The Annals of Statistics
1976, Vol. 4, No. 6, 1200-1209

MAXIMUM LIKELIHOOD ESTIMATION OF A COMPOUND POISSON PROCESS[1]

BY LÉOPOLD SIMAR

CORE

The problem of estimating the compounding distribution of a compound Poisson process from independent observations of the compound process has been analyzed by Tucker (1963). A maximum likelihood method is proposed. The existence, uniqueness and convergence of the resulting estimator are derived. One obtains practical solutions by means of a very simple algorithm which is briefly described. A numerical example is presented in the risk business framework.

1. Introduction and summary. This study analyzes the problem of estimating the compounding distribution of a compound Poisson process.

The compound Poisson process is encountered in many contexts. In renewal theory, it represents the distribution of the counts of a Poisson process of known rate during the renewal intervals, where the renewal distribution is the compounding distribution of the process. Similarly, in the M/G/1 queue with Poisson arrivals of known rate, the number of arrivals during an arbitrary service period has a compound Poisson distribution. Here the compounding distribution is the service time distribution. Furthermore, the compound Poisson process may be more realistic than the Poisson process in modelling distribution of insurance claims over fixed periods. (Proneness models: Seal [5]).

Tucker [9] has estimated the moments of the compounding distribution in order to find an estimator. By contrast, in the present paper a maximum likelihood procedure is proposed. In Section 2 the problem is formalized while in Section 3 the existence, the uniqueness and the convergence of the maximum likelihood (M.L.) estimator are derived. An algorithm for its computation is proposed in Section 4. A practical application in the risk business framework is in Section 4.3.

2. Definition and notation. Consider a random variable $\tilde{\Theta}$ with values on the nonnegative integers and having a compound ("mixed") Poisson distribution. Thus $\Pr(\tilde{\Theta} = i) = \pi_i$ is of the form

$$\pi_i = \int_0^\infty \frac{x^i e^{-x}}{i!} \, dF(x) = \frac{\mu_i}{i!}\,, \tag{2.1}$$

Received July 1974; revised April 1976.

[1] Preliminary works were supported at CORE (Université de Louvain) by the Belgian Government through the "Programme National d'Impulsion à la Recherche en Informatique" under contract N° I. 14 bis/6 and at Cornell University by a NATO research fellowship.

AMS 1970 *subject classifications.* Primary 62G05; Secondary 62M99.

Key words and phrases. Compound Poisson process, mixed Poisson process, maximum likelihood method, risk business.

where $F(x)$ is any cumulative distribution function (cdf) supported by $[0, \infty)$ with $F(+\infty) = 1$ and

$$\mu_i \equiv \int_0^\infty x^i e^{-x}\, dF(x)\,, \qquad i = 0, 1, \cdots . \tag{2.2}$$

The cdf F is called the compounding (mixing) distribution of the process. Note that μ_i (as defined in equation (2.1)) is not the usual power moment of F but rather the ith power moment of the measure $e^{-x}\, dF(x)$. However, the moment point μ of F is defined to be the sequence $\{\mu_0, \mu_1, \cdots\}$.

The problem of estimating $F(x)$ from independent observations on $\tilde{\Theta}$ makes sense since the compound Poisson process is known to be identifiable in terms of $F(x)$ (Teicher [7]). In particular, this means that F is uniquely determined by its moment point μ.

In a sample of n independent observations on $\tilde{\Theta}$ define $m_i = m_i(n)$ as the number of observations equal to i, and let $\alpha_i = \alpha_i(n) = m_i/n$. It is known that for all $i \geqq 0$ $E[\alpha_i(n)] = \mu_i/i!$ so that $\alpha_i(n) \to \mu_i/i!$ with probability one as $n \to \infty$. (Strong law of large numbers.) Further $\sum \alpha_i = 1$ and $\sum \mu_i/i! = 1$.

The logarithm of the likelihood function can be written

$$\log L = \sum m_i \log p_i = \sum m_i \log \mu_i + \text{constant}.$$

The problem on hand is to determine the cdf F (if any) which maximizes the function $\sum m_i \log \mu_i$ or (since n is given) the function:

$$\Phi = \sum \alpha_i \log \mu_i = \textstyle\sum_{h=1}^{q} \alpha_{i_h} \log \mu_{i_h} \tag{2.3}$$

where $q = q(n) \leqq n$ is the number of indices i_h where $\alpha_i > 0$, and $0 \leqq i_1 < i_2 < \cdots < i_q = N$. Thus $N = N(n)$ is the largest value of $\tilde{\Theta}$ observed in a particular sample.

3. Maximum likelihood estimator.

3.1. *Existence of a solution.* In order to prove the existence of such an F we first prove the existence of a positive measure of mass $\leqq 1$ which maximizes Φ.

Let $\mathscr{F}$ be the set of positive measures on R_+ with total mass less than or equal to 1. The expression (2.2) defines a mapping of $\mathscr{F}$ to $R_+{}^q$ which associates to each $F \in \mathscr{F}$ a q-tuple of moments $(\mu_{i_1}, \cdots, \mu_{i_q})$. Let M_q, the moment space, be the image of $\mathscr{F}$.

We claim that M_q is convex and compact. Indeed $\mathscr{F}$ is weakly compact (Loève [2], page 179). For each $h = 1, \cdots, q$, the function $x^{i_h}e^{-x}$ is continuous and tends to zero at infinity. Therefore, from the extended Helly-Bray lemma (Loève [2], page 181), every sequence of points in M_q contains a subsequence which converges to a point of M_q. Consequently M_q is compact. Convexity is obvious.

Since M_q is convex and compact and since Φ is strictly concave in $(\mu_{i_1}, \cdots, \mu_{i_q})$ it follows that on M_q the function Φ assumes its maximum value at a unique point to be denoted by $\hat{\mu} = (\hat{\mu}_{i_1}, \cdots, \hat{\mu}_{i_q})$. Note that this point is well defined and in principle known though it may not be easily computed.

Each point in M_q and thus $\hat{\mu}$ is realized by at least one positive measure of mass $\leqq 1$. Let $d\hat{F}$ be such a measure and suppose that $d\hat{F}$ has a total mass $1 - \eta$ with $\eta > 0$. Replacing $\hat{F}$ by $\hat{F} + \eta\delta(1)$ where $\delta(y)$ represents a unit mass at a point y, all the $\hat{\mu}_{i_h}$ $(h = 1, \cdots, q)$ strictly increase, so that Φ also increases, a contradiction to the hypothesis that $\hat{\mu}$ is optimal in M_q. (Note that Φ cannot be equal to $-\infty$ at $\hat{\mu}$, otherwise the point in M_q realized by $\delta(1)$ would be better.)

Consequently, *any positive measure $d\hat{F}$ corresponding to the unique optimal point $\hat{\mu}$ has a total mass equal to one.* (Note that working with positive measures of mass $\leqq 1$ is needed in order to guarantee the compactness.)

3.2. *Uniqueness of the solution.*

UNIQUENESS THEOREM. *There is a unique maximum likelihood estimator $\hat{F}$. It is supported by the set of points $\{x_1, x_2, \cdots, x_r\}$ $(0 \leqq x_1 < x_2 < \cdots < x_r)$ where*

(1) *if $x_1 = 0$ then $r \leqq [(N+2)/2]$, while if $x_1 > 0$, $r \leqq [(N+1)/2]$. Further*
(2) *$r \leqq q$.*

PROOF. Let F be any cdf with moment point μ in M_q. Consider $(1-\varepsilon)\hat{F} + \varepsilon F$ for ε in $(0, 1)$. The value of Φ at this point is $\sum_{h=1}^{q} \alpha_{i_h} \log((1-\varepsilon)\hat{\mu}_{i_h} + \varepsilon\mu_{i_h})$. Moreover Φ is stictly concave and maximal for $\varepsilon = 0$, thus taking the derivative with respect to ε:

$$\sum_{h=1}^{q} \frac{\alpha_{i_h}}{\hat{\mu}_{i_h}} (\mu_{i_h} - \hat{\mu}_{i_h}) \leqq 0 . \tag{3.1}$$

Equality holds in (3.1) if $(1-\varepsilon)\hat{F} + F$ is still a distribution function for some $\varepsilon < 0$.

Taking F as the distribution entirely concentrated at a point x, we have $\mu_{i_h} = x^{i_h}e^{-x}$ so that (3.1) becomes

$$\sum_{h=1}^{q} \frac{\alpha_{i_h}}{\hat{\mu}_{i_h}} x^{i_h}e^{-x} \leqq 1 \qquad \text{for all} \quad x \geqq 0 . \tag{3.2}$$

Integrating the left-hand side of (3.2) with respect to $\hat{F}$ we find that

$$\int_0^\infty \sum_{h=1}^{q} \frac{\alpha_{i_h}}{\hat{\mu}_{i_h}} x^{i_h}e^{-x}\, d\hat{F}(x) = 1 , \qquad \text{since} \quad \sum_{h=1}^{q} \alpha_{i_h} = 1 .$$

This shows that the measure $d\hat{F}$ is supported by a finite set of points such that (3.2) holds with the equality sign. Let $0 \leqq x_1 < \cdots < x_r$ be these points. The following lemma yields an upper bound on r. Here, $\beta_h = \alpha_{i_h}/\hat{\mu}_{i_h}$.

LEMMA 3.1. *Let a sequence of real coefficients β_h, $h = 1, \cdots, q$ be such that*

$$\sum_{h=1}^{q} \beta_h x^{i_h} \leqq e^x \qquad \textit{for all} \quad x \geqq 0 \quad \textit{where} \quad 0 \leqq i_1 < \cdots < i_q = N .$$

Let $0 \leqq x_1 < \cdots < x_r$ be the points where equality holds.

(1) *If $x_1 = 0$ then $r \leqq [(N+2)/2]$, while if $x_1 > 0$, $r \leqq [(N+1)/2]$.*
(2) *Further $r \leqq q$.*

The proof of this lemma can be found in the Appendix.

Consider now a fixed $\hat{F}$. The $\hat{\mu}_{i_h}$ are unique and well defined (but not necessarily computable) by (2.3). The β_h are also well defined, since the α_{i_h} and the i_h follow directly from the observations. Thus the number r and the points $x_1, \cdots, x_r$ are in principle also known.

Let p_j denote the mass of $d\hat{F}$ at the point x_j ($j = 1, \cdots, r$; $p_j \geqq 0$; $\sum_{j=1}^{r} p_j = 1$). From (2.2) we have

$$\sum_{j=1}^{r} p_j x_j^{i_h} e^{-x_j} = \hat{\mu}_{i_h}, \qquad h = 1, \cdots, q. \tag{3.3}$$

We know that there is at least one solution $\{p_i\}$ to this system. Since the rank of the matrix $\{x_j^{i_h}, j = 1, \cdots, r;\ h = 1, \cdots, q\}$ is equal to min (r, q), the system (3.3) has a unique solution provided $r \leqq q$. From the second part of Lemma 3.1, the latter is true.

Consequently, the uniqueness theorem is proven.

3.3. *Convergence of the maximum likelihood estimator.* Let F^0 denote the true (but unknown) compounding distribution with moment point μ and let $\hat{F}_n$ denote the M.L. estimator of F^0 with moment point ν_n, obtained by using n observations. (The trivial case $F^0(x) = \delta(0)$ is excluded, since $\alpha_0(n) = 1$ and $\hat{F}_n(x) = \delta(0)$ for all n.)

CONVERGENCE THEOREM. *As $n \to \infty$, the sequence $\{\hat{F}_n\}$ of* M.L. *estimators weakly converges with probability one to the true distribution F^0.*

PROOF. Since ν_n denotes the M.L. moment point, we have, by (3.2):

$$\sum_{i=0}^{\infty} \frac{\alpha_i(n)}{\nu_{i,n}} \xi_i \leqq 1 \qquad \text{for any moment point} \quad \xi. \tag{3.4}$$

From the weak compactness theorem we can draw from $\{\hat{F}_n\}$ a subsequence weakly converging to a df, say F, of total mass less than or equal to one. According to the Helly-Bray lemma $\lim_{n\to\infty} \nu_{i,n} = \nu_i$ with ν_i corresponding to F. We must prove that $F = F^0$, that is (see Section 2), $\nu = \mu$.

Since for $i \geqq 0$, with probability one $\lim_{n\to\infty} \alpha_i(n) = \mu_i/i! > 0$ we have from (3.4) for all fixed K and for large n:

$$\sum_{i=0}^{K} \frac{\alpha_i(n)}{\nu_{i,n}} \xi_i < 1 \qquad \text{for any} \quad \xi \quad \text{with probability one.}$$

Consequently

$$\sum_{i=0}^{K} \frac{\mu_i}{i!} \frac{\xi_i}{\nu_i} \leqq 1 \qquad \text{for any moment point} \quad \xi.$$

Note that $\nu_i \neq 0$ for all i, otherwise the sum would be equal to $+\infty$ which is impossible since for all n it must be less than one. Since the last relation is true for all K, we have

$$\sum_{i=0}^{\infty} \frac{\mu_i}{i!} \frac{\xi_i}{\nu_i} \leqq 1 \qquad \text{for any moment point} \quad \xi. \tag{3.5}$$

Consider now the function

$$\psi(\xi) = \sum_{i=0}^{\infty} \frac{\mu_i}{i!} \log \frac{\xi_i}{\mu_i}$$

defined at any moment point ξ, such that $\sum_{i=0}^{\infty} \xi_i/i! \leqq 1$. It is a strictly concave function and since $\sum_{i=0}^{\infty} \mu_i/i! = 1$, from Jensen's inequality we have $\psi(\xi) \leqq 0$ with equality if and only if $\xi = \mu$. If $F \neq F^0$, that is, $\nu \neq \mu$ then $\psi(\nu) < 0$, and the directional derivative of $\psi(\xi)$ at the point ν in the direction of μ would be strictly positive, i.e.,

$$\sum_{i=0}^{\infty} \frac{\mu_i}{i!} \frac{\mu_i - \nu_i}{\nu_i} > 0 .$$

As an easy computation shows, this would however contradict (3.5) with $\xi = \mu$. Consequently $\nu = \mu$. Therefore, with probability one, any convergent subsequence of $\{F_n\}$ weakly converges to F^0, and the convergence theorem is thus proven.

4. Practical applications.

4.1. *An algorithm for computing the* M.L. *estimator*. We now present an algorithm for computing $\hat{F}_n$ from the data $\alpha_i = m_i/n$. Let S_k be the unit simplex of dimension k:

$$S_k = \{p_j, j = 1, \cdots, k \mid p_j \geqq 0, \sum_{j=1}^{k} p_j = 1\} .$$

Let $(x_1, \cdots, x_k) \in R_+{}^k$, and let $P = \{(p_j, x_j), j = 1, \cdots, k\}$. The problem on hand is to determine a point $P \in S_k \otimes R_+{}^k$ that maximizes the function

$$\Phi = \sum_{i=0}^{N} \alpha_i \log \sum_{j=1}^{k} p_j x_j{}^i e^{-x_j} . \tag{4.1}$$

A solution P is defined as "admissible" if k is less than or equal to the bounds given in Lemma 3.1, i.e., letting $0 \leqq x_1 < \cdots < x_k$

(1) if $x_1 = 0$ then $k \leqq [(N + 2)/2]$ while if $x_1 > 0$ $k \leqq [(N + 1)/2]$.
(2) $k \leqq q$ where q is the number of α_i different from zero.

From an arbitrary point P, with moment point μ, the gradient (directional derivative) of the objective function Φ in the direction of an other point Q with moment point ν is given by:

$$\nabla_{P\to Q} = \sum_{i=0}^{N} \frac{\alpha_i}{\mu_i} \nu_i - 1 .$$

This gradient is maximal if Q is concentrated at the point x^* which maximizes $\sum_{i=0}^{N} \alpha_i/(\mu_i) x^i e^{-x}$ on $[0, \infty)$. Therefore from a given point P the optimal direction is towards $\delta(x^*)$. The corresponding gradient is then

$$\nabla_{P\to\delta(x^*)} = \max_{x\geqq 0} \sum_{i=0}^{N} \frac{\alpha_i}{\mu_i} x^i e^{-x} - 1 . \tag{4.2}$$

Now we can describe the algorithm.

Step 0 (*Initialization*). Choose an arbitrary *but admissible* solution which is not the unit mass at the origin.

Step 1. From the fixed support $x_1, \cdots, x_k$ select $\{p_i\}$ to optimize (4.1) on S_k. This gives a point P in $S_k \otimes R_+{}^k$ with moments $(\mu_0, \cdots, \mu_n)$ such that

$$\sum_{h=1}^{q} \frac{\alpha_{i_h}}{\mu_{i_h}} x_j{}^{i_h} e^{-x_j} \leqq 1 , \qquad j = 1, \cdots, k . \tag{4.3}$$

(If it were not the case, $(1 - \varepsilon)P + \varepsilon\delta(x_j)$ would be better than P for $\varepsilon \in (0, 1]$ and P could not be a solution of Step 1.) Moreover at each point x_j that has a positive probability the *equality holds* in (4.3) since

$$\sum_{j=1}^{k} p_j \sum_{h=1}^{q} \frac{\alpha_{i_h}}{\mu_{i_h}} x_j{}^{i_h} e^{-x_j} = 1 .$$

Step 2. From this point P we compute x^* and the optimal direction by (4.2). If the maximal gradient thus computed is equal to zero we stop the algorithm. If not, we have to compute the optimal point in this direction. Therefore we add the point x^* to the preceding support and go back to Step 1.

If necessary at each iteration, after Step 1, we transform the solution to get an admissible one as shown in the next section. Note that the desired point achieves the minimum in the minimax problem:

$$\min_\mu \max_{x \geqq 0} \left[\sum_{i=0}^{N} \frac{\alpha_i}{\mu_i} x^i e^{-x} - 1 \right] .$$

(This minimax being equal to 0.)

This algorithm was found quite satisfactory from a practical point of view. Furthermore since it uses the inherent structure of the problem to facilitate the search of a good direction, it is more attractive than most heuristic procedure. In the next section we prove that in the (eventual) transformation of the current solution of Step 1 to get an admissible one the objective function Φ does not decrease. Since the computation in Step 2 gives a direction along which the gradient is strictly positive, it is clear that the optimization of Step 1 (with the new support) will strictly increase the objective function. Therefore at each iteration Φ strictly increases. This however does not guarantee the convergence toward $\Phi_{\max}$ which seems actually difficult to prove but it shows the relevance of the procedure.

4.2. *Admissibility conditions.* Let $P = \{(p_j, x_j), j = 1, \cdots, k\}$ be the current solution of the optimization of Step 1, where we only consider the points x_j, $j = 1, \cdots, k$ having a strictly positive p_j. Since the initial solution was chosen to be an admissible one and since we add only one point to the support at each iteration of the algorithm (in Step 2) it must be noted that the admissibility conditions can only be violated by one point more than the bound (1) and (2) given in Lemma 3.1.

4.2.1. *Condition* (1) *is violated.* Let P the inadmissible solution of Step 1 be supported by $0 \leqq x_1 < \cdots < x_{m+1}$ where $m = [(N+2)/2]$ if $x_1 = 0$, $m = [(N+1)/2]$ if $x_1 > 0$.

Let μ be the moment point of P. Consider now the *lower* principal representations of $\{\mu_0, \cdots, \mu_N\}$ relative to $[0, \infty)$ (Karlin and Studden [1] page 45 and page 157). This point, say Q, is supported by $0 \leqq y_1 < \cdots < y_m$ and can be computed solving an algebraic moment problem (Mammana [3]). Let ν be the moment point of Q. Obviously $\nu_j = \mu_j$ for $j \leqq N$ but it is well known, since x^j for $j \geqq N+1$ is strictly convex relative to the Tchebycheff family x^i, $i = 0, 1, \cdots, N$, that $\nu_j < \mu_j$ for all $j \geqq N+1$. (Markov–Krein theorem [1], page 157). Therefore the total mass of Q is less than one. Indeed:

$$\sum_{j=1}^{m} q_j = \sum_{k=0}^{\infty} \frac{\nu_k}{k!} < \sum_{k=0}^{\infty} \frac{\mu_k}{k!} = \sum_{j=1}^{m+1} p_j = 1 .$$

(The last equality holds since P is a solution of Step 1.) Now if we normalize the solution Q into Q^* so that

$$q_j^* = \frac{q_j}{\sum_{j=1}^{m} q_j} > q_j ,$$

the corresponding moments strictly increase, i.e., $\nu_k^* > \nu_k$. Therefore $\sum_{i=0}^{N} \alpha_i \log \nu_i^* > \sum_{i=0}^{N} \alpha_i \log \nu_i = \sum_{i=0}^{N} \alpha_i \log \mu_i$ hence the solution Q^* is admissible and is better than P in the sense that the corresponding value of the likelihood function Φ has strictly increased. Now we go back to Step 1 with the new support $0 \leqq y_1 < \cdots < y_m$.

4.2.2. *Condition* (2) *is violated.* In this case the solution P of Step 1 is supported by $(q+1)$ points. Consider now x_j, $j = 1, \cdots, q+1$ and μ_{i_h}, $h = 1, \cdots, q$ as fixed constants. The following system (S) of $(q+1)$ equations in $(q+1)$ unknowns p_j, $j = 1, \cdots, q+1$.

$$\sum_{j=1}^{q+1} p_j x_j^{i_h} e^{-x_j} = \mu_{i_h} , \qquad h = 1, \cdots, q , \tag{4.4}$$

$$\sum_{j=1}^{q+1} p_j = 1 \tag{4.5}$$

has at least one solution (which corresponds to the point P). The linear combination (of coefficients α_{i_h}/μ_{i_h}, $h = 1, \cdots, q$) of the q equations (4.4) is

$$\sum_{h=1}^{q} \frac{\alpha_{i_h}}{\mu_{i_h}} \sum_{j=1}^{q+1} p_j x_j^{i_h} e^{-x_j} = \sum_{h=1}^{q} \frac{\alpha_{i_h}}{\mu_{i_h}} \mu_{i_h} = 1 . \tag{4.6}$$

Since the equality holds in (4.3) at each point x_j, $j = 1, \cdots, q+1$ of the support of P, the left member of (4.6) is precisely $\sum_{j=1}^{q+1} p_j$ so that (4.6) is exactly (4.5). Consequently the rank of the system (S) is equal to the rank of the system (4.4) which is exactly q. (See Section 3.2, the system (3.3).) Therefore all the solutions $p_1, \cdots, p_{q+1}$ of the system (S) lies in a subspace of dimension one (a straight line) in the unit simplex S_{q+1}. In particular, there exists a solution at the frontier of S_{q+1} (one of the intersections of that straight line with this frontier) which cancels one of the p_j. Such a point, say Q, can be easily found. Considering only the points in the support of Q which have a strictly positive probability, we have now an admissible solution and since the values of μ_{i_h}, $h = 1, \cdots, q$ are the same, Φ has not changed. Now we go to Step 2 of the algorithm.

4.3. *A numerical example.* From Thyrion [8], we took the number of accident claims submitted in a single year to La Royale Belge Insurance Company out of 9,461 policies covering both "business" and "tourist" automobiles. (See Table 2, column 1).

After six iterations or passes through Step 1 and Step 2, and one solution of an algebraic moment problem, we obtained the following results. (Execution time was less than two seconds on a IBM 370/155.) Table 1 shows that there are essentially two classes of customers: about 76% of them with a low Poisson risk $\lambda \cong 0.089$ and about 24% with a risk $\lambda \cong 0.58$. Table 2, column 2 shows the compound probabilities computed with these estimators. Seal ([5], page 16, Table 2.3) proposes three other fittings for the same data. To summarize, it is found that a Poisson fit (1 parameter) is very bad while a negative binomial fit (2 parameters) is judged poor. Finally the Thyrion's fitting (3 parameters, mixed Poisson) is said to be adequate.

The latter fit is in fact very close to the M.L. fit given in Table 2. This is not surprising since the M.L. estimator derived here turns out to be essentially a 3 parameters distribution (see Table 1), but it shows the power of our procedure since it starts without this a priori restriction.

TABLE 1
Maximum likelihood solution

Abscissas x_i	Probabilities p_i, $i = 1$ to 4
0.08854	0.75997
0.58020	0.23617
3.17606	0.00370
3.66871	0.00016

TABLE 2
$N = 7$ *and* $n = 9{,}461$

Observed frequencies	Computed probabilities
$\alpha_0 = 0.82867$	$P_0 = 0.82793$
$\alpha_1 = 0.13920$	$P_1 = 0.13880$
$\alpha_2 = 0.02526$	$P_2 = 0.02579$
$\alpha_3 = 0.00444$	$P_3 = 0.00524$
$\alpha_4 = 0.00148$	$P_4 = 0.00131$
$\alpha_5 = 0.00042$	$P_5 = 0.00051$
$\alpha_6 = 0.00042$	$P_6 = 0.00024$
$\alpha_7 = 0.00011$	$P_7 = 0.00011$
sum = 1.00	sum = 0.99993

It is difficult to compare all the fits together. Nevertheless, computing the errors $|\alpha_i - p_i|$ for all i we observe that the M.L. fit of Table 2 produces a uniformly smaller error for all i greater than zero than the Poisson and the negative binomial fits and Thyrion's fit has larger errors for five of the eight differences.

In Simar [6], a comparison is made between Tucker's procedure [9] and the M.L. estimator developed above, where it was found that for large n, the two procedures are nearly identical. Nevertheless, using simulation techniques it was concluded that the maximum likelihood method is better in the sense that the sampling distribution of the error of estimation, defined as

$$\max_{x \in [0,\infty)} |F_n(x) - F^0(x)|$$

was significantly less for the M.L. estimator than for Tucker's estimator. This is a crude result, but it allows us to hope that such a result can be proven analytically.

Acknowledgment. This paper is a part of my doctoral thesis [6]. I would like to express my thanks to J. F. Mertens, my thesis advisor, for his valuable aid throughout this work. I wish also to thank S. Pollock and J. P. Vial for their help in revising the paper.

APPENDIX

PROOF OF LEMMA 3.1. Consider the function $f(x) = \sum_{h=1}^{q} \beta_h x^{i_h} - e^x$ with $0 \leqq i_1 < \cdots < i_q = N$.

Part (1) *of the lemma.* The functions x^j $(j = 0, \cdots, N)$ together with the function e^x form an extended Tchebycheff system $\{u_0, u_1, \cdots, u_N, u_{N+1}\}$ on $[0, \infty)$ (see Karlin and Studden [1] page 6) and thus (see [1], page 24) no nonzero linear combination $\phi(x) = C_0 + C_1 x + \cdots + C_N x^N + C_{N+1} e^x$ can have more than $(N + 1)$ zeros counting multiplicities. Hence, if $\phi(x) \leqq 0$ for all $x \geqq 0$, then either $\phi(0) = 0$ and there are only $[N/2]$ positive zeros or $\phi(0) < 0$ and there are at most $[(N + 1)/2]$ positive zeros.

Part (2) *of the lemma.* Consider an entire function $\phi(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ with real coefficients. Let r be the number of positive zeros of $\phi(x)$ counting multiplicities and let s be the number of changes of sign in the sequence $a_0, a_1, \cdots$ ignoring the zero terms. Then (Pólya–Szegö [4], page 48, problems 38 and 40) $r \leqq s$ and further if s is finite then $s—r$ is an even integer.

Now apply this to the function

$$f(x) = \sum_{j=0}^{\infty} a_j x^j = \sum_{h=1}^{q} \beta_h x^{i_h} - e^x$$

with $f(x) \leqq 0$ for $x \geqq 0$. One has

$$a_{i_h} = \beta_h - \frac{1}{(i_h)!} \qquad \text{for} \quad h = 1, \cdots, q$$

$$a_j = -\frac{1}{j!} < 0 \qquad \text{otherwise.}$$

Let q' denote the number of $j \geqq 0$ with $a_j > 0$, that is, the number $q' \leqq q$ of indices $h = 1, \cdots, q$ for which $\beta_h > 1/(i_h)!$. It follows that $f(x)$ has at most $2q'$ zeros in $(0, \infty)$ counting multiplicities; that is, at most q' positive zeros.

Consider the case where also $f(0) = 0$ which means that $i_1 = 0$ and $\beta_{i_1} = 1$, while $a_0 = 1$. Then by the same reasoning there can be at most $q'' \leqq q - 1$ positive zeros where q'' denotes the number of indices $h = 2, \cdots, q$ with $\beta_h > 1/(i_h)!$.

REFERENCES

[1] Karlin, S. J. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Wiley, New York.

[2] Loève, M. (1963). *Probability Theory.* Van Nostrand, New York.

[3] Mammana, C. (1954). Sul problema algebraico dei momenti. *Ann. Scuola Norm. Sup. Pisa* **8** 133–140.

[4] Pólya, G. and Szegö (1925). *Aufgaben und Lehrsätze aus der Analysis*, **2**. Springer, Berlin.

[5] Seal, H. L. (1969). *Stochastic Theory of a Risk Business.* Wiley, New York.

[6] Simar, L. (1974). The estimation of the compounding distribution of a compound Poisson process. CORE Discussion Paper 7404, Louvain.

[7] Teicher, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

[8] Thyrion, P. (1961). Contribution à l'étude des bonus pour non sinistre en assurance automobile. *Astin Bull.* **1** 142–162.

[9] Tucker, H. G. (1963). An estimate of the compounding distribution of a compound Poisson distribution. *Theor. Probability Appl.* **8** 195–200.

CORE
De Croylaan 54
B-3030 Heverlee
Belgium