

# APPROXIMATIONS OF THE GENERALISED POISSON FUNCTION

LAURI KAUPPI and PERTTI OJANTAKANEN  
Helsinki

## I. *Setting the problem*

One of the basic functions of risk theory is the so-called generalised Poisson function  $F(x)$ , which gives the probability that the total amount of claims  $\xi$  does not exceed some given limit  $x$  during a year (or during some other fixed time period). For  $F(x)$  is obtained the well known expansion

$$F(x) = \sum_{k=0}^{\infty} \frac{e^{-n} n^k}{k!} S^{k*}(x), \quad (1)$$

where  $n$  is the expected number of claims during this time period and  $S^{k*}(x)$  is the  $k$ :th convolution of the distribution function  $S(z)$  of the size of one claim. The formula (1) is, however, much too inconvenient for numerical computations and for most other applications. One of the main problems of risk theory, which is still partly open, is to find suitable methods to compute, or at least to approximate, the generalised Poisson function.

A frequently used approximation is to replace  $F(x)$  by the normal distribution function having the same mean and standard deviation as  $F$  as follows:

$$F(x) \approx \Phi \left( \frac{x - n\alpha_1}{\sqrt{n} \alpha_2} \right), \quad (2)$$

where  $\alpha_1$  and  $\alpha_2$  are the first zero-moments of  $S(z)$ :

$$\alpha_i = \int_0^{\infty} z^i d S_M(z). \quad (3)$$

$S_M(z)$  is here again the distribution function of the size of one claim. To obtain more general results a reinsurance arrangement is assumed under which the maximum net retention is  $M$ . Hence the

portfolio on the company's own retention is considered. If the reinsurance is of Excess of Loss type, then

$$S_M(z) = \begin{cases} S(z) & \text{when } z < M \\ \mathbf{1} & \text{when } z \geq M, \end{cases}$$

where  $S(z)$  is the distribution function of the size of one total claim.

The normal approximation is very simple to use and it gives a good review of the interdependence of the different variables involved in the risk process. In accordance with the central limit theorem of probability calculus the normal approximation tends asymptotically to the function  $F$  when  $n \rightarrow \infty$ . Unfortunately, however, especially for small values of  $n$  and when the risk distribution is very heterogeneous it does not always give satisfactory accuracy. Another drawback of the normal approximation is its poor applicability in the Polya cases, where the formula (1) is further generalised to allow for certain fluctuation of the basic probabilities (and of  $n$ ).

Another, and in practice much more accurate, approximation formula is given by Esscher. This formula, although somewhat cumbersome, allows the computations of the numerical values of  $F$ , but it is not very suitable for giving a good general review of the mutual dependence of the different variables.

Recently two methods of calculating  $F$  by means of electronic computers have been presented. A Swedish team has performed computations by means of a formula based on the inversion of the characteristic function. Another method based on the Monte Carlo method has been developed by some Finnish actuaries.

Our work consists of parallel computations made by the normal approximation, Esscher formula and Monte Carlo method. Comparison of the results gives an illustration of the goodness of fit of the different approximation methods. Further analysis of the mutual deviations of the different results reveals certain clear regularities. They made it possible to propose some corrections to the normal approximation, which seem to improve the goodness of fit in an essential degree without losing too much in simplicity. This method was applied also to the Swedish material mentioned above.

The basic material of the study consists of claims statistics concerning industrial fire insurance and third party motor insurance furnished by Suomen Vahinkovakuutuksen Tilastokeskus (The Statistical Centre of The Non-Life Insurance of Finland). At first, these two classes of insurance were treated separately, until the results obtained proved this to be unnecessary. The execution of the study followed the pattern familiar from the natural sciences: first a number of experiments and then an analysis of the results. One generalised Poisson function derived from the basic material can be regarded as an experiment in this sense. The results were obtained by computing the material by both the Monte Carlo method and the Esscher formula as well as the normal approximation. It was possible to increase the number of the experiments by altering the fundamental distribution by giving different values to the expected number of claims on one hand — and by reinsurance on the other hand. The computations relate to the net retention of the company, supposing that the top risks are cut by an Excess of Loss reinsurance under which the maximum net retention  $M$  is given values corresponding to those applied as a rule in practice. The values corresponding to  $M = \infty$ , i.e. a nonreinsured portfolio, are also computed. The computations are thus carried out both with “well behaving” claim distributions including only limited risks as well as “dangerous” distributions including very large risks in order that the results should represent the circumstances of claims usual in practice on the company’s net retention and also an extremely dangerous one. In this manner 40 distributions were obtained, each of which can be regarded as a different generalised Poisson function. Half of the distributions represent cases of industrial fire insurance and the balance of third party motor insurance.

## 2. *Analysing the results obtained by the Esscher formula*

When analysing the results it was found that the ruin limits corresponding to a certain probability  $\varepsilon = 1 - F(x)$ , which were computed by the Esscher formula, deviated from those obtained by means of the normal approximation, the larger deviations arising with the greater values of the skewness

$$\gamma_1 = \frac{\alpha_3}{\sqrt{n} (\alpha_2)^{3/2}}$$

The observed correlation between  $x_\varepsilon$  and  $\gamma_1$  is seen quite clearly from Figure 1. Values of the expression  $\frac{x_\varepsilon}{y_\varepsilon} - 1$  as a function of  $\gamma_1$  are drawn in the figure on a double logarithmic scale with different ruin probabilities,  $x_\varepsilon$  denoting a standardised variable in accordance with the Esscher formula and  $y_\varepsilon$  denoting the corresponding normal variable. It is seen that the pairs of points corresponding to one value of  $\varepsilon$  are placed quite nicely on one straight line, the lines corresponding to different values of  $\varepsilon$  being parallel. The figure gives the following equation between  $x_\varepsilon$  and  $y_\varepsilon$

$$x_\varepsilon = y_\varepsilon (1 + C_\varepsilon \cdot \gamma_1^{0.98}) \quad (4)$$

where  $C_\varepsilon$  is a constant depending on  $\varepsilon$  only and having the values 0.16, 0.30 and 0.44, for  $\varepsilon$  0.05, 0.01 and 0.001. In Figure 1 the

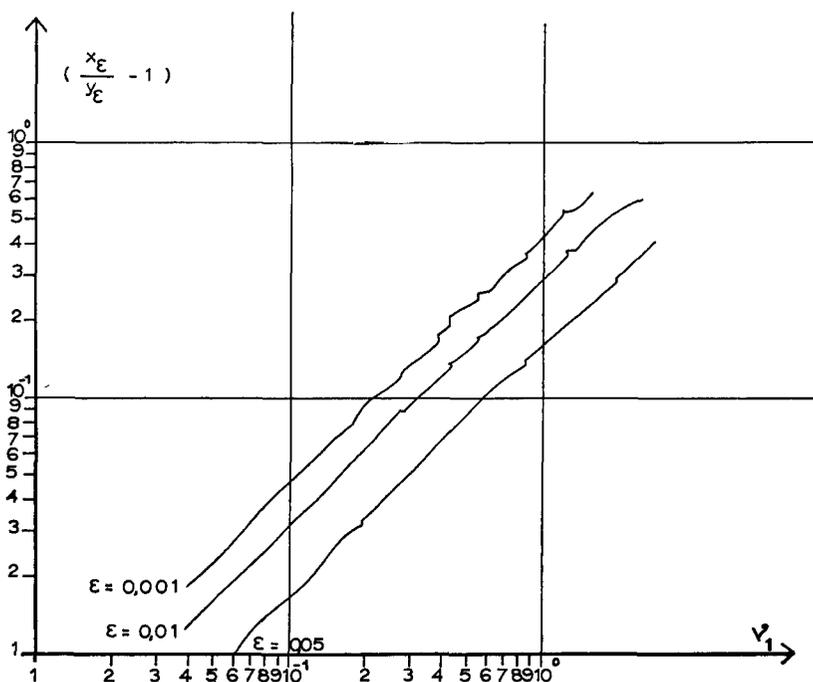


Fig. 1.

results obtained from both the industrial fire insurance and the third party motor insurance are included, which were, even concerning the numerical values of the constant  $C_\varepsilon$ , the same in spite of the great difference in primary distributions.

### 3. Comparison with the Monte Carlo method

The computations were also carried out with the Monte Carlo method (Pesonen: "Solvency measurement", Congress of Actuaries 1964, Hovinen: "A method to compute convolution", Congress of Actuaries 1964, "A Procedure to Compute the Values of the Generalised Poisson Function", ASTIN Colloquium 1965). In accordance with this method the run of an insurance company corresponding to every distribution was simulated in the 10.000 "observation years". The value of  $\varepsilon$  corresponding to each value of the variable  $\varepsilon$  is the ratio of the ruined cases to the total number of years (= 10.000). The number obtained in this manner is a random quantity, to which is attached the usual random inaccuracy, for which, however, the upper bounds can be estimated. The proportional amount of this inaccuracy increases when  $\varepsilon$  decreases and the method was not (for 10.000 observation years) more suitable, when  $\varepsilon < 0.001$ .

When analysing the results it appeared that as to the analytic form, the expression

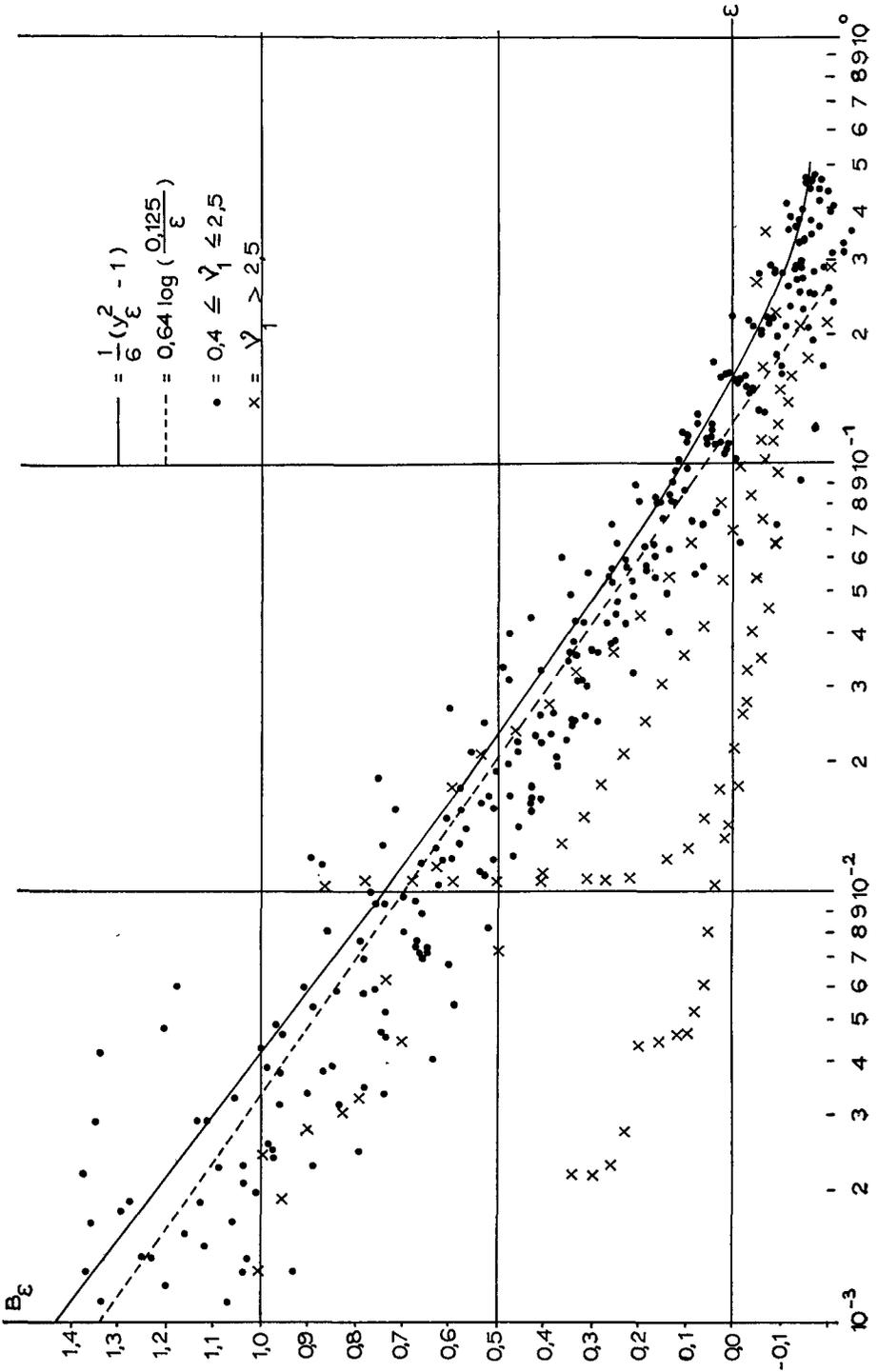
$$x_\varepsilon = y_\varepsilon + B_\varepsilon \cdot \gamma_1 \quad (5)$$

seemed to be suited to approximate the results computed by the Monte Carlo method,  $B_\varepsilon$  being a quantity depending only on  $\varepsilon$ . In this expression  $B_\varepsilon$  is to be determined by experiment. If  $B_\varepsilon$  is solved from expression (5), we have

$$B_\varepsilon = \frac{x_\varepsilon - y_\varepsilon}{\gamma_1}.$$

The above mentioned inaccuracy, being a consequence of the Monte Carlo method, is also reflected in the numbers  $B_\varepsilon$ , and the lesser  $\gamma_1$  is, the greater the inaccuracy.

Figure 2 shows values which have been computed using the results of  $B_\varepsilon$  as the function of  $\varepsilon$ , drawn on a semilogarithmic scale.



The points corresponding to the  $\gamma_1$  values 0.4 — 2.5 and  $> 2.5$  are distinguished.

The following observations may be made:

1<sup>o</sup>. When  $\gamma_1 < 0.4$ , the inaccuracy of  $B_\epsilon$  as a consequence of the Monte Carlo method is so great that it has not been possible to draw any conclusions. In these cases, however, even a rather large relative deviation in  $B_\epsilon$  does not give rise to any very significant absolute error in the approximation obtained by formula (5).

2<sup>o</sup>. When  $0.4 \leq \gamma_1 \leq 2.5$ , the  $(\epsilon, B_\epsilon)$  points respectively fall within a certain range of the diagonal of the field, it is not unreasonable to regard them as lying on a straight line. The deviations from this hypothesis can, at least partly, be explained by the above mentioned inaccuracy of the Monte Carlo method.

3<sup>o</sup>. When  $\gamma_1 > 2.5$ , it is not possible to find a satisfactory approximation procedure which is based on  $\gamma_1$  alone.

If the equation of  $B_\epsilon$  is determined from Figure 2 and put into expression (5) a formula

$$\boxed{x_\epsilon = y_\epsilon + 0.64 \gamma_1 \log \left( \frac{0.125}{\epsilon} \right)} \tag{5}'$$

is obtained. Even if it includes some anomalies in marginal values, this can be regarded as an acceptable approximation formula within reasonable limits of accuracy in the area of practical significance.

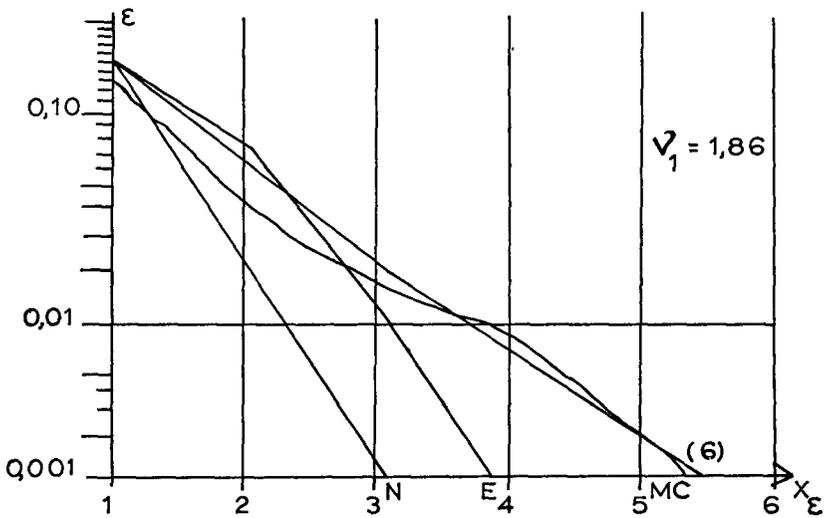
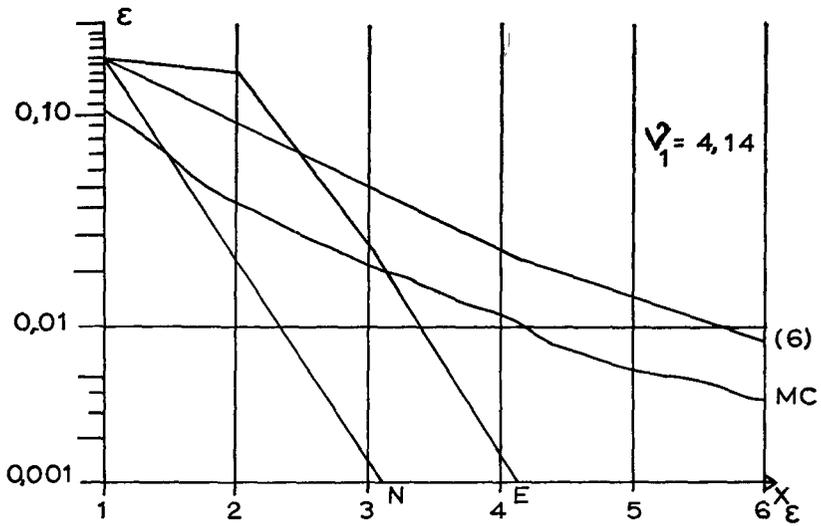
It is immediately seen that, if in formula (4) the exponent 0.98 is rounded off to 1 the formula is in fact the same as (5)' provided that

$$C_\epsilon = \frac{0.64}{y_\epsilon} \log \left( \frac{0.125}{\epsilon} \right)$$

4. *The formula derived theoretically*

Had this paper been prepared a few months ago, the formula (5)' would have been offered as a final approximation formula, but at this stage of our study Mr. Loimaranta put forth the formula

$$x_\epsilon = y_\epsilon + \frac{\gamma_1}{6} (y_\epsilon^2 - 1) + o \left( \frac{1}{n} \right), \tag{6}$$



E = Edgeworth

MC = Monte Carlo

(6) = Formula (6)

N = Normal distribution

Fig. 3. — Fig. 4.

where  $o(1/n)$  denotes an expression approaching to zero, when  $n \rightarrow \infty$ , in the same way as  $1/n$ . This formula is an application of a more general expansion due to Cornish and Fisher and present<sup>e</sup> by Kendall in his book "The Advanced Theory of Statistics" (Part 1, page 158). This series has been derived from the Edgeworth expansion

$$\epsilon = 1 - \Phi(x) + \frac{\gamma_1}{6} \Phi^{(3)}(x) + \dots \tag{7}$$

by solving  $x$  from this equation as a function of  $\epsilon$  and expressing  $\epsilon$  as a function of  $y$  from  $1 - \epsilon = \Phi(y)$  and then developing  $x$  as an expansion of  $y$ . Comparing values computed directly by means of the Edgeworth expansion (7) and by means of its inversion formula (6) it was quite surprisingly noticed that the latter gave essentially closer approximations. This is seen in the comparisons given in Figures 3-6 where the values computed by formula (6), formula (7) and the Monte Carlo method are compared in graphs drawn on a probability scale. It can be seen clearly that when  $\gamma_1$  is  $> 0.5$ , the Edgeworth expansion does not give accurate values for  $\epsilon$  (even the order of magnitude being uncertain), when  $x > 3$ . On the other hand, the curves in accordance with the formula (6) seem to run at least approximately in the right area. Only if  $\gamma_1$  is very large (the case  $\gamma_1 = 4.14$  in Figure 3) are the results inaccurate.

In Figure 2 the curve  $(\epsilon, B_\epsilon)$  is drawn according to formula (6). The experimental points seem to fall more below rather than above the curve corresponding to the formula (6).

To examine the accuracy of formula (6) theoretically, the remainder term  $o(1/n)$  must be analysed. The next term of the expansion is

$$\frac{\gamma_2}{24} (y_\epsilon^3 - 3y_\epsilon) - \frac{\gamma_1^2}{36} (2y_\epsilon^3 - 5y_\epsilon), \tag{8}$$

where  $\gamma_2$  is the excess  $\alpha_4/\alpha_2^2 n$  (cf. (3)). Naturally, nothing definite regarding general applicability can be implied from the magnitude of expression (8) because when  $\gamma_1$  is fixed and finite,  $\gamma_2$  can even be infinite.

By appealing purely to theory hardly any certain conclusions about the influence of the expression can be drawn, as the expansion

sion on which the formula (6) is based is divergent. However, it seems that when including expression (8) into the approximation according to the formula (6) the accuracy as compared with the results computed by the Monte Carlo method in general increases. There are however exceptions. On the other hand, within the limits of the material here examined the expression (8) brings the approximation according to the formula (6) closer towards the values computed by the Esscher formula in every case.

Table 1 includes all the results of the distributions computed by the Monte Carlo method, the Esscher formula and the formula (6) as well as the values of the remainder term (8). The distributions are arranged in increasing order of the values of  $y_1$ . The accuracy of the formula (6) proved to be about the same as that of the Esscher formula.

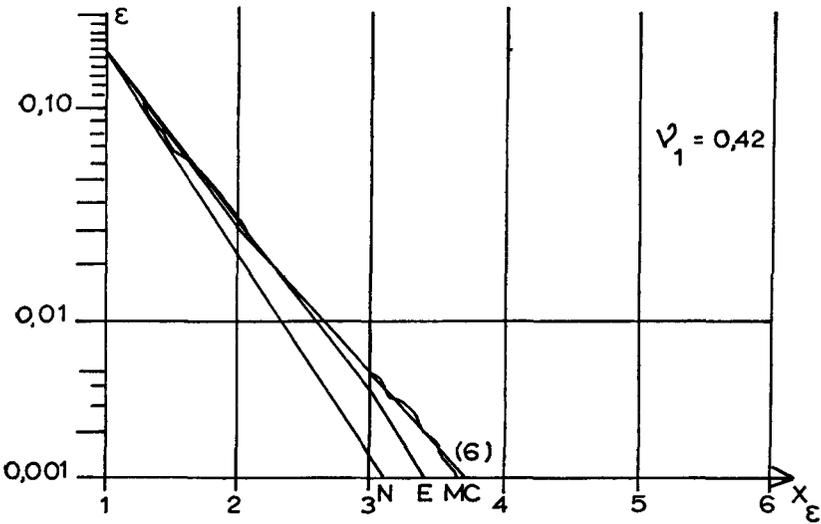
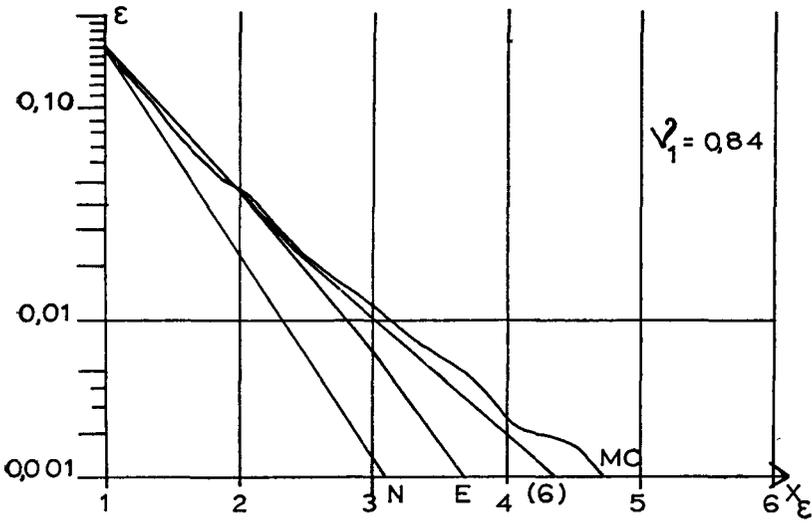
It is interesting to notice that formula (6) is in fact the same as formula (4), provided the exponent 0.98 is rounded off to 1 and if

$$C_\varepsilon = \frac{y_\varepsilon^2 - 1}{6y_\varepsilon}.$$

When computed from this formula the values 0.16, 0.32 and 0.46 for  $\varepsilon = 0.05$ , 0.01 and 0.001 are found for the constant  $C_\varepsilon$ , which deviate only slightly from the "empirical" values mentioned above.

##### 5. Comparison with a Swedish material

To check the reliability of formulae (5)' and (6) they were applied to risk distributions which a Swedish team of actuaries had used for computation of the function  $F(x)$  by means of a method based on the inversion of the Fourier transforms (Bohman and Esscher "To Compute the Distribution Function when the Characteristic Function is known", *Skandinavisk Aktuarietidskrift* 1964). This comparison is illustrated in Figure 7 which is constructed as Figure 2. The figure shows that the above mentioned results can also be applied to these distributions. It should be observed that no reinsurance was assumed. Hence the distributions were of a very heterogeneous type for which the approximation methods can be expected to be least suitable. However, the compatibility seems to be fairly satisfactory except for large



E=Edgeworth

MC=Monte Carlo

(6)=Formula (6)

N=Normal distribution

Fig. 5. — Fig. 6.

values of  $\gamma_1$ . In this region, as in the Finnish material for the great values of  $\gamma_1$ , larger deviations appear and the formula may not be applicable in this area, confirmation of the earlier conclusion.

It is interesting to observe that the figure also includes Polya cases, where fluctuation of the basic probabilities is assumed. No significant differences between Poisson and Polya cases can be noted. In fact a closer examination reveals that the formula (6) is even more suited for the Polya cases than for the Poisson cases as can be seen from the table below. In this table  $k$  is the Polya constant ( $k = \infty$  corresponds the Poisson case). The first two rows are outside the area of suitability (because  $\gamma_1$  is so great).

LIFE INSURANCE B  
Values of  $x_e$  computed by the formula (6)

$n$	$k$	$\gamma_1$	Correct value $x = 2.00$	Correct value $x = 3.00$
100	20	3.45	3.26	3.89
	$\infty$	3.85	3.48	4.12
500	20	1.20	2.03	3.05
	$\infty$	1.72	2.22	3.19
1000	20	0.77	2.02	3.01
	$\infty$	1.22	2.06	3.05
10000	20	0.45	2.01	3.01
	$\infty$	0.38	2.01	3.01

### 6. Comparison and criticism of the results

The study has given three approximation formulae to deal with the generalised Poisson function. The first, formula (4), is an endeavour to approximate the Esscher formula, the second, formula (5)' gives results computed by the Monte Carlo method and the third, formula (6), was deduced theoretically, without however paying close regard to accuracy.

Formula (4) is valid as an approximation in the same area as the Esscher formula, at least no significant deviations are to be found. As to the formula (5)' it can be deduced in the distributions observed herein that statistically it gives a somewhat better result than formula (6). As regards formula (6) it is clearly more accurate

Table 1

Comparison of  $x$ -values obtained by means of Monte Carlo method, Esscher formula and formula (6)

$I$  = Industrial fire insurance       $M$  = Third party motor insurance

Branch	$n$	$M$	$\gamma_1$	$\varepsilon = 0.01$		$\gamma_\varepsilon = 2.326$		$\varepsilon = 0.001$		$\gamma_\varepsilon = 3.091$	
				Monte Carlo	Esscher	(6)	(8)	Monte Carlo	Esscher	(6)	(8)
$I$	10 000	$\bar{Z} + 1/2\sigma$	0.0387	2.33	2.35	2.35		3.09	3.15	3.15	
$I$	10 000	$\bar{Z} + 2\sigma$	0.0603	2.33	2.37	2.37		3.04	3.18	3.18	
$M$	10 000	$\bar{Z} + 5\sigma$	0.0702	2.35	2.38	2.38		3.16	3.19	3.19	
$I$	2 000	$\bar{Z} + 1/2\sigma$	0.0874	2.38	2.39	2.39		3.25	3.21	3.22	
$I$	10 000	$\bar{Z} + 5\sigma$	0.0886	2.34	2.39	2.39		3.16	3.22	3.22	
$M$	10 000	$\bar{Z} + 10\sigma$	0.0951	2.35	2.38	2.40		3.09	3.23	3.23	
$M$	10 000	$\bar{Z} + 18\sigma$	0.1220	2.34	2.42	2.42		3.09	3.26	3.26	
$I$	2 000	$\bar{Z} + 2\sigma$	0.1349	2.42	2.43	2.43		3.37	3.29	3.28	
$M$	2 000	$\bar{Z} + 5\sigma$	0.1570	2.47	2.44	2.44		3.50	3.31	3.32	
$I$	500	$\bar{Z} + 1/2\sigma$	0.1732	2.43	2.45	2.45		3.48	3.33	3.34	-0.01
$M$	10 000	$\infty$	0.1866	2.39	2.46	2.46		3.17	3.36	3.36	+0.01
$I$	2 000	$\bar{Z} + 5\sigma$	0.1953	2.53	2.47	2.47		3.52	3.36	3.37	-0.01
$M$	2 000	$\bar{Z} + 10\sigma$	0.2127	2.45	2.48	2.48		3.50	3.38	3.39	-0.01
$I$	500	$\bar{Z} + 2\sigma$	0.2694	2.53	2.52	2.52	-0.01	3.65	3.45	3.47	-0.02
$M$	2 000	$\bar{Z} + 18\sigma$	0.2729	2.54	2.53	2.53	-0.01	3.50	3.47	3.48	-0.01
$M$	500	$\bar{Z} + 5\sigma$	0.3144	2.75	2.55	2.56	-0.01	3.66	3.52	3.54	-0.02
$I$	100	$\bar{Z} + 1/2\sigma$	0.3879	2.62	2.60	2.61	-0.02	3.60	3.60	3.64	-0.04
$I$	500	$\bar{Z} + 5\sigma$	0.3912	2.60	2.59	2.61	-0.02	3.85	3.63	3.65	-0.04
$M$	2 000	$\infty$	0.4174	2.60	2.63	2.63		3.64	3.71	3.69	+0.03
$M$	500	$\bar{Z} + 10\sigma$	0.4266	2.77	2.63	2.64	-0.02	3.98	3.67	3.70	-0.04
$I$	10 000	$\infty$	0.5410	2.72	2.70	2.72	-0.01	3.81	3.84	3.86	-0.01
$M$	500	$\bar{Z} + 18\sigma$	0.5470	2.83	2.72	2.73	-0.02	4.10	3.83	3.87	-0.04
$I$	100	$\bar{Z} + 2\sigma$	0.6037	2.73	2.74	2.77	-0.04	3.94	3.88	3.95	-0.10
$M$	100	$\bar{Z} + 5\sigma$	0.7033	2.80	2.80	2.84	-0.05	3.60	4.01	4.09	-0.11
$M$	500	$\infty$	0.8366	3.06	2.95	2.94	+0.01	4.62	4.34	4.28	+0.12
$I$	20	$\bar{Z} + 1/2\sigma$	0.8674	2.81	2.89	2.96	-0.09	4.32	4.19	4.33	-0.22
$I$	100	$\bar{Z} + 5\sigma$	0.8746	2.84	2.92	2.97	-0.08	4.11	4.22	4.34	-0.17
$M$	100	$\bar{Z} + 10\sigma$	0.9527	2.86	2.97	3.03	-0.09	4.17	4.32	4.45	-0.18
$I$	2 000	$\infty$	1.2092	3.19	3.16	3.21	-0.07	4.87	4.71	4.82	-0.06
$M$	100	$\bar{Z} + 18\sigma$	1.2217	2.86	3.15	3.22	-0.11	4.10	4.68	4.83	-0.19
$I$	20	$\bar{Z} + 2\sigma$	1.3500	3.22	3.20	3.32	-0.21	4.68	4.75	5.02	-0.48
$M$	20	$\bar{Z} + 5\sigma$	1.5727	3.40	3.34	3.48	-0.25	4.82	5.00	5.33	-0.54
$M$	100	$\infty$	1.8564	3.75	3.71	3.69	+0.02	5.37	—	5.74	+0.57
$I$	20	$\bar{Z} + 5\sigma$	1.9557	3.49	3.55	3.76	-0.39	5.20	—	5.88	-0.85
$M$	20	$\bar{Z} + 10\sigma$	2.1304	3.44	3.66	3.89	-0.44	5.45	—	6.13	-0.90
$I$	500	$\infty$	2.4178	3.85	—	4.10	-0.30	6.38	—	6.53	-0.23
$M$	20	$\bar{Z} + 18\sigma$	2.732	4.02	4.10	4.33	-0.57	—	—	6.99	-0.97

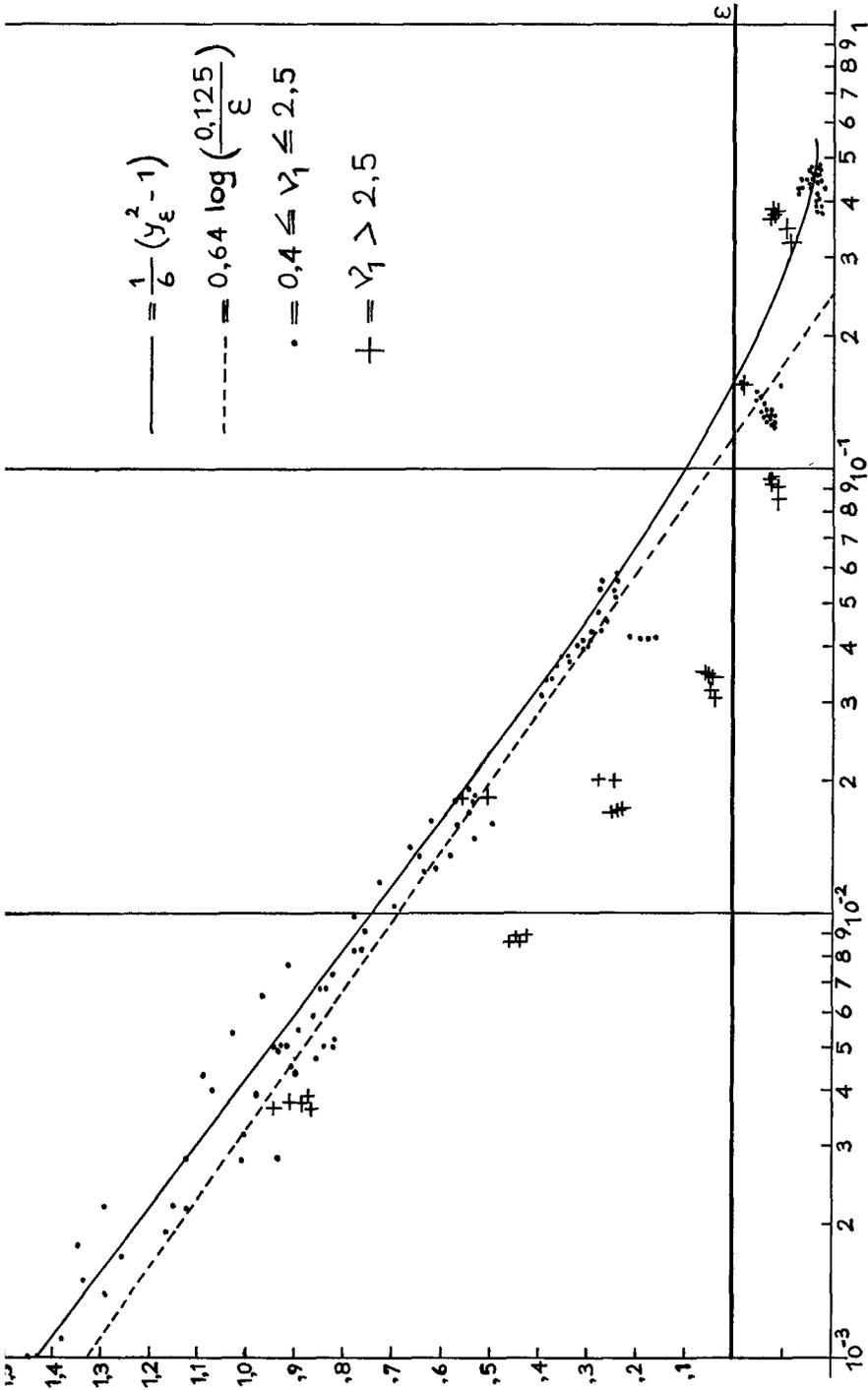


Fig. 7.

than the Edgeworth expansion; in fact it is suitable for use on the same area as the Esscher formula and almost equal in accuracy.

The study outlined above has revealed that it is possible, based on the normal distribution, to develop an improved method of approximating the generalised Poisson function. The use of the method itself is very simple. The values of  $y_e$ , which are obtained from tables concerning the normal distribution, have to be altered in the manner of the formulae (4), (5)' or (6). The mapping out of the area of suitability is still an open question, but the material available was, in fact, already fairly abundant and represented different types of risk distributions. It has, without exception, established the suitability for at least approximate calculations to be acceptable, provided the skewness  $\gamma_1 < 2.5$ . Some other materials have proved that the reliability of formula (6) can become questionable if  $\gamma_1 > 1.5$ .