On Maximisation of the Likelihood for the Generalised Gamma Distribution

Angela Noufaily & M.C. Jones

Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

Abstract

We explore computational aspects of likelihood maximisation for the generalised gamma distribution. We improve on earlier work by being the first to formulate the iterative approach to solution of the likelihood score equations in such a way that the individual equations involved are uniquely solvable. We observe that the resulting algorithm is well-behaved and competitive with the application of standard optimisation procedures. Most importantly, we argue that, far from being problematic as a number of authors have suggested, the generalised gamma distribution is actually particularly amenable to maximum likelihood estimation, by the standards of general three- or more-parameter distributions.

KEY WORDS: Broyden-Fletcher-Goldfarb-Shanno algorithm; iterative solution; Nelder-Mead algorithm

1. INTRODUCTION

The generalised gamma (GG) distribution (Stacy, 1962) has recently seen a great increase in practical application and interest (e.g. Cox, Chu, Schneider & Muñoz, 2007, and references therein). This useful distribution for positive univariate continuous data has canonical density

$$f(t) = \{\Gamma(\alpha/\beta)\}^{-1}\beta t^{\alpha-1}\exp(-t^{\beta}), \quad t > 0,$$

where α and β are shape parameters each of which we take to be positive. In practice, one also includes a scale parameter $\theta > 0$ through fitting the scaled density $\theta^{-1}f(\theta^{-1}t)$. The GG distribution encompasses both the gamma ($\beta = 1$) and Weibull ($\beta = \alpha$) distributions — and hence also the exponential distribution — as special cases and the lognormal and normal distributions as limiting cases.

Early attempts at developing iterative schemes for maximum likelihood (ML) estimation of the parameters of the GG distribution (Parr & Webster, 1965, Stacy & Mirham, 1965, Hager & Bain, 1970, Stacy, 1973, Prentice, 1974, Lawless, 1980, Wingo, 1987, Cohen & Whitten, 1988, Wong, 1993) were not entirely successful, a consistent theme being the lack of guaranteed unique solvability of one or more of the equations involved. This seems to have inculcated in some researchers that ML estimation is not a good option for the GG distribution and recourse has therefore been made to a variety of alternative ad hoc estimation schemes. For example, Huang & Hwang (2006) claim that Prentice's (1974) version of ML estimation is "still quite complicated", Gomes, Combes & Dussauchoy (2008) mention the "many difficulties described in the literature to estimate the parameters" and bemoan the "complexity" of the ML method, and Song (2008) cites "the difficulty of estimating and computing its parameters" as a "major impediment" to use of the GG distribution. Cooray & Ananda (2008) use their perception that "there is difficulty in developing inference procedures with the generalized gamma distributions, especially the maximum likelihood parameter estimation" to justify eschewing the model in terms of a simpler alternative.

The purpose of this paper is to help rehabilitate ML estimation as a perfectly reasonable method — in a computational sense — for estimation of the parameters of the GG distribution. We make four specific contributions. First, building on Lawless (1980), we formulate the iterative approach to solution of the likelihood score equations in such a way that the individual

equations involved are uniquely solvable. Second, we provide bounds on the values of the ML estimates. Third, we show both that this iterative approach is successful and that it is competitive with, or even a little better than, standard optimisation procedures applied to the GG distribution. Fourth, and importantly, we show that ML estimation is perfectly viable with any of these techniques in the following senses. The GG likelihood surface, while not unimodal, appears to comprise a strong global maximum together with very few local maxima which are much smaller than, and far away from, the global maximum. Also, while the different maximisation methods sometimes fail to find the global maximum on a given single run, they succeed sufficiently often that an approach consisting of running any one (or a mix) of the maximisation methods mentioned above just a few times from different starting points is virtually guaranteed to locate that distinct global maximum. It seems to us that, far from being especially problematic, the GG distribution is actually particularly amenable to ML estimation, by the standards of general three- or more-parameter distributions. Note that our comments apply only to the computational task of maximising the likelihood and not to the inferential quality of the estimates obtained, although in our limited experience and that of others, the latter appear to be reasonable; for some discussion of them, see Lawless (2003).

2. SOLVING THE SCORE EQUATIONS

Let $T_1, ..., T_n$ be a random sample taken from the GG distribution. We extend the work done by Lawless (1980) — who fixed values of one parameter — to solve the score equations fully in a simple and yet efficient way taking all parameters to be unknown. First, take logs of the data and work with $Y_i = \log T_i,\ i = 1, ..., n$. Reparametrise the distribution of Y so that its density is

$$f(y) = \frac{k^{k-\frac{1}{2}}}{\sigma\Gamma(k)} \exp\left\{\sqrt{k}w - ke^{w/\sqrt{k}}\right\}, \qquad y \in \mathbb{R},$$

where $w = \sigma^{-1}(y - \mu)$. Here, $-\infty < \mu < \infty$ is a location parameter, $\sigma > 0$ a scale parameter and $k > 0$ a shape parameter. In terms of the original parameterisation, $\mu = \log\theta + \beta^{-1}\log(\alpha/\beta)$, $\sigma = 1/\sqrt{\alpha\beta}$ and $k = \alpha/\beta$. The invariance of ML estimation means, of course, that ML estimates of the original sets of parameters can be obtained from ML estimates of the new parameters via these equations.

2.1 The Equations to be Solved

The log likelihood is

$$n\left\{-\log\sigma+\left(k-\frac{1}{2}\right)\log k-\log\Gamma(k)+\sqrt{k}\frac{(\overline{Y}-\mu)}{\sigma}-k\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)S_0\right\}$$

where $\bar{Y}$ is the sample mean and we define

$$S_j\equiv\frac{1}{n}\sum_{i=1}^{n}Y_i^j\exp\left(\frac{Y_i}{\sigma\sqrt{k}}\right);\quad j=0,1,2.$$

Differentiating the loglikelihood with respect to μ, σ and k in turn yields the score equations

$$0=\frac{n\sqrt{k}}{\sigma}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)S_0-1\right\},\tag{2.1}$$

$$0=\frac{n\sqrt{k}}{\sigma^2}\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)(S_1-\mu S_0)-\frac{\sigma}{\sqrt{k}}-\overline{Y}+\mu\right\},\tag{2.2}$$

and

$$0=n\left\{\exp\left(\frac{-\mu}{\sigma\sqrt{k}}\right)\left(\frac{S_1-\mu S_0}{2\sigma\sqrt{k}}-S_0\right)+\log k+1-\frac{1}{2k}-\psi(k)+\frac{\overline{Y}-\mu}{2\sigma\sqrt{k}}\right\}.\tag{2.3}$$

As in Lawless (1980), (2.1) yields

$$\mu=\sigma\sqrt{k}\log S_0,\tag{2.4}$$

an explicit expression for μ in terms of k and σ. Using this, we can reduce (2.2) to

$$R(\sigma)\equiv\frac{S_1}{S_0}-\overline{Y}-\frac{\sigma}{\sqrt{k}}=0.\tag{2.5}$$

Again following Lawless, we think of this as an equation in σ for any μ and k and solve it numerically. Finally, using (2.4) and (2.5), we find that (2.3) reduces to

$$T(k)\equiv\log k-\psi(k)-\frac{L}{\sqrt{k}}=0\tag{2.6}$$

where $L=(\mu-\overline{Y})/\sigma$. This equation for k, given values for μ and σ is also to be solved numerically. We show in the next subsection that (2.5) and (2.6) are

well-behaved equations with unique solutions in σ and k, respectively. This is the achievement that makes a contribution beyond previous approaches to solving the score equations in this context.

2.2 Proof That Equations (2.5) and (2.6) Have Unique Solutions

In what follows, disregard the absurd case where all Ys are equal.

It is easy to show that $R(\sigma)$ is monotone decreasing:

$$\frac{\partial R(\sigma)}{\partial \sigma} = \frac{1}{\sqrt{k}} \left(\frac{S_1^2 - S_0 S_2}{\sigma^2 S_0^2} - 1 \right);$$

both terms inside the brackets are negative, the first one by the Cauchy Schwartz inequality. Also, $\lim_{\sigma \to 0} R(\sigma) = Y_{\max} - \overline{Y} > 0$ and $\lim_{\sigma \to \infty} R(\sigma) = -\infty$, so there must be precisely one value of $\sigma > 0$ (for any μ, k) for which $R(\sigma) = 0$.

In fact, we can speed up our search a little by using the fact that this root will lie in a certain interval. This is because the interpretation of S_1/S_0 as a mean of Y values implies that $S_1/S_0 < Y_{\max}$ and hence that if σ_0 is the root of $R(\sigma) = 0$, then it satisfies

$$0 < \sigma_0 = \sqrt{k} \left(\left. \frac{S_1}{S_0} \right|_{\sigma_0} - \overline{Y} \right) < \sqrt{k} \{ Y_{\max} - \overline{Y} \}. \tag{2.7}$$

We now turn our attention to (2.6). How does this behave as a function of k for fixed μ and σ, and hence L? For small k, $\psi(k) \sim -k^{-1}$ which goes to (minus) infinity faster than either of the other two terms and hence $\lim_{k \to 0} T(k) = \infty$. For large k, $\psi(k) \sim \log k - (2k)^{-1}$ (Abramowitz & Stegun, 1965, p.259) and so $\lim_{k \to \infty} T(k) = 0$. This limit is reached from either the positive or negative side, depending on the sign of $\overline{Y} - \mu$, since $-L/\sqrt{k}$ is the dominant term. In fact, provided the current values of μ and σ satisfy (2.4),

$$\exp\left(\frac{\mu}{\sqrt{k}\sigma} \right) = \text{average}(\exp(Y)) > \exp(\text{average}(Y)) = \exp\left(\frac{\overline{Y}}{\sqrt{k}\sigma} \right),$$

the first equality by (2.4), the inequality being Jensen's inequality. Here, the average is taken over the distribution that selects Y uniformly from the

set $Y_1/(\sqrt{k}\sigma)$, ..., $Y_n/(\sqrt{k}\sigma)$. It follows that $\mu > \overline{Y}$, $L > 0$, and the limit is reached from the negative side.

Here is a proof that the equation $T(k) = 0$ has a single root in $0 < k < \infty$. Rewrite (2.6) as $\tau(k) = L$, say, where

$$\tau(k) \equiv \sqrt{k}\{\log(k) - \psi(k)\}.$$

Then, Theorem 1 of Alzer (1997) says that $\tau(k)$ is strictly completely monotonic on $(0, \infty)$. (Alzer's theorem actually applies more widely to $k^\alpha\{\log(k) - \psi(k)\}$ for any $\alpha \leq 1$.) In particular, this means that $\tau(k)$ is strictly monotonically decreasing (from ∞ down to 0) and since $L > 0$ there must be a unique solution to the equation of interest.

Here, too, an interval can be specified within which the solution, k_0 say, is contained. The inequalities

$$\frac{1}{2k} < \log(k) - \psi(k) < \frac{1}{k}$$

(Alzer, 1997) imply that $(2k_0)^{-1} < k_0^{-1}L < k_0^{-1}$ and, rearranging, we obtain

$$\frac{1}{4L^2} < k_0 < \frac{1}{L^2}. \tag{2.8}$$

2.3 The Iterative Algorithm

We have now proved that each of the equations (2.5) and (2.6) has a unique root, the latter provided that $L > 0$. We will solve (2.4), (2.5) and (2.6) simultaneously and iteratively to obtain maximum likelihood estimates, $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$, of μ, σ and k, respectively. Our suggested algorithm is the following:

1. Set the iteration number i to 0 and obtain an initial guess for $L = L_0 > 0$.

2. Set $i = i + 1$.

3. For given L_{i-1}, compute $\widehat{k}_i$ by solving (2.6) using either the bisection method or the Newton Raphson algorithm (we have used the former).

4. Replace the obtained $\widehat{k}_i$ in $R(\sigma)$ to compute $\widehat{\sigma}_i$ by solving (2.5) using the bisection method or the Newton Raphson algorithm (we have used the latter).

5. Substitute $\widehat{k}_i$ and $\widehat{\sigma}_i$ into (2.4) to obtain the corresponding $\widehat{\mu}_i$.

6. Use these estimates to obtain L_i and to compute the value of the log likelihood function.

7. Repeat steps (2), (3), (4), (5) and (6) until desired accuracy of the likelihood is achieved.

Note the order of steps 3 and 4 to guarantee the positivity of L_i.

When implementing the above algorithm in R (R Development Core Team, 2009), we added some computational devices to avoid crashes caused by large parameter values. The first of these was for $k > 171$, to employ Stirling's formula (Abramowitz & Stegun, 1965, p.257) as an approximation for $\Gamma(k)$. The second is to compute S_1/S_0 in (2.5) by first multiplying numerator and denominator by $\exp(-(\sigma\sqrt{k})^{-1}Y_{\max})$ where σ and k are the current values of those parameters. A similar trick applies to S_0 in (2.4).

We note that our theoretical results above are still partial in the sense that they do not guarantee convergence of our algorithm nor uniqueness of the ML estimate. These issues will be explored numerically in Section 3.

2.4 Bounds on Maximum Likelihood Estimates

Let $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$ denote ML estimates of μ, σ and k. Then, the inequalities obtained in Section 2.2 for use at intermediate stages of the ML algorithm also apply to the ML estimates themselves. We therefore have the reassurance that

$$\overline{Y} < \hat{\mu} < Y_{\max}. \tag{2.9}$$

From (2.8),

$$\frac{1}{4}\frac{\hat{\sigma}^2}{(\hat{\mu}-\overline{Y})^2} < \hat{k} < \frac{\hat{\sigma}^2}{(\hat{\mu}-\overline{Y})^2}$$

and from (2.7),

$$0 < \hat{\sigma}^2 < \hat{k}(Y_{\max}-\overline{Y})^2$$

which combine to give

$$\hat{k}(\hat{\mu}-\overline{Y})^2 < \hat{\sigma}^2 < \hat{k}\min\left[\{2(\hat{\mu}-\overline{Y})\}^2, (Y_{\max}-\overline{Y})^2\right]. \tag{2.10}$$

3. PERFORMANCE OF OUR ALGORITHM

Based on the iterative algorithm we proposed in Section 2.3, we developed a program in R (R Development Core Team, 2009) that computes the ML estimates of a generalized gamma distribution. The aim in this section is to compare it with other optimization methods available within the R function `optim`, namely the Nelder-Mead method introduced by Nelder & Mead (1965) and the Broyden-Fletcher-Goldfarb-Shanno method (BFGS) explained in Nocedal & Wright (1999).

We simulated 100 generalized gamma data sets within R by taking the $1/\beta$'th power of gamma variates generated by the R function `rgamma` which uses the methods of Ahrens & Dieter (1974, 1982). The parameter sets associated with each dataset were themselves selected at random. We also simulated another set of "initial parameters" consisting of 100 triplets (k_0, σ_0, and μ_0). For every data set, we ran the three mentioned methods (our program, Nelder-Mead and BFGS) starting from each set of initial parameters. Thus, we ran every method 10,000 times in total. The parameters k and σ are positive and a reasonable range for their values would be around the interval (0,6); therefore, we used a gamma distribution with scale parameter 1 and shape parameter 2 to simulate k and σ for both sets of "real" and initial parameters. The parameter μ can be positive or negative; therefore, we used a standard normal distribution to simulate its values, again for both sets of real and initial parameters. (Note that, for our program, k_0 is not needed if $L_0 > 0$; if, by the random mechanisms above, $L_0 < 0$, we set $k_0 = 1/L_0^2$ and proceed from Step 4.) This exercise was repeated with the same set of real and initial parameters for $n = 200$ and $n = 500$.

For each of the three algorithms we specified two stopping criteria, the relative change in log-likelihood from one iteration to the next (reltol) and the maximum number of iterations (maxit). For our program, we set reltol to be 10^{-8} and maxit to be 2000 whereas for both BFGS and Nelder-Mead, we set reltol to be 10^{-16} and maxit to be 10^9. In our program, we also set the root-finding methods to stop when they reached a relative accuracy of 10^{-4} and to report an error if more than 1000 iterations were needed for convergence. These values were chosen so that each of the three algorithms take an approximately equal time to produce their results.

We considered a program to have failed to reach the global maximum of the likelihood if one of the following occurred:

1. The program crashed and reported an error.

2. A "non-global maximum" is attained rather than the global one. In most cases, this only means that the program has been stopped by reaching maxit or reltol when an increase in the number of iterations allowed or a decrease in the value of reltol would solve the problem (at a cost, of course, in time taken). In a few cases, the non-global maximum appears to be a local maximum of the likelihood surface.

3. The rendered value of the maximum likelihood is, essentially, equal to negative infinity (-Inf).

With regard to item 2, the global maximum is taken to be the maximum likelihood attained for a data set over the 300 times its likelihood was maximised (by each of 3 methods from 100 sets of initial parameter values); the global maximum was said to be achieved in another run if the maximised likelihood value was within 0.01 of the overall maximum. Table 1 presents the total number of failure of the programs together with a closeup on the number of times each of the mentioned reasons contributes to the failures.

Table 1: Number of times each method fails to reach the maximal likelihood.

	Our Program	BFGS	Nelder-Mead
$n = 200$			
Total number of failures	135	915	116
Total number of "non-global maxima"	134	827	23
Total number of reported errors	1	23	14
Total number of -Inf	0	65	79
$n = 500$			
Total number of failures	129	799	127
Total number of "non-global maxima"	129	711	32
Total number of reported errors	0	24	21
Total number of -Inf	0	64	74

It is gratifying to observe little difference in performance between the two sample sizes. Overall, our program and the Nelder-Mead algorithm attain the global maximum likelihood the highest proportion of times (98.7% and 98.9% when $n = 200$), a little ahead of the BFGS algorithm (90.9% when $n = 200$). When $n = 500$, our program never crashed nor went to "$-$infinity"; Nelder-Mead displayed a different pattern of failures. Local maxima of the likelihood, on the rare occasions they were observed, are both far from the global maximum and have much smaller values of the likelihood.

4. THE SIMPLE LINEAR REGRESSION CASE

In this section, we revisit the work of Sections 2 and 3 for the simple linear regression case in which the data are (X_i, T_i), $i = 1, ..., n$, and the X's are a univariate covariate. We thus model μ not as a constant but in terms of the covariate and two parameters to be estimated, through $\mu = a + bX$. Note that both $a, b \in \mathbb{R}$. On the original scale, this means that the scale parameter θ is modelled as a log-linear function of the covariate. This is a much abbreviated account; for details, see the forthcoming Ph.D. thesis of the first author.

We find it convenient to reparametrize from a and b to $a' = a/\sigma$, $b' = b/\sigma$. Also, write

$$S_{jl} = \frac{1}{n} \sum_{i=1}^{n} Y_i^j X_i^l \exp\left(\frac{Y_i}{\sigma\sqrt{k}} - \frac{b'}{\sqrt{k}} X_i \right); \quad j, l = 0, 1, 2.$$

Then, it turns out that the score equation associated with differentiation with respect to a' results in the following analogue of (2.4):

$$a' = \sqrt{k} \log(S_{00}), \tag{4.1}$$

an expression for a' in terms of b', σ and k. Using (4.1), the score equation associated with b' reduces to

$$B_R(b') = \frac{S_{01}}{S_{00}} - \overline{X} = 0 \tag{4.2}$$

where $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$, an equation in b' for any k and σ. It is straightforward to show that there exists precisely one value of b' for which $B_R(b') = 0$;

$\partial B_R(b')/\partial b' > 0$ by the Cauchy-Schwartz inequality once more, and $B_R(b')$ is bounded by $X_{\min} - \overline{X} < 0$ and $X_{\max} - \overline{X} > 0$. Therefore, we can solve (4.2) numerically to find its root b'. The analogue of (2.5) turns out to be

$$R_R(\sigma) = \frac{S_{10}}{S_{00}} - \overline{Y} - \frac{\sigma}{\sqrt{k}} = 0. \tag{4.3}$$

Replacement of S_1/S_0 in (2.5) by S_{10}/S_{00} in (4.3) makes essentially no difference to our argument which shows that there is a unique value of $\sigma > 0$ which solves $R_R(\sigma) = 0$ for any a', b', k. Finally, the analogue of (2.6) is

$$T_R(k) \equiv \log(k) - \psi(k) - \frac{L_R}{\sqrt{k}} = 0 \tag{4.4}$$

where $L_R = k^{-1/2}\left\{(\overline{Y}/\sigma) - (a' + b'\overline{X})\right\}$. Again, Jensen's inequality can be used to show that $L_R > 0$ provided a', b' and σ satisfy (4.1), and hence (4.4) also has a single root in k in those circumstances.

Our algorithm becomes:

1. Set the iteration number i to 0 and obtain an initial guess for $L_R = L_{R,0} > 0$ and $b' = b'_0$.

2. Set $i = i + 1$.

3. For given $L_{R,i-1}$, compute $\widehat{k}_i$ by solving (4.4) using either the bisection method or the Newton Raphson algorithm (we have used the former).

4. Replace the obtained $\widehat{k}_i$ in $R_R(\sigma)$ to compute $\widehat{\sigma}_i$ by solving (4.3) using the bisection method or the Newton Raphson algorithm (we have used the latter).

5. Replace the obtained $\widehat{k}_i$ and $\widehat{\sigma}_i$ in $B_R(b')$ to compute $\widehat{b'}_i$ by solving (4.2) using the bisection method or the Newton Raphson algorithm (we used the former after experiments with the latter led to too many program failures).

6. Substitute $\widehat{k}_i$, $\widehat{\sigma}_i$ and $\widehat{b'}_i$ into (4.1) to obtain the corresponding $\widehat{a'}_i$.

7. Use these estimates to obtain $L_{R,i}$ and to compute the value of the log likelihood function.

8. Repeat steps (2), (3), (4), (5), (6) and (7) until desired accuracy of the likelihood is achieved.

9. Set $\widehat{a} = \widehat{\sigma}\widehat{a'}$, $\widehat{b} = \widehat{\sigma}\widehat{b'}$.

We repeated the experiment of Section 3 with the addition of the extra parameter b being generated, like a, from the standard normal distribution. For the regression situation, we also tested an R implementation of the Prentice (1974) approach to ML estimation provided in the VGAM package of T.W. Yee. However, we were unable to make this program work anything like as well as any of the others and so we have removed it once more from our comparisons. Results are given in Table 2.

Table 2: Number of times each method fails to reach the maximal likelihood in the simple regression case.

	Our Program	BFGS	Nelder-Mead
$n = 200$			
Total number of failures	320	1035	220
Total number of "non-global maxima"	290	1035	179
Total number of reported errors	30	0	0
Total number of -Inf	0	0	41
$n = 500$			
Total number of failures	168	766	136
Total number of "non-global maxima"	140	766	86
Total number of reported errors	28	0	0
Total number of -Inf	0	0	50

Compared with Table 1, each method, unsurprisingly, has an increased number of reported errors, but the situation is still very good. For example, when $n = 200$, our program still has a 96.8% success rate, Nelder-Mead a 97.8% success rate, and BFGS an 89.7% success rate. It should be admitted that it was the less successful version of our program that uses the Newton-Raphson algorithm at Step 5 that was matched with the other programs for speed in the regression case, the purely bisection version reported on in Table 1 typically taking several times longer.

5. CONCLUSIONS

Our underlying thesis is that, computationally, maximum likelihood estimation for p-parameter parametric distributions, where p is small to moderate, say $p = 3, 4, 5, 6$, is generally straightforward. One should take care that one is working with a sensible parameterisation in which the parameters are clearly identifiable (which often corresponds to being clearly interpretable). And yes, away from the very special cases of exponential families and simple location-scale models, likelihoods are not concave and may have local maxima. The existence of more than one substantial local maximum and/or of local maxima close to the global maximum can cause problems. But, in our experience, the GG distribution is one for which neither of these circumstances arises — GG likelihoods appear to have a clear global maximum with any local maxima being much smaller and distant. Reasonable maximisation algorithms therefore find the global maximum 90% of the time or more, typically, and much more of the time in some cases. This leads to the simple strategy of running, say, our program or the BFGS method (or even a mix of the two) from a small number, m, of randomly chosen sets of starting values resulting in almost guaranteed location of the global ML estimate. For example, $m = 5$ independent Bernoulli trials with $p = 0.9$ give $p(0) = 0.00001$, $m = 10$ such give $p(0) = 0.0000000001$, where $p(0)$ is the probability that the maximum likelihood value obtained in the runs corresponding to the 'trials' is not the global maximum. Hence our claim that, despite other claims in the literature to the contrary, the GG distribution is actually one for which ML estimation is — in a computational sense — quite straightforward and reliable.

REFERENCES

Abramowitz, M. & Stegun, I.A. (eds) (1965) *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover, New York.

Ahrens, J.H. & Dieter, U. (1974) Computer methods for sampling from gamma, beta, Poisson and binomial distributions. *Computing*, 12, 223–246.

Ahrens, J.H. & Dieter, U. (1982) Generating gamma variates by a modified rejection technique. *Communications of the ACM*, 25, 47–54.

Alzer, H. (1997) On some inequalities for the gamma and psi functions. *Mathematics of Computation*, 66, 373–389.

Cohen, A.C. & Whitten, B.J. (1988) *Parameter Estimation in Reliability and Life Span Models*. Dekker, New York.

Cooray, K. & Ananda, M.A.A. (2008) A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics – Theory and Methods*, 37, 1323–1337.

Cox, C., Chu, H., Schneider, M.F. & Muñoz, A. (2007) Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, 26, 4352–4374.

Gomes, O., Combes, C. & Dussauchoy, A. (2008) Parameter estimation of the generalized gamma distribution. *Mathematics and Computers in Simulation*, 79, 955–963.

Hager, H.W. & Bain, L.J. (1970) Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association*, 65, 1601–1609.

Huang, P.H. & Hwang, T.Y. (2006) On new moment estimation of parameters of the generalized gamma distribution using its characterization. *Taiwanese Journal of Mathematics*, 10, 1083–1093.

Lawless, J.F. (1980) Inference in the generalized gamma and log gamma distributions. *Technometrics*, 22, 409–419.

Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, Second Edition. Wiley, Hoboken, NJ.

Nelder, J.A. & Mead, R. (1965) A simplex algorithm for function minimization. *Computer Journal*, 7, 308–313.

Nocedal, J. & Wright, S.J. (1999) *Numerical Optimization*. Springer.

Parr, V.B. & Webster, J.T. (1965) A method for discriminating between failure density functions used in reliability predictions. *Technometrics*, 7, 1–10.

Prentice, R.L. (1974) Log-gamma model and its maximum likelihood estimation. *Biometrika*, 61, 539–544.

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`

Song, K. (2008) Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing. *IEEE Transactions on Image Processing*, 17, 1233–1250.

Stacy, E.W. (1962) A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33, 1187–1192.

Stacy, E.W. (1973) Quasimaximum likelihood estimators for two-parameter gamma distributions. *IBM Journal of Research and Development*, 17, 115–124.

Stacy, E.W. & Mirham, G.A. (1965) Parameter estimation for a generalized gamma distribution. *Technometrics*, 7, 349–358.

Wingo, D.R. (1987) Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation. *IEEE Transactions on Reliability*, 36, 586–590.

Wong, J.Y. (1993) Readily obtaining the maximum-likelihood estimates of the 3-parameter generalized gamma-distribution of Stacy and Mirham. *Microelectronics and Reliability*, 33, 2243–2251.