On approximations for the distribution of a heterogeneous risk portfolio

BJØRN SUNDT, Oslo*

On Approximations for the Distribution of a Heterogeneous Risk Portfolio

1 The Setup and Some Previously Studied Approximations

Let $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N$ be independent random variables taking integer values in the range $[0, 1, \ldots, R]$, and introduce

$$f_i(x) = Pr(\tilde{x}_i = x). \qquad (i=1,\ldots,N;\ x=0,1,\ldots,R)$$

It is assumed that $f_i 0)$ is significant for all i, and we want to find the discrete density of the sum $\tilde{y} = \sum_{i=1}^{N} \tilde{x}_i$, that is,

$$g(y) = Pr(\tilde{y} = y) = \left(\mathop{*}_{i=1}^{N} f_i \right)(y).$$

This density can of course be computed exactly by convoluting the f_i's. However, this task could be very time-consuming if N is large and if the $\tilde{x}_i$'s can take more than a few values.

An often used approximation is to assume that $\tilde{y}$ is the sum of a random number $\tilde{n}$ of independent and identically distributed random variables $\tilde{w}_1, \ldots, \tilde{w}_{\tilde{n}}$ with common density

$$h(w) = Pr(\tilde{w} = w), \qquad (w = 1, 2, \ldots, R)$$

independent of $\tilde{n}$. One usually puts

$$h(w) = \frac{1}{\lambda} \sum_{i=1}^{N} f_i(w)$$

with

$$\lambda = \sum_{i=1}^{N} (1 - f_i(0)),$$

and for the density

$$\pi_n = Pr(\tilde{n} = n)$$

* The present research was performed while the author was staying at the Laboratory of Actuarial Mathematics, University of Copenhagen.

of $\tilde{n}$ one usually applies the Poisson density

$$\pi_n = \frac{\lambda^n}{n!} e^{-\lambda} \qquad (n=0,1,\ldots)$$

(cf. Gerber (1979), Jewell and Sundt (1981)). This approximation gives exact match for the first moment of $\tilde{y}$, but Var $\tilde{y}$ in the approximation is greater than the exact value.
Jewell and Sundt (1981) suggested to replace the Poisson assumption by a binomial one, that is,

$$\pi_n = \binom{M}{n} \pi^n (1-\pi)^{M-n}, \qquad (n=0,1,\ldots,M)$$

where the parameters π and M are chosen so as to match exactly the mean and approximately the variance of $\tilde{y}$. The match of the variance is only approximate as M has to be an integer. One gets

$$M \approx \frac{\left(\sum_{i=1}^{N} E\tilde{x}_i\right)^2}{\sum_{i=1}^{N} E^2\tilde{x}_i}$$

$$\pi = \frac{\lambda}{M}.$$

From Schwartz's Inequality we see that $M \leqslant N$. The compound binomial approximation can of course be reformulated as an approximation to g by k^{M*}, where the discrete density k is given by

$$k(x) = \begin{cases} 1-\pi & (x=0) \\ \pi h(x). & (x=1,2,\ldots,R) \end{cases} \tag{1}$$

We see that if $f_i = f$ independent of i, then $k=f$, $M=N$, and $g=f^{N*}$, that is, the approximation is exact, whereas the compound Poisson approximation is never exact.
When considering the compound binomial approximation as k^{M*}, one must admit that the approximation looks a bit unnatural. However, this does not exclude that it could work well in practice. In the numerical example given by Jewell and Sundt (1981) and reproduced in Section 5 of the present paper, the approximation gives very satisfactory results.

2 A Natural Approximation

We introduce the "average" density

$$f(x)=\frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

and get

$$h(w)=\frac{f(w)}{1-f(0)}$$

$$\lambda=N(1-f(0)).$$

Thus we see that the compound Poisson approximation depends on $f_1,\ldots,f_N$ only through f. Therefore the approximation may be considered as consisting of two steps:

i) Approximate $\mathop{*}\limits_{i=1}^{N} f_i$ by f^{N*}.

ii) Approximate f^{N*} by a compound Poisson distribution with Poisson parameter λ and severity distribution h.

A natural question now is of course: Wouldn't it be better to omit the second step and approximate g by f^{N*}? We note that like k^{M*} this approximation is exact in the special case when $f_1=f_2=\ldots=f_N$.

The approximation f^{N*} can be given a natural motivation related to the theory of experience rating (cf. e.g. Norberg (1979)). We know that there are differences between the policies, but consider them as random. To each policy i there is connected a random parameter $\tilde{\theta}_i$ containing the risk characteristics of that particular policy. It is assumed that the conditional density

$$\phi(x|\theta)=Pr(\tilde{x}_i=x|\tilde{\theta}_i=\theta)$$

is independent of i and that $\tilde{\theta}_1,\tilde{\theta}_2,\ldots,\tilde{\theta}_N$ are independent and identically distributed; we denote their common distribution by U. The unconditional density of $\tilde{x}_i$ is

$$\phi(x)=Pr(\tilde{x}_i=x)=\int \phi(x|\theta)dU(\theta),$$

and in the present model the density of $\tilde{y}$ is

$$\gamma(y)=Pr(\tilde{y}=y)=\phi^{N*}(y).$$

In our original setup it was assumed that $\tilde{x}_1\ldots,\tilde{x}_N$ were independent with discrete densities $f_1,\ldots,f_N$. It is now natural to interpret $f_i(x)$ as $\phi(x|\theta_i)$, where θ_i denotes the value of $\tilde{\theta}_i$.

A reasonable estimator of U is the empirical distribution $\hat{U}$ given by

$$\hat{U}\{A\}=\frac{1}{N}\#\{i\colon \tilde{\theta}_i\in A\},$$

and we estimate $\phi(x)$ by

$$\hat{\phi}(x)=\int \phi(x|\theta)d\hat{U}(\theta)=\frac{1}{N}\sum_{i=1}^{N}\phi(x|\tilde{\theta}_i)$$

and $\gamma(y)$ by

$$\hat{\gamma}(y)=\hat{\phi}^{N*}(y).$$

But then the value of $\hat{\phi}(x)$ is

$$\frac{1}{N}\sum_{i=1}^{N}\phi(x|\theta_i)=\frac{1}{N}\sum_{i=1}^{N}f_i(x)=f(x).$$

Thus our estimate of $\gamma(y)$ is $f^{N*}(y)$, which we previously suggested as approximation to $g(y)$.

3 Recursive Computation

For the compound binomial approximation Jewell and Sundt (1981) recommended that g should be computed by the recursive method

$$g(y)=\begin{cases}(1-\pi)^M & (y=0)\\ \left(\dfrac{\pi}{1-\pi}\right)\displaystyle\sum_{x=1}^{\min(y,R)}\left[(M+1)\frac{x}{y}-1\right]h(x)g(y-x), & (y=1,2,\ldots,MR)\end{cases}$$

described by Panjer (1981). Insertion of (1) gives

$$g(y)=\begin{cases}k(0)^M & (y=0)\\ \dfrac{1}{k(0)}\displaystyle\sum_{x=1}^{\min(y,R)}\left[(M+1)\frac{x}{y}-1\right]k(x)g(y-x), & (y=1,\ldots,MR)\end{cases}$$

and in particular we see that

$$f^{N*}(y)=\begin{cases}f(0)^N & (y=0)\\ \dfrac{1}{f(0)}\displaystyle\sum_{x=1}^{\min(y,R)}\left[(N+1)\frac{x}{y}-1\right]f(x)f^{N*}(x-y). & (y=1,\ldots,NR)\end{cases}\tag{2}$$

This recursive method for the computation of the N-th convolution of a discrete density f has been described by De Pril (1985), and it can be used to compute f^{N*} as an approximation to g.

4 Associated Functions

We introduce the tail

$$G^c(y) = Pr(\tilde{y} > y) = \sum_{x=y+1}^{\infty} g(x) = 1 - \sum_{x=0}^{y} g(x)$$

and the stop-loss premium

$$\bar{G}(y) = E(\tilde{y} - y)^+ = \sum_{x=y}^{\infty} G^c(x) = E\tilde{y} - \sum_{x=0}^{y-1} G^c(x)$$

of the distribution of $\tilde{y}$. If the values of g are known, we easily compute $G^c(y)$ and $\bar{G}(y)$ recursively by

$$G^c(y) = \begin{cases} 1 & (y = -1) \\ G^c(y-1) - g(y) & (y = 0, 1, \ldots) \end{cases}$$

$$\bar{G}(y) = \begin{cases} E\tilde{y} & (y = 0) \\ \bar{G}(y-1) - G^c(y-1), & (y = 1, 2, \ldots) \end{cases}$$

and approximated values are found by inserting approximations to $g(y)$ in these recursions.

It has been shown by Bühlmann et al. (1977) that the stop-loss premiums found from the compound Poisson approximation are always greater than or equal to the exact values. As this result in particular holds for identically distributed $\tilde{x}_i$'s, we have that the compound Poisson approximation always gives stop-loss premiums greater than or equal to those found from the approximation f^{N*}. Furthermore, the stop-loss premiums found from the compound Poisson approximation are always greater than or equal to the ones found from the approximation k^{M*}. This can be seen in the following way. Consider M independent and identically distributed claim amounts with common discrete density k. We approximate the aggregate distribution of these claim amounts by the compound Poisson distribution with Poisson parameter

$$\lambda' = M(1 - k(0))$$

and discrete severity density

$$h'(w) = \frac{k(w)}{1 - k(0)}. \qquad (w = 1, 2, \ldots, R)$$

By the above mentioned result from Bühlmann et al. (1977) the stop-loss premiums of this compound distribution are greater than or equal to those of k^{M*}. It is easily seen that $\lambda' = \lambda$ and $h' = h$, and thus the proposition is proved.

In the numerical example in Section 5 the stop-loss premiums found from f^{N*} are always greater than or equal to the exact ones, and it is tempting to conclude that this approximation always gives an upper bound to the exact stop-loss premiums. However, such a result does not hold in general as can be seen from the counter-example with $N=2$ displayed in Table 1.

Table 1

y	$f_1(y)$	$f_2(y)$	$\bar{G}(y)$	
			exact	f^{N*}-approximation
0	0.5714	0.5	1.0714	1.0714
1	0.2857	0.5	0.3571	0.3584
2	0.1429	0	0.0714	0.0663
3	0	0	0	0.0051
4	0	0	0	0

5 A Numerical Example

To illustrate the approximations we use a numerical example due to Gerber (1979) and studied by Jewell and Sundt (1981). We consider a portfolio consisting of $N=31$ policies, and the random values $\tilde{x}_i$ are either 0 or a «face value» c_j with probability $1-q_i$ and q_i respectively, as shown in Table 2.

Table 2. Number of policies with indicated q_i and c_j

q_i	face values c_j				
	1	2	3	4	5
0.03	2	3	1	2	–
0.04	–	1	2	2	1
0.05	–	2	4	2	2
0.06	–	2	2	2	1

In Table 3 we show the average density f, and in Table 4 Var $\tilde{y}$ for the exact distribution and the three approximations.

Table 3. Average density f

x	0	1	2	3	4	5
$31\,f(x)$	29.60	0.06	0.35	0.43	0.36	0.20

Table 4

Exact and approximated variance of $\tilde{y}$ (differing digits underlined)

	Exact value	Approximations		
		Poisson	f^{N^*}	k^{M^*}
Var $\tilde{y}$	15.3003	$\underline{16.0900}$	$15.\underline{4397}$	$15.3\underline{146}$

Furthermore, we have $\lambda = 1.4$, $M = 26$, and $\pi = 0.0538462$.
By the recursive method (2) we computed the density $g(y)$ for the approximation f^{N^*}. From these values we found the tail $G^c(y)$ and the stop-loss premium $\bar{G}(y)$. The computed values are given in Tables 5–7, Appendix, compared to the corresponding values taken from Jewell and Sundt (1981) for the exact distribution, the Poisson approximation, and k^{M^*}. In Figures 1–3, Appendix, we show the percentage error in each approximation for the function of interest. Like the other two approximations, f^{N^*} does not give a particularly good approximation to g; it fluctuates above and below the exact density in about the same manner as k^{M^*}. When ranking the three approximations to $g(y)$ for the different y's, we see that $f^{N^*}(y)$ tends to be the second best approximation, whereas the compound Poisson approximation and $k^{M^*}(y)$ alternate having the first and third place.
For the tail $G^c(y)$, the general impression is that k^{M^*} gives the best approximation, and that the approximation based on f^{N^*} performs better than the compound Poisson approximation. However, this ranking does not hold uniformly; for $y = 6$, the compound Poisson is best and k^{M^*} worst.
For the stop-loss premiums $\bar{G}(y)$, the compound Poisson approximation always gives the greatest error, but as pointed out in Section 4, it has the advantage that it is analytically shown that it will always give an upper bound for the exact stop-loss premium. Except for $y \leq 2$ k^{M^*} gives smaller errors than f^{N^*}.

6 Conclusion

In the present paper we have discussed f^{N*} as approximation to g as an alternative to the compound Poisson approximation and k^{M*}. It is argued that f^{N*} has a better intuitive appeal than the two other approximations. We have not performed any profound analytical comparison of the three approximations. It is of course dangerous to base any firm conclusions on one single numerical example. However, for approximations to G^c and $\bar{G}$, it is the impression that f^{N*} usually performs better than the compound Poisson approximation, but worse than k^{M*}.

References

Bühlmann, H., Gagliardi, B., Gerber, H. U., and Straub, E. (1977), Some inequalities for stop-loss premiums. *ASTIN Bulletin* **9**, 75–83.

De Pril, N. (1985). Recursive evaluation of convolutions of identically distributed integer-valued random variables. To be published in *ASTIN Bulletin*.

Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory*. R. D. Irwin, Inc., Homewood.

Jewell, W. S., and Sundt, B. (1981). Improved approximations for the distribution of a heterogeneous risk portfolio. *Bulletin of the Association of Swiss Actuaries*, 221–239.

Norberg, R. (1979). The credibility approach to experience rating. *Scandinavian Actuarial Journal*, 181–221.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22–26.

Bjørn Sundt
Storebrand-Norden
P.O. Box 1380, Vika
N-0114 Oslo 1
Norway

Appendix

Table 5

Exact values and approximations to total sum density
(differing digits underlined)

y	g(y) = Pr{ỹ=y}: Exact result	Approximations: Poisson	Approximations: f^{N*}	Approximations: k^{M*}
0	0.23819	0.24660	0.23869	0.23714
1	0.01473	0.01480	0.01500	0.01504
2	0.08773	0.08675	0.08795	0.08818
3	0.11318	0.11122	0.11282	0.11313
4	0.11071	0.11040	0.11220	0.11256
5	0.09633	0.09286	0.09471	0.09507
6	0.06155	0.06101	0.06259	0.06291
7	0.06902	0.06543	0.06700	0.06732
8	0.05482	0.05458	0.05567	0.05589
9	0.04315	0.04132	0.04187	0.04197
10	0.03011	0.03058	0.03069	0.03071
11	0.02353	0.02331	0.02315	0.02311
12	0.01828	0.01834	0.01804	0.01797
13	0.01251	0.01315	0.01273	0.01265
14	0.00871	0.00922	0.00875	0.00866
15	0.00591	0.00650	0.00605	0.00596
16	0.00415	0.00460	0.00419	0.00411
17	0.00272	0.00318	0.00283	0.00277
18	0.00174	0.00212	0.00184	0.00179
19	0.00112	0.00141	0.00119	0.00115
20	0.00071	0.00094	0.00076	0.00073
30	$3.09434\cdot 10^{-6}$	$8.63294\cdot 10^{-6}$	$4.57655\cdot 10^{-6}$	$3.98500\cdot 10^{-6}$
40	$3.53514\cdot 10^{-9}$	$36.4155\cdot 10^{-9}$	$9.89290\cdot 10^{-9}$	$7.37055\cdot 10^{-9}$

Table 6
Exact values and approximations to tail
(differing digits underlined)

	$G^C(y) = \Pr\{\tilde{y}>y\}$			
		Approximations		
y	Exact result	Poisson	f^{N*}	k^{M*}
0	0.76181	0.75340	0.76131	0.76286
1	0.74707	0.73861	0.74631	0.74782
2	0.65934	0.65185	0.65837	0.65964
3	0.54615	0.54063	0.54555	0.54651
4	0.43544	0.43023	0.43334	0.43395
5	0.33912	0.33737	0.33864	0.33888
6	0.27757	0.27637	0.27605	0.27597
7	0.20855	0.21094	0.20904	0.20865
8	0.15373	0.15636	0.15337	0.15276
9	0.11058	0.11504	0.11150	0.11079
10	0.08048	0.08446	0.08081	0.08008
11	0.05695	0.06115	0.05766	0.05696
12	0.03866	0.04281	0.03962	0.03899
13	0.02615	0.02966	0.02689	0.02635
14	0.01744	0.02044	0.01813	0.01769
15	0.01153	0.01394	0.01208	0.01173
16	0.00738	0.00934	0.00789	0.00762
17	0.00467	0.00617	0.00506	0.00485
18	0.00292	0.00404	0.00321	0.00306
19	0.00181	0.00263	0.00202	0.00192
20	0.00110	0.00169	0.00126	0.00118
30	$3.49840\cdot10^{-6}$	$12.4621\cdot10^{-6}$	$5.76662\cdot10^{-6}$	$4.87524\cdot10^{-6}$
40	$3.10833\cdot10^{-9}$	$45.5298\cdot10^{-9}$	$10.37457\cdot10^{-9}$	$7.42541\cdot10^{-9}$

Table 7
Exact values and approximations to stop-loss premiums
(differing digits underlined)

	$\bar{G}(y) = E[(\tilde{y}-y)^+]$			
	Exact result	Approximations		
y		Poisson	f^{N*}	k^{M*}
0	4.49000	4.49000	4.49000	4.49000
1	3.72819	3.73660	3.72869	3.72714
2	2.98112	2.99799	2.98237	2.97932
3	2.32179	2.34614	2.32401	2.31968
4	1.77563	1.80551	1.77846	1.77317
5	1.34019	1.37527	1.34512	1.33922
6	1.00106	1.03790	1.00648	1.00034
7	0.72350	0.76153	0.73044	0.72437
8	0.51495	0.55059	0.52139	0.51572
9	0.36122	0.39423	0.36802	0.36296
10	0.25064	0.27919	0.25652	0.25217
11	0.17017	0.19472	0.17572	0.17209
12	0.11322	0.13357	0.11806	0.11513
13	0.07456	0.09076	0.07844	0.07614
14	0.04840	0.06110	0.05155	0.04979
15	0.03096	0.04065	0.03342	0.03210
16	0.01943	0.02671	0.02134	0.02037
17	0.01205	0.01737	0.01346	0.01276
18	0.00738	0.01120	0.00840	0.00791
19	0.00446	0.00716	0.00519	0.00485
20	0.00265	0.00453	0.00316	0.00293
30	$7.25353 \cdot 10^{-6}$	$29.7953 \cdot 10^{-6}$	$12.72764 \cdot 10^{-6}$	$10.5809 \cdot 10^{-6}$
40	$5.72441 \cdot 10^{-9}$	$101.020 \cdot 10^{-9}$	$20.92164 \cdot 10^{-9}$	$14.6686 \cdot 10^{-9}$

Figure 1

Percentage error in approximations to density $g(y)$ versus y

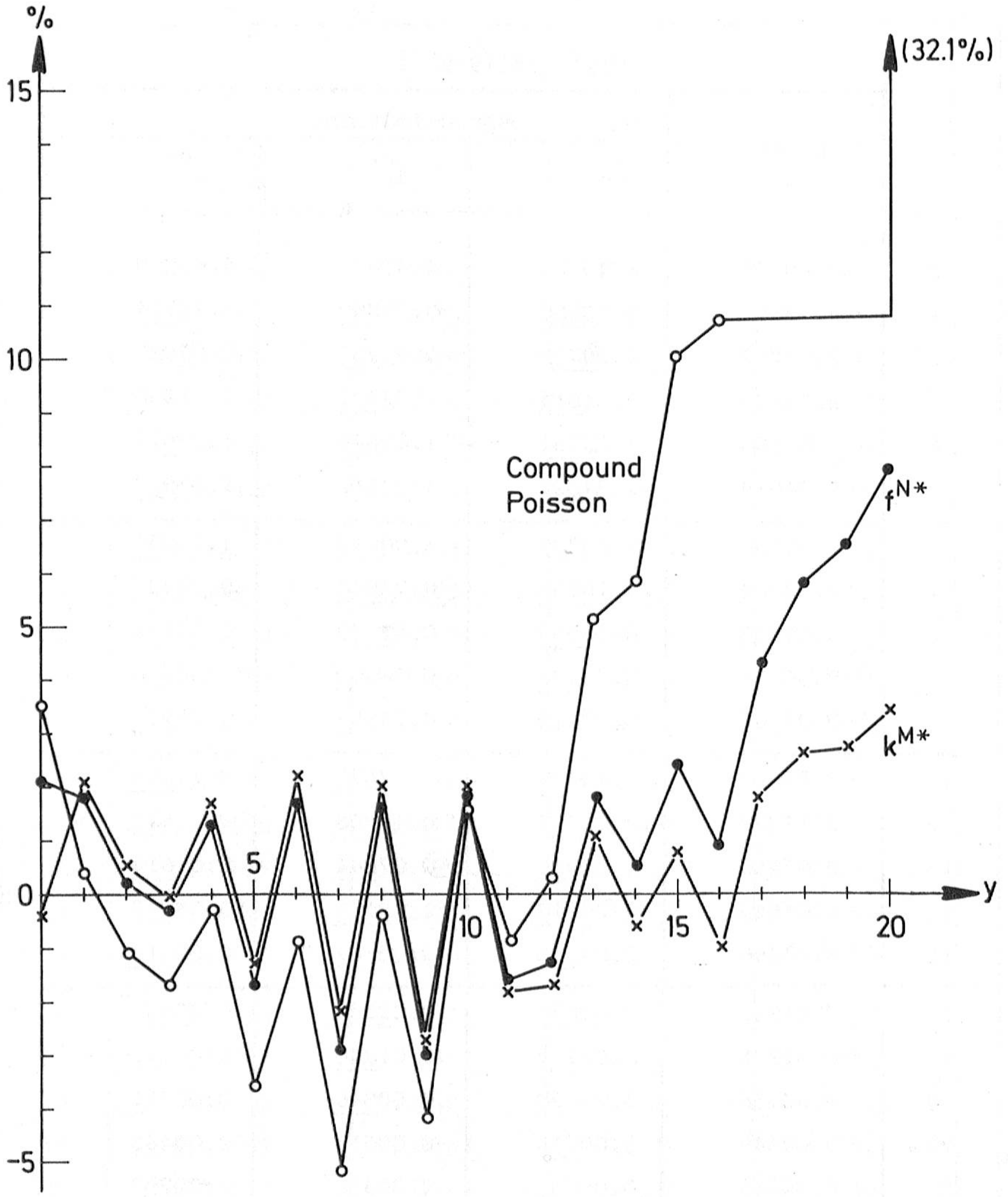


Figure 2
Percentage error in approximations to complementary distribution
$G^c(y)$ versus y

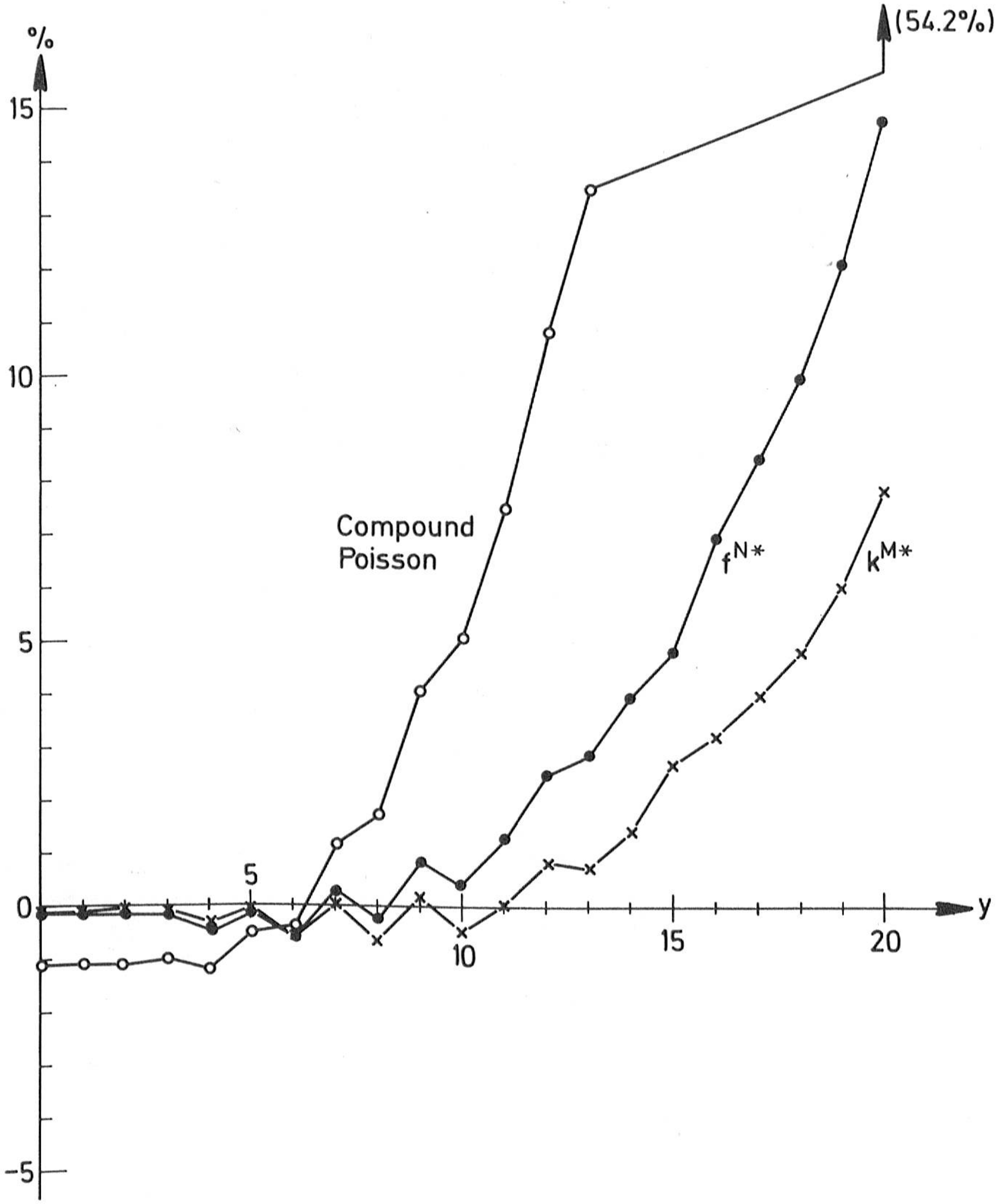
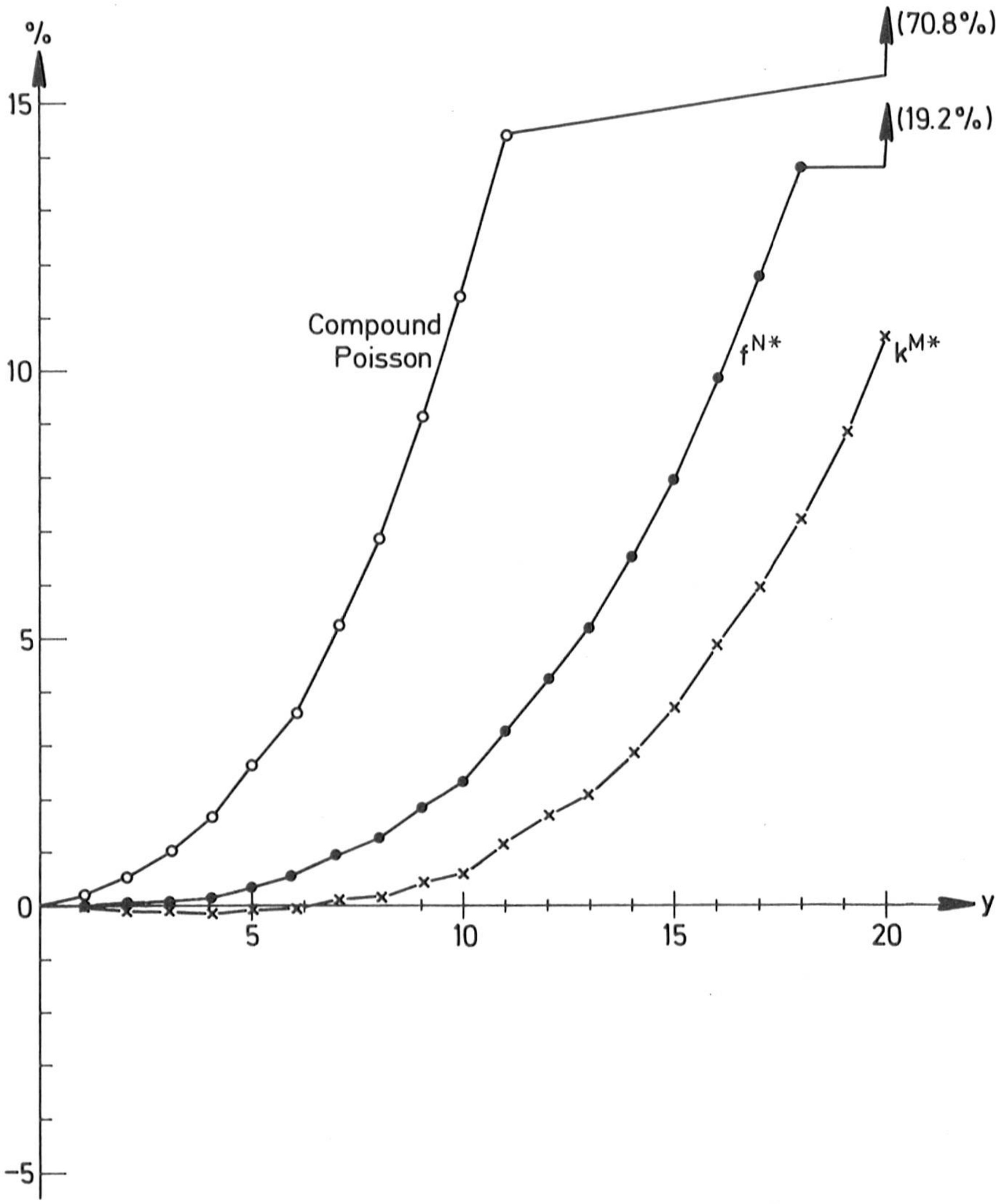


Figure 3
Percentage error in approximations to stop-loss premium $\bar{G}(y)$ versus y



Abstract

We consider N independent risks $\tilde{x}_1, \ldots, \tilde{x}_N$ with discrete densities

$$f_i(x) = Pr(\tilde{x}_i = x) \qquad (x = 0, 1, \ldots, R)$$

and approximate the discrete density of $\sum_{i=1}^{N} \tilde{x}_i$ by the N-th convolution of $f = N^{-1} \sum_{i=1}^{N} f_i$. This approximation is compared to two previously studied approximations. A numerical example is given.

Zusammenfassung

Wir betrachten N unabhängige Risiken $\tilde{x}_1, \ldots, \tilde{x}_N$ mit diskreten Dichten

$$f_i(x) = Pr(\tilde{x}_i = x) \qquad (x = 0, 1, \ldots, R)$$

und approximieren die diskrete Dichte von $\sum_{i=1}^{N} \tilde{x}_i$ durch die N-te Faltung von $f = N^{-1} \sum_{i=1}^{N} f_i$. Diese Approximation wird mit zwei früher untersuchten Approximationen verglichen und es wird ein numerisches Beispiel angeführt.

Résumé

L'article considère N risques indépendants $\tilde{x}_1, \ldots, \tilde{x}_N$ à densités discrètes

$$f_i(x) = Pr(\tilde{x}_i = x) \qquad (x = 0, 1, \ldots, R)$$

et établit une approximation de la densité discrète de $\sum_{i=1}^{N} \tilde{x}_i$ par la convolution d'ordre N de $f = N^{-1} \sum_{i=1}^{N} f_i$. Cette approximation est ensuite comparée à deux estimations étudiées par le passé. L'auteur présente un exemple numérique.