PARAMETER ESTIMATION OF THE GENERALIZED GAMMA DISTRIBUTION

GOMES O.*, COMBES C.**, DUSSAUCHOY A.***
**PRISMa, Bât Nautibus, Université Lyon 1, 43 bd du 11 novembre 1918, 69622 Villeurbanne cedex- France - ophelie.gomes@laposte.net*
**EURISE – Faculté de Sciences, Université Jean Monnet, 23 rue du Docteur Paul Michelon, 42023 Saint-Etienne cedex 2 - France - *combes@univ-st-etienne.fr*
**** PRISMa, Bât Nautibus, Université Lyon 1, 43 bd du 11 novembre 1918, 69622 Villeurbanne cedex- France - alain-d@bat710.univ-lyon1.fr*

Abstract: This article focuses on the parameter estimation of the generalized gamma distribution. Because of many difficulties described in the literature to estimate the parameters, we propose here a new estimation method. The algorithm associated to this heuristic method is implemented in Splus. We validate the resulting routine on the particular cases of the generalized gamma distribution.

Keyword: Shifted generalized gamma distribution, generalized gamma distribution, parameter estimation, Chi-squared goodness-of-fit test.

1 Introduction

The generalized gamma distribution is a younger distribution (1962) than the normal distribution (1774). It was introduced by Stacy [15] in order to combine the power of two distributions: the Gamma distribution and the Weibull distribution. The generalized gamma distribution is a popular distribution because it is extremely flexible. This distribution is also convenient because it includes as special cases several distributions: the exponential distribution, the LogNormal distribution, the Weibull distribution, the Levy distribution…
These interests are nevertheless in contradiction with the difficulties in estimating the parameters. This topic was dealt in many papers but the complexity of the results proves that this topic is still an opened item.
This paper proposes a new heuristic approach in parameter estimation of the generalized gamma distribution using an iterative method. This routine was implemented in Splus software.
In the section 2, we describe the characteristic of the generalized gamma distribution and give some application areas. An overview of literature on the parameter estimation of the generalized gamma distribution is presented in section 3. The section 4 deals with the proposed heuristic method called algorithm I.T.E.V. In section 5, we apply the resulting routine on known generalized gamma distribution in order to validate the estimation method. We terminate with a conclusion and some perspectives.

2 Generalized gamma distribution

Historically, the generalized gamma distribution was described for the first time by Stacy in 1962 [15, 2]. However, this distribution would have appeared in the literature since 1924-1925 with Amoroso [6], 150 years after that the normal distribution appeared.
Let ***X*** a random variable. If ***X*** follows a generalized gamma distribution with parameters *a*, *l* and *c*, the probability density function (p.d.f.) is given by:

$$f(x) = \begin{cases} \dfrac{|c| a^{lc} x^{lc-1} e^{-(ax)^c}}{\Gamma(l)} & if \quad x > 0 \\ 0 & if \quad x \le 0 \end{cases}$$

Where:

- $c \in \Re$ and $l > 0$ are *shape* parameters,
- $a \in \Re$ is *scale* parameter,
- $\Gamma(.)$ denotes the gamma function

With: $E[X] = \dfrac{\Gamma\left(l + \frac{1}{c}\right)}{a \quad \Gamma(l)}$

$$V[X] = \frac{\Gamma\left(l+\frac{2}{c}\right)\ \Gamma(l) - \left[\Gamma\left(l+\frac{2}{c}\right)^2\right]}{a^2\ \Gamma(l)^2}$$

the *r*th moment of *X (r=1, 2, …)* is

$$E[X^r] = \begin{cases} \dfrac{1}{a^r}\dfrac{\Gamma\left(\dfrac{cl+r}{c}\right)}{\Gamma(l)}, & r/c > -l \\ \infty, & otherwise \end{cases}$$

Figure 1 and 2 give some example of the density function for a generalized gamma distribution where parameters a=1 and l=4 and c taking various values (the pictures was generated using Splus software [7]).

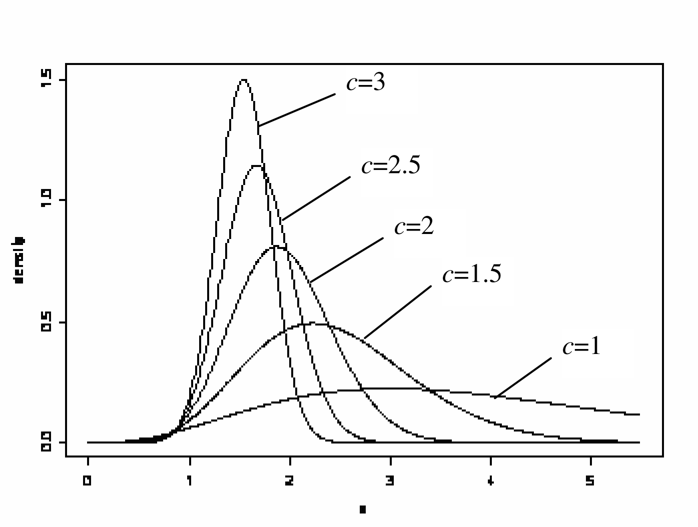

Figure 1: density function for a generalized gamma distribution whith parameter c=1, 1.5, 2, 2.5, 3 from right to left

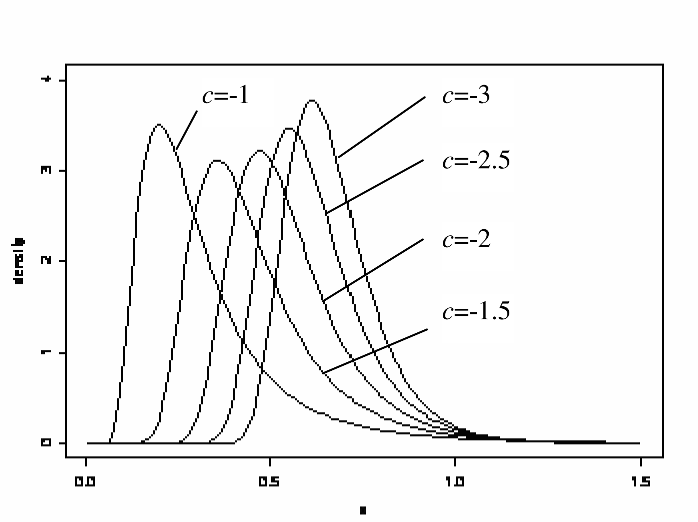

Figure 2: density function for a generalized gamma distribution whith parameter c=-1, -1.5, -2, -2.5, -3 from left to right

The generalized gamma distribution is extremely flexible. It is able to mimic several density function shapes (figure 1 & 2). In addition, it includes as special cases the Exponential distribution, the Gamma distribution, the Rayleigh distribution, the Half Normal distribution, the Chi-square distribution, the Levy distribution, the Weibull distribution and the Log-normal distribution in limit case (l tends to infinite) [3, 6, 13, 15]. Finally, for some negative ***c*** parameters, some of the moments don't exist depending on the ***c*** parameter value [15]; we have the same behaviour for the α-stables distribution [9]. As a consequence, the generalized gamma distribution is used in many fields such as health costs [10] (the distribution is used to examine regression modelling using the generalized gamma distribution), civil engineering [12] (it is used as a flood frequency analysis model) or economics [6] (it is used in various income distributions modelling).

Even if the generalized gamma distribution has particular cases such as exponential or gamma distribution, the parameter estimation of this distribution is not as easy as for an exponential or a gamma distribution. We will see in the next paragraph that parameter estimation of a generalized gamma distribution is still an opened topic.

3 Literature review of the parameter estimation for a generalized gamma distribution

Parameter estimation for the generalized gamma distribution was widely treated in literature. The authors Manning and Bsu [10] give names of researchers who studied the parameter estimation of the generalized gamma distribution. However, difficulties still persist [3].
If parameter estimation is easy for a gamma distribution, the estimation becomes complex when the parameter c is added in the density function. The difficulty increases also with the fact that different sets of parameters conduce to same density function [6, p156].
Maximum likelihood method was treated by Hirose [3], Stacy [15], Kleiber and Kotz [6] and Lawless [8]. Moment method was treated by Kleiber and Kotz [6]. Others authors such as Cohen and Whitten [2] or Hwang and Huang [4] tried to define new estimators.
The moment method is difficult to use because of the fact that different sets of parameters conduce to same density function. Maximum Likelihood method is also difficult to use since three likelihood equations need to be solved simultaneously [3, 4]. These two methods were derived in order to find a way to estimate the parameter of the generalized gamma distribution [3, 4]. But the results are not attractive due to their complexity. Consequently, it seems that there is no simple and straightforward method to estimate the parameters of a generalized gamma distribution.
However, some heuristics were proposed avoiding all difficulties that are present using the moment or the Maximum Likelihood methods. Stacy gives a graphic method [15] but it is difficult to implement it. Cohen and Whitten propose an iterative method [2] that loops on the parameter c and uses reference tables. If, in the past, this method conduces to approximate values for the parameters, nowadays, this iterative method could give better results with the computer's power.
Some statistic softwares have also integrated the parameter estimation of the generalized gamma distribution. The use of Splida [14], extension of Splus, generates lots of errors regarding the parameter estimation for a generalized gamma distribution. Hyfran [5] proposes maximum likelihood and moment methods. The number of errors is weak, but sometimes, the given parameters are surprising and are not realistic.
In this paragraph, we describe how the parameter estimation of the generalized gamma distribution is difficult. It has been treated by many authors and incorporated in software. However some difficulties still persist in the estimation. These difficulties are present in the literature and in the current software as well.
In the next paragraph, we present a new routine for parameter estimation of a generalized gamma distribution. Our objective is not to propose a new theoretical estimation of the parameters of the generalized gamma distribution but to propose an easy method to fit a generalized gamma distribution on given samples.

4 Algorithm I.T.E.V.: Iteration Transformation Estimation Validation

The previous paragraph illustrated why the estimation of the parameters for a generalized gamma distribution is still an opened topic. Because of the difficulties in applying the moment and maximum likelihood methods, we will not propose a new extension of these methods.
In the proposed approach, we use goodness-of-fit tests (GOFT) that measure the degree of agreement between the distribution of a data sample and a theoretical distribution. In all cases, a statistic test is compared against a known critical value to accept or to reject the hypothesis that the sample is from the postulated distribution. The chi-squared test is a nonparametric GOFT method. Various goodness-of-fit tests exist; quantile, Kolmogorov-Smirnov (depending on the sample size), Log-likelihood [3], AIC [6, p158] and probability plots could also be used.
The p.d.f. of the generalized gamma distribution is defined only for positive values of x (see definition of the p.d.f. of the generalized gamma distribution in section 2). Consequently, the algorithm we present must be applied only on samples for which all the values are positives.
To construct our routine, we use the fact that:

- Let ***X*** a random variable. If ***X*** follows a generalized gamma distribution with parameters a, l and c, then $\boldsymbol{Y}=\boldsymbol{X}^{c}$ follows a gamma distribution with parameters $A= a^c$ and l [6],
- A and l can be estimated and computed by

- $A = \dfrac{E[Y]}{VAR[Y]}$
- $l = \dfrac{E[Y]^2}{VAR[Y]}$

- Computer's power has increased over many years. We consider that today they are fast enough to be able to perform many loops,
- The value of the *c* parameter must keep realistic because of performance issue. Indeed, in the proposed method, we use iterations on parameter c. So, we need to determine the beginning and the end of the loop. This interval must be finite and must be constructed in function of the use of the algorithm. It is also constrained by the computer's power.

Let ***X*** the sample for which we want to fit a generalized gamma distribution with parameters *a*, *l* and *c*. The principle of the algorithm is:

1. Loop on parameter *c* in function of the defined interval
2. Transform sample to $\boldsymbol{Y}=\boldsymbol{X}^c$
3. Estimate the parameters *A* and *l* of the random variable ***Y*** which is supposed to follow a gamma distribution with parameter $A= a^c$ and *l*
4. Perform a Chi-squared goodness-of-fit test between the sample ***Y*** and the empirical distribution *gamma(A,l)*.
5. Keep the *pvalue* of the previous test in a table with the associated parameters of the generalized gamma distribution: $a=A^{1/c}$, *l* and *c*
6. At the end of the loop, return the parameters *a, l* and *c* for which the pvalue score of the chi-square test is maximum

In this initial algorithm, there is no restriction for the parameter *l*. However, *Γ(l)* expression tend to the infinite when parameter *l* increase. To avoid infinite values, we have added a control on the value of *Γ(l)* expression. We have also defined a function "*validatePoints*" that tests that the parameters *a*, *l* and *c* given by the estimation are realistic regarding the sample we try to fit. Therefore, two verifications have been added in each loop before step 4 (Perform a Chi-squared goodness-of-fit test):

- *Γ(l)* must exist: this condition will depend on the language used to implement the routine,
- "*validatePoints*": in function of the sample (x_1, x_i , x_n) of ***X*** and the parameters $a=A^{1/c}$, *l* and *c* estimated, three points of the density function are calculated: $f(x_{min})$, $f(x_{max})$ and $f(x_{mean})$ where *f* represents the density function of the generalized gamma distribution, x_{min}, x_{max} and x_{mean} are respectively the minimum, the maximum and the mean of the sample (x_1,.., x_i,…, x_n) of ***X***. We check that the three points calculated are not infinite neither equal to an ε value. First and third quantile points can also be used for this validation.

If one of the two tests failed, we do not perform the chi-square test and we return a *pvalue* equal to 0 for the combination *a*, *l* and *c* tested.
Finally the algorithm looks like:

```
S ← sample
pvalueTable ← table of double
For c between c1 and c2
        Stemp ←S^c
        A ←mean(Stemp)/var(Stemp)
        a ← A^(1/p)
        l ← mean(Stemp)²/var(Stemp)
        if (gamma(l) is finite and  validatePoints(a, l,  c)=true)  then
                pvalueTable [i] ← pvalue of the chi-square test between Stemp and gamma(A, l)
        else
                pvalueTable [i] ←0
        end if
End loop
Return parameters for which the pvalue is maximum.
```

Algorithm 1: basis of the algorithm I.T.E.V.(Iteration Transformation Estimation Validation)

The most important point in this algorithm is performance. Indeed, loops can take many memory resources and slow down the parameter estimation process. So, the choice in the implementation of the loop and the choice for the values c_1 and c_2 are important. Furthermore, the design of the loop depends on the precision needed for the parameter *c*. So, there is a compromise to find between precision and performance.

For instance, let's suppose that the precision required is a tenth for parameter *c*., there is two ways to implement the loop on the parameter *c*. The first implementation is the one presented in Algorithm 1: one single loop between c_1 and c_2 with step equal to 0.1. Another approach could be:

- First loop will go trough all unit *c* value: *1, 2, 3, 4*, etc. At the end of the first loop, the *c* parameter for which the chi-square test is maximum is kept: it is c_3.
- Second loop will give the precision requested by looping around the c_3 parameter: c_3*-0.9*, c_3*-0.8,…,* c_3*+0.1, …,* c_3*+0.9*

This kind of loop is very powerful in term of performance compared to one single loop between c_1 and c_2 values with step equal to 0.1. As an example, if c_1=-20 and c_2=20 the performances are:

- Number of iterations for one single loop: 400
- Number of iterations for two loops: 80

However, if performance is good by using two loops, there is a risk to find a local maximum.
This algorithm was implemented in Splus software and in Java language.
Table 1 presents a performance comparison between Splus and Java implementation. The estimation method is applied on the same samples in Java and Splus software. We capture the time of estimation in Table 1. In this comparison test, we have tested the two ways to implements the loop on parameter *c*. G.G. Full corresponds to the parameter estimation of the generalized gamma distribution for one single loop with c_1=-20 and c_2=-20 values with step equal to 0.1. G.G. Small corresponds to the parameter estimation of the generalized gamma distribution for two loops with c_1=-20 and c_2=-20.

	Java		Splus	
Sample size	G.G. Full	G.G. Small	G.G. Full	G.G. Small
60	<1s	<1s	2s	1s
100	<1s	<1s	2s	1s
500	1s	1s	3s	1s
1,000	2s	1s	3s	1s
5,000	7s	2s	8s	1s
10,000	15s	3s	19s	4s

Table 1: Performance comparison for the parameter estimation of the generalized gamma distribution. The comparison is between Splus and Java implementation and between the two implementations of the loop on parameter *c*: one single loop with step equal to 0.1 (G.G. Full) and two loops (G.G. Small)

Loops in Splus are known to not be performant because of a bad management of the memory [11]. However, in table 4, we see that the time to estimate the parameters of a generalized gamma distribution for a sample of size 5,000 is less than 10 seconds with Splus implementation.
We presented in this paragraph a new routine to estimate the parameters of the generalized gamma distribution. This method is based on the loop on the parameter *c*. We have seen in table 1 that this method is acceptable regarding the performance aspect. In the next paragraph, we apply the method on samples that follow distributions known as special cases of the generalized gamma distribution in order to validate the algorithm.

5 Validation of the algorithm I.T.E.V on known generalized gamma distribution

In the previous paragraph, we introduced a new estimation method for the parameters of the generalized gamma distribution. We will now validate the routine by applying it on the distributions that are particular cases of the generalized gamma distribution. These particular cases are listed by Stacy [15], Kleiber & Kotz [6] and Samorodnitsky & Taqq [13].
We use Splus software to generate the samples and to implement the algorithm. For each distribution to be tested, we generate five samples of size 10,000 on which we apply the algorithm. The results of the tests are listed in tables 2 to 10.

	c	*a*	*l*
Theory	1	1	1
Test 1	1	0.99	0.99
Test 2	1	1.03	1.03
Test 3	1	1	1
Test 4	1	1.03	1.02
Test 5	1	1.01	1

Table 2: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for an exponential(1) distribution.

	c	*a*	*l*

Theory	1	1	3
Test 1	1	1.01	3.04
Test 2	1	1	2.99
Test 3	1	0.97	2.94
Test 4	1	1	2.98
Test 5	1	1.01	3.03

Table 3 : Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a gamma(1,3) distribution

	c	*a*	*l*
Theory	1	0.5	10
Test 1	1	0.5	10.07
Test 2	0.9	0.81	12.22
Test 3	0.9	0.82	12.28
Test 4	1	0.51	10.17
Test 5	1	0.5	10.05

Table 4 : Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Chi-square(20) distribution

	c	*a*	*l*
Theory	2	2	1
Test 1	2.3	1.73	0.78
Test 2	2.1	1.91	0.92
Test 3	2.2	1.8	0.84
Test 4	2	2.01	1
Test 5	2.1	1.9	0.92

Table 5: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Rayleigh distribution.

	c	*a*	*l*
Theory	2	1/3	1
Test 1	1.9	0.36	1.1
Test 2	1.9	0.35	1.08
Test 3	1.9	0.36	1.09
Test 4	2	0.33	0.99
Test 5	2	0.33	1

Table 6: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Weibull distribution of parameters 2 and 3.

	c	*a*	*l*
Theory	2	0.71	0.5
Test 1	2	0.71	0.5
Test 2	2.1	0.68	0.46
Test 3	2	0.7	0.5
Test 4	2.1	0.69	0.47
Test 5	1.9	0.76	0.57

Table 7: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Half-normal(0,1) distribution.

	c	*a*	*l*
Theory	-1	2	0.5
Test 1	-1.1	2.24	0.43
Test 2	-1	1.95	0.51
Test 3	-1	2.01	0.5
Test 4	-1	2.08	0.48
Test 5	-1	2.2	0.51

Table 8: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Levy distribution.

	c	a	l
Theory	Limit case when l tend to infinite		
Test 1	0.1	1E20	98
Test 2	-0.1	1E-20	98
Test 3	-0.1	1E-20	98
Test 4	0.1	1E20	101
Test 5	-0.1	1E-20	101

Table 9: Results of the algorithm implemented in Splus and applied on samples of size 10,000 for a Log-normal(0,1) distribution.

The results in table 9 present that the log-normal distribution can be fitted by two generalized gamma distribution. This result has been validated by some calculations not presented here because the demonstration of this result is out of the article scope.
All the tests presented above were controlled drawing both theoretical density function with estimated generalized gamma density function. The figure 3 presents an example of the result for the log-normal distribution.

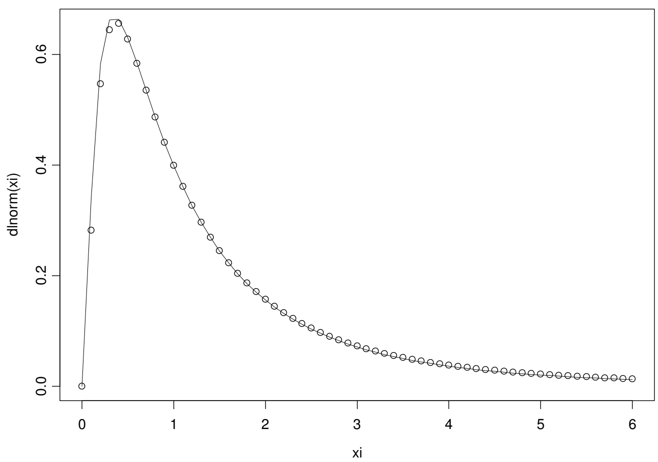

Figure 3: Log-normal density function versus generalized gamma distribution density function with parameter c=0.1, a=1e20, l=100

After validating the algorithm in term of performance, we proved in this paragraph that the estimation method is correct. Indeed, by applying it on known special cases of the generalized gamma distribution, we have the expected parameters. Cohen and Whitten have proposed earlier a routine that already looped on parameter *c*. We have reused this approach and we have associated to it the power of the computer and a variable transformation.

6 Conclusion

The estimation of the parameters for the generalized gamma distribution is treated in many articles. Moments and maximum likelihood methods don't provide a unique and simple solution. These methods were derived by some authors but the results are very complex. Some others methods, not based on the moment or the maximum likelihood methods, were proposed: Stacy proposed a graphic method and Cohen &Whitten a routine that loops on the parameter *c*. We use this last idea to define our routine. We use also the power transformation in order to have a gamma distribution. Then, we use the chi-square test between the sample to be fitted and the gamma distribution.
The new algorithm proposed is powerful, provides the expected values when it is applied on particular cases of the generalized gamma distribution. The only constraint is in the values of the sample we want to fit. Indeed, the generalized gamma distribution is defined for positive values. Consequently, the algorithm designed in this

article requires also positive values in the sample to be fitted. To remove this inconvenient, we can use the four-parameter generalized gamma distribution.
There are two perspectives for this work. The first one is to replace the Chi-squared test by another goodness-of-fit test like AIC or the log-likelihood test in order to confirm that the estimation method is independent from the goodness-of-fit test used. The second one is to replace Splus language by another one in order to confirm that the estimation method is independent from the language.

7 Bibliography

[1] BOBEE, B. (1988). **"***The generalized method of moments applied to the log-Pearson 3 distribution". *Journal of Hydraulic Engineering, ASCE, 114(8): 899-909

[2] COHEN A.C., WHITTEN B. J., "*Parameter estimation in reliability and life span models"*, Marcel Dekker, New York, 1988

[3] HIROSE H., *"Maximum likelihood parameter estimation by model augmentation with application to the extended four-parameter generalized gamma distribution*" [on line] available on http://search2.libi.kyutech.ac.jp/TR/pdf/csse1.pdf consulted the 12/10/04

[4] HWANG T-Y, HUANG P-H, *"On new moment estimation of parameters of the generalized gamma distribution using it's characterization"* [on line] available on http://www.stat.lanl.gov/MMR2004/Extended%20Abstracts/TYHwang.pdf consulted the

[5] Hyfran [on line] available on http://www.inrs-ete.uquebec.ca/activites/groupes/chaire_hydrol/pdf/HYFRAN-ficheA2004.pdf consulted the 11/11/04

[6] KLEIBER C., KOTZ S., *"Statistical size distributions in economics and actuarial sciences",* wiley interscience: 2003, ISBN 0-471-15064-9

[7] KRAUSE A., OLSON M., "*The basics of S and S-PLUS*", Second edition, Statistics and computing Edition Springer, ISBN 0-387-98961-7. 2001

[8] LAWLESS J.F., *"Statistical models and methods for lifetime data",* Wiley-Interscience, New-York:1982

[9] LEVY-VEHEL J., WALTER C., *"Les marchés fractals, Efficience, ruptures et tendances sur les marchés financiers",* 1ère édition, Paris : finance Puf, 2002, 194 pages

[10] MANNING W.G., BASU A., "*Modeling Costs with Generalized Gamma Regression*" [on line] available on http://www.src.uchicago.edu/prc/pdfs/mannin02.pdf consulted the 8/3/2005

[11] PACIOREK C., *"Memory management in S"* [on line] available on http://www.biostat.harvard.edu/~paciorek/computing/Memoery_management_in_S.html consulted on 7/23/2005

[12] PHAM T., ALMHANA J., "*The Generalized Gamma Distribution: its Hazard Rate and Stress-Strenght Mode*l", IEEE transactions on reliability, vol 44, n° 3, September 1995.

[13]SAMORODNITSKY Gennady, TAQQU Murad S., *"Stable Non-Gaussian Random Processes",* Chapman and Hall/CRC 2000 ISBN 0-412-05171-0 Page 10

[14] Splida, *"Releases notes"* [on line] available on http://www.public.iastate.edu/~wqmeeker/splida/ReleaseNotesSV4.txt consulted the 11/11/04

[15] STACY E.W, MIHRAN G.A., "*Parameter estimation for a generalized gamma distributions"*, technometrics, Vol 7, N°3, Aout 1965