

# STATISTICAL ANALYSIS OF NATURAL EVENTS IN THE UNITED STATES

BY CHARLES LEVI<sup>1</sup>

*Compagnie Transcontinentale de Réassurance, Paris*

AND CHRISTIAN PARTRAT

*Institut de Statistique, Université Pierre et Marie Curie, Paris*

## ABSTRACT

A statistical analysis is performed on natural events which can produce important damages to insurers. The analysis is based on hurricanes which have been observed in the United States between 1954 et 1986.

At first, independence between the number and the amount of the losses is examined. Different distributions (Poisson and negative binomial for frequency and exponential, Pareto and lognormal for severity) are tested. Along classical tests as chi-square, Kolmogorov-Smirnov and non parametric tests, a test with weights on the upper tail of the distribution is used: the Anderson – Darling test.

Confidence intervals for the probability of occurrence of a claim and expected frequency for different potential levels of claims are derived.

The Poisson Log-normal model gives a very good fit to the data.

## KEYWORDS

Catastrophe risk; fitting models; frequency; severity; XL treaties.

## 1. INTRODUCTION

The United States of America are regularly hit by different types of natural events. Hurricanes affect the east part of the United States, tornadoes the middle one. Hailstorms and winter freeze may take place all over the United States. Earthquakes are observed in some specific zones as California (for example 1906 and 1989 San Francisco quakes).

These events cause very important losses. On the average the insured losses represent 4% of the premium income in classes as fire and multiperils for homeowners, farmowners and commercial risks.

<sup>1</sup> Presented at the 21th Astin Colloquium, New-York, November 15–17, 1989.

A very important hurricane may induce a 8 billion US \$ insured loss which would represent 20 % of the premium income of these classes for one year. This percentage is even higher for an insurance company located in hurricane prone zones (Texas, Florida, Georgia, ...).

Direct insurers and reinsurers (underwriting non proportional treaties) must estimate their exposure in order to define an adequate reinsurance coverage.

The topic of the study is to get some results on the loss amount and frequency distributions of these events. In order to do homogeneous analysis, the study has been realized on a sample of hurricanes affecting the United States.

ISO keeps in its data base all losses (natural events) since 1949 whose amount exceeds 1 million US \$ (5 millions US \$ after 1982). Three factors explain the evolution of the losses amount from 1949: inflation, the number of

TABLE 1  
HURRICANES EXCEEDING 30 MILLIONS \$

Year	Frequency	First loss	Second loss	Other losses		
1986	0					
1985	5	39,7	582,0	439,9	47,1	83,9
1984	1	41,2				
1983	1	893,1				
1982	1	192,0				
1981	0					
1980	1	106,2				
1979	2	216,7	1243,4			
1978	0					
1977	0					
1976	1	52,8				
1975	1	351,6				
1974	1	36,2				
1973	0					
1972	1	431,5				
1971	1	57,3				
1970	1	1602,1				
1969	1	822,2				
1968	0					
1967	1	260,1				
1966	1	58,7				
1965	1	6299,9				
1964	3	814,9	137,2	203,8		
1963	0					
1962	0					
1961	2	1263,5	53,7			
1960	1	1313,0				
1959	2	118,4	167,8			
1958	1	70,1				
1957	1	503,7				
1956	1	64,8				
1955	2	529,9	87,8			
1954	3	2465,4	317,9	2753,9		

people having the coverage against hurricanes in their insurance policy, demographic evolution.

These three factors have been taken into account in the trending of the losses (FRIEDMAN, 1987) in order to get an homogeneous data base in 1987 US \$. Nevertheless as the indexation coefficients for the first years were close to 100 and those for the years 1954 to 1982 were lower than 30, the observed period of time has been shortened to 33 years (1954 to 1986). During these years 37 hurricanes have been observed (cost of each hurricane in 1987 US \$ exceeding 30 millions).

2. HYPOTHESIS

Consider  $N$  the random variable (r.v.) of the yearly loss frequency  $N(x_0)$  the r.v. of the losses exceeding  $x_0$ , with  $x_0$  fixed. Let  $X_i$  be the amount of the loss  $i$  and  $\underline{X} = (X_1, \dots, X_N)$  the r.v. of the yearly loss amounts; the distribution of each  $X_i$  is supposed continuous.

$K$  observations years ( $K = 33$ ) are available. They produce a realization  $(n_k, \underline{x}^{(k)})_{k=1, \dots, K}$  of a  $K$ -sample  $(N_1, \underline{X}^{(1)}), \dots, (N_K, \underline{X}^{(K)})$  of  $(N, \underline{X})$ .

Two hypothesis are made

(H1)  $N$  and  $(X_1, X_2, \dots)$  are independent random variables

(H2)  $X_1, X_2, \dots$  are i.i.d. random variables.

(H1) may be partly checked looking at the 25 years for which at least one loss has been observed. The grouping of the first losses in three classes gives the following contingency table (into parenthesis theoretical frequencies in case of independence).

Yearly frequency \ Loss	Loss			Total
	$\leq 200$	$200 < \leq 1000$	$> 1000$	
1	9 (7,92)	6 (6,48)	3 (3,60)	18
2	1 (1,76)	2 (1,44)	1 (0,80)	4
3 and over	1 (1,32)	1 (1,08)	1 (0,60)	3
Total	11	9	5	25

Chi-square independence test gives an observed  $x_{obs}^2 = 1,23$  which for the significance level ( $P$ -value) is  $\hat{\alpha} = P(x_4^2 > x_{obs}^2) = 0,87$ . So (H1) is accepted.

**Remark:** A grouping of yearly frequencies in two classes in order to follow the

$$\text{Cochran criterion } \left[ \frac{n_i \cdot n \cdot j}{n} \geq 1 \forall (i, j) \text{ and } \frac{n_i \cdot n \cdot j}{n} \geq 5 \text{ for at least 80\% of } (i, j) \right]$$

would lead to the same conclusion.

For (H2) independence of  $X_1$  and  $X_2$ , and identical distribution of  $X_1, X_2$  and  $X_3$  are checked using three non parametric tests: Kendall, Spearman and Kruskal-Wallis (GIBBONS, 1974).

— Independence of  $X_1$  and  $X_2$

For the 7 years during which at least 2 losses have been observed, the Kendall tau statistic can be written as follows:

$$T = \frac{2}{7(7-1)} \sum_{\substack{1 \leq i < j \leq K \\ n_i, n_j \geq 2}} A_{ij} \quad \text{with}$$

$$A_{ij} = \begin{cases} 1 & (X_1^{(j)} - X_1^{(i)})(X_2^{(j)} - X_2^{(i)}) > 0 \\ 0 & \text{if } (X_1^{(j)} - X_1^{(i)})(X_2^{(j)} - X_2^{(i)}) = 0 \\ -1 & (X_1^{(j)} - X_1^{(i)})(X_2^{(j)} - X_2^{(i)}) < 0 \end{cases}$$

The critical region for Kendall test at level  $\alpha = 0,20$  is  $W_\alpha = \{|T| > 0,4286\}$ . The observed tau being  $T = -0,333$ , independence between  $X_1$  et  $X_2$  can be assumed for any reasonable level.

Let  $R_k$  be the rank of  $X_1^{(k)}$  among the 7 observations (ordered increasingly) for which  $n_k \geq 2$  and  $S_k$  be the rank of  $X_2^{(k)}$ , the Spearman rho statistic is

$$R = \frac{\sum_k (R_k - \bar{R})(S_k - \bar{S})}{\sqrt{\sum_k (R_k - \bar{R})^2} \sqrt{\sum_k (S_k - \bar{S})^2}}$$

The critical region for Spearman test with a level  $\alpha$  of 0,20 is  $W_\alpha = \{|R| > 0,536\}$ , observed rho is computed at  $-0,464$  so the conclusion is the same as for Kendall test.

— Identical distribution of  $X_1, X_2$  and  $X_3$

Let  $F_i$  ( $i = 1, 2, 3$ ) be the cumulative distribution function (c.d.f.) of  $X_i$ , only years when at least  $i$  losses have occurred being selected:  $\{X_i^{(k)}: k \text{ with } n_k \geq i\}$ .

The null hypothesis  $F_1 = F_2 = F_3$  is tested against the alternative  $\exists i, j: F_i \neq F_j$  by the Kruskal-Wallis test. Under the assumption that loss amounts  $(X_i^{(k)})_{i \geq 3}$  are identically distributed, we have a  $m_1$ -sample of  $X_1$  ( $m_1 = 25$ ), a  $m_2$ -sample of  $X_2$  ( $m_2 = 7$ ) and a  $m_3$ -sample of  $X_3$  ( $m_3 = 5$ ). These samples are assumed to be independent.

Let  $M = \sum_{i=1}^3 m_i, R_i$  the sum of ranks of the  $i$ th sample observations in

the combined (increasingly) ordered configuration of the  $M$  observations with  $\sum_{i=1}^3 R_i = \frac{M(M+1)}{2} : R_1 = 482, R_2 = 128, R_3 = 93$ . Under the null

hypothesis  $E(R_i) = \frac{m_i(M+1)}{2} \forall i [E(R_1) = 475, E(R_2) = 133, E(R_3) = 95]$ ,

the Kruskal-Wallis statistic

$$KW = \frac{12}{M(M+1)} \sum_{i=1}^3 \frac{1}{m_i} \left[ R_i - \frac{m_i(M+1)}{2} \right]^2$$

is free (its distribution is independent of the common  $F_i$  distribution). Asymptotically ( $m_i \rightarrow +\infty \forall i$ )  $KW$  is chi-squared distributed with 2 degrees of freedom. This asymptotic distribution is used in practice when  $m_i \geq 5 \forall i$ . Here the critical region for the Kruskal-Wallis test  $\{KW > c\}$  has a significance level  $\hat{\alpha} = P(x_2^2 > KW) = 0,97$  (the observed  $KW$  statistic having a value of 0,054).

**Remark :**

1. If the size of the third sample  $m_3 (= 5)$  seems too small to use the asymptotic distribution of  $KW$ , it is still possible to test  $F_1 = F'$  against  $F_1 \neq F'$  [ $F'$  being the c.d.f. of  $X_i (i \geq 2)$ ] with a  $m_1$ -sample of  $F_1 (m_1 = 25)$  and a  $m'$ -sample of  $F' (m' = 12)$ . In this case the Kruskal-Wallis test is the Mann-Whitney-Wilcoxon test and has a significance level  $\hat{\alpha} = 0,82$ .

2. Under the assumption of the  $X_i$ 's independence the Kruskal-Wallis test may be used to check the hypothesis  $(H1)$ : no effect of the yearly loss frequency upon their amount:

Considering the yearly loss amounts  $\{X_1^{(k)} : k \text{ with } n_k = 1\}$  for years when exactly one hurricane occurs,  $\{X_i^{(k)} = k \text{ with } n_k = 2; i = 1, 2\}$  for years with two hurricanes and  $\{X_i^{(k)} : k \text{ with } n_k \geq 3; i = 1, 2, \dots, n_k\}$  for years with more than two hurricanes as independent samples with respective sizes  $m_1 = 18, m_2 = 8, m_3 = 11$  of distributions  $G_1, G_2, G_3$ , the Kruskal-Wallis test of the null hypothesis  $G_1 = G_2 = G_3$  gives a significance level  $\hat{\alpha} = 0,89$  (observed  $KW = 0,25$ ).

Hereafter  $(H1)$  and  $(H2)$  will be assumed to be true.  $X$  will be the random variable parent of  $X_i$  and  $F_X$  its c.d.f. (assumed to be continuous.)

3. LOSS FREQUENCY

The realization  $(n_1, \dots, n_K)$  of the  $K$ -sample  $(N_1, \dots, N_K)$  from  $N$  is given in

the following Table 2. Let  $\bar{n} = \frac{1}{K} \sum_k n_k$  and  $\hat{\sigma}_n^2 = \frac{1}{K} \sum_k (n_k - \bar{n})^2$ .

Different distributions fitting the loss frequency are examined.

— Poisson distribution  $\mathcal{P}(\lambda)$  ( $\lambda > 0$ )

with  $P_\lambda(N = n) = e^{-\lambda} \lambda^n/n!$  ( $n \in \mathbb{N}$ ),  $E(N) = V(N) = \lambda$ ,  $\hat{\lambda} = \bar{n}$  is the maximum likelihood estimator (M.L.E.) of  $\lambda$ . A confidence interval at a level of at least  $(1-\alpha)$  for  $\lambda$  is  $[\hat{\lambda}_i, \hat{\lambda}_s]$  with

$$\hat{\lambda}_i = \frac{1}{2K} x^2 \quad (\alpha/2) \quad \text{and} \quad \hat{\lambda}_s = \frac{1}{2K} x^2 \quad (1-\alpha/2).$$

$$2 \sum_k n_k \qquad \qquad \qquad 2 \left( \sum_k n_k + 1 \right)$$

In these expressions  $x^2(\alpha/2)$  and  $x^2(1-\alpha/2)$  are the  $\alpha/2$  and  $(1-\alpha/2)$  fractiles of the chi-square distribution.

— Negative binomial  $N\mathcal{B}(r, p)$  ( $r > 0, p \in ]0,1[$ )

with  $P_{r,p}(N = n) = \frac{\Gamma(r+n)}{\Gamma(r)n!} p^r (1-p)^n$  ( $n \in \mathbb{N}$ ),  $E(N) = \frac{r(1-p)}{p}$  and

$V(N) = \frac{r(1-p)}{p^2} > E(N)$ ; the estimation of  $(r, p)$  by the M.L.E. or by the moments requires that the condition  $\hat{\sigma}_n^2 > \bar{n}$  is fulfilled.

From the frequencies by year of hurricanes, we have  $\bar{n} = \frac{37}{33} = 1,12121$

and  $\hat{\sigma}_n^2 = 1,0762$ . So a fit by a negative binomial distribution is impossible.

TABLE 2  
YEARLY FREQUENCY OF HURRICANES EXCEEDING 30 MILLIONS \$

Yearly frequency $i$	Observed freq. $v_i$	Theoretical freq. $K\hat{p}_i$	$\frac{(v_i - K\hat{p}_i)^2}{K\hat{p}_i}$
0	8	10,75	0,703
1	18	12,06	2,926
2	4	6,76	1,127
3	2	2,53	0,111
4	0	} 0,90	0,011
5	1		
6 and over	0		
Total	33	33	4,878 = $\chi_{obs}^2$

The  $\chi^2$  goodness-of-fit test to a Poisson distribution with  $\hat{\lambda} = \bar{n}$ ,  $\hat{p}_i = P_i(N = i)$  and  $c = 5$  classes gives (see Table 2) a significance level  $\hat{\alpha}$  fulfilling condition

$$P(x_{c-2}^2 > x_{\text{obs}}^2) \leq \hat{\alpha} \leq P(x_{c-1}^2 > x_{\text{obs}}^2),$$

belonging to the interval [0,18; 0,30].

So the fit of  $N$  to a Poisson distribution  $\mathcal{P}(\lambda)$  is accepted with for  $\lambda$ :

$$\text{M.L.E. } \hat{\lambda} = 1,12121$$

Confidence interval at a level at least 0,98

$$[0,73736; 1,63005]$$

**Remark :**

1. The M.L.E. of  $\lambda$  obtained from grouped data (5 classes) is  $\hat{\lambda} = 1,09866$ , so to state precisely the chi-square test gives a significance level  $\hat{\alpha} = P(\chi_3^2 > 4,866) = 0,18$ .

2. The fit of a Poisson distribution to that kind of event frequency can be checked with the distribution (see Table 3) of the frequency by year of all the north atlantic hurricanes which approached the United States from 1899 to 1986 (meteorological data, US Department of commerce):

$$\bar{n} = 1,7045, \quad \hat{\sigma}_n^2 = 1,8218, \quad \hat{\alpha} \in [0,72; 0,84].$$

TABLE 3  
YEARLY FREQUENCY OF ALL NORTH ATLANTIC HURRICANES

Yearly frequency $i$	Observed freq. $v_i$	Theoretical freq. $K\hat{p}_i$	$\frac{(v_i - K\hat{p}_i)^2}{K\hat{p}_i}$
0	16	16,00	0,000
1	28	27,28	0,019
2	23	23,25	0,003
3	14	13,21	0,047
4	3	5,63	1,229
5	2	2,63	0,714
6	2		
7 and over	0		
Total	88	88	2,012 = $\chi_{\text{obs}}^2$

4. LOSS AMOUNT

Loss amounts are assumed to be i.i.d. random variables. Let

$n = \sum_{k=1}^K n_k (= 37)$ , a realization  $(x_1, \dots, x_n)$  of a  $n$ -sample  $(X_1, \dots, X_n)$  of  $X$  is obtained; all losses are over 30.

The aim of the following lines is to estimate the probability  $P(X \geq x_0) = 1 - F_X(x_0)$  that a loss amount exceeds  $x_0$  and to derive a confidence interval at a level  $1 - \alpha (= 0,98)$ .

### 1. Non parametric estimation

Let  $X_{(1)} \leq \dots \leq X_{(n)}$  be the ordered sample corresponding to  $(X_1, \dots, X_n)$  and  $(F_n^*(x))_{x \in \mathbb{R}^+}$  the empirical c.d.f.

Considering  $X_{(0)} = 30$ ,  $X_{(n+1)} = +\infty$ , we have, for  $k = 0, \dots, n$ ,

$$F_n^*(x) = \frac{k}{n} \quad \text{if } X_{(k)} < x \leq X_{(k+1)}.$$

The statistic  $1 - F_n^*(x_0) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[x_0, +\infty[}(X_i)$  is an unbiased consistent esti-

mator of  $1 - F_X(x_0)$ .

Furthermore if  $D_n(1 - \alpha)$  is the  $(1 - \alpha)$  fractile  $[D_n(0,98) = 0,244$  for  $n = 37]$  of the Kolmogorov-Smirnov statistic  $D_n = \sup_{x \in \mathbb{R}^+} |F_n^*(x) - F_X(x)|$  associated

to the sample, if we let, for  $x \in \mathbb{R}^+$ ,

$$\begin{aligned} I_n(x) &= \max [1 - F_n^*(x) - D_n(1 - \alpha), 0] \\ S_n(x) &= \min [1 - F_n^*(x) + D_n(1 - \alpha), 1] \end{aligned}$$

the band  $([I_n(x), S_n(x)])_{x \in \mathbb{R}^+}$  is a level  $(1 - \alpha)$  confidence band for  $1 - F_X(x_0)$  meaning that

$$P[I_n(x) \leq 1 - F_X(x) \leq S_n(x) \forall x \in \mathbb{R}^+] = 1 - \alpha.$$

The table with the values of  $1 - F_n^*(x)$ ,  $I_n(x)$  and  $S_n(x)$  for  $k = 0, \dots, n$  and  $x_{(k)} < x \leq x_{(k+1)}$  is presented in Appendix 1.

Joining with segments the points

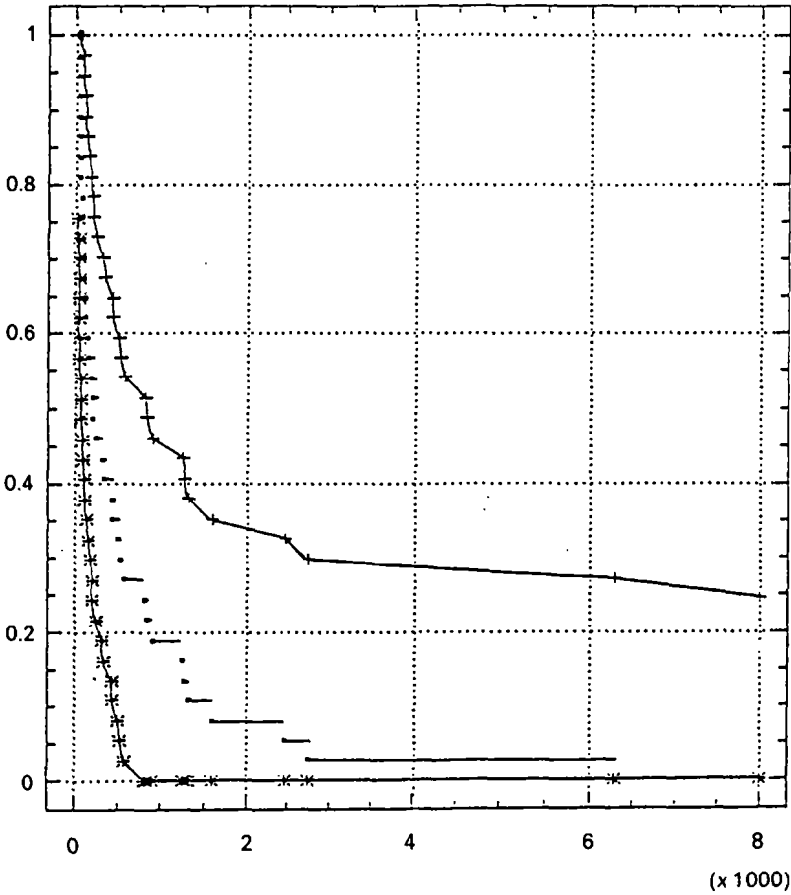
$$\left( x_{(k)}, \min \left( \frac{n-k+1}{n} + D_n(1 - \alpha), 1 \right) \right)_{k=0, \dots, n+1}$$

for the superior envelope and

$$\left( x_{(k)}, \max \left( \frac{n-k}{n} - D_n(1 - \alpha), 0 \right) \right)_{k=0, \dots, n+1}$$

for the inferior envelope, a confidence band  $(B_x)_{x \in \mathbb{R}^+}$  containing the first one  $([I_n(x), S_n(x)])_{x \in \mathbb{R}^+}$  and graphically easier to draw is derived. Graph 1 shows the plot of  $1 - F_n^*(x)$  and  $B_x$  for  $30 \leq x \leq 8000$ .





GRAPH 1. Plot of  $1 - F_n^*(x)$  and  $B_x$ .

**2. Parametric family of distributions**

A graphical approach and the value of significance levels of goodness-of-fit tests based on the empirical c.d.f. (D'AGOSTINO and STEPHENS, 1986) are used to test the fit of observations to a family  $\mathcal{F} = \{F(x; \theta) : \theta \in \Theta\}$  of parametric distributions ( $\theta$  varying in an open subset  $\Theta$  of  $\mathbb{R}^q$ ).

For the graphical procedure ( $Q-Q$  plot) following results are applied: for  $r = 1, \dots, n$

$$E[F_X(X_{(r)})] = \frac{r}{n+1}; \quad V[F_X(X_{(r)})] = \frac{r(n-r+1)}{(n+1)^2(n+2)} = O(1/n).$$

For  $n \geq 30$  (a generally accepted level) a realization of  $F_X(X_{(r)})$  is very likely close to  $\frac{r}{n+1}$ . So it is possible to write  $F_X(x_{(r)}) \simeq \frac{r}{n+1}$  for  $r = 1, \dots, n$ .

By an adequate transformation, depending of the examined family, the procedure is equivalent to estimate whether  $n$  points are roughly on a straight line.

Let  $\hat{\theta}$  be the M.L.E. of  $\theta$  in the hypothesis  $F_X \in \mathcal{F}$ , the goodness-of-fit test is based on Anderson-Darling statistic

$$\begin{aligned} \hat{A}_n^2 &= n \int_0^{+\infty} \frac{[F_n^*(x) - F(x; \hat{\theta})]^2}{F(x; \hat{\theta}) [1 - F(x; \hat{\theta})]} dF(x, \hat{\theta}) \\ &= -n - \frac{1}{n} \sum_{r=1}^n (2r-1) \{ \text{Log } F(X_{(r)}; \hat{\theta}) + \text{Log } [1 - F(X_{(n-r+1)}; \hat{\theta})] \}. \end{aligned}$$

This statistic gives one of the globally most powerful tests (D'AGOSTINO and STEPHENS, 1986). Moreover it is an adequate statistic of the here studied

problem because of the weight  $\left( \text{factor } \frac{1}{1 - F(x; \hat{\theta})} \right)$  given to the tail of the distribution.

In order to compare with other tests, Kolmogorov-Smirnov statistic will be computed:

$$\hat{D}_n = \text{Sup}_{x \in \mathbb{R}^+} |F_n^*(x) - F(x, \hat{\theta})| = \max(\hat{D}_n^+, \hat{D}_n^-) \text{ with}$$

$$\hat{D}_n^+ = \max_{r=1, \dots, n} \left[ \frac{r}{n} - F(X_{(r)}; \hat{\theta}) \right] \text{ and } \hat{D}_n^- = \max_{r=1, \dots, n} \left[ F(X_{(r)}; \hat{\theta}) - \frac{r-1}{n} \right].$$

Let  $\hat{T}_n$  be one of these two test statistics and  $\hat{T}_{n,x}$  its value for the realization  $x = (x_1, \dots, x_n)$  of  $(X_1, \dots, X_n)$ . The distribution of  $\hat{T}_n$  under the null hypothesis  $H_0: F_X \in \mathcal{F}$  depends generally only on  $n$  and the examined family. Thus a significance level  $\hat{\alpha}(x) = P^{H_0}[\hat{T}_n > \hat{T}_{n,x}]$  may be computed from the table of this distribution.

**Remark:** It is not advisable to compare the fit of two families of distributions to the observations by a simple comparison of their  $\hat{T}_{n,x}$ . Indeed the same deviation has not the same likelihood to be reached under  $H_0$ . For example considering  $\hat{D}_n$ , for  $n = 37$  and  $\hat{D}_{n,x} = 0,165$ :  $P^{H_0}(\hat{D}_n > \hat{D}_{n,x}) = 0,24$  if  $\mathcal{F} = \{F_0\}$  has only one distribution (fully specified),  $P^{H_0}(\hat{D}_n > \hat{D}_{n,x}) = 0,08$  if  $\mathcal{F}$  is the exponential distributions family,  $P^{H_0}(\hat{D}_n > \hat{D}_{n,x}) = 0,15$  if  $\mathcal{F}$  is the log-normal distributions family.

The histogram of the observations suggest to choose a dissymmetrical distribution. Successively exponential, Pareto and log-normal distributions will be tried: for these distributions there are statistical tables which give the goodness-of-fit significance levels  $\hat{\alpha}(x)$ .

2a. Exponential distribution  $\varepsilon(\beta; 30)$

With  $\beta > 0$ , this distribution has the following density and c.d.f.

$$f_{\beta, 30}(x) = \beta e^{-\beta(x-30)} \mathbb{1}_{[30, +\infty[}(x)$$

$$F_{\beta, 30}(x) = 1 - e^{-\beta(x-30)} \quad (x \geq 30) \quad \text{so} \quad -\text{Log}[1 - F_{\beta, 30}(x)] = \beta(x - 30).$$

Let  $(Y_1, \dots, Y_n)$  be a  $n$ -sample of  $\{\varepsilon(\beta, 30) : \beta > 0\}$  and  $(y_1, \dots, y_n)$  its realization.

\* the  $n$  points  $\left( y_{(r)} - 30, -\text{Log} \left( 1 - \frac{r}{n+1} \right) \right) \quad r = 1, \dots, n$  are roughly on a straight line going through  $(0, 0)$  with a positive slope (the slope of an adjusted line on these points gives if necessary a graphical estimation of  $\beta$ ).

\* M.L.E. of  $\beta$  is  $\hat{\beta} = \frac{n}{\sum_{i=1}^n (y_i - 30)}$ , M.L.E. of  $1 - F_{\beta, 30}(x_0)$  is

$$1 - F_{\hat{\beta}, 30}(x_0) = e^{-\hat{\beta}(x_0 - 30)} \quad \text{for } x_0 \geq 30.$$

\* a level  $(1 - \alpha)$  confidence interval with symmetric risks is

$$\text{for } \beta : \left[ \hat{\beta} \frac{x_{2n}^2(\alpha/2)}{2n} ; \hat{\beta} \frac{x_{2n}^2(1-\alpha/2)}{2n} \right]$$

for  $1 - F_{\beta, 30}(x_0)$ :

$$\left\{ \exp \left\{ \frac{-\hat{\beta} x_{2n}^2(1-\alpha/2)}{2n} (x_0 - 30) \right\} ; \exp \left\{ \frac{-\hat{\beta} x_{2n}^2(\alpha/2)}{2n} (x_0 - 30) \right\} \right\}$$

as  $\frac{2n\beta}{\hat{\beta}}$  is  $x_{2n}^2$  distributed (d.f.  $2n$ ).

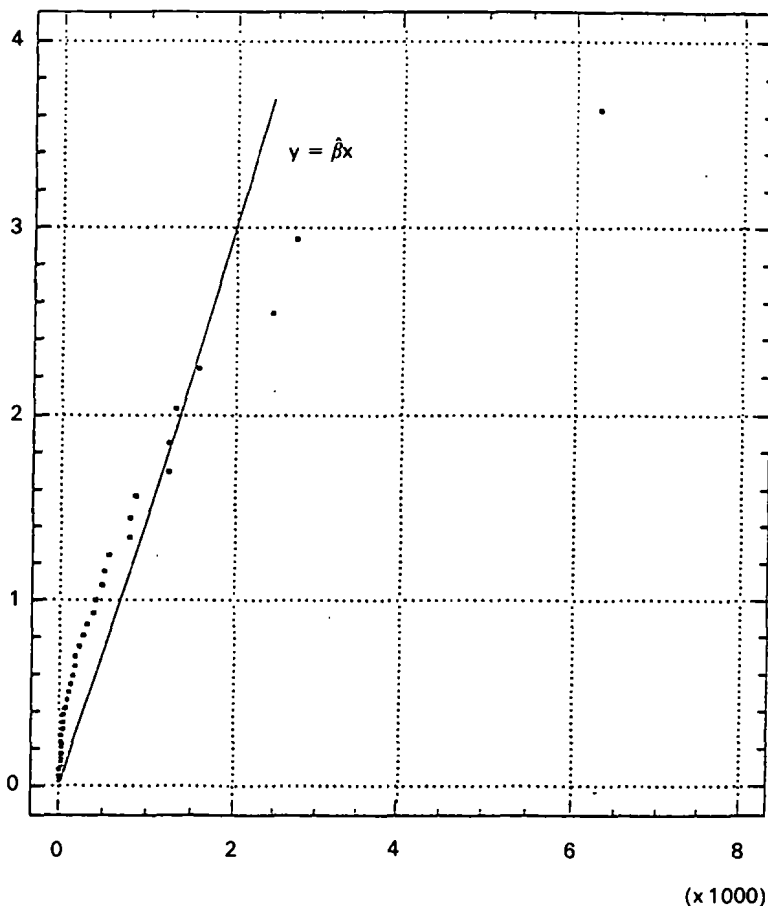
The graphical procedure applied to the 37-sample  $(x_1, \dots, x_n)$  of  $X$  in graph 2 rejects in a first approach a fit to an exponential distribution: the tail of this distribution is too light to take into account the observed amounts of loss.

With  $\hat{\beta} = \frac{1}{638,2} = 0,00157$ , the significance levels of the goodness-of-fit tests

corroborate the lack of fit of the exponential distribution to the data :

$$\hat{A}_n^2 = 5,98054 \quad \hat{\alpha}(x) \ll 0,0025$$

$$\hat{D}_n = 0,2599 \quad \hat{\alpha}(x) \ll 0.005.$$



GRAPH 2. Plot of the points  $\left( x_{(r)} - 30, -\text{Log} \left( 1 - \frac{r}{n+1} \right) \right) r = 1, \dots, n$ .

### 2b. Pareto distribution $P(\gamma; 30)$

With  $\gamma > 0$  this distribution has the following density and c.d.f.

$$g_{\gamma, 30}(x) = \frac{\gamma 30^\gamma}{x^{\gamma+1}} \mathbb{1}_{]30, +\infty[}(x),$$

$$G_{\gamma, 30}(x) = 1 - \left( \frac{30}{x} \right)^\gamma \text{ so } -\text{Log} [1 - G_{\gamma, 30}(x)] = \gamma \text{Log} \frac{x}{30} \quad (x \geq 30).$$

Let  $(Y_1, \dots, Y_n)$  be a  $n$ -sample of  $\{P(\gamma; 30) : \gamma > 0\}$  and  $(y_1, \dots, y_n)$  its realization.

\* the  $n$  points  $\left( \text{Log} \frac{y(r)}{30}, -\text{Log} \left( 1 - \frac{r}{n+1} \right) \right) r = 1, \dots, n$  are roughly on a straight line going through  $(0, 0)$  with a positive slope (the slope of an adjusted line on these points gives if necessary a graphical estimation of  $\gamma$ ).

\* M.L.E. of  $\gamma$  is  $\hat{\gamma} = \frac{n}{\sum_{i=1}^n \text{Log} \frac{Y_i}{30}}$ , M.L.E. of  $1 - G_{\gamma,30}(x_0)$  is

$$\widehat{1 - G_{\gamma,30}(x_0)} = \left( \frac{30}{x_0} \right)^{\hat{\gamma}} \text{ for } x_0 \geq 30.$$

\* a level  $(1 - \alpha)$  confidence interval with symmetric risks is

$$\text{for } \gamma : \left[ \hat{\gamma} \frac{x_{2n}^2(\alpha/2)}{2n}; \hat{\gamma} \frac{x_{2n}^2(1-\alpha/2)}{2n} \right]$$

$$\text{for } 1 - G_{\gamma,30}(x_0) : \left[ \left( \frac{30}{x_0} \right)^{\hat{\gamma} \frac{x_{2n}^2(1-\alpha/2)}{2n}}; \left( \frac{30}{x_0} \right)^{\hat{\gamma} \frac{x_{2n}^2(\alpha/2)}{2n}} \right]$$

as  $\frac{2n\gamma}{\hat{\gamma}}$  is  $x_{2n}^2$  distributed.

Graph 3 shows that the  $n$  points  $\left( \text{Log} \frac{x(r)}{30}, -\text{Log} \left( 1 - \frac{r}{n+1} \right) \right)$

$r = 1, \dots, n$  are not roughly on a straight line. Pareto distribution has a too heavy tail for the observed amounts of loss.

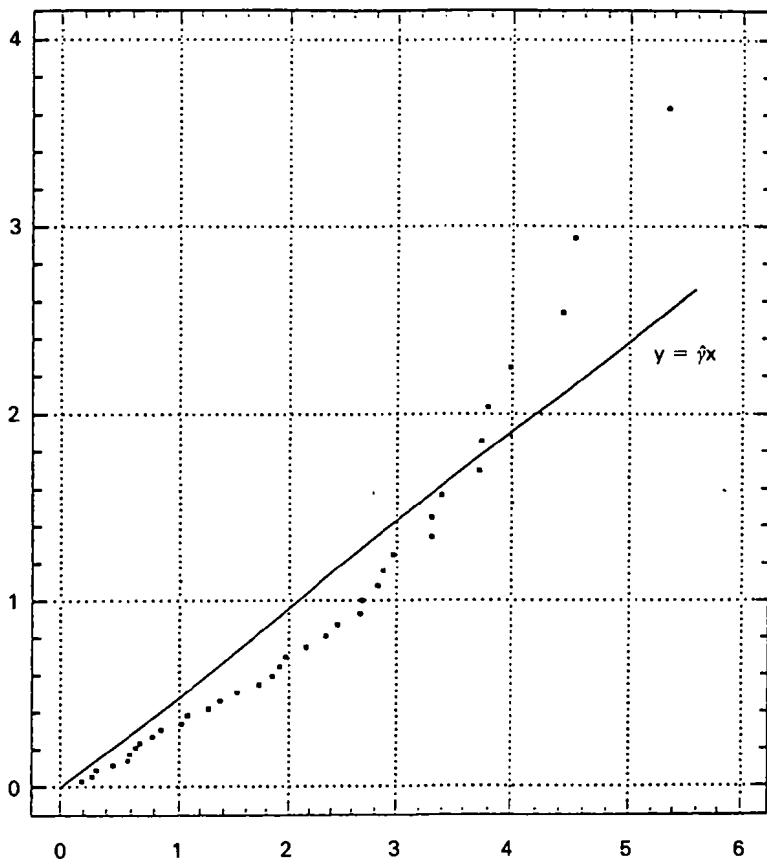
With  $\hat{\gamma} = 0,465141$  the test statistics can be computed as follows

$$\hat{A}_n^2 = 1,56365 \text{ with a significance level } \hat{\alpha}(x) = 0,025$$

$$\hat{D}_n = 0,14586 \text{ with a significance level } \hat{\alpha}(x) = 0,16 .$$

Comparing the two significance levels demonstrates the interest of  $\hat{A}_n^2$  relatively to  $\hat{D}_n$ . The fit to a Pareto distribution is rejected by  $\hat{A}_n^2$  (tail of the distribution) though such a fit seems to be acceptable with  $\hat{D}_n$ , taking into account the small number of observations.

The fit to a Pareto distribution being rejected, the lower and upper limits of the confidence interval are not computed.



GRAPH 3. Plot of the points  $\left( \text{Log} \frac{x_{(r)}}{30}, -\text{Log} \left( 1 - \frac{r}{n+1} \right) \right) r = 1, \dots, n.$

### 2c. Log-normal distribution $\text{Log} N(\mu, \sigma; 30)$

With  $\mu \in \mathbb{R}$  and  $\sigma > 0$ , a random variable  $Y$  is log-normally distributed if  $\text{Log}(Y-30)$  is normally distributed  $N(\mu, \sigma)$ . Its density is

$$h_{\mu, \sigma, 30}(x) = \frac{1}{\sqrt{2\pi} \sigma (x-30)} \exp \left\{ -\frac{1}{2\sigma^2} [\text{Log}(x-30) - \mu]^2 \right\} \mathbb{1}_{]30, +\infty[}(x).$$

Let  $\Phi$  be the c.d.f. of  $N(0, 1)$ , the c.d.f. of the log-normal distribution can be written ( $x \geq 30$ )

$$H_{\mu, \sigma, 30}(x) = \Phi \left[ \frac{\text{Log}(x-30) - \mu}{\sigma} \right] \text{ therefore } \Phi^{-1}[H_{\mu, \sigma, 30}(x)] = \frac{\text{Log}(x-30) - \mu}{\sigma}.$$

Let  $(Y_1, \dots, Y_n)$  be a  $n$ -sample of  $\{\text{Log } N(\mu, \sigma, 30) : \mu \in \mathbb{R}, \sigma > 0\}$

\* The  $n$  points  $\left( \text{Log } (y_{(r)} - 30), \Phi^{-1} \left( \frac{r}{n+1} \right) \right) r = 1, \dots, n$  are roughly on a straight line with a positive slope.

\* M.L.E. of  $(\mu, \sigma)$  is  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \text{Log } (Y_i - 30)$

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n [\text{Log } (Y_i - 30) - \hat{\mu}]^2}$$

M.L.E. of  $1 - H_{\mu, \sigma, 30}(x_0)$  is  $\widehat{1 - H_{\mu, \sigma, 30}}(x_0) = 1 - \Phi \left[ \frac{\text{Log } (x_0 - 30) - \hat{\mu}}{\hat{\sigma}} \right]$

\* the way to derive a confidence interval for  $1 - H_{\mu, \sigma, 30}(x_0)$  is explained later.

Graph 4 shows a very good fit of the log-normal distribution to the 37 observations. It is corroborated by the values of the test statistics computed with  $\hat{\mu} = 5.19853$  and  $\hat{\sigma} = 1.74297$ :

$$\hat{A}_n^2 = 0.26265 \text{ with a significance level } \hat{\alpha}(x) = 0.70$$

$$\hat{D}_n = 0.07939 \text{ with a significance level } \hat{\alpha}(x) \geq 0.15.$$

The values of  $\widehat{1 - H_{\mu, \sigma, 30}}(x_0)$  for  $x_0$  varying from 100 to 8000 are presented in Appendix 2 (column 1) and plotted in Graph 5.

2d. *Confidence interval for*  $1 - \Phi \left[ \frac{\text{Log } (x_0 - 30) - \mu}{\sigma} \right]$

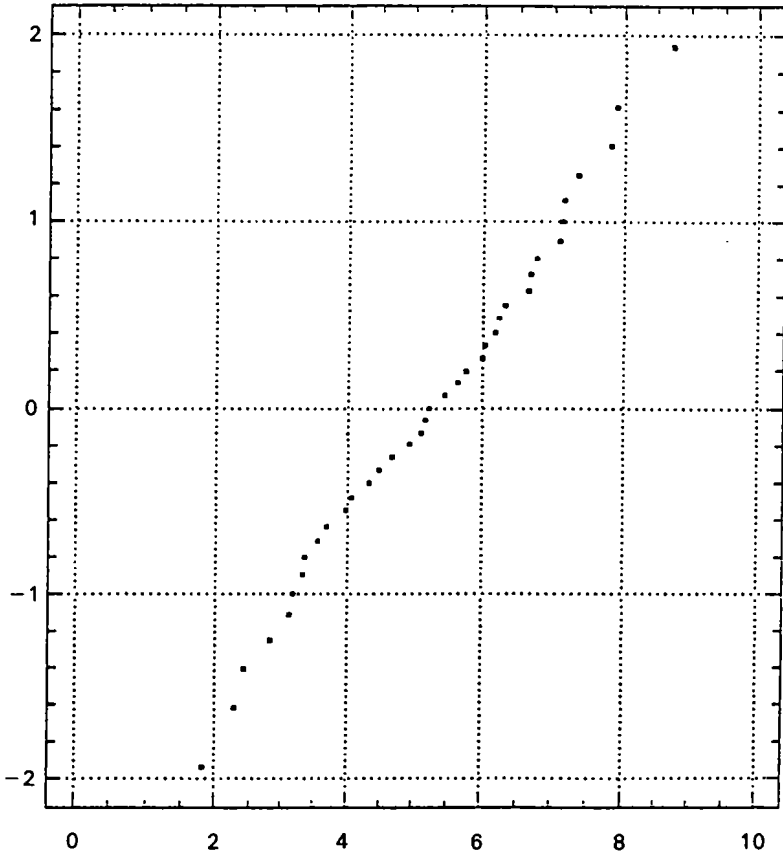
As the size of the sample ( $n = 37$ ) is too small to use the confidence interval derived from the asymptotic normality of  $(\hat{\mu}, \hat{\sigma})$  and the  $\delta$ -method, the non-central Student distribution and its table (RESNIKOFF and LIEBERMAN, 1957) are to be used.

Let  $Y_i = \text{Log } (X_i - 30)$  for  $i = 1, \dots, n$ ,  $\bar{Y} = \frac{1}{n} \sum_i Y_i$  and

$$S_Y^2 = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2, (Y_1, \dots, Y_n) \text{ is a } n\text{-sample of } N(\mu, \sigma).$$

So  $\frac{\sqrt{n}}{\sigma} [\text{Log } (x_0 - 30) - \bar{Y}]$  is distributed as  $N \left[ \frac{\sqrt{n}}{\sigma} \{\text{Log } (x_0 - 30) - \mu\}, 1 \right]$

and  $\frac{(n-1) S_Y^2}{\sigma^2}$  is  $x_{n-1}^2$  distributed. These two random variables being inde-



GRAPH 4. Plot of the points  $\left( \text{Log}(x_{(r)} - 30), \Phi^{-1} \left( \frac{r}{n+1} \right) \right) r = 1, \dots, n$ .

pendent, the distribution of  $\frac{\sqrt{n} [\text{Log}(x_0 - 30) - \bar{Y}]}{S_Y}$  is a  $t_{n-1}, \sqrt{n} [\text{Log}(x_0 - 30) - \mu]/\sigma$

non-central Student distribution with  $(n - 1)$  degrees of freedom and centrality parameter  $\sqrt{n} [\text{Log}(x_0 - 30) - \mu]/\sigma$ .

In a more general way the  $\gamma$ -fractile  $t_{v,\delta}(\gamma)$  of a Student distribution  $t_{v,\delta}$  with  $v$  d.f. and centrality parameter  $\delta$  is, for fixed  $v$  and  $\gamma$ , a strictly increasing continuous function of  $\delta$  noted  $C_{v,\gamma}$  with  $P[t_{v,\delta} < C_{v,\gamma}(\delta)] = \gamma \forall \delta$ .

Let  $C_{v,\gamma}^{-1}(t)$  be its reciprocal function: for fixed  $t \in \mathbb{R}$ ,  $\delta = C_{v,\gamma}^{-1}(t)$  is the only solution of the equation,  $d$  being the unknown:  $P[t_{v,d} < t] = \gamma$ . From that it follows



$$\left[ \frac{1}{\sqrt{n}} C_{n-1, 1-\alpha/2}^{-1} \left( \frac{\sqrt{n}}{S_Y} [\text{Log}(x_0 - 30) - \bar{Y}] \right), \right. \\ \left. \frac{1}{\sqrt{n}} C_{n-1, \alpha/2}^{-1} \left( \frac{\sqrt{n}}{S_Y} [\text{Log}(x_0 - 30) - \bar{Y}] \right) \right]$$

is a level  $(1 - \alpha)$  confidence interval for  $\frac{\text{Log}(x_0 - 30) - \mu}{\sigma}$  with symmetric risks.

For  $1 - \Phi \left[ \frac{\text{Log}(x_0 - 30) - \mu}{\sigma} \right]$  the lower and upper limits of the confidence interval are

$$1 - \Phi \left\{ \frac{1}{\sqrt{n}} C_{n-1, \alpha/2}^{-1} \left( \frac{\sqrt{n}}{S_Y} [\text{Log}(x_0 - 30) - \bar{Y}] \right) \right\} \text{ and} \\ 1 - \Phi \left\{ \frac{1}{\sqrt{n}} C_{n-1, 1-\alpha/2}^{-1} \left( \frac{\sqrt{n}}{S_Y} [\text{Log}(x_0 - 30) - \bar{Y}] \right) \right\}.$$

From the fractiles of the Resnikoff-Lieberman table, it is possible to compute this interval for  $n = 37$  and  $1 - \alpha = 0,98$  (by linear interpolation and with a limited accuracy) only for  $x_0 \geq 1500$ . So it seems to be preferable to use the following approximation of fractile  $t_{n-1, \delta}(\gamma)$  (VAN EEDEN, 1961):

(1)  $t_{n-1, \delta}(\gamma) \simeq t_{n-1}(\gamma) + h(\delta)$  with

$$h(\delta) = \delta + \frac{\delta}{4(n-1)} (1 + 2q^2 + q\delta) + \frac{\delta}{96(n-1)^2} [3(4q^4 + 12q^2 + 1) + \\ + 6(q^3 + 4q)\delta - 4(q^2 - 1)\delta^2 - 3q\delta^3]$$

and with  $t_{n-1}(\gamma)$  and  $q$  being the  $\gamma$ -fractiles of the (central) Student distribution and of the normal distribution  $N(0, 1)$ .

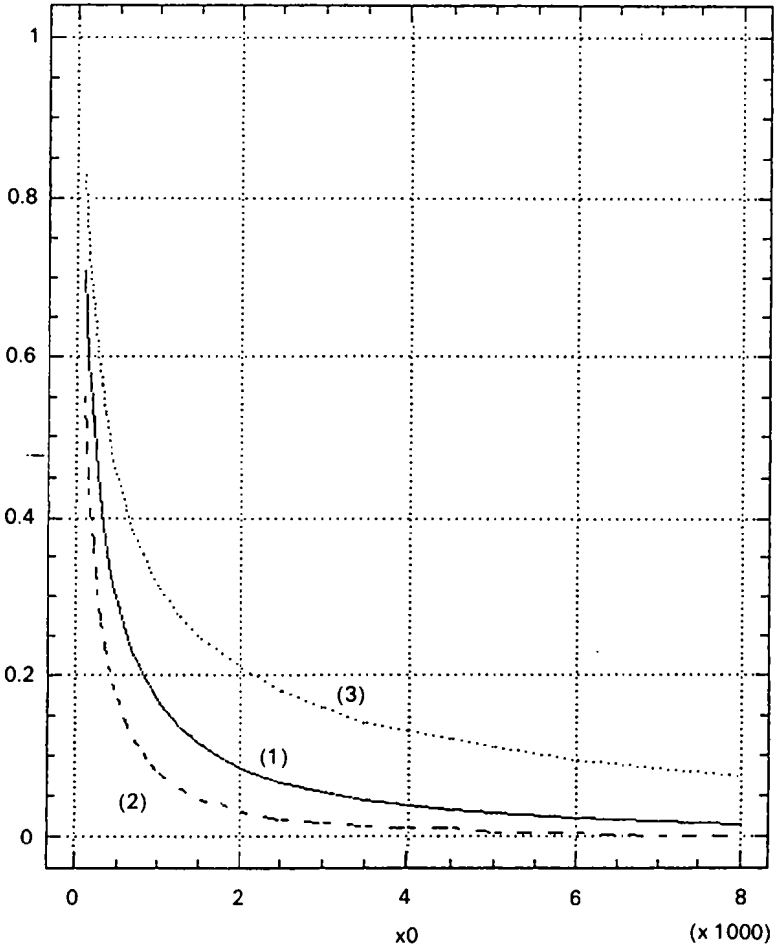
Let  $t_0 = \frac{\sqrt{n}}{s_y} [\text{Log}(x_0 - 30) - \bar{y}]$ , the approximation (1) provides  $C_{n-1, \gamma}^{-1}(t_0)$

as solution of the equation ( $\delta$  being the unknown):  $t_{n-1}(\gamma) - t_0 + h(\delta) = 0$ . This equation can be numerically solved using the Newton-Raphson algorithm

[ a starting value could be  $\delta_0 = t_0 - t_{n-1}(\gamma)$ , obtained by neglecting the terms

$$\frac{1}{n-1} \text{ and } \frac{1}{(n-1)^2} \text{ in (1)} ] .$$

Appendix 2 shows in columns 2 and 3 the lower and upper limits of the level 0,98 confidence interval for  $1 - H_{\mu, \sigma, 30}(x_0)$ . These limits are plotted in Graph 5.



GRAPH 5. Plot of  $\widehat{1 - H_{\mu, \sigma, 30}(x_0)}$  (curve 1), lower limit (curve 2) and upper limit (curve 3) of the confidence interval for  $1 - H_{\mu, \sigma, 30}(x_0)$ . The log-normal case.

5. FREQUENCY OF LOSSES WITH AN AMOUNT  $\geq X_0$

Let, for fixed  $x_0 \geq 30$ ,  $N(x_0)$  the r.v. of the yearly frequency of losses exceeding  $x_0$ . Using the same notations as before and considering that the r.v.  $N$  has a Poisson  $\mathcal{P}(\lambda)$  distribution, under  $(H_1)$  and  $(H_2)$ , the following results are obtained.

**Theorem :**

- a)  $N(x_0)$  is Poisson distributed with parameter  $\lambda(x_0) = \lambda[1 - F_X(x_0)]$ .
- b) If the distribution of  $X$  belongs to the family  $\mathcal{F} = \{F(x; \theta) : \theta \in \Theta\}$ , the M.L.E. of  $\lambda(x_0)$  is  $\widehat{\lambda}(x_0) = \hat{\lambda}[1 - F(x_0; \hat{\theta})]$ .
- c) If  $[\hat{\lambda}_l, \hat{\lambda}_s]$  and  $[I(x_0), S(x_0)]$  are confidence intervals for  $\lambda$  and  $1 - F_X(x_0)$  at a level of at least  $(1 - \alpha/2)$ ,  $[\hat{\lambda}_l I(x_0), \hat{\lambda}_s S(x_0)]$  is a confidence interval for  $\lambda(x_0)$  at a level of at least  $(1 - \alpha)$ .

**Proof :**

- a) Direct calculation.
- b) Because of the independence hypothesis and invariance of the M.L.E.
- c)  $P[\hat{\lambda}_l I(x_0) \leq \lambda(x_0) \leq \hat{\lambda}_s S(x_0)] \geq$

$$P[\hat{\lambda}_l \leq \lambda \leq \hat{\lambda}_s, I(x_0) \leq 1 - F_X(x_0) \leq S(x_0)]$$

and the result with the Bonferroni inequality  $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$  for any two events  $A$  and  $B$ . It is worthwhile to note that a direct use of the independence frequency-amount would give a level  $\geq (1 - \alpha/2)^2 = 1 - \alpha + \frac{\alpha^2}{4}$

very close to  $(1 - \alpha)$

These results applied to the frequency and amounts of hurricanes give in the same way as for  $1 - F_X(x_0)$  but at a level  $1 - \alpha = 0.96$  :

- \* In the non parametric case (Appendix 1)
  - an estimation of  $\lambda(x_0)$ :  $\hat{\lambda}[1 - F_n^*(x_0)]$  (Column 4)
  - a confidence band  $([\hat{\lambda}_l I_n(x), \hat{\lambda}_s S_n(x)])_{x \in \mathbb{R}^+}$  for  $\lambda(x_0)$  such as

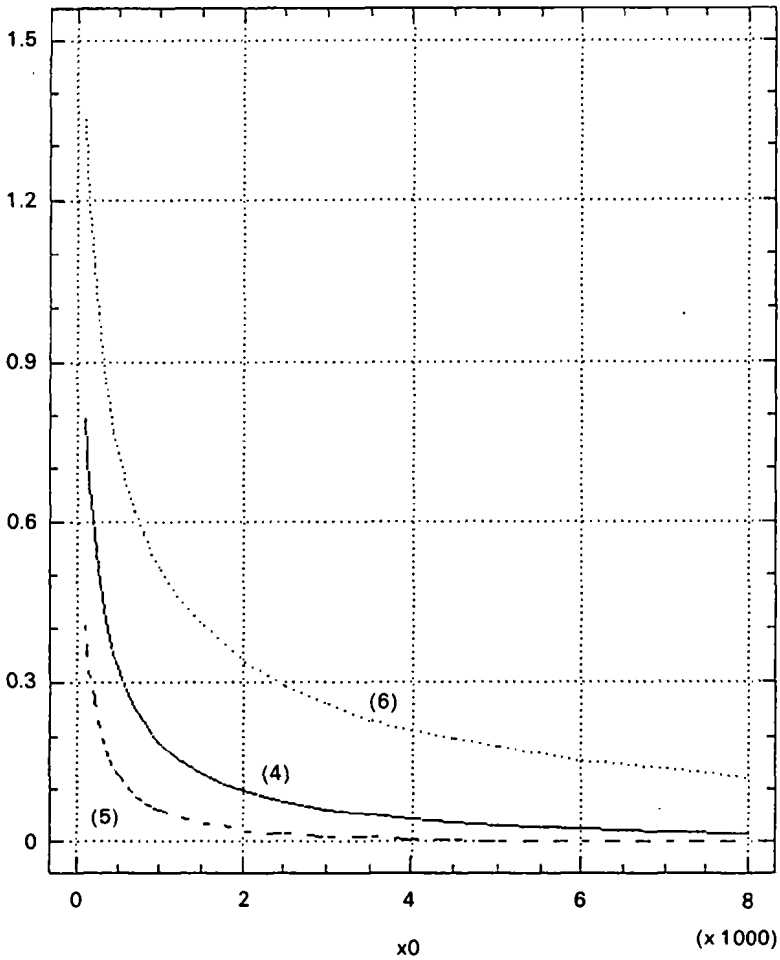
$$P[\hat{\lambda}_l I_n(x) \leq \lambda(x) \leq \hat{\lambda}_s S_n(x) \forall x \in \mathbb{R}^+] \geq 1 - \alpha.$$

The values  $\hat{\lambda}_l I_n(x)$  and  $\hat{\lambda}_s S_n(x)$  are shown in Columns 5 and 6.

- \* In the log normal case (Appendix 2)
  - the M.L.E.  $\widehat{\lambda}(x_0)$  of  $\lambda(x_0)$  (Column 4)
  - the upper and lower limits of a confidence interval for  $\lambda(x_0)$  (Columns 5 and 6).

Graph 6 shows a plot of these values.

In conclusion Table 4 shows for the values of  $x_0$  for which observations are available, in order to judge of the goodness-of-fit: the M.L.E.  $\widehat{\lambda}(x_0)$  derived from the model, the empirical mean ( $\bar{n}$ ) and variance ( $\hat{\sigma}_n^2$ ) of the yearly frequency of losses exceeding  $x_0$ , the empirical distribution of the frequencies



GRAPH 6. Plot of  $\widehat{\lambda}(x_0)$  (curve 4), lower limit (curve 5) and upper limit (curve 6) of the confidence interval for  $\lambda(x_0)$ . The log-normal case.

(columns obs.), to compare with the theoretical distribution derived from the Poisson log-normal model (column theor.).

Empirically the fit of the model seems very satisfactory.

#### CONCLUDING REMARKS

These results do not seem to be exclusive for hurricanes in the United States. So they could be used to modelize the frequency and amount distributions of natural events of any kind in the United States (for examples tornadoes) and even world wide.



## APPENDIX I

Values of:	$1 - F_n^*(x)$	: Col. 1
	$I_n(x)$	: Col. 2
	$S_n(x)$	: Col. 3
	$\hat{\lambda}[1 - F_n^*(x)]$	: Col. 4
	$\hat{\lambda}_i I_n(x)$	: Col. 5
	$\hat{\lambda}_s S_n(x)$	: Col. 6

$k$	$x_{(k)} < x \leq x_{(k+1)}$	(1)	(2)	(3)	(4)	(5)	(6)	
0	30,0	36,2	1,000	0,756	1,000	1,121	0,557	1,630
1	36,2	39,7	0,973	0,729	1,000	1,091	0,538	1,630
2	39,7	41,2	0,946	0,702	1,000	1,061	0,518	1,630
3	41,2	47,1	0,919	0,675	1,000	1,030	0,498	1,630
4	47,1	52,8	0,892	0,648	1,000	1,000	0,478	1,630
5	52,8	53,7	0,865	0,621	1,000	0,970	0,458	1,630
6	53,7	57,3	0,838	0,594	1,000	0,940	0,438	1,630
7	57,3	58,7	0,811	0,567	1,000	0,909	0,418	1,630
8	58,7	64,8	0,784	0,540	1,000	0,879	0,398	1,630
9	64,8	70,1	0,757	0,513	1,000	0,848	0,378	1,630
10	70,1	83,9	0,730	0,486	0,974	0,818	0,358	1,587
11	83,9	87,8	0,703	0,459	0,947	0,788	0,338	1,543
12	87,8	106,2	0,676	0,432	0,920	0,758	0,318	1,499
13	106,2	118,4	0,649	0,405	0,893	0,727	0,298	1,455
14	118,4	137,2	0,622	0,378	0,866	0,697	0,278	1,411
15	137,2	167,8	0,595	0,351	0,839	0,667	0,259	1,367
16	167,8	192,0	0,568	0,324	0,812	0,636	0,239	1,323
17	192,0	203,8	0,541	0,297	0,785	0,606	0,219	1,279
18	203,8	216,7	0,514	0,270	0,758	0,576	0,199	1,235
19	216,7	260,1	0,486	0,242	0,730	0,545	0,179	1,191
20	260,1	317,9	0,459	0,215	0,703	0,515	0,159	1,147
21	317,9	351,6	0,432	0,188	0,676	0,485	0,139	1,103
22	351,6	431,5	0,405	0,161	0,649	0,454	0,119	1,056
23	431,5	439,9	0,378	0,134	0,622	0,424	0,099	1,015
24	439,9	503,7	0,351	0,107	0,595	0,394	0,079	0,970
25	503,7	529,9	0,324	0,080	0,568	0,364	0,059	0,926
26	529,9	582,0	0,297	0,053	0,541	0,333	0,039	0,882
27	582,0	814,9	0,270	0,026	0,514	0,303	0,019	0,838
28	814,9	822,2	0,243	0,000	0,487	0,273	0,000	0,794
29	822,2	893,1	0,216	0,000	0,460	0,242	0,000	0,750
30	893,1	1243,4	0,189	0,000	0,433	0,212	0,000	0,706
31	1243,4	1263,5	0,162	0,000	0,406	0,182	0,000	0,662
32	1263,5	1313,0	0,135	0,000	0,379	0,152	0,000	0,618
33	1313,0	1602,1	0,108	0,000	0,352	0,121	0,000	0,574
34	1602,1	2465,4	0,081	0,000	0,325	0,091	0,000	0,530
35	2465,4	2753,9	0,054	0,000	0,298	0,061	0,000	0,486
36	2753,9	6299,9	0,027	0,000	0,271	0,030	0,000	0,442
37	6299,9	$+\infty$	0,000	0,000	0,244	0,000	0,000	0,398

APPENDIX 2  
LOG NORMAL CASE

- Estimation of  $1 - F_X(x_0)$ : Col. 1
- Lower and upper limits of the confidence interval for  $1 - F_X(x_0)$ : Col. 2 and 3
- Estimation of  $\lambda(X_0)$ : Col. 4
- Lower and upper limits of the confidence interval for  $\lambda(X_0)$ : Col. 5 and 6.

$x_0$	(1)	(2)	(3)	(4)	(5)	(6)
100	0,707	0,55	0,83	0,793	0,41	1,35
150	0,593	0,44	0,73	0,665	0,32	1,19
200	0,514	0,36	0,66	0,577	0,27	1,08
250	0,455	0,31	0,61	0,511	0,23	0,99
300	0,409	0,27	0,565	0,459	0,20	0,92
350	0,372	0,24	0,53	0,417	0,175	0,86
400	0,341	0,21	0,50	0,382	0,16	0,81
450	0,315	0,19	0,47	0,353	0,14	0,77
500	0,292	0,17	0,45	0,327	0,13	0,73
600	0,255	0,14	0,41	0,286	0,105	0,67
700	0,226	0,12	0,38	0,254	0,09	0,62
800	0,203	0,105	0,355	0,228	0,08	0,58
900	0,184	0,09	0,33	0,206	0,07	0,54
1000	0,168	0,08	0,31	0,188	0,06	0,51
1250	0,137	0,06	0,28	0,153	0,045	0,45
1500	0,115	0,045	0,25	0,129	0,035	0,41
1750	0,098	0,04	0,23	0,110	0,03	0,375
2000	0,085	0,03	0,21	0,096	0,02	0,34
2500	0,067	0,02	0,18	0,075	0,015	0,295
3000	0,054	0,015	0,16	0,061	0,01	0,26
3500	0,045	0,01	0,14	0,051	0,01	0,23
4000	0,038	0,01	0,13	0,043	0,005	0,21
4500	0,033	0,01	0,12	0,037	0,005	0,195
5000	0,029	0,005	0,11	0,032	0,0	0,18
6000	0,022	0,005	0,095	0,025	0,0	0,155
7000	0,018	0,0	0,085	0,020	0,0	0,14
8000	0,015	0,0	0,075	0,017	0,0	0,12

## ACKNOWLEDGMENT

We would like to thank Dr. T. MACK and the referees for their helpful comments on the first version of this paper.

## REFERENCES

- D'AGOSTINO, R. B. and STEPHENS, M. A. (1986) *Goodness-of-fit techniques*. Marcel Dekker.
- FRIEDMAN, D. G. (1987) U.S. hurricanes and wind storms *A technical briefing*. Insurance and Reinsurance research group Ltd. London.
- GIBBONS, J. D. (1971) *Non parametric Statistical Inference*. Mac Graw Hill.
- HOGG, R. V. and KLUGMAN, S. A. (1984) *Loss Distributions*. Wiley.
- JOHNSON, N. L. and KOTZ, S. (1970) *Continuous Univariate Distributions*. Houghton.
- PATRICK, G. (1980) Estimating casualty insurance loss amount distributions. *Proceedings of the casualty Actuarial Society*. Vol. LXVII.
- RAMLAU-HANSEN, H. (1988) A Solvency study in non-life insurance. Part I. Analysis of fire, windstorm and glass claims. *Scandinavian Actuarial Journal*, 3-34.
- RESNIKOFF, G. J. and LIEBERMAN, G. J. (1957) *Tables of the non-central t-distribution*. Stanford University Press.
- TIAGO DE OLIVEIRA, J. (1977) Statistical methodology for large claims. *ASTIN Bulletin* 9, 1-9.
- U.S. DEPARTMENT OF COMMERCE (1987) Tropical cyclones of the north atlantic ocean 1871-1986. *Historical climatology series*, 6-2.
- VAN EEDEN, C. (1961) Some approximations to the percentage points of the non-central *t*-distribution. *Revue de l'Institut International de Statistique* 29, 4-31.

CHARLES LEVI

*Compagnie Transcontinentale de Réassurance, 15 rue Louis le Grand,  
75002 Paris, France.*

CHRISTIAN PARTRAT

*Institut de Statistique, Université Pierre et Marie Curie,  
4 Place Jussieu, 75252 Paris Cedex 05, France.*