Published in final edited form as:

Commun Stat Simul Comput. 2017 ; 46(5): 3833–3840. doi:10.1080/03610918.2015.1024857.

Letter to the editor comments on Groparu-Cojocaru and Doray (2013)

Fadoua Balabdaoui^{1,*} and Saonli Basu^{2,†}

¹CEREMADE, Université Paris-Dauphine, 75775, CEDEX 16, Paris, France

²Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

Abstract

Although estimating the five parameters of an unknown Generalized Normal Laplace (GNL) density by minimizing the distance between the empirical and true characteristic functions seems appealing, the approach cannot be advocated in practice. This conclusion is based on extensive numerical simulations in which a fast minimization procedure delivers deceiving estimators with values that are quite far away from the truth. These findings can be predicted by the very large values obtained for the true asymptotic variances of the estimators of the five parameters of the true GNL density.

Keywords

characteristic function; Generalized Normal Laplace; least squares; Pólya frequency function

In a paper recently published in this journal, Groparu-cojocaru and Doray (2013) minimize an L_2 distance between the empirical and true characteristic functions to estimate $\theta = (\mu, \sigma, \alpha, \beta, \rho) \in \mathbb{R} \times (0, \infty)^4$, the five-dimensional parameter of a Generalized Normal Laplace (GNL) distribution. We recall that a random variable X follows such a distribution if there exist $Z \sim N(0, 1)$, $Y_1 \sim \Gamma(\rho, 1)$ and $Y_2 \sim \Gamma(\rho, 1)$, which are mutually independent such that

$$X = \rho\mu + \sqrt{\rho\sigma}Z + \frac{1}{\alpha}Y_1 - \frac{1}{\beta}Y_2.$$

Unless $\rho = 1$, in which case the distribution specializes to the Normal-Laplace distribution; see e.g. Reed (2006), it is not possible to write the density of X in a closed form. This hinders for example the use of likelihood maximization to estimate the parameters. As an alternative approach, one can search for an estimate that makes the empirical and true characteristic functions, Φ_n and $\Phi(\cdot, \theta)$ respectively, close in some specified sense. Here, Φ_n is based on n i.i.d. $X_1, \dots, X_n$ from an unknown GNL distribution. Let $u(\cdot, \theta)$ and $v(\cdot, \theta)$

*Corresponding author. fadoua@ceremade.dauphine.fr.

†The research was partially supported by NIH grant R01-DA033958.

denote the real and imaginary parts of $\Phi(\cdot, \theta)$. Also, let U_n et V_n be the real and imaginary parts of Φ_n , that is $U_n(t) = n^{-1} \sum_{j=1}^n \cos(X_j t)$ and $V_n(t) = n^{-1} \sum_{j=1}^n \sin(X_j t)$. It is known that

$$v(t, \theta) = \exp\left(-\frac{\rho\sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \cos\left[\rho\left(\mu t + \arctan\left(\frac{t}{\alpha}\right) - \arctan\left(\frac{t}{\beta}\right)\right)\right]$$

and

$$v(t, \theta) = \exp\left(-\frac{\rho\sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 + \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \sin\left[\rho\left(\mu t + \arctan\left(\frac{t}{\alpha}\right) - \arctan\left(\frac{t}{\beta}\right)\right)\right]$$

for $t \in \mathbb{R}$; see Groparu-cojocaru and Doray (2013), page 1990.

For given real points $\{t_1, \dots, t_k\}$, Groparu-cojocaru and Doray (2013) define the vectors

$$Z_n = (U_n(t_1), \dots, U_n(t_k), V_n(t_1), \dots, V_n(t_k))$$

and

$$Z(\theta) = (u(t_1, \theta), \dots, u(t_k, \theta), v(t_1, \theta), \dots, v(t_k, \theta))$$

and consider minimizing the L_2 distance

$$d_n(\theta) = [Z_n - Z(\theta)]^T Q(\theta) (Z_n - Z(\theta)).$$

The matrix Q is chosen later in the paper to be equal to $\sum_n^{-1}$, a convergent estimate of the true variance-covariance matrix of Z_n (assuming that it is non-singular). To choose the numbers t_i , $i \in \{1, \dots, k\}$, Groparu-cojocaru and Doray (2013) compute the respective variances of $U_n(t) = \text{Re}(\Phi_n(t))$ and $V_n(t) = \text{Im}(\Phi_n(t))$ and their covariance. The authors write on page 1993

“As $t \rightarrow 0$, the processes $\text{Re}(\Phi_n(t))$ and $\text{Im}(\Phi_n(t))$ are uncorrelated and $(\text{Re}(\Phi_n(t)), \text{Im}(\Phi_n(t)))$ becomes a good estimate of $(\text{Re}(\Phi(t)), \text{Im}(\Phi(t)))$. Therefore, the values of t should be chosen near 0”.

As discussed below, such a choice is not always recommendable. In case the matrix $Q(\theta)$ is taken to be the identity, the obtained estimator is denoted by $\hat{\theta}$ in Groparu-cojocaru and Doray (2013). In this case, the estimator $\hat{\theta}$ simply minimizes

$$d_n(\theta) \sum_{j=1}^k \left\{ [U_n(t_j) - u(t_j, \theta)]^2 + [V_n(t_j) - v(t_j, \theta)]^2 \right\}.$$

As mentioned above, Groparu-cojocaru and Doray (2013) have also considered the estimator minimizing the weighted L_2 distance where $Q(\theta) = \sum_n^{-1}$, the best one in terms of efficiency. However, when implementing the corresponding minimization procedure, we have run into numerical instabilities probably due to the ill-conditioning of the matrix Σ_n . For the purpose of this note, we shall focus on this simpler estimator $\hat{\theta}_n$, that is, the one corresponding to $Q(\theta)$ equal to the identity, and shall compare our results to the ones reported by Groparu-cojocaru and Doray (2013) in their tables. For the true parameter is $\theta_0 = (3, 2, 2, 4, 4)$, the authors report the mean values of $\hat{\theta}$ (but not the standard deviation) based on 100 replications, where in each replication a sample of size $n = 500$ (Table 1) and $n = 1000$ (Table 2) was generated from the true GNL distribution with the parameter θ_0 . In the tables were also indicated the different choices of t_i , $1 \leq i \leq k$ labeled as Case A, B, ..., F. In Case E, where $\{t_1, \dots, t_{10}\} = \{10^{-4}, 2 \times 10^{-4}, \dots, 9 \times 10^{-4}, 10^{-3}\}$, the estimator $\hat{\theta}$ seems to perform better compared to the other Cases (A, B, C, and D), where the t_i 's were not taken to be very close to 0.

To replicate the findings of Groparu-cojocaru and Doray (2013), we have selected Case E and have implemented two programs based on gradient free optimization routines:

- in R: using the R-function “optim” which uses the “L-BFGS-B” method.
- in C++: using the Nelder-Mead (NM) simplex minimization algorithm available from the GSL library.

Because the minimization is performed over a 5-dimensional space, both codes were fed with a very large number of starting values. One chosen scheme was to generate these starting values uniformly from a given 5-dimensional rectangle. In an alternative scheme, we fi generate ρ from a standard exponential distribution. Then, we use the following facts that if $X \sim \text{GNL}$ with parameter $(\mu, \sigma, \alpha, \beta, \rho)$, then

$$\frac{E(X)}{\rho} = \mu + \frac{1}{\alpha} - \frac{1}{\beta}, \quad \frac{\text{Var}(X)}{\rho} = \sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}.$$

A random vector of dimension 3, $v = (v_1, v_2, v_3)$, is then generated from a Dirichlet distribution with parameter $(1/3, 1/3, 1/3)$ independently of ρ . If $\bar{X}_n$ and S_n^2 are the empirical mean and variance of the sample at hand, then starting values for $\mu, \sigma, \alpha, \beta$ (given the realization of ρ) are taken to be

$$\sigma = \sqrt{V_1 \frac{S_n^2}{\rho}}, \quad \alpha = 1 / \sqrt{V_2 \frac{S_n^2}{\rho}}, \quad \beta = 1 / \sqrt{V_3 \frac{S_n^2}{\rho}}, \quad \mu = \frac{\bar{X}_n}{\rho} - \frac{1}{\alpha} - \frac{1}{\beta}.$$

It is of course tempting not to restrict oneself to the sample mean and variance and estimate the higher order moments. However, this comes with the price of dealing with larger variances. Moreover, the corresponding 5-dimensional system in μ , σ , α , β and ρ is non-linear and may yield non-permissible solutions for the non-negative parameters.

In order to achieve enough accuracy, we took the number of initial values to be equal to 10,000. We found that the minimization results depend quite heavily on the chosen initial values and that the obtained estimates are quite far from the truth. To keep this letter to a reasonable length, we only report here the results obtained for the larger sample size $n = 1000$. We have used the minimization procedure implemented in C++ based on the NM simplex algorithm as explained above. The minimization procedure should stop when one of the following conditions is true:

- A minimum has been found to within the user-specified precision (tolerance limit).
- A user-specified maximum number of iterations has been reached (number of iterations).
- An error has occurred.

The handling of these conditions is under user control. To produce the results of Table 1, we have generated the starting values according to the fi scheme, from the uniform distribution $U(-4, 4) \times U(0, 3) \times U(0, 3) \times U(0, 5) \times U(0, 5)$. The results of Table 2 correspond to generating the starting values according to the second scheme described above. The parameters of the exponential and Dirichlet distributions do not seem to have an affect on the main conclusions. For both schemes, the number of replications was taken to be equal to 100, that is, samples of size $n = 1000$ were generated 100 times from the true GNL distribution with parameter $\theta_0 = (3, 2, 2, 4, 4)$. Also, the tolerance limit and maximal number of iterations of the NM procedure were taken to be equal to 10^{-5} and 100,000 respectively. If we restrict attention to the mean values obtained, neither the results of Table 1 nor those of Table 2 of the present note seem to match with those reported in Table 2 of Groparu-cojocaru and Doray (2013) for estimating the parameters, especially μ , σ^2 , α and β for Case E. Also, note that our estimates for μ , σ^2 and β exhibit a very large variation. For complete transparency, the code we used to obtain the results of Table 1 and 2 can be found at <http://www.biostat.umn.edu/~saonli>. The user can also find some explanations as to how to execute the C++ program.

The rather deceiving numerical results obtained above can be explained theoretically by the fact that choosing the grid $\{10^{-4}, 2 \times 10^{-4}, \dots, 10^{-3}\}$ yields an estimate with a large variance. To compute the asymptotic variance of $\hat{\theta}$, one can use the formula given by Heathcote (1977), page 258. For a given non-increasing weight function G , Heathcote (1977) considers the minimizer of the distance

$$D_G^n(\theta) = \int_{\mathbb{R}} |\Phi_n(t) - \Phi(t, \theta)|^2 dG(t)$$

where $|z|$ is the modulus of $z \in \mathbb{C}$. Let $\hat{\theta}_n^G$ denote the minimizer. As the weight function is allowed to be discrete, minimizing d_n is equivalent to minimizing D_n^G with G the uniform distribution on $\{t_1, \dots, t_k\}$. Consistency of $\hat{\theta}_n^G$ and asymptotic normality follow under the usual regularity conditions. Taking for granted that those conditions are satisfied as this is not diffi to establish, it turns out that we have for the estimator $\hat{\theta}$ of Groparu-cojocaru and Doray (2013)

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, \Lambda^{-1} \Sigma \Lambda^{-1})$$

where Σ is the variance-covariance matrix of the vector $(K_1(\theta_0), \dots, K_5(\theta_0))$ where for $r \in \{1, \dots, 5\}$

$$K_r(\theta_0) = \frac{1}{k} \sum_{j=1}^k \left[(\cos(t_j X) - v(t_j, \theta)) \frac{\partial v(t_i, \theta)}{\partial \theta_r} \Big|_{\theta_0} + (\sin(t_j X) - v(t_j, \theta)) \frac{\partial v(t_j, \theta)}{\partial \theta_r} \Big|_{\theta_0} \right]$$

that is, for $(r, s) \in \{1, \dots, 5\} \times \{1, \dots, 5\}$

$$\begin{aligned} \Sigma_{rs} &= \frac{1}{k^2} \sum_{j=1}^k \sum_{i=1}^k \text{cov}(\cos(Xt_i), \cos(Xt_j)) \frac{\partial u(t_i, \theta)}{\partial \theta_r} \Big|_{\theta_0} \frac{\partial u(t_j, \theta)}{\partial \theta_s} \Big|_{\theta_0} \\ &\quad + \frac{1}{k^2} \sum_{j=1}^k \sum_{i=1}^k \text{cov}(\cos(Xt_i), \sin(Xt_j)) \left[\frac{\partial u(t_i, \theta)}{\partial \theta_r} \Big|_{\theta_0} \frac{\partial v(t_j, \theta)}{\partial \theta_s} \Big|_{\theta_0} + \frac{\partial u(t_i, \theta)}{\partial \theta_s} \Big|_{\theta_0} \frac{\partial v(t_j, \theta)}{\partial \theta_r} \Big|_{\theta_0} \right] \\ &= \frac{1}{k^2} \sum_{j=1}^k \sum_{i=1}^k \text{cov}(\sin(Xt_i), \sin(Xt_j)) \frac{\partial v(t_i, \theta)}{\partial \theta_r} \Big|_{\theta_0} \frac{\partial v(t_j, \theta)}{\partial \theta_s} \Big|_{\theta_0}. \end{aligned} \quad (1)$$

Also,

$$\Lambda_{rs} = \frac{1}{k} \sum_{j=1}^k \left[\frac{\partial u(t_j, \theta)}{\partial \theta_r} \Big|_{\theta_0} \frac{\partial u(t_i, \theta)}{\partial \theta_s} \Big|_{\theta_0} + \frac{\partial v(t_i, \theta)}{\partial \theta_r} \Big|_{\theta_0} \frac{\partial v(t_j, \theta)}{\partial \theta_s} \Big|_{\theta_0} \right].$$

for $(r, s) \in \{1, \dots, 5\} \times \{1, \dots, 5\}$. The covariances of the trigonometric terms showing up in (1) can be computed using the expressions given in page 257 of Heathcote (1977) and which follow from the well-known identities

$$\cos(x) = 2^{-1} (e^{ix} + e^{-ix}) \quad \text{and} \quad \sin(x) = (2i)^{-1} (e^{ix} - e^{-ix}).$$

The partial derivatives of $(t, \theta) \mapsto u(t, \theta)$ and $(t, \theta) \mapsto v(t, \theta)$ with respect to θ_r , $r = 1, \dots, 5$ can be found in Appendix. We have used the *LinearAlgebra* library of Maple to compute the product $\Lambda^{-1} \Sigma \Lambda^{-1}$ to fi that for $\theta_0 = (3, 2, 2, 2, 4)$ and $t_i = 10^{-4}i$, $i = 1, \dots, 10$

$$\Lambda^{-1}\Sigma\Lambda^{-1} = \begin{pmatrix} 6.9228 \times 10^7 & -3.4876 \times 10^5 & 1.1294 \times 10^7 & -8.0864 \times 10^8 & 1.9525 \times 10^7 \\ -3.4876 \times 10^5 & 70817.8877 & 3.37900 \times 10^5 & 6.7757 \times 10^6 & 12008.07809 \\ 1.1294 \times 10^7 & 3.3790 \times 10^5 & 1.0577 \times 10^7 & -1.118 \times 10^8 & -2.0428 \times 10^6 \\ -8.0864 \times 10^8 & 6.7757 \times 10^6 & 1.118 \times 10^8 & 9.5546 \times 10^9 & 2.2586 \times 10^8 \\ 1.9525 \times 10^7 & 12008.07809 & -2.0428 \times 10^6 & 2.2586 \times 10^8 & 6.0286 \times 10^6 \end{pmatrix}$$

A quick calculation shows that one should have a sample size at least equal to $n = 2.659 \times 10^{12}$ to have $\hat{\mu} - 3 \in [-0.01, 0.01]$ with probability 0.95! Hence, although convergence at $1/n$ is theoretically guaranteed, the very large samples required makes its use not reasonable for any practical purpose.

How to choose the weight function G without too much loss of efficiency has been the main question raised by Heathcote (1977). In the example of estimating a GNL density, it appears that the choice $t_i = 10^{-4}i$, $i \in \{1, \dots, 10\}$ yields estimates of very large variations. At the same time, it is not known whether compared to other configurations A, B, C and D, this choice would be better. Having said this, choosing the t_i 's close to 0 is not always something to recommend in general. In the unidimensional case, Heathcote (1977) finds very elegant sufficient conditions for the weight function G so that $\hat{\theta}^G$ achieves a reasonable relative efficiency. Those conditions are stated in terms of $\rho_1(t, \theta)$ and $\rho_2(t, \theta)$, the respective correlations of $\cos(tX)$ and $\sin(tX)$ and $\log f(\cdot, \theta)/\theta$. The rationale behind studying these quantities is the following: if one would like $\hat{\theta}^G$ to have a variance that is not very large compared to $1/I(\theta_0)$, the inverse of the Fisher information, the real and imaginary parts of the estimated characteristic function should be well correlated with the score function $\log f(\cdot, \theta)/\theta$. This correlation can be more easily studied through the components $\rho_1(t, \theta)$ and $\rho_2(t, \theta)$. The weight function G should be chosen so that it puts enough weight in a region around the value of t where these correlations are high. More explicitly, Heathcote (1977) shows that in case there exists a constant $c \in (0, 1]$ such that $|\rho_1(t, \theta)|, |\rho_2(t, \theta)| \in [c, 1]$ for $t \in I$ where $G(t) > 0$ for all $t \in I$ then the relative efficiency of the estimator $\hat{\theta}_n^G$ is c^2 .

This makes the link to the question raised by Groparu-cojocaru and Doray (2013) as to how to choose the numbers t_i 's. Heathcote (1977) considers a nice example on the symmetric stable law with index parameter $\theta \in (0, 2]$, whose characteristic function is given by $\Phi(t, \theta) = \exp(-|t|^\theta)$, $t \in \mathbb{R}$. When $\theta \in (0, 2)$, then one can show that $\rho_1(t, \theta) \rightarrow 0$ as $t \rightarrow 0$. Hence, for this range of θ , it may not be a good idea to choose a weight function G that assigns substantial weight to points near the origin. In this example, it appears that the choice $t_i = 10^{-4}i$, $i \in \{1, \dots, 10\}$ would have been not a good option. In multivariate estimation problems like the one considered by Groparu-cojocaru and Doray (2013), extending the sufficient conditions Groparu-cojocaru and Doray (2013) seems to be nontrivial. We believe that the additional complexity stems from the fact that one has to deal with the inverse of the Fisher information matrix. This inversion makes the link between the variance terms of the components of the Maximum Likelihood estimator and those of the estimator $\hat{\theta}^G$ obvious. Thus, the problem remains open.

To conclude this letter, we think that the interesting problem of estimating the 5 parameters of a GNL density via minimizing an L_2 distance between the empirical and true characteristic functions seems to be hard to implement in practice. The impracticality of the approach is mainly due to the fact very large sample sizes are needed to see convergence. The other point is the difficulty found in implementing the minimization method which appears to be quite sensitive with respect to the choice of the starting value. It is interesting to note that in case one is ready to assume that ρ is an integer, then it is not difficult to show that the corresponding GNL density is a Pólya frequency (PF) using the characterization by Schoenberg (1951) of such special densities. Thus, using the fact that any PF is log-concave (see also Schoenberg (1951)), estimating the true GNL density via the nonparametric log-concave Maximum Likelihood estimator may be appealing, see e.g. Dümbgen and Rufibach (2009). One may even use this nonparametric approach to obtain a smoother estimate of the characteristic function (instead of the empirical one) which could yield a better moment estimates of the unknown parameters.

Appendix

The partial derivatives of u and v with respect to θ . In the following, we write

$$\Psi(t) = \rho \left(\mu t + \arctan\left(\frac{t}{\alpha}\right) - \arctan\left(\frac{t}{\beta}\right) \right).$$

Then, for all $t \in \mathbb{R}$ and $\theta \in \mathbb{R} \times (0, \infty)^4$ we have that

$$\begin{aligned} \frac{\partial u(t, \theta)}{\partial \theta_1} &= \frac{\partial u(t, \theta)}{\partial \mu} \\ &= -\rho t \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \sin(\Psi(t)) \\ \frac{\partial u(t, \theta)}{\partial \theta_2} &= \frac{\partial u(t, \theta)}{\partial \sigma} = -\rho \sigma t^2 u(t, \theta) \\ \frac{\partial u(t, \theta)}{\partial \theta_3} &= \frac{\partial u(t, \theta)}{\partial \alpha} \\ &= \frac{\rho t}{t^2 + \alpha^2} \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \left\{ \frac{t}{\alpha} \cos(\Psi(t)) + \sin(\Psi(t)) \right\} \\ \frac{\partial u(t, \theta)}{\partial \theta_4} &= \frac{\partial u(t, \theta)}{\partial \beta} \\ &= \frac{\rho t}{t^2 + \beta^2} \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \left\{ \frac{t}{\beta} \cos(\Psi(t)) + \sin(\Psi(t)) \right\} \\ \frac{\partial u(t, \theta)}{\partial \theta_5} &= \frac{\partial u(t, \theta)}{\partial \rho} \\ &= -\exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \times \\ &\quad \left\{ -\frac{\sigma t^2 + \log(1 + t^2/\alpha^2) + \log(1 + t^2/\beta^2)}{2} \cos(\Psi(t)) + \Psi(t) \sin(\Psi(t)) \right\} \end{aligned}$$

and

$$\begin{aligned}
\frac{\partial v(t, \theta)}{\partial \theta_1} &= \frac{\partial v(t, \theta)}{\partial \mu} \\
&= \rho t \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \cos(\Psi(t)) \\
\frac{\partial v(t, \theta)}{\partial \theta_2} &= \frac{\partial v(t, \theta)}{\partial \sigma} = -\rho \sigma t^2(t, \theta) \\
\frac{\partial v(t, \theta)}{\partial \theta_3} &= \frac{\partial v(t, \theta)}{\partial \alpha} \\
&= \frac{\rho t}{t^2 + \alpha^2} \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \left\{\frac{t}{\alpha} \sin(\Psi(t)) + \cos(\Psi(t))\right\} \\
\frac{\partial v(t, \theta)}{\partial \theta_4} &= \frac{\partial v(t, \theta)}{\partial \beta} \\
&= \frac{\rho t}{t^2 + \beta^2} \exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \left\{\frac{t}{\beta} \sin(\Psi(t)) + \cos(\Psi(t))\right\} \\
\frac{\partial v(t, \theta)}{\partial \theta_5} &= \frac{\partial v(t, \theta)}{\partial \beta} \\
&= -\exp\left(-\frac{\rho \sigma^2 t^2}{2}\right) \left[\frac{\alpha^2 \beta^2}{(\alpha^2 + t^2)(\beta^2 + t^2)}\right]^{\rho/2} \times \\
&\quad \left\{-\frac{\sigma t^2 + \log(1 + t^2/\alpha^2) + \log(1 + t^2/\beta^2)}{2} \sin(\Psi(t)) + \Psi(t) \cos(\Psi(t))\right\}
\end{aligned}$$

References

- Dümbgen L, Rufibach K. Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli*. 2009; 15:40–68.
- Groparu-cojocaru I, Doray LG. Inference for the Generalized Normal Laplace distribution. 2013; 42:1989–1997.
- Heathcote CR. The integrated squared error estimation of parameters. *Biometrika*. 1977; 64:255–264.
- Reed, WJ. *Advances in distribution theory, order statistics, and inference*. Stat. Ind. Technol., Birkhäuser Boston; Boston, MA: 2006. The normal-Laplace distribution and its relatives; p. 61-74.
- Schoenberg IJ. On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J. Analyse Math*. 1951; 1:331–374.

Table 1

Mean and standard deviation of the components of the estimator $\mathcal{J}$ computed using the Nelder-Mead Simplex minimization algorithm with tolerance and maximal number of iterations equal to 10^{-5} and 100,000 respectively. The number of replications was taken to be equal to 100, and the starting values were generated according to the first scheme. See text for more details.

Parameter	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$d_n(\hat{\theta})$
Mean	93.362	1350114	2.707	406.053	3.525	4.469×10^{-22}
Standard deviation	845.697	1348136	6.123	3953.239	4.827	6.590×10^{-22}

Table 2

Mean and standard deviation of the components of the estimator $\hat{\theta}$ computed using the Nelder-Mead Simplex minimization algorithm with tolerance and maximal number of iterations equal to 10^{-5} and 100,000 respectively. The number of replications was taken to be equal to 100 and the starting values were generated according to the second scheme. See text for more details.

Parameter	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$d_n(\hat{\theta})$
Mean	32.667	11904.89	26.150	2964.35	1.769	5.233×10^{-22}
Standard deviation	83.077	59858.84	116.692	12262.05	1.426	7.301×10^{-22}