

Estimation of $P(Y < X)$ for the Lévy distribution

Hosein Najarzadegan * Saman Babaii † Sadegh Rezaei ‡ Saralees Nadarajah § ¶

Abstract

Three point estimators and two interval estimators of $P(Y < X)$ are derived when X and Y are independent Lévy random variables. Their performance with respect to relative biases, relative mean squared errors, coverage probabilities, and coverage lengths is assessed by simulation studies and a real data application.

Keywords: Bayesian estimator, Bootstrap confidence interval, Lévy distribution, Maximum likelihood estimator, Uniformly minimum variance unbiased estimator.

2000 AMS Classification: 62G05

Received: 08.11.2014 *Accepted:* 23.03.2015 *Doi:* 10.15672/HJMS.2015599877

1. Introduction

Let X be a Lévy random variable with scale parameter σ_x . Then the probability density function (pdf) and the cumulative distribution function (cdf) of X are:

$$f(x, \sigma_x) = \sqrt{\frac{\sigma_x}{2\pi}} x^{-\frac{3}{2}} \exp\left(-\frac{\sigma_x}{2x}\right)$$

and

$$F(x, \sigma_x) = 2 \left[1 - \Phi\left(\sqrt{\frac{\sigma_x}{x}}\right) \right],$$

respectively, for $x > 0$ and $\sigma_x > 0$, where $\Phi(\cdot)$ denotes the standard normal cdf. According to O'Reilly and Rueda [28], $\frac{1}{X}$ is a gamma random variable with shape parameter $\frac{1}{2}$ and scale parameter $\frac{2}{\sigma}$. Lévy distribution has no moments.

*Department of Statistics, Amirkabir University of Technology, Tehran, IRAN

†Department of Statistics, Amirkabir University of Technology, Tehran, IRAN

‡Department of Statistics, Amirkabir University of Technology, Tehran, IRAN

§School of Mathematics, University of Manchester, Manchester M13 9PL, UK, Email: mbbssn2@manchester.ac.uk

¶Corresponding Author.

Lévy distribution and the more general alpha-stable distribution have received applications in many areas, including dispersive transport in disordered semiconductors, stock and stock-indexes returns, linear dynamical systems, income distribution, stochastic artificial neural networks, many-particle quantum systems, oil pricing time-series, distributions of stochastic payoff variations, real traffic flow, satellite magnetic field measurements, models for circular data, models of asset trading, directed percolation with incubation times, earthquake slip spatial distributions, models for financial markets with central regulation, long correlation times in supermarket sales, edge turbulence of fusion devices, network traffic behavior in switched Ethernet systems, modeling individual behavior in a large marine predator, evolutionary programming using mutations, distribution of marks in high school, fractal structures, models for fish locomotion, distribution of economical indices, south Spain seismic series, geophysical data analysis, supermarket sales, velocity difference in systems of vortex elements, currency exchange market, random field models for geological heterogeneity, structural reorganization in rice piles, and wave scattering from self-affine surfaces. Three of the most recent applications relate to daily price fluctuations in the Mexican financial market index (Alfonso *et al.* [1]), observations of anomalous diffusion (Sagi *et al.*, 2012), and bistable systems (Srokowski [33]).

In the stated areas, it is of interest to estimate the probability $R = P(Y < X)$ when X and Y are independent Lévy random variables. For example, X and Y could represent: stock returns for two different commodities; oil prices in two different countries; traffic at two different locations; earthquake magnitudes at two different locations; marks at two different high schools; and, so on.

Estimation of $P(Y < X)$ is widely known as stress-strength modeling: if X denotes the stress that a system is subjected to and Y the strength of the system then $P(Y < X)$ is the probability of the failure of the system. Many papers have investigated estimation of $P(Y < X)$ when X and Y arise from a specific distribution. For details, see Awad and Gharraf [4], Surles and Padgett [34] for the case X, Y are Burr distributed; Constantine *et al.* [11], Ismail *et al.* [20] for the case X, Y are gamma distributed; Obradovic *et al.* [27] for the case X, Y are geometric-Poisson distributed; Babayi *et al.* [5] for the case X, Y are generalized logistic distributed; Kundu and Raqab [22] for the case X, Y are generalized Rayleigh distributed; Saracoglu *et al.* [32] for the case X, Y are Gompertz distributed; Nadar *et al.* [25] for the case X, Y are Kumaraswamy distributed; Downtown [14], Reiser and Guttman [30] for the case X, Y are normal distributed; Genc [17] for the case X, Y are Topp-Leone distributed; McCool [24] for the case X, Y are Weibull distributed. There are also semiparametric and nonparametric methods for estimating $P(Y < X)$. Kotz *et al.* [21] provide an excellent review of known work.

There has not been much work on the estimation of $R = P(Y < X)$ when X and Y are independent Lévy random variables. The only paper we are aware of is Ali and Woo [3]. But the estimators given in Ali and Woo [3] are not those for R . A related paper by Ali *et al.* [2] studies the distribution of $X/(X + Y)$.

In this note, we provide point as well as interval estimators for $R = P(Y < X)$. The point estimators considered are: maximum likelihood estimator, uniformly minimum variance unbiased estimator (UMVUE) and Bayes estimator taken as the mean of the posterior distribution of R given suitable priors. The interval estimators considered are: asymptotic maximum likelihood estimator and bootstrap based percentile estimator. The performance of these estimators is assessed by simulation studies as well as by a real data application.

2. Point estimators of R

In this section, we give three point estimators for R . Their performances are compared by a simulation study in Section 5.1. Throughout, we suppose X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m are independent random samples from the Lévy distribution with scale parameters σ_x and σ_y , respectively.

2.1. Maximum likelihood estimator of R . The maximum likelihood estimators of σ_x and σ_y are:

$$\hat{\sigma}_x = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

and

$$\hat{\sigma}_y = \frac{m}{\sum_{j=1}^m \frac{1}{Y_j}},$$

respectively. Ali and Woo [3] show that:

$$R = \frac{2}{\pi} \sin^{-1} \frac{1}{\sqrt{1 + \frac{\sigma_y}{\sigma_x}}}.$$

Thus, the maximum likelihood estimator of R follows by the invariance property:

$$(2.1) \quad \hat{R} = \frac{2}{\pi} \sin^{-1} \frac{1}{\sqrt{1 + \frac{\hat{\sigma}_y}{\hat{\sigma}_x}}}.$$

2.2. UMVUE of R . To find the UMVUE of R , we use results in Ismail *et al.* [20]. It is easy to see that $\left(\sum_{i=1}^n 1/X_i, \sum_{j=1}^m 1/Y_j\right)$ is complete and sufficient for (σ_x, σ_y) . Let

$$\Phi(X, Y) = \begin{cases} 1, & \text{if } \frac{1}{X} < \frac{1}{Y}, \\ 0, & \text{if } \frac{1}{X} > \frac{1}{Y}. \end{cases}$$

Then, one can see that $\Phi(X, Y)$ is an unbiased estimator of R . It follows by Lehmann-Scheffe theorem (see page 369 in Casella [8]) that:

$$\tilde{R} = E \left(\Phi(X, Y) \left| \sum_{i=1}^n \frac{1}{X_i}, \sum_{j=1}^m \frac{1}{Y_j} \right. \right)$$

is an UMVUE. Since $\frac{1}{X}, \frac{1}{Y}, \sum_{i=1}^n \frac{1}{X_i}, \sum_{j=1}^m \frac{1}{Y_j}$ are gamma random variables, we have from Ismail *et al.* [20] that

$$(2.2) \quad \tilde{R} = \begin{cases} \int_0^1 F_{W_2} \left(\frac{U}{V} w_1 \right) f_{W_1}(w_1) dw_1, & \text{if } U \leq V, \\ \int_0^{\frac{V}{U}} F_{W_2} \left(\frac{U}{V} w_1 \right) f_{W_1}(w_1) dw_1 + 1 - F_{W_1} \left(\frac{V}{U} \right), & \text{if } U > V, \end{cases}$$

where $W_1 \sim \text{Beta}\left(\frac{1}{2}, \frac{n-1}{2}\right)$ and $W_2 \sim \text{Beta}\left(\frac{1}{2}, \frac{m-1}{2}\right)$ are beta random variables, $U = \sum_{i=1}^n \frac{1}{X_i}$ and $V = \sum_{j=1}^m \frac{1}{Y_j}$. If $W \sim \text{Beta}(a, b)$ then its cdf is the incomplete beta function ratio defined by $I_w(a, b) = \int_0^w t^{a-1}(1-t)^{b-1} dt / B(a, b)$, where $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ denotes the beta function. So, (2.2) can be expressed as

$$\tilde{R} = \begin{cases} \frac{1}{B(1/2, (n-1)/2)} \int_0^1 I_{Uw/V} \left(\frac{1}{2}, \frac{m-1}{2}\right) w^{-1/2} (1-w)^{(n-3)/2} dw, \\ \text{if } U \leq V, \\ \\ \frac{1}{B(1/2, (n-1)/2)} \int_0^{V/U} I_{Uw/V} \left(\frac{1}{2}, \frac{m-1}{2}\right) w^{-1/2} (1-w)^{(n-3)/2} dw \\ + 1 - I_{V/U} \left(\frac{1}{2}, \frac{n-1}{2}\right), \\ \text{if } U > V. \end{cases} \tag{2.3}$$

An alternative expression using the series expansion

$$I_w(a, b) = \frac{w^a}{B(a, b)} \sum_{k=0}^{\infty} \frac{(1-b)_k w^k}{(a+k)k!},$$

where $(e)_k = e(e+1) \cdots (e+k-1)$ denotes the ascending factorial, is

$$\tilde{R} = \begin{cases} \frac{1}{B(1/2, (n-1)/2) B(1/2, (m-1)/2)} \\ \cdot \sum_{k=0}^{\infty} \frac{((3-m)/2)_k}{(k+1/2) k!} B\left(k+1, \frac{n-1}{2}\right) \left(\frac{U}{V}\right)^{k+1/2}, \\ \text{if } U \leq V, \\ \\ \frac{1}{B(1/2, (n-1)/2) B(1/2, (m-1)/2)} \\ \cdot \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{((3-m)/2)_k ((3-n)/2)_\ell}{(k+1/2) (k+\ell+1) k! \ell!} \left(\frac{V}{U}\right)^\ell \\ + 1 - \frac{1}{B(1/2, (n-1)/2)} \sum_{k=0}^{\infty} \frac{((3-n)/2)_k}{(k+1/2) k!} \left(\frac{V}{U}\right)^{k+1/2}, \\ \text{if } U > V. \end{cases} \tag{2.4}$$

This expression can be used to compute measures like the variance, skewness and kurtosis of \tilde{R} . For example, using equation (6.455.1) in Gradshteyn and Ryzhik [18], one can show

that

$$\begin{aligned}
E(\tilde{R}^2) &= \frac{1}{B^2(1/2, (n-1)/2) B^2(1/2, (m-1)/2)} \\
&\quad \cdot \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{((3-m)/2)_k ((3-m)/2)_\ell}{(k+1/2) k! (\ell+1/2) \ell!} \\
&\quad \cdot B\left(k+1, \frac{n-1}{2}\right) B\left(\ell+1, \frac{n-1}{2}\right) \\
&\quad \cdot I(k+\ell+1, -k-\ell-1) \\
&+ \frac{1}{B^2(1/2, (n-1)/2) B^2(1/2, (m-1)/2)} \\
&\quad \cdot \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{((3-m)/2)_k}{(k+1/2) k!} \\
&\quad \cdot \frac{((3-n)/2)_\ell ((3-m)/2)_p ((3-n)/2)_q}{(k+\ell+1) \ell! (p+1/2) (p+q+1) p! q!} J(\ell+q, -\ell-q) \\
&+ 1 + \frac{1}{B^2(1/2, (n-1)/2)} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{((3-n)/2)_k}{(k+1/2) k!} \\
&\quad \cdot \frac{((3-n)/2)_\ell}{(\ell+1/2) \ell!} J(k+\ell+1, -k-\ell-1) \\
&+ \frac{2}{B(1/2, (n-1)/2) B(1/2, (m-1)/2)} \\
&\quad \cdot \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{((3-m)/2)_k}{(k+1/2) k!} \\
&\quad \cdot \frac{((3-n)/2)_\ell}{(k+\ell+1) \ell!} J(\ell, -\ell) \\
&- \frac{2}{B^2(1/2, (n-1)/2) B(1/2, (m-1)/2)} \\
&\quad \cdot \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{p=0}^{\infty} \frac{((3-m)/2)_k}{(k+1/2) k!} \\
&\quad \cdot \frac{((3-n)/2)_\ell ((3-n)/2)_p}{(k+\ell+1) \ell! (p+1/2) p!} J\left(\ell+p+\frac{1}{2}, -\ell-p-\frac{1}{2}\right) \\
&- \frac{2}{B(1/2, (n-1)/2)} \sum_{k=0}^{\infty} \frac{((3-n)/2)_k}{(k+1/2) k!} J\left(k+\frac{1}{2}, -k-\frac{1}{2}\right),
\end{aligned}$$

where

$$\begin{aligned}
I(\alpha, \beta) &= \frac{2^{\alpha+\beta} \sigma_x^{n/2} \sigma_y^{m/2} \Gamma(\alpha+\beta+(m+n)/2)}{(\sigma_x+\sigma_y)^{\alpha+\beta+(m+n)/2} (\alpha+n/2) \Gamma(m/2) \Gamma(n/2)} \\
&\quad \cdot {}_2F_1\left(1, \alpha+\beta+\frac{m+n}{2}; \alpha+\frac{n}{2}+1; \frac{\sigma_x}{\sigma_x+\sigma_y}\right)
\end{aligned}$$

and

$$J(\alpha, \beta) = \frac{2^{\alpha+\beta} \Gamma(\alpha+n/2) \Gamma(\beta+m/2)}{\sigma_x^\alpha \sigma_y^\beta \Gamma(m/2) \Gamma(n/2)} - I(\alpha, \beta),$$

where

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$$

and

$${}_2F_1(a, b; c; x) = \sum_{k=0}^\infty \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}$$

denote the gamma and Gauss hypergeometric functions, respectively. So, the variance of \tilde{R}^2 is $E(\tilde{R}^2) - R^2$.

2.3. Bayes estimator of R . Suppose the scale parameters, σ_x and σ_y , have the following gamma priors:

$$\sigma_x \sim \Gamma\left(\frac{r_1}{2}, \lambda_1\right) \quad \text{and} \quad \sigma_y \sim \Gamma\left(\frac{r_2}{2}, \lambda_2\right).$$

There are several reasons why we have chosen gamma priors: i) the resulting posterior pdfs of σ_x and σ_y ,

$$\sigma_x|x \sim \Gamma\left(\frac{n+r_1}{2}, \lambda_1 + \frac{1}{2}u\right) \quad \text{and} \quad \sigma_y|y \sim \Gamma\left(\frac{m+r_2}{2}, \lambda_2 + \frac{1}{2}v\right),$$

where $u = \sum_{i=1}^n \frac{1}{X_i}$ and $v = \sum_{i=1}^m \frac{1}{Y_i}$, belong to the same class; ii) According to Felsenstein [16], assuming a prior distribution “of rates such as a gamma distribution or lognormal distribution has deservedly been popular”; iii) According to Lambert *et al.* [23], gamma priors are “the most common used prior distribution for variance parameters, not least because it is used in many of the examples provided with the WinBUGS software”; iv) According to page 69 in Congdon [10], there has “been considerable debate about appropriate priors for variance and precision parameters ... the most common option is a gamma”; v) According to Dorfman and Karali [13], the gamma prior “on the error variance term is a standard one”.

If we suppose σ_x and σ_y are independent then the joint posterior pdf of σ_x and σ_y is:

$$\begin{aligned} f(\sigma_x, \sigma_y|x, y) &= \sigma_x^{\frac{r_1+n}{2}-1} \frac{\left(\frac{1}{2}u + \lambda_1\right)^{\frac{r_1+n}{2}}}{\Gamma\left(\frac{r_1+n}{2}\right)} \exp\left(-\sigma_x \left[\frac{1}{2}u + \lambda_1\right]\right) \\ &\quad \cdot \exp\left(-\sigma_y \left[\frac{1}{2}v + \lambda_2\right]\right) \sigma_y^{\frac{r_2+m}{2}-1} \frac{\left(\frac{1}{2}v + \lambda_2\right)^{\frac{r_2+m}{2}}}{\Gamma\left(\frac{r_2+m}{2}\right)}. \end{aligned}$$

Thus, the posterior pdf of R is:

$$f_R(r|x, y) = C \frac{\cot\left(\frac{\pi}{2}r\right)^{r_2+m-1} \left[1 + \cot\left(\frac{\pi}{2}r\right)^2\right]}{\left[\left(\frac{1}{2}u + \lambda_1\right) + \left(\frac{1}{2}v + \lambda_2\right) \cot\left(\frac{\pi}{2}r\right)^2\right]^{\frac{n+m+r_1+r_2}{2}}},$$

where

$$C = \pi \frac{\left(\frac{1}{2}u + \lambda_1\right)^{\frac{r_1+n}{2}} \left(\frac{1}{2}v + \lambda_2\right)^{\frac{r_2+m}{2}}}{\Gamma\left(\frac{r_1+n}{2}\right) \Gamma\left(\frac{r_2+m}{2}\right)} \Gamma\left(\frac{r_1+r_2+m+n}{2}\right).$$

Under the mean squared error loss function, the Bayes estimator of R is:

$$\begin{aligned}
 \widehat{R}_{Bayes} &= \int_0^1 r f_R(r|x, y) dr \\
 (2.5) \quad &= C \int_0^1 r \frac{\cot\left(\frac{\pi}{2}r\right)^{r_2+m-1} \left[1 + \cot\left(\frac{\pi}{2}r\right)^2\right]}{\left[\left(\frac{1}{2}u + \lambda_1\right) + \left(\frac{1}{2}v + \lambda_2\right) \cot\left(\frac{\pi}{2}r\right)^2\right]^{\frac{n+m+r_1+r_2}{2}}} dr.
 \end{aligned}$$

Analytical solutions for the above integral are not available.

3. Interval estimators of R

In this section, we give two interval estimators for R . Their performances are compared by a simulation study in Section 5.2. Throughout, we suppose X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m are independent random samples from the Lévy distribution with scale parameters σ_x and σ_y , respectively.

3.1. Asymptotic confidence interval. For large sample sizes, a confidence interval for R can be obtained based on maximum likelihood estimation. For this purpose, we first obtain an asymptotic distribution of the maximum likelihood estimators, $\widehat{\sigma}_x$ and $\widehat{\sigma}_y$.

1. Theorem. If $n \rightarrow \infty$ and $m \rightarrow \infty$ such that $\frac{n}{m} \rightarrow p$ then

$$(\sqrt{n}(\widehat{\sigma}_x - \sigma_x), \sqrt{m}(\widehat{\sigma}_y - \sigma_y)) \rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2\sigma_x^2 & 0 \\ 0 & 2\sigma_y^2 \end{pmatrix}.$$

Proof. The proof is straightforward using asymptotic normality of $\widehat{\sigma}_x$ and $\widehat{\sigma}_y$. □

The asymptotic distribution of \widehat{R} can now be easily deduced.

2. Theorem. If $n = m$ and $n \rightarrow \infty$ then

$$\sqrt{n}(\widehat{R} - R) \rightarrow N(0, D),$$

where

$$D = \frac{4\sigma_x\sigma_y}{\pi^2(\sigma_x + \sigma_y)^2}.$$

Hence, a 95 percent asymptotic confidence interval for R is

$$(3.1) \quad \left(\widehat{R} - 1.96\sqrt{\frac{D}{n}}, \widehat{R} + 1.96\sqrt{\frac{D}{n}} \right).$$

Proof. Follows by the delta method (see pages 33-35 of Davison [12]). □

3.2. Bootstrap confidence interval. Bootstrap confidence intervals are useful for small sample sizes. Here, we propose a percentile based bootstrap confidence interval due to Efron [15]. It can be constructed by the following scheme:

- (1) From the samples X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m , compute the maximum likelihood estimates, $\hat{\sigma}_x$ and $\hat{\sigma}_y$;
- (2) Using $\hat{\sigma}_x$, generate a bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$ and similarly using $\hat{\sigma}_y$ generate a bootstrap sample $Y_1^*, Y_2^*, \dots, Y_m^*$. The inversion method was used to generate samples: this entails inverting the standard normal cdf and routines for this inversion are widely available even in pocket calculators. From the samples $X_1^*, X_2^*, \dots, X_n^*$ and $Y_1^*, Y_2^*, \dots, Y_m^*$, compute the maximum likelihood estimate of R , say \hat{R}^* ;
- (3) Repeat step 2, B times, giving the estimates, say $\hat{R}_1^*, \hat{R}_2^*, \dots, \hat{R}_B^*$, of R ;
- (4) Compute the empirical cdf, say $\hat{G}(\cdot)$, of $\hat{R}_1^*, \hat{R}_2^*, \dots, \hat{R}_B^*$. Then an approximate 95 percent confidence interval of R is

$$(3.2) \quad \left[\hat{G}^{-1}(0.025), \hat{G}^{-1}(0.975) \right],$$

where $\hat{G}^{-1}(\cdot)$ denotes the inverse function of $\hat{G}(\cdot)$.

Another bootstrap based interval is the bootstrap- t confidence interval for R . We shall not consider this here as it performed similarly to the percentile based bootstrap confidence interval.

4. A real data application

As mentioned in Section 1, one application of the Lévy distribution is to model stock index data. Here, we discuss such an application.

The data are S&P/IFC (Standard & Poor's / International Finance Corporation) global daily price indices in United States dollars for Egypt and South Africa, the two largest economies in Africa. The data cover the period from the 1st of January 1996 to the 31st of October 2008. The data were obtained from the database Datastream.

Following common practice, daily log returns were computed as first order differences of logarithms of daily price indices. Let X denote the daily log returns from South Africa and Y the daily log returns from Egypt. Some summary statistics for the data on X are: range = 0.078847, first quartile = 0.020640, median = 0.026720, and third quartile = 0.034270. Some summary statistics for the data on Y are: range = 0.086333, first quartile = 0.017030, median = 0.024120, and third quartile = 0.036920. The sample size for both data sets is 153.

The Lévy distribution was fitted to the data on X and Y by the method of maximum likelihood. We obtained the estimates $\hat{\sigma}_x = 0.02392927$ and $\hat{\sigma}_y = 0.01898238$. The chisquare and Kolmogorov-Smirnov tests for the fit to the log returns from South Africa gave the p -values 0.061 and 0.063. The chisquare and Kolmogorov-Smirnov tests for the fit to the log returns from Egypt gave the p -values 0.051 and 0.077. Since the Kolmogorov-Smirnov test assumes that the fitted distribution gives the "true" parameter values, the p -values were computed using Monte Carlo simulation.

Using the fitted estimates of σ_x and σ_y , we were able to compute $R = P(X < Y)$ using the three point estimation methods. For the maximum likelihood method, we obtained $\hat{R} = 0.5367768$. For the UMVUE, we obtained $\hat{R} = 0.5368976$. For the Bayes method, we obtained $\hat{R} = 0.5367891$. It is remarkable that all three estimates are identical up to the first three decimal places. We took $\lambda_1 = \lambda_2 = 1$ and $r_1 = r_2 = 1$ for the Bayes method. Other choices gave similar results.

Using the fitted estimates of σ_x and σ_y , we were also able to compute $R = P(X < Y)$ using the two interval estimation methods. Using the asymptotic method, we obtained the 95 percent confidence interval (0.4866747, 0.5868788). Using the bootstrap method, we obtained the 95 percent confidence interval (0.4969806, 0.577812). The coverage length is smaller for the bootstrap method. We took $B = 500$ for the bootstrap method. Other choices gave similar results.

Both the confidence intervals contain $R = 0.5$ as a real value. Hence, there is no evidence that the daily log returns differ significantly between South Africa and Egypt. Further statistical analysis of the data set can be found in Nadarajah *et al.* [26].

5. Simulation studies

5.1. Simulation study for point estimators of R . Here, we perform a simulation study to compare the performances of the maximum likelihood estimator, the UMVUE and the Bayes estimator of R . The performance was assessed in terms of relative biases and relative mean squared errors. The following scheme was used:

- (1) Generate ten thousand samples of $\{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m\}$;
- (2) Compute the estimators, (2.1), (2.2) and (2.5), for each of the ten thousand samples, say R_{1i}, R_{2i}, R_{3i} for $i = 1, 2, \dots, 10000$. (2.2) and (2.5) were computed using the function `integrate` in R (R Development Core Team [29]);
- (3) Compute the relative biases for the three estimators as

$$\text{Bias}_j = \frac{1}{10000} \sum_{i=1}^{10000} (R_{ji} - R) / R$$

for $j = 1, 2, 3$;

- (4) Compute the relative mean squared errors for the three estimators as

$$\text{MSE}_j = \frac{1}{10000} \sum_{i=1}^{10000} (R_{ji} - R)^2 / R$$

for $j = 1, 2, 3$.

We repeated this scheme for $m = n = 2, 3, \dots, 100$ and $(\sigma_x, \sigma_y) = (1, 1), (1, 2), (1, 5), (2, 2), (2, 5), (5, 5)$. For the Bayes estimator, we took $\lambda_1 = \lambda_2 = 1$ and $r_1 = r_2 = 1$, as in Section 4. Plots of the relative biases, $\text{bias}_1, \text{bias}_2$ and bias_3 , versus n are shown in Figure 1. Plots of the relative mean squared errors, $\text{MSE}_1, \text{MSE}_2$ and MSE_3 , versus n are shown in Figure 2. The red line in Figure 1 represents the relative biases being zero.

The following observations can be drawn from Figures 1 and 2:

- (1) the magnitudes of the relative biases and relative mean squared errors generally decrease to zero with increasing n . Also the relative biases appear to take both positive and negative values when $\sigma_x = \sigma_y$;
- (2) the relative biases for (2.1), (2.2) and (2.5) appear not too different when $\sigma_x = \sigma_y$;
- (3) the relative biases for (2.1) and (2.5) appear generally positive when $\sigma_x < \sigma_y$;
- (4) the relative biases for (2.1) and (2.2) appear smallest when $\sigma_x < \sigma_y$;
- (5) the relative biases for (2.5) appear largest when $\sigma_x < \sigma_y$;
- (6) the relative mean squared errors appear smallest, second smallest and largest for (2.5), (2.1) and (2.2), respectively, for small n ;
- (7) the relative biases and relative mean squared errors for all three estimators appear reasonable for all n and parameter values.

We have presented results for limited choices of (σ_x, σ_y) and for only one choice of $(\lambda_1, \lambda_2, r_1, r_2)$. But the results were the same for a wide range of other choices for

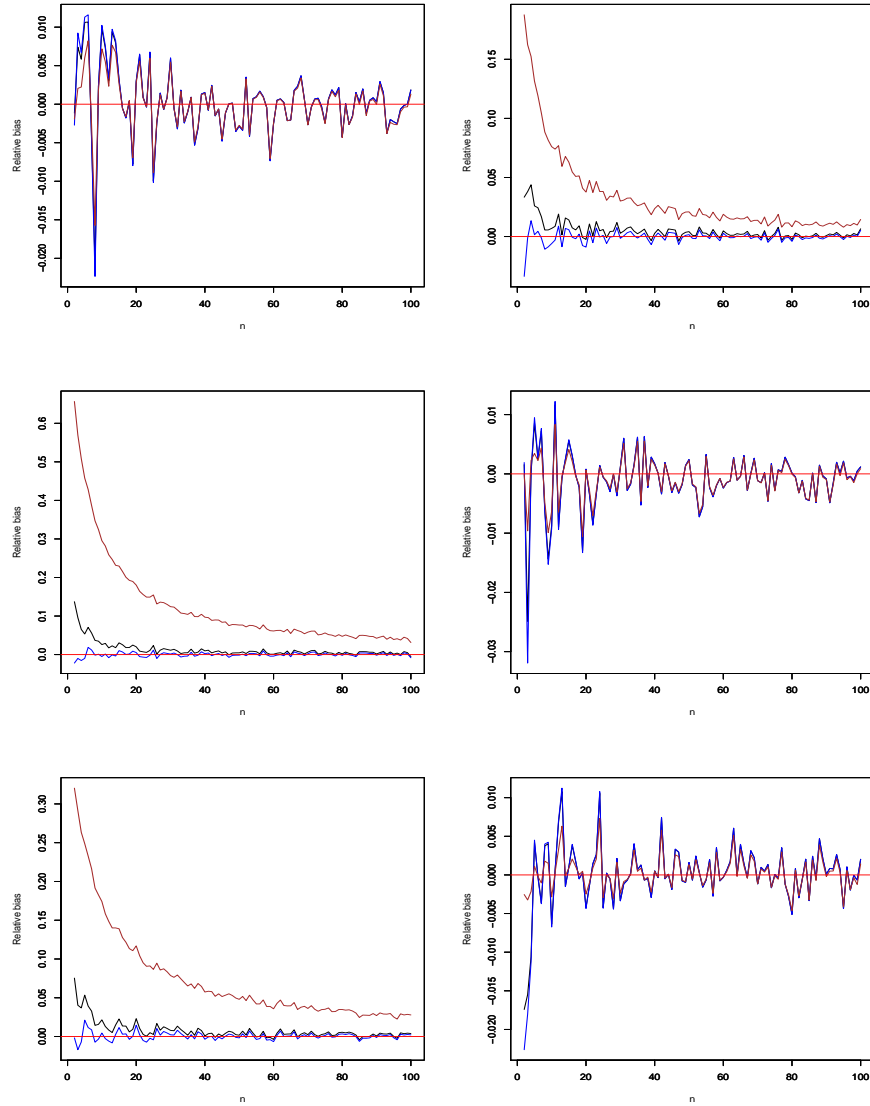


Figure 1. Relative biases of (2.1) in black, (2.2) in blue and (2.5) in brown. Top left is for $(\sigma_x, \sigma_y) = (1, 1)$, top right is for $(\sigma_x, \sigma_y) = (1, 2)$, middle left is for $(\sigma_x, \sigma_y) = (1, 5)$, middle right is for $(\sigma_x, \sigma_y) = (2, 2)$, bottom left is for $(\sigma_x, \sigma_y) = (2, 5)$, and bottom right is for $(\sigma_x, \sigma_y) = (5, 5)$.

(σ_x, σ_y) and $(\lambda_1, \lambda_2, r_1, r_2)$, including choices where $\lambda_1 \neq \lambda_2$ and $r_1 \neq r_2$. Similar results were also obtained when the gamma priors were replaced by non informative priors. In

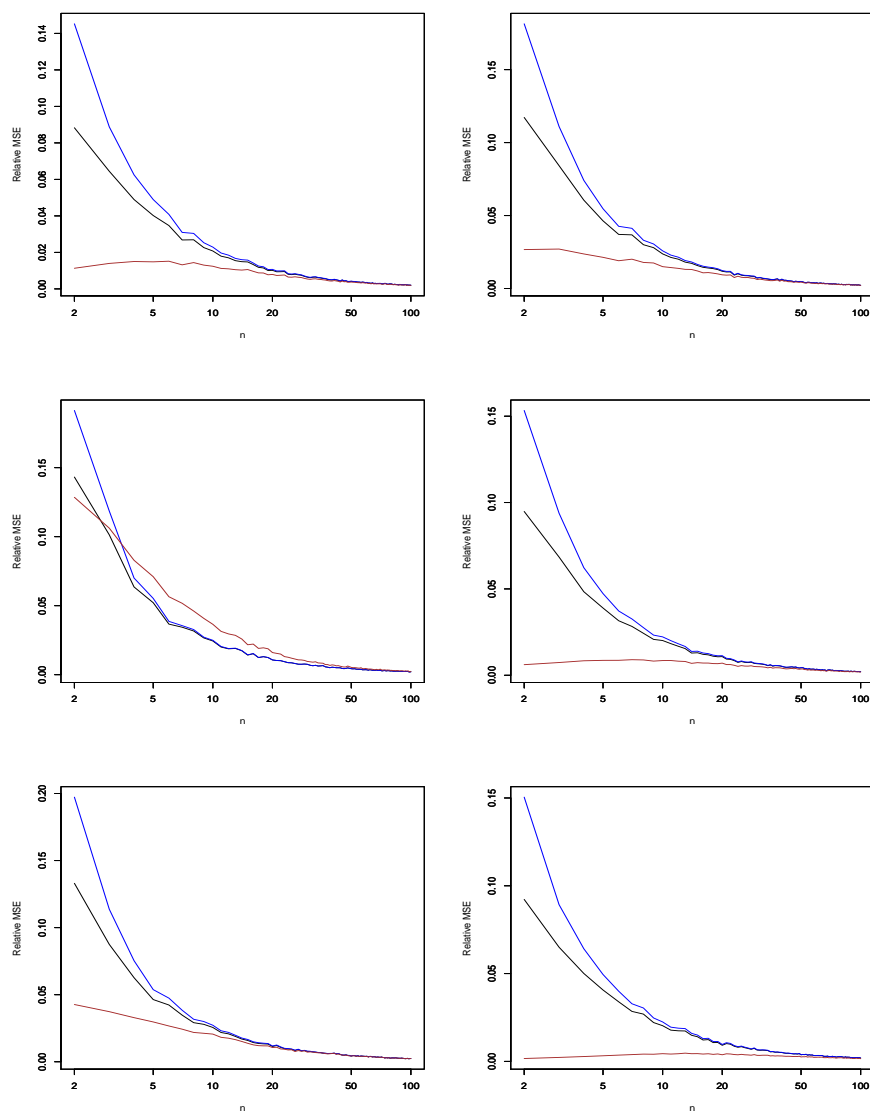


Figure 2. Relative mean squared errors of (2.1) in black, (2.2) in blue and (2.5) in brown. Top left is for $(\sigma_x, \sigma_y) = (1, 1)$, top right is for $(\sigma_x, \sigma_y) = (1, 2)$, middle left is for $(\sigma_x, \sigma_y) = (1, 5)$, middle right is for $(\sigma_x, \sigma_y) = (2, 2)$, bottom left is for $(\sigma_x, \sigma_y) = (2, 5)$, and bottom right is for $(\sigma_x, \sigma_y) = (5, 5)$.

particular, the magnitude of the relative biases generally decreased to zero with increasing n , the relative mean squared errors generally decreased to zero with increasing n ,

the relative biases for all three estimators appeared reasonable for all n , and the relative mean squared errors for all three estimators appeared reasonable for all n .

5.2. Simulation study for interval estimators of R . Here, we perform a simulation study to compare the performances of the asymptotic maximum likelihood and percentile based bootstrap confidence intervals for R . The performance was assessed in terms of coverage probabilities and coverage lengths. The following scheme was used:

- (1) Generate ten thousand samples of $\{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m\}$;
- (2) Compute the confidence intervals, (3.1) and (3.2), for each of the ten thousand samples, say (L_{1i}, U_{1i}) and (L_{2i}, U_{2i}) for $i = 1, 2, \dots, 10000$;
- (3) Compute the coverage probabilities for the two intervals as

$$P_j = \frac{1}{10000} \sum_{i=1}^{10000} I \left\{ L_{ji} < \frac{2}{\pi} \sin^{-1} \frac{1}{\sqrt{1 + \frac{\sigma_y}{\sigma_x}}} < U_{ji} \right\}$$

for $j = 1, 2$;

- (4) Compute the coverage lengths for the two intervals as

$$L_j = \frac{1}{10000} \sum_{i=1}^{10000} (U_{ji} - L_{ji})$$

for $j = 1, 2$.

We repeated this scheme for $m = n = 1, 2, \dots, 100$ and $(\sigma_x, \sigma_y) = (1, 1), (1, 2), (1, 5), (2, 2), (2, 5), (5, 5)$. For the bootstrap confidence interval, we took $B = 500$, as in Section 4. Plots of the coverage probabilities, P_1 and P_2 , versus n are shown in Figure 3. Plots of the coverage lengths, L_1 and L_2 , versus n are shown in Figure 4. The red line in Figure 3 represents the 95 percent nominal level.

The following observations can be drawn from Figures 3 and 4:

- (1) coverage probabilities generally approach the nominal level with increasing n and coverage lengths generally decrease with increasing n ;
- (2) coverage probabilities for (3.2) appear closer to the nominal level for all $n < 40$. Thereafter (3.1) and (3.2) appear to perform equally well.

We have presented results for limited choices of (σ_x, σ_y) and for only one choice of B . But the results were the same for a wide range of other choices for (σ_x, σ_y) and $B > 500$. In particular, the coverage probabilities generally approached the nominal level with increasing n and the coverage lengths generally decreased with increasing n .

6. Conclusions

In this note, we have studied estimation of $R = P(Y < X)$ when X and Y are independent Lévy random variables. We have considered three different point estimators for R : maximum likelihood estimator, UMVUE and Bayes estimator. We have considered two different interval estimators for R : asymptotic maximum likelihood estimator and bootstrap based percentile estimator.

Among the three point estimators, the Bayes estimator has the smallest relative mean squared errors but also the largest relative biases. The maximum likelihood estimator and the UMVUE have the smallest relative biases. But they do not have the smallest relative mean squared errors.

Among the two interval estimators, the bootstrap estimator has better coverage probabilities for small n . Both estimators perform equally well for all sufficiently large n .

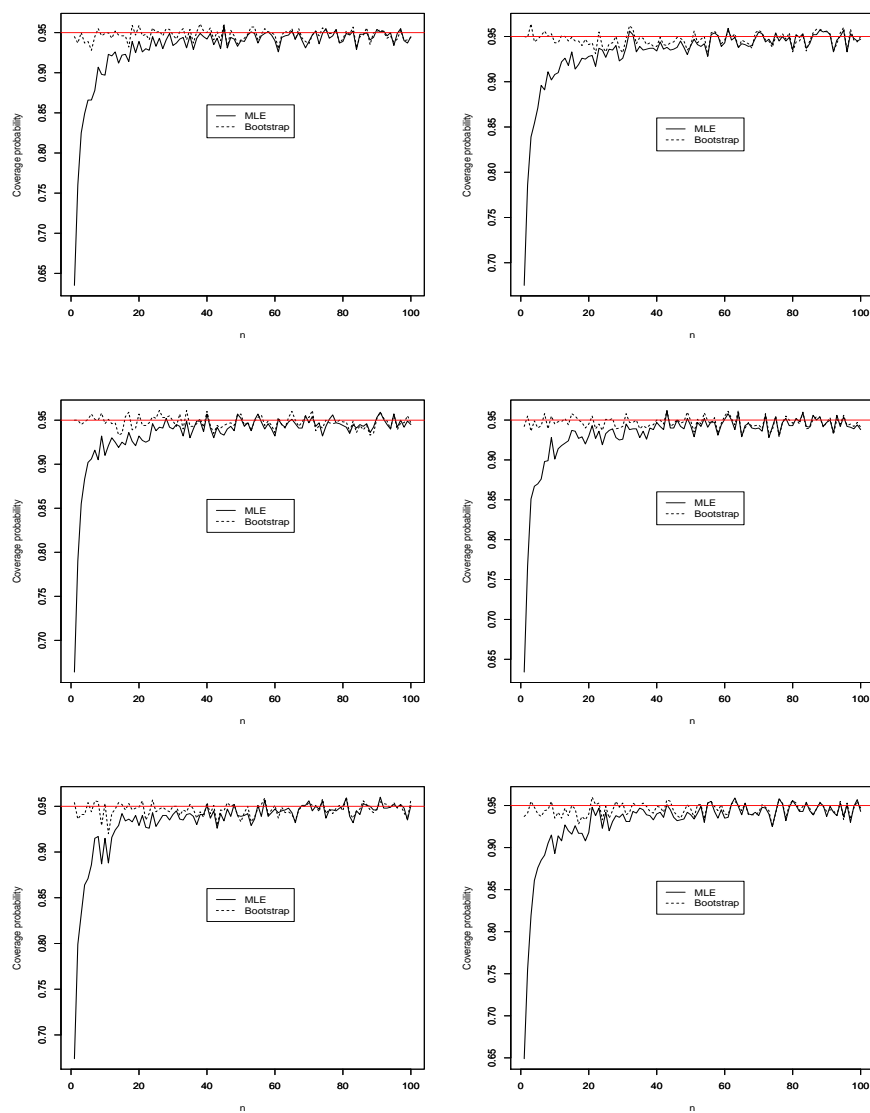


Figure 3. Coverage probabilities of (3.1) and (3.2). Top left is for $(\sigma_x, \sigma_y) = (1, 1)$, top right is for $(\sigma_x, \sigma_y) = (1, 2)$, middle left is for $(\sigma_x, \sigma_y) = (1, 5)$, middle right is for $(\sigma_x, \sigma_y) = (2, 2)$, bottom left is for $(\sigma_x, \sigma_y) = (2, 5)$, and bottom right is for $(\sigma_x, \sigma_y) = (5, 5)$.

In Sections 5.1 and 5.2, we have taken $m = n$ for simplicity. But the stated observations were the same when $m \neq n$.

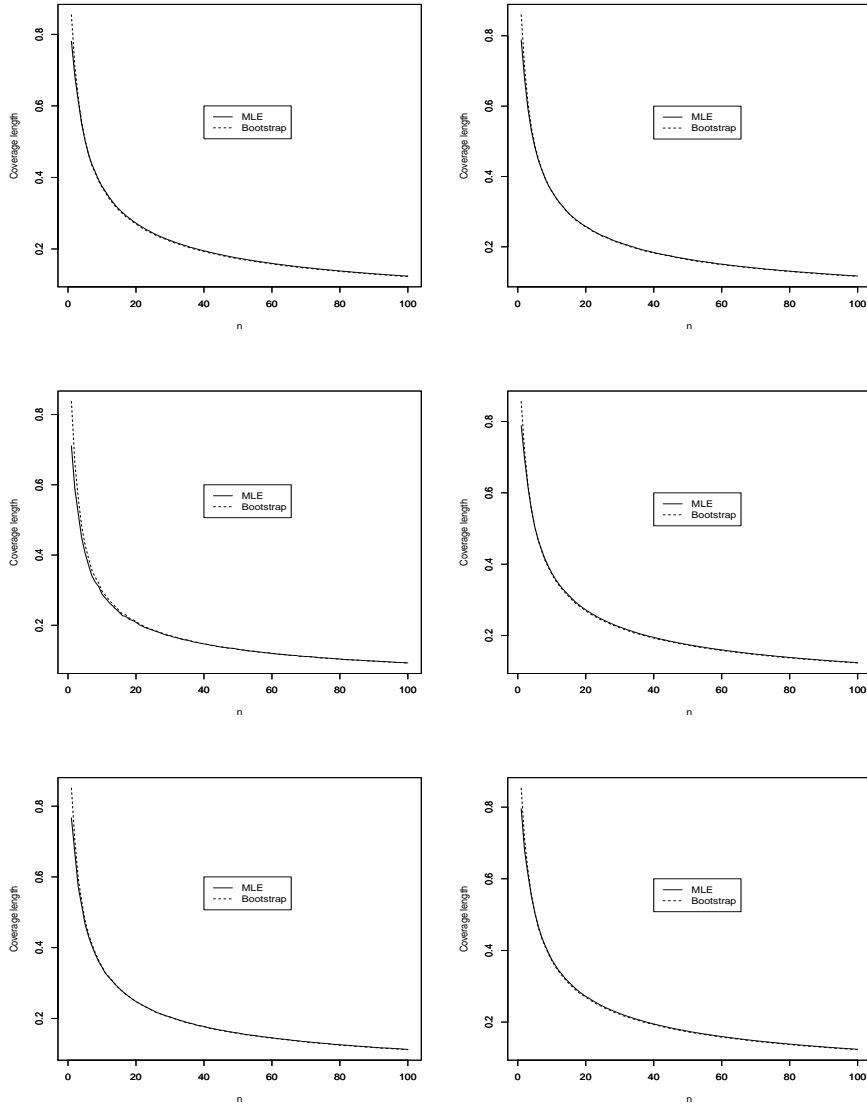


Figure 4. Coverage lengths of (3.1) and (3.2). Top left is for $(\sigma_x, \sigma_y) = (1, 1)$, top right is for $(\sigma_x, \sigma_y) = (1, 2)$, middle left is for $(\sigma_x, \sigma_y) = (1, 5)$, middle right is for $(\sigma_x, \sigma_y) = (2, 2)$, bottom left is for $(\sigma_x, \sigma_y) = (2, 5)$, and bottom right is for $(\sigma_x, \sigma_y) = (5, 5)$.

This is the first time estimation of $R = P(Y < X)$ for Lévy random variables has been studied in a comprehensive manner. Previously only maximum likelihood estimation of R has been considered for Lévy random variables.

A more comprehensive study of the estimation of $R = P(Y < X)$ for Lévy random variables could consider other point as well as interval estimators. These could include Bayesian highest posterior density intervals (Chen and Shao [9]), interval estimators based on the signed log-likelihood ratio due to Barndorff-Nielsen [6], interval estimators based on the modified signed log-likelihood ratio due to Barndorff-Nielsen [7], and robust estimators based on the theory of bounded influence M -estimators (Greco and Ventura [19]).

Acknowledgments

The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

References

- [1] Alfonso, L., Mansilla, R., Terrero-Escalante, C.A. *On the scaling of the distribution of daily price fluctuations in the Mexican financial market index*, Physica A—Statistical Mechanics and Its Applications, **391**, 2990-2996, 2012.
- [2] Ali, M.M., Nadarajah, S., Woo, J.S. *On the ratio $X/(X+Y)$ for Weibull and Lévy distributions*, Journal of the Korean Statistical Society, **34**, 11-20, 2005.
- [3] Ali, M.M., Woo, J.S. *Inference on reliability $P(X < Y)$ in the Lévy distribution*, Mathematical and Computer Modeling, **41**, 965-971, 2005.
- [4] Awad, A.M., Gharraf, M.K. *Estimation of $P(Y < X)$ in the Burr case: A comparative study*, Communications in Statistics - Simulation and Computation, **15**, 389-403, 1986.
- [5] Babayi, S., Khorram, E., Tondro, F. *Inference of $R = P[X < Y]$ for generalized logistic distribution*, Statistics, **48**, 862-871, 2014.
- [6] Barndorff-Nielsen, O.E. *Inference on full and partial parameters, based on the standardized signed log-likelihood ratio*, Biometrika, **73**, 307-322, 1986.
- [7] Barndorff-Nielsen, O.E. *Modified signed log-likelihood ratio*, Biometrika, **78**, 557-563, 1991.
- [8] Casella, G. *Statistical Inference*, second edition. Duxbury Press, 2001.
- [9] Chen, M.-H., Shao, Q.-M. *Monte Carlo estimation of Bayesian credible and HPD intervals*, Journal of Computational and Graphical Statistics, **8**, 69-92, 1999.
- [10] Congdon, P. *Bayesian Statistical Modelling*. John Wiley and Sons, New York, 2007.
- [11] Constantine, K., Tse, S.K., Karson, M. *Estimation of $P(Y < X)$ in the gamma case*, Communications in Statistics - Simulation and Computation, **15**, 365-388, 1986.
- [12] Davison, A.C. *Statistical Models*. Cambridge University Press, Cambridge, 2003.
- [13] Dorfman, J.H., Karali, B. *Do farmers hedge optimally or by habit? A Bayesian partial-adjustment model of farmer hedging*, Journal of Agricultural and Applied Economics, **42**, 791-803, 2010.
- [14] Downtown, F. *The estimation of $\Pr(Y < X)$ in the normal case*, Technometrics, **15**, 551-558, 1973.
- [15] Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, **38**, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1982.
- [16] Felsenstein, J. *Taking variation of evolutionary rates between sites into account in inferring phylogenies*, Journal of Molecular Evolution, **53**, 447-455, 2001.
- [17] Genc, A.I. *Estimation of $P(X > Y)$ with Topp-Leone distribution*, Journal of Statistical Computation and Simulation, **83**, 326-339, 2013.
- [18] Gradshteyn, I.S., Ryzhik, I.M. *Table of Integrals, Series, and Products*, sixth edition. Academic Press, San Diego, 2000.
- [19] Greco, L., Ventura, L. *Robust inference for the stress-strength reliability*, Statistical Papers, **52**, 773-788, 2011.
- [20] Ismail, R., Jeyaratnam, S., Panchapakesan, S. *Estimation of $\Pr[X > Y]$ for gamma distributions*, Journal of Statistical Computation and Simulation, **26**, 253-267, 1986.

- [21] Kotz, S., Lumelskii, Y., Pensky, M. *The Stress-Strength Model and Its Generalizations: Theory and Applications*. World Scientific, Singapore, 2003.
- [22] Kundu, D., Raqab, M.Z. *Estimation of $R = P[Y < X]$ for three-parameter generalized Rayleigh distribution*, Journal of Statistical Computation and Simulation, **85**, 725-739, 2015.
- [23] Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., Jones, D.R. *How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS*, Statistics in Medicine, **24**, 2401-2428, 2005.
- [24] McCool, J.I. *Inference on $P(Y < X)$ in the Weibull case*, Communications in Statistics - Simulation and Computation, **20**, 129-148, 1991.
- [25] Nadar, M., Kyzylaslan, F., Papadopoulos, A. *Classical and Bayesian estimation of $P(Y < X)$ for Kumaraswamy's distribution*, Journal of Statistical Computation and Simulation, **84**, 1505-1529, 2014.
- [26] Nadarajah, S., Chan, S., Afuecheta, E. *Extreme value analysis for emerging African markets*, Quality and Quantity, **48**, 1347-1360, 2014.
- [27] Obradovic, M., Jovanovic, M., Milosevic, B., Jevremovic, V. *Estimation of $P\{X \leq Y\}$ for Geometric-Poisson model*, Hacettepe Journal of Mathematics and Statistics, in press, 2015.
- [28] O'Reilly, F.J., Rueda, H. *A note on the fit for the Lévy distribution*, Communications in Statistics—Theory and Methods, **27**, 1811-1821, 1998.
- [29] R Development Core Team. *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria, 2015.
- [30] Reiser, B., Guttman, I. *A comparison of three point estimators for $P(Y < X)$ in the normal case*, Computational Statistics and Data Analysis, **5**, 59-66, 1987.
- [31] Sagi, Y., Brook, M., Almog, I., Davidson, N. *Observation of anomalous diffusion and fractional self-similarity in one dimension*, Physical Review Letters, **108**, Article Number 093002, 2012.
- [32] Saracoglu, B., Kaya, M.F., Abd-Elfattah, A.M. *Comparison of estimators for stress-strength reliability in the Gompertz case*, Hacettepe Journal of Mathematics and Statistics, **38**, 339-349, 2009.
- [33] Srokowski, T. *Multiplicative Lévy noise in bistable systems*, European Physical Journal, B, **85**, Article Number 65, 2012.
- [34] Surles, J.G. and Padgett, W.J. *Inference for $P(Y < X)$ in the Burr type X model*, Journal of Applied Statistical Science, **7**, 225-238, 1998.