledgements

We gratefully acknowledge grants from Universidad de Córdoba (Colombia) and the Mobility Program of the Universidad Industrial de Santander (Colombia). We thank two anonymous referees for helpful comments.

[Recibido: junio de 2012 — Aceptado: abril de 2013]

References

Akaike, H. (1974), 'A new look at statistical model identification', *IEEE Transaction on Automatic Control* **AC-19**(6), 716–722.

Arellano-Valle, R. B., Gómez, H. W. & Quintana, F. (2004), 'A new class of skew-normal distributions', *Communications in Statistics – Theory and Methods* **33**, 1465–1480.

Arellano-Valle, R. B., Gómez, H. W. & Quintana, F. (2005), 'Statistical inference for a general class of asymmetric distributions', *Journal of Statistical Planning and Inference* **128**, 427–443.

Arnold, B., Gómez, H. & Salinas, H. (2009), 'On multiple constraint skewed models', *Statistics* **43-3**, 279–293.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**, 171–178.

Barndorff-Nielsen, O. (1983), 'On a formula for the distribution of the maximum likelihood estimator', *Biometrika* **70**(2), 343–365.

Castillo, E. & Hadi, A. (1995), 'A method for estimating parameters and quantiles of distributions of continuous random variables', *Computational Statistics and Data Analysis* **20**(4), 421–439.

Chiogna, M. (1998), 'Some results on the scalar skew-normal distribution', *Journal of the Italian Statistical Society* **1**, 1–14.

Durrans, S. R. (1992), 'Distributions of fractional order statistics in hydrology', *Water Resources Research* **28-6**, 1649–1655.

Efron, B. (1979), 'Bootstrap methods: another look at the Jackknife', *Annals of Statistics* **7**, 1–26.

Efron, B. (1982), 'The Jackknife, the Bootstrap, and other Resampling Plans', *CBMS 38, SIAM-NSF* .

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Eugene, N., Lee, C. & Famoye, F. (2002), 'Beta-normal distribution and its applications', *Communications in Statistics-Theory and Methods* **31**, 497–512.

Farias, R., Moreno, G. & Patriota, A. (2009), 'Reducción de modelos en la presencia de parámetros de perturbación', *Revista Colombiana de Estadística* **32**(1), 99–121.

Fernández, C. & Steel, M. F. J. (1998), 'On Bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association* **93**, 359–371.

Gómez, H., Venegas, O. & Bolfarine, H. (2007), 'Skew-symmetric distributions generated by the distribution function of the normal distribution', *Environmetrics* **18**, 395–407.

Gupta, A. K., Chang, F. C. & Huang, W. J. (2002), 'Some skew-symmetric models', *Random Operators Stochastic Equations* **10**, 113–140.

Gupta, D. & Gupta, R. (2008), 'Analyzing skewed data by power normal model', *Test* **17**, 197–210.

Henze, N. (1986), 'A probabilistic representation of the skew-normal distribution', *Scandinavian Journal of Statistics* **13**, 271–275.

Lehmann, E. L. (1953), 'A graphical estimation of mixed Weibull parameter in life testing electron tubes', *Technometrics* **1**, 389–407.

Mudholkar, G. S. & Hutson, A. D. (2000), 'The epsilon-skew-normal distribution for analyzing near-normal data', *Journal of Statistical Planning and Inference* **83**, 291–309.

O'Hagan, A. & Leonard, T. (1976), 'Bayes estimation subject to uncertainty about parameter constraints', *Biometrika* **63**, 201–203.

Pewsey, A. (2000), 'Problems of inference for Azzalini's skewnormal distribution', *Journal of Applied Statistics* **27-7**, 859–870.

Pewsey, A., Gómez, H. & Bolfarine, H. (2012), Likelihood based inference for distributions of fractional order statistics., *in* 'II Jornada Internacional de Probabilidad y Estadística', Pontificia Universidad Católica del Perú.

R Development Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Roberts, C. (1966), 'A correlation model useful in the study of twins', *Journal of the American Statistical Association* **61**, 1184–1190.

Sen, P. & Singer, J. (1993), *Large Sample Methods in Statistics: An Introdution with Applications*, Chapman and Hall, New York.

Sen, P., Singer, J. & Pedroso de Lima, A. (2010), *From Finite Sample to Asymptotic Methods in Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Severini, T. (1998), 'An approximation to the modified profile likelihood function', *Biometrika* **85**, 403–411.

Severini, T. A. (2000), *Likelihood Methods in Statistics*, Oxford University press.