DOI 10.1007/s10958-016-3035-3
Journal of Mathematical Sciences, Vol. 218, No. 3, October, 2016

ON ROBUST ALGORITHM FOR FINDING MAXIMUM LIKELIHOOD ESTIMATION OF THE GENERALIZED INVERSE GAUSSIAN DISTRIBUTION*

I. Yaroshenko[1]

In this paper, we propose robust numerical methods for finding the maximum likelihood estimation of the generalized inverse Gaussian distribution. A comparative analysis of the existing algorithms and the results of numerical experiments are presented. Special attention is paid to reproducibility of the tests.

1. Introduction

In big data processing, data mining, there exists a contradiction between the accuracy of inference and the speed (performance) requirements: the more rapidly the data are processed, the more rough the techniques used and, hence, the less accurate the inference. At the same time, modern data mining techniques based on probability mixture models substantially use iterative numerical algorithms realizing advanced statistical procedures, say, EM-algorithm-type techniques [11]. Therefore, there is an urgent need for high-performance numerical procedures of statistical analysis. At the same time, numerical methods used in the analysis of big data, e.g., in the algorithms of situational analysis, are required to be highly reliable. They must calculate with certainty the correct result within a limited time frame.

The present paper proposes a robust numerical method for finding the maximum likelihood estimation of the *generalized inverse Gaussian distribution*, which was first proposed by Good [9]. The notion of the generalized hyperbolic distribution as a GIG-variance-mean mixture of normal distributions was introduced by Barndorff-Nielsen [5, 6]. Both families are already widely used in many fields including financial mathematics.

The links between the subfamilies of the generalized inverse Gaussian distributions and the subfamilies of the generalized inverse hyperbolic distribution are presented in simplified form in a work by Paolella [13].

The density of the generalized inverse Gaussian distribution is given by

$$f_{GIG}(x;\lambda,\chi,\psi)=\left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}}\frac{1}{2K_\lambda\left(\sqrt{\chi\psi}\right)}x^{\lambda-1}\exp\left\{-\frac{1}{2}(\chi x^{-1}+\psi x)\right\},\quad x>0. \tag{1}$$

Here $\lambda\in\mathbb{R}$,

$$\begin{aligned}&\chi>0,\psi\geqslant 0,\quad \text{if } \lambda<0,\\ &\chi>0,\psi>0,\quad \text{if } \lambda=0,\\ &\chi\geqslant 0,\psi>0,\quad \text{if } \lambda>0,\end{aligned}$$

$K_\lambda(z)$ is the modified Bessel function of the third kind with index λ,

$$K_\lambda(z)=\int\limits_0^\infty \exp(-z\cosh y)\cosh(\lambda y)dy,\quad z\in\mathbb{C},\ \operatorname{Re} z>0.$$

[1] Lomonosov Moscow State University, Moscow, Russia, e-mail: ilyayaroshenko@gmail.com

* Research supported by the Russian Scientific Foundation, project 14–11–00364.

 1072-3374/16/2183-0354

As $\chi \downarrow 0$ or $\psi \downarrow 0$, an uncertainty arises in (1) , so the cases of $\chi = 0$ and $\psi = 0$ that correspond to the *gamma distribution* and the *inverse gamma distribution* are considered separately as special cases of the *generalized gamma distribution* with a fixed power parameter.

2. The generalized inverse Gaussian distribution

2.1. Families of asymptotic distributions

As $\chi \downarrow 0$, the generalized inverse Gaussian distribution turns into the gamma distribution with the parameters $(\lambda, 2/\psi)$

$$\lim_{\chi \downarrow 0} f_{GIG}(x; \lambda, \chi, \psi) = f_{\Gamma}(x; \lambda, 2/\psi), \quad x > 0,\ \lambda > 0,\ \psi > 0,$$

$$f_{\Gamma}(x; \kappa, \beta) = \frac{1}{\beta^{\kappa}\Gamma(\kappa)} x^{\kappa-1} \exp\left\{-\frac{x}{\beta}\right\};$$

as $\psi \downarrow 0$, it turns into the inverse gamma distribution with the parameters $(-\lambda, \chi/2)$

$$\lim_{\psi \downarrow 0} f_{GIG}(x; \lambda, \chi, \psi) = f_{I\Gamma}(x; -\lambda, \chi/2), \quad x > 0,\ \lambda < 0,\ \chi > 0,$$

$$f_{I\Gamma}(x; \kappa, \beta) = \frac{\beta^{\kappa}}{\Gamma(\kappa)} x^{-\kappa-1} \exp\left\{-\frac{\beta}{x}\right\}.$$

The gamma distribution and the inverse gamma distribution are special cases of the generalized gamma distribution with power parameters $\nu = 1$ and $\nu = -1$ respectively.

The numerical methods for finding maximum likelihood estimation of the generalized gamma distribution with a fixed power parameter ν are presented in the final part of the present paper.

2.2. The proper generalized inverse Gaussian distribution

This section of the paper concerns the algorithm for finding maximum likelihood estimation of the generalized inverse Gaussian distribution. The algorithm suggests the use of a range of nonstandard functions. The formulas for calculating such functions are presented in the next section of this paper.

When $\psi > 0$ and $\chi > 0$, it is possible to substitute variables

$$\eta = \sqrt{\frac{\chi}{\psi}}, \qquad \omega = \sqrt{\psi\chi},$$

$$f_{GIG}(x; \lambda, \omega, \eta) = \frac{1}{2\eta K_{\lambda}(\omega)} \left(\frac{x}{\eta}\right)^{\lambda-1} \exp\left\{-\frac{\omega}{2}\left(\left(\frac{x}{\eta}\right)^{-1} + \frac{x}{\eta}\right)\right\}, \quad x > 0, \tag{2}$$

$f_{GIG}(x; \lambda, \omega, \eta)$ is the density function of *the proper generalized inverse Gaussian distribution.* Then the log-likelihood function is

$$\frac{L_{GIG}(\lambda, \omega, \eta)}{n} = -\lambda \log \eta - \log(2K_{\lambda}(\omega)) + (\lambda - 1)\bar{x}_{\sim} - \frac{\omega}{2}\left(\bar{x}_{-1}\eta + \frac{\bar{x}_1}{\eta}\right), \tag{3}$$

where

$$\bar{x}_{-1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}, \qquad \bar{x}_{\sim} = \frac{1}{n}\sum_{i=1}^{n}\log x_i, \qquad \bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4}$$

$(\bar{x}_{-1}, \bar{x}_{\sim}, \bar{x}_1)$ is the minimal sufficient statistic in the case of the generalized inverse Gaussian distributions. The following relations between the harmonic mean, geometric mean, and arithmetic mean are well known:

$$\bar{x}_{-1}^{-1} \leqslant \exp \bar{x}_{\sim} \leqslant \bar{x}_1, \tag{5}$$

and hence

$$\bar{x}_{-1}\bar{x}_1 \geqslant 1. \tag{6}$$

Equality in (5) (and (6)) occurs if and only if $x_1 = \cdots = x_n$.

Jørgensen [10] proves a number of important theorems that allow one to make the transition from the three-dimensional optimization problem to the unidimensional optimization problem. The following statement is a direct consequence of these theorems.

Suppose we are given the $(\bar{x}_{-1}, \bar{x}_{\sim}, \bar{x}_1)$ statistic and $\bar{x}_{-1}\bar{x}_1 > 1$. Then the maximum likelihood for this statistic can be attained either for the gamma or inverse gamma distributions or for the proper inverse Gaussian distribution. In case it is attained for the proper inverse Gaussian distribution, the maximum likelihood estimation $\hat{\lambda}$ is calculated as

$$\hat{\lambda} = \operatorname*{argmax}_{\lambda\in(-u,u)} \frac{L_{GIG}(\lambda, \hat{\omega}_\lambda, \hat{\eta}_\lambda)}{n}, \qquad u = \frac{\bar{x}_{-1}\bar{x}_1}{\bar{x}_{-1}\bar{x}_1 - 1}, \tag{7}$$

where the pair $(\hat{\omega}_\lambda, \hat{\eta}_\lambda)$ is the maximum likelihood estimation for fixed λ. The $\hat{\omega}_\lambda$, $\hat{\eta}_\lambda$ estimations are calculated as

$$\hat{\omega}_\lambda = D_\lambda^{-1}(\bar{x}_{-1}\bar{x}_1), \tag{8}$$

$$\hat{\eta}_\lambda = \sqrt{\frac{\bar{x}_1}{\bar{x}_{-1}}}\sqrt{\frac{R_{-\lambda}(\hat{\omega}_\lambda)}{R_\lambda(\hat{\omega}_\lambda)}}, \tag{9}$$

where $D_\lambda^{-1}(s)$ is the inverse function for $D_\lambda(\omega)$,

$$D_\lambda(\omega) = \frac{K_{\lambda+1}(\omega)K_{\lambda-1}(\omega)}{K_\lambda(\omega)^2}, \tag{10}$$

$$R_\lambda(\omega) = \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}. \tag{11}$$

It was proved by Jørgensen [10] that:

1. $D_\lambda(\omega)$ is strictly decreasing on $\mathbb{R}_+$.
2. $L_{GIG}(\lambda) = L_{GIG}(\lambda, \hat{\omega}_\lambda, \hat{\eta}_\lambda)$ is strictly convex on $(-u, u)$.

Hence, the transition to the unidimensional problem is completely correct. The function $D_\lambda^{-1}(s)$ can be calculated as the solution to the equation $D(\omega) - s = 0$.

Jørgensen proposed the use of the Newton–Raphson method to calculate $\hat{\omega}_\lambda$. Without formal proof of convergence, he notes that this method works well with a equivalent equation $\log(D_\lambda(e^y)-1)-\log(s-1) = = 0$, $y = \log\omega$. Due to the fact that at small ω exponential overflow takes place in calculating $K_\lambda(\omega)$, Jørgensen pays special attention to asymptotic representations of function $D_\lambda(\omega)$ for various λ. He proposes tabulation as a way of likelihood maximization of $L_{GIG}(\lambda)$.

Implementation of asymptotic representations of $D_\lambda(\omega)$ at small ω has shown that they are quite inaccurate while exponential overflow takes place in calculating $K_\lambda(\omega)$. In this connection, the following section contains a description of evaluation of a range of functions, including $D_\lambda(\omega)$.

The Newton–Raphson method and the equivalent equation proposed by Jørgensen clearly do not make it feasible to attain full machine precision while solving the equation $D(\omega) - s = 0$. Moreover, robustness of the Newton–Raphson method as applied to the present task needs to be researched further. That is why we propose to replace this algorithm with a modification of the TOMS748 [2] algorithm for finding the root of a one-variable function. The TOMS748 algorithm performs search of the root of a function on a given interval. Since the root is searched for $\omega > 0$, the interval shall be $[m, M]$, where m

is the smallest representable normalized value that is not zero, and M is the largest representable value that is not infinity. The TOMS748 algorithm is not efficient enough on intervals that have limits with a large exponent difference (in terms of floating-point numbers representation). In the cases of m and M the exponent difference is at its maximum, and the algorithm can perform thousands of iterations in order to attain full machine precision. The main point of the modification of the TOMS748 algorithm is the following: in case the iterations of the algorithm are not efficient enough, binary search is to be used. However, the average shall not be considered as the arithmetic mean but as the number that divides the set of numbers of the given interval into equal parts. Following this approach, the $[1, 4]$ interval shall be divided by 2. Such addition of binary search allows to eliminate uncertainty by one bit. The described modification can be found in the `findRoot` function of the D standard library.

In order to find the maximum of $L_{GIG}(\lambda)$ on the interval $(-u, u)$, we propose to use the standard algorithm for unidimensional maximization using quadratic interpolation. This algorithm is described in detail by Brent [7]. Numerous math packages contain Brent's algorithm. However, it is necessary to make sure that the algorithm allows to define the two constants that determine tolerance — both absolute and relative.

The proposed algorithm for finding maximum likelihood estimation is presented in the statistical package [18].

2.3. Functions

The algorithms for calculating $D_\nu(x)$, $\sqrt{\frac{R_\nu(x)}{R_{-\nu}(x)}}$, and $\log K_\nu(x)$ are very similar to the algorithm used for $K_\nu(x)$, which is presented in [8, 16, 17]. A brief presentation of the algorithm for $K_\nu(x)$ can be found in [14]. In the following, we will only mention the most important features specific to the required functions.

According to the algorithm for $K_\nu(x)$:

$$K_\nu(x) = \sqrt{\pi}\frac{1}{\sqrt{x}}\exp(-x)\frac{1}{1+S},$$

$$R_\nu(x) = \frac{1}{x}\left[\nu + \frac{1}{1} + x + \left(\nu^2 - \frac{1}{4}\right)\frac{z_0}{z_1}\right], \qquad |\nu| \leqslant \frac{1}{2},$$

where $\frac{z_0}{z_1}$ is the continued fraction CF2

$$\frac{z_0}{z_1} = \frac{1}{b_1+}\frac{a_2}{b_2+}\cdots$$

with

$$b_n = 2(n+x),$$

$$a_{n+1} = -\left[(n+1/2)^2 - \nu^2\right].$$

CF2 converges rapidly at $x \geqslant 2$.

For $x \leqslant 2$ the following series are used:

$$K_\nu(x) = \sum_{k=0}^{\infty} c_k f_k, \quad \frac{x}{2}K_{\nu+1}(x) = \sum_{k=0}^{\infty} c_k h_k, \qquad |\nu| \leqslant \frac{1}{2}.$$

The sequences $\{c_k\}$, $\{f_k\}$, $\{h_k\}$ and the sum S can be found in the works referenced earlier.

From the recurrence formula

$$K_{\nu+1}(x) = K_{\nu-1}(x) + \frac{2\nu}{x}K_\nu(x)$$

stable formulas can be obtained for $x \geqslant 2$

$$R_{\nu+1}(x) = \frac{2\nu}{x} + \frac{1}{R_\nu(x)}, \tag{12}$$

and for $x \leqslant 2$

$$\frac{x}{2} R_{\nu+1}(x) = \nu + \frac{\frac{x^2}{4}}{\frac{x}{2} R_\nu(x)}. \tag{13}$$

Then, with regard to (10) and (11),

$$D_\nu(x) \leftarrow \begin{cases} D_{-\nu}(x), & \nu < 0, \\ R_\nu(x)\frac{2}{x}(\frac{x}{2}R_\nu(x) - \nu), & x < 2,\ 0 \leqslant \nu \leqslant \frac{1}{2}, \\ R_\nu(x)(R_\nu(x) - \frac{2\nu}{x}), & x \geqslant 2,\ 0 \leqslant \nu \leqslant \frac{1}{2}, \\ \dfrac{\frac{x}{2}R_\nu(x)}{\frac{x}{2}R_{\nu-1}(x)}, & x < 2,\ \nu > \frac{1}{2}, \\ \dfrac{R_\nu(x)}{R_{\nu-1}(x)}, & x \geqslant 2,\ \nu > \frac{1}{2}, \end{cases} \tag{14}$$

and

$$\sqrt{\frac{R_\nu(x)}{R_{-\nu}(x)}} \leftarrow \begin{cases} \dfrac{1}{\sqrt{\dfrac{R_{-\nu}(x)}{R_\nu(x)}}}, & \nu < 0, \\ \sqrt{\dfrac{\frac{x}{2}K_{\nu+1}(x)}{\frac{x}{2}K_{\nu+1}(x) - \nu K_\nu(x)}}, & x < 2,\ 0 \leqslant \nu \leqslant \frac{1}{2}, \\ \sqrt{\dfrac{R_\nu(x)}{R_\nu(x) - \frac{2\nu}{x}}}, & x \geqslant 2,\ 0 \leqslant \nu \leqslant \frac{1}{2}, \\ \frac{2}{x}\sqrt{\frac{x}{2}R_\nu(x)}\sqrt{\frac{x}{2}R_{\nu-1}(x)}, & x < 2,\ \nu > \frac{1}{2}, \\ \sqrt{R_\nu(x)}\sqrt{R_{\nu-1}(x)}, & x \geqslant 2,\ \nu > \frac{1}{2}. \end{cases} \tag{15}$$

Analogously for $\log K_\mu(x)$:

$$\log K_\mu(x) \leftarrow \begin{cases} \log K_{-\mu}(x), & \mu < 0, \\ \log K_\nu(x), & x < 2,\ 0 \leqslant \mu \leqslant \frac{1}{2}, \\ \log \frac{1}{1+S} - \frac{\log x}{2} - x, & x \geqslant 2,\ 0 \leqslant \mu \leqslant \frac{1}{2}, \\ \log K_\nu(x) + n(\log 2 - \log x) + \sum\limits_{i=0}^{n-1} \log\left(\frac{x}{2}R_{\mu+i}(x)\right), & x < 2,\ \mu > \frac{1}{2} \\ \log \frac{1}{1+S} - \frac{\log x}{2} - x + \sum\limits_{i=0}^{n-1} \log R_{\mu+i}(x), & x \geqslant 2,\ \mu > \frac{1}{2}, \end{cases} \tag{16}$$

where ν and n are such that $\nu = \mu - n$, $n \in \mathbb{N}$, $|\nu| \leqslant \frac{1}{2}$. The `frexp` function of the C standard library can be used to optimize calculation of sums and differences of logarithms. It extracts the exponent from the floating point number and allows one to replace the sum or the difference of logarithms with a logarithm of product or ratio respectively without exponential overflow.

2.4. Comparison with other algorithms and tests

Synthetic tests were carried out on a given set of statistics determined in (4). In order to attain representative coverage, it is convenient to use the parametrization

$$(\bar{x}_1, \bar{x}_{-1}, \bar{x}_\sim) = \left(\bar{x}_\alpha \bar{x}_\beta, \frac{\bar{x}_\beta}{\bar{x}_\alpha}, \log \bar{x}_\gamma\right), \tag{17}$$

Table 1. Test results

Feature	Jørgensen–Yaroshenko	NR Nelder–Mead	Proper Nelder–Mead
% of failures	0	15 %	4 %
% of better results	92 % / 51 %	8 %	49 %
Max time	1.77 ms	1.36 ms	6.03 ms
Mean time	329 μs	137 μs	555 μs
Median time	261 μs	264 μs	317 μs
Min time	87 μs	33 μs	97 μs

where

$$\bar{x}_\alpha > 0, \quad \bar{x}_\beta > 1, \quad \frac{\bar{x}_\alpha}{\bar{x}_\beta} < \bar{x}_\gamma < \bar{x}_\alpha \bar{x}_\beta. \tag{18}$$

The following set of statistics was used in testing:

$$\{(\bar{x}_\alpha, \bar{x}_\beta)_i\} = \{2^i\}_{i=-8}^{8} \times \{1 + 2^i\}_{i=-8}^{8}, \tag{19}$$

for each element of which 64 uniformly distributed numbers $\bar{x}_\gamma$ were chosen, according to (18). The total number of tests was $17 \times 17 \times 64 = 18496$.

The Broyden–Fletcher–Goldfarb–Shanno algorithm and the Nelder–Mead [12] simplex method are considered to be the most efficient algorithms for likelihood function maximization.

The BFGS algorithm requires partial derivative $\frac{\partial K_\lambda(\omega)}{\partial \omega}$ to be calculated; $\frac{K_{\lambda+h}(\omega) - K_{\lambda-h}(\omega)}{2h}$ can be used as the approximate derivative. The Ridders method [15] produces more accurate results, but it requires more function calls for $K_\lambda(\omega)$. At the same time, neither of the ways of approximation of partial derivative are robust since they require a fixed scale parameter regulating the value of h, which actually varies significantly.

In the comparison with the Nelder–Mead algorithm we used its implementation from Numeric Recipes (NR) [14], which was optimized for three-dimensional space. In testing, attention should be paid to the algorithm stopping criterion. As in numerical oprimization packages, the algorithm in NR is discontinued in case the values of the likelihood function across all simplex points are similar enough. For the likelihood function, which is rather sloping in the vicinity of maximum, it is necessary to additionally require maximum absolute difference of standard projections of simplex points. The limit of the number of iterations of the algorithm was at 4000. The following points served as the initial simplex on $\mathbb{R}^3_{(\lambda,\eta,\omega)}$ space:

$$(0,1,1), \quad (0,1,2), \quad (0,2,1), \quad (0,2,2), \quad (1,1,1).$$

For all tested algorithms, the tolerance constants were equal to the square root of the machine epsilon `1.49e-08`. It was assumed that the algorithm produces a better result when the value of the likelihood function at a calculated point was higher as compared to the result of the other algorithm. If the maximum absolute difference of likelihood functions exceeded 10^{-5}, the test of the algorithm with a lesser value of the likelihood function was considered to have failed.

It is important to note several points. First, in the case of the proper Nelder–Mead algorithm a significant increase in the number of iterations and a significant reduction of the tolerance constants had almost no influence on the failed test count. Second, the values of parameters that were calculated in the failed tests differed greatly from the required values.

Table 1 presents test results for the Jørgensen–Yaroshenko algorithm, the Nelder–Mead algorithm and its modification. In Fig. 1 the distribution graphs for failed tests of the proper Nelder–Mead algorithm are given.

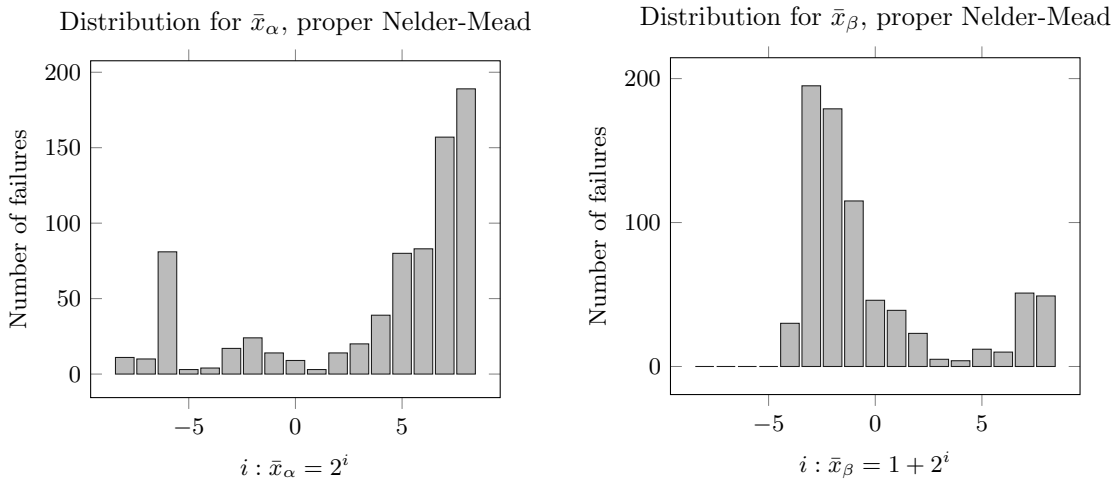

Fig. 1. Distribution for failed tests.

3. The generalized gamma distribution with a fixed power

Maximum likelihood estimation of the generalized gamma distribution with a fixed power is quite straightforward. However, it makes sense to present an efficient numerical method as such basic algorithms are sometimes simply not given due attention.

The density of the generalized gamma distribution is given by

$$f_{G\Gamma}(x;\nu,\kappa,\sigma) = \frac{|\nu|}{\sigma\Gamma(\kappa)}\left(\frac{x}{\sigma}\right)^{\kappa\nu-1}\exp\left\{-\left(\frac{x}{\sigma}\right)^{\nu}\right\}, \quad x>0;$$

the log-likelihood function is

$$\frac{L_{G\Gamma}(\nu,\kappa,\sigma)}{n} = \log\frac{|\nu|}{\Gamma(\kappa)} + (\kappa\nu-1)\bar{x}_{\sim} - \kappa\log\sigma^{\nu} - \frac{\bar{x}_\nu}{\sigma^\nu},$$

where

$$\bar{x}_\nu = \frac{1}{n}\sum_{i=1}^{n} x_i^\nu, \qquad \bar{x}_{\sim} = \frac{1}{n}\sum_{i=1}^{n}\log x_i.$$

The likelihood equations have the form (power parameter ν is fixed)

$$\kappa = \text{logmdigammaInverse}\left(\log(\bar{x}_\nu) - \nu\bar{x}_{\sim}\right),$$

$$\sigma = \left(\frac{\bar{x}_\nu}{\kappa}\right)^{\frac{1}{\nu}},$$

where logmdigammaInverse(c) is the inverse function for logmdigamma$(\kappa) = \log(\kappa) - \psi(\kappa)$, and $\psi(\kappa)$ is the digamma function.

The logmdigammaInverse(c) function is calculated as the solution to the logmdigamma$(\kappa) - c = 0$ equation. In order to solve this equation, it is recommended that the TOMS748 [2] algorithm be used. When possible, TOMS748 uses inverse cubic interpolation, otherwise, parabolic interpolation and the secant method are used.

The following inequality [3] holds

$$\frac{1}{2}y < \text{logmdigamma}\left(\frac{1}{y}\right) < y, \quad y>0,$$

which is matched by parametrization $\kappa = \frac{1}{y}$ for searching for the root on the interval $[c, 2c]$

$$\kappa = \text{logmdigammaInverse}(c) = \frac{1}{\text{findRoot}_{[c,2c]}\left\{y \to \text{logmdigamma}\left(\frac{1}{y}\right) - c\right\}}. \tag{20}$$

The number of iterations required for convergence of the present algorithm generally does not depend on c. Actually, the quantity of floating-point numbers on the interval $[c, 2c]$ is much the same for each c. In order to calculate logmdigammaInverse(c) when double-precision numbers are used, on average six and not more than nine function calls for logmdigamma(c) are required so that full machine precision can be attained. In the neighborhood of zero logmdigammaInverse(c) $\sim \frac{1}{2c}$, and in the neighborhood of infinity logmdigammaInverse(c) $\sim \frac{1}{c}$, which follows from the limits that can be found in [4]. An actual implementation of logmdigammaInverse(c) will be included in the D standard library starting from summer 2015.

Since $\psi(\kappa)$ asymptotically tends to $\log(\kappa)$, logmdigamma(κ) should be implemented independently. Formulas (6.3.18) and (6.3.6) from a work by Abramowitz and Stegun [1] may be useful for its implementation. There are an actual implementations of logmdigamma(κ) in the D standard library and the `statmod` package for the R programming language.

REFERENCES

1. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York (1965).
2. G. E. Alefeld, F. A. Potra, and Yixun Shi, "Algorithm 748: enclosing zeros of continuous functions," *ACM Trans. Math. Softw.*, **21**, 327–344 (1995).
3. H. Alzer, "On some inequalities for the gamma and psi functions," *Math. Comput.*, **66**, 373–389 (1997).
4. G. D. Anderson, R. W. Barnard, K. C. Richards, M. K. Vamanamurthy, and M. Vuorinen, "Inequalities for zero-balanced hypergeometric functions," *Trans. Am. Math. Soc.*, **347**, 1713–1723 (1995).
5. O. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size," *Proc. R. Soc. London Math. Phys. Sci.*, **353**, No. 1674, 401–419 (1977).
6. O. Barndorff-Nielsen, *Hyperbolic Distributions and Distributions on Hyperbolae*, University of Aarhus, Aarhus (1977).
7. R. P. Brent, *Algorithms for Minimization without Derivatives*, Dover Publications, New York (1973).
8. J. B. Campbell, "On Temme's algorithm for the modified Bessel function of the third kind," *ACM Trans. Math. Softw.*, **6**, No. 4, 581–586 (1980).
9. I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, **40**, No. 3–4, 237–264 (1953).
10. B. Jørgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution*, University of Aarhus, Aarhus (1980).
11. V. Yu. Korolev, *Probabilistic and Statistical Methods for the Decomposition of Volatility of Chaotic Processes*, Moscow State University Publishing House, Msocow (2011).
12. J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, **7**, 308–313 (1965).
13. M. Paolella, *Intermediate Probability: A Computational Approach*, Wiley, New York (2007).
14. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York (2007).

15. C. J. F. Ridders, "Accurate computation of $F'(x)$ and $F'(x)F''(x)$," *Adv. Eng. Softw.*, **4**, No. 2, 75–76 (1978).

16. N. M. Temme, "On the numerical evaluation of the modified Bessel function of the third kind," *J. Comput. Phys.*, **19**, No. 3, 324–337 (1975).

17. I. J. Thompson and A. R. Barnett, "Modified Bessel functions $I_\nu(z)$ and $K_\nu(z)$ of real order and complex argument, to selected accuracy," *Comput. Phys. Commun.*, **47**, No. 2–3, 245–257 (1987).

18. I. Yaroshenko, *Atmosphere: statistical package*, `http://9il.github.io/atmosphere/` (2015).