NOTE CONCERNING THE DISTRIBUTION FUNCTION OF THE TOTAL LOSS EXCLUDING THE LARGEST INDIVIDUAL CLAIMS

HANS AMMETER
Zurich, Switzerland

The theory of extreme values is a special branch of mathematical statistics and was mainly treated by E. J. Gumbel [4] *). This theory has only been applied in a few cases to problems in the insurance business. The first practical application to insurance known to the author of the present paper is due to A. Thépaut who has invented a new reinsurance system called ECOMOR [5]. According to this system the reinsurer covers the excess risk for the m largest claims and the ceding company retains an amount equal to the $(m + 1)$ largest claim. The credit for having pointed out the importance of the theory of extreme values belongs to R. E. Beard [1]. Recently E. Franckx [3] has found a most remarkable result by disclosing the general form of the distribution for the largest claim occurring in a certain accounting period.

The present paper starts from the consideration that not only is the distribution of major claims, which might be eliminated by means of reinsurance, of interest to an insurer but also the distribution of the remaining total loss after excluding the largest claims. The nature of this distribution is important not only in connection with stability and security, but also for statistical investigations of the observed claim ratio. The credibility of such an investigation might be greatly improved if a suitable number of major claims were excluded. To simplify matters, the present paper considers the case where only the largest claim is excluded.

The numerical computations for the distribution in question raise some difficulties and will not be discussed at the moment. The purpose of this paper is to find expressions for the first moments of the distribution function which are susceptible to

*) Numbers in [ ] refer to the list of references.

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

numerical computations. All formulas are based on the Poisson risk process. The more general compound Poisson processes should hardly give rise to additional difficulties.

Finally the results will be illustrated by means of the Pareto distribution which seems most appropriate to represent major claims and which leads to some remarkable results.

I. THE GENERAL FORM OF THE DISTRIBUTION OF THE TOTAL LOSS, WHEN THE LARGEST CLAIM IS EXCLUDED

In his paper [3] mentioned above *E.* Franckx has defined the distribution function $\Phi\ (m)$ of the largest claim m by the expression

$$\Phi\ (m) = Q\ [S\ (m)]. \tag{1}$$

In this formula

$$Q\ (s) = \sum_{r=0}^{\infty} s^r\ q_r$$

is the generating function of the distribution of the number of claims r, i.e. q_r is the probability that exactly r claims occur during the observed period. Furthermore $S\ (m)$ means the distribution function of the individual claim amounts. If one starts from the Poisson formula the distribution of the number of claims r

$$q_r = \frac{e^{-t}\ t^r}{r\ !} \tag{2}$$

the distribution function of the largest claim m is—as already shown by Franckx—defined by applying the generating function of the Poisson distribution (2)

$$Q\ (s) = e^{-t\ (1-s)} \tag{2'}$$

and

$$\Phi\ (m) = e^{-t\ [1-S\ (m)]} \tag{3}$$

where t is the expected number of claims.

The frequency function of the largest claim m can be derived from formula (3) by differentiation

$$\varphi\ (m) = t\ s\ (m)\ e^{-t\ [1-S\ (m)]}. \tag{4}$$

Starting from this formula the moments M_k of the distribution of the largest claim (3) can be determined

$$M_h = \int_0^\infty m^k \varphi(m)\, dm. \tag{5}$$

To find the frequency function $f(x, t, -1)$ of the total loss excluding the largest claim, a possible solution consists in developing the right hand side of formula (4) in the following way:

$$\varphi(m) = e^{-t} s(m) \left[\frac{t}{1!} + \frac{2t^2 S(m)}{2!} + \frac{3t^3}{3!} S^2(m) + \dots\dots\dots \right]$$

$$= \sum_{r=1}^{\infty} \frac{e^{-t}\, t^r}{r!} r\, s(m)\, S^{r-1}(m). \tag{4'}$$

From this formula the frequency function $f_m(x, t, -1)$ of the remaining total loss x can be derived. If the excluded largest claim lies in the interval m, $m + dm$, this equation is

$$f_m(x, t, -1) = \sum_{r=1}^{\infty} \frac{e^{-t}\, t^r}{r!} r\, s(m)\, [s_m(x)]^{*(r-1)} \tag{6}$$

In this formula $s_m(x)$ denotes the frequency function of claims truncated at the point m, i.e.

$$s_m(x) = s(x) \qquad \text{when } x \leq m$$

and $$s_m(x) = 0 \qquad \text{when } x > m$$

and $$[s_m(x)]^{*(r-1)}$$

is the $(r-1)^{th}$ convolution of $s_m(x)$. Attention must be focussed on the fact that $s(x)$ is first truncated and then convoluted and not first convoluted and then truncated. That is to say that

$$[s_m(x)]^{*(r-1)}\, dx$$

is the probability that the total amount of claims x, after excluding the largest claim m, when r claims occur lies in the interval x, $x + dx$.

The frequency function $f(x, t, -1)$ for any largest claim m may be derived from formula (6) by means of integration over m

$$f(x, t, -1) = \int_0^\infty f_m(x, t, -1)\, dm =$$

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

$$= \int_0^\infty \sum_{r=1}^\infty \frac{e^{-t} t^r}{r!} r \, s(m) \, [s_m(x)]^{*(r-1)} \, dm. \qquad (7)$$

The analogy with the well-known formula for the frequency function $f(x, t)$ of the total claim x is striking, for

$$f(x, t) = \sum_{r=1}^\infty \frac{e^{-t} t^r}{r!} s^{*r}(x).$$

Formula (7) thus gets the plausible interpretation that the convoluted power $s^{*r}(x)$ is replaced by the expression

$$r \, s(m) \, [s_m(x)]^{*(r-1)}$$

with subsequent integration over m.

II. THE MEAN OF THE DISTRIBUTION OF THE TOTAL LOSS, WHEN THE LARGEST CLAIM IS EXCLUDED

The unknown mean may be designated by $\mu_1^{(-1)}$ and is given by the integral

$$\mu_1^{(-1)} = \int_0^\infty x f(x, t, -1) \, dx.$$

Since the total loss x in the formula (7) only occurs in the convoluted power

$$[s_m(x)]^{*(r-1)},$$

it might be helpful to determine first the mean

$$\int_0^\infty x \, [s_m(x)]^{*(r-1)} \, dx,$$

which can be derived from the simple mean value

$$S_{m,1} = \int_0^m x \, s_m(x) \, dx = \int_0^m x \, s(x) \, dx.$$

The necessary link between these mean values may be found by considering the corresponding characteristic functions

$$\sigma_m(z) = \int_0^\infty e^{izx} \, s_m(x) \, dx \qquad \text{and}$$

$$\sigma_m^{r-1}(z) = \int_0^\infty e^{izx} \, [s_m(x)]^{*(r-1)} \, dx$$

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

From these formulas the following relation may be found by differentiation over t and then putting $t = 0$:

$$\int_0^\infty x \; [s_m(x)]^{*(r-1)} \, dx = (r-1) \, S_{m,1} \, S(m)^{r-2}. \tag{8}$$

Formula (8) may hence be denoted as

$$\mu_1^{(-1)} = \int_0^\infty \sum_{r=1}^\infty \frac{e^{-t} \, t^r}{r\,!} \, r \, (r-1) s\,(m) \, S_{m,1} \; S\,(m)^{r-2} \; dm =$$

$$= \int_0^\infty t^2 \, s(m) \;\; S_{m,1} \; e^{-t\,[1-S(m)]} \; dm. \tag{9}$$

III. THE STANDARD DEVIATION OF THE DISTRIBUTION OF THE TOTAL LOSS, WHEN THE LARGEST CLAIM IS EXCLUDED

The second moment about the origin of the distribution (7) may be denoted as $\mu_2^{(-1)}$ and is given by the equation

$$\mu_2^{(-1)} = \int_0^\infty x^2 \; f\,(x, t, -1) \;\; dx$$

The standard deviation can thus be calculated by using the well-known formula

$$\sigma^{(-1)} = \sqrt{\mu_2^{(-1)} - [\mu_1^{(-1)}]^2}$$

The second moment is calculated in an analogous manner to the mean. From the corresponding characteristic functions the following relation is first obtained:

$$\int_0^\infty x^2 \, [s_m(x)]^{*(r-1)} \, dx = (r-1) \, (r-2) \, S_{m,1}^2 \;\; S(m)^{r-3} + (r-1) \; S_{m,2} \, S(m)^{r-2},$$

where

$$S_{m,\,2} = \int_0^m x^2 \;\; s_m\,(x) \;\; dx = \int_0^m x^2 \;\; s\,(x) \;\; dx.$$

In a similar way as for the mean the following formula is then determined:

$$\mu_2^{(-1)} = \int_0^\infty \sum_{r=1}^\infty \frac{e^{-t} \, t^r}{r} \, s\,(m) \left[r\,(r-1) \;\, (r-2) \;\, S_{m,1}^2 \; S\,(m)^{r-3} + \right.$$

$$+ r(r-1) S_{m,2} \overset{r-2}{S}(m) \Big] dm = \int_0^\infty t\, s(m)\, e^{-t\,[1-S(m)]} \Big[(t\, S_{m,1})^2 + t\, S_{m,2} \Big] dm. \tag{10}$$

IV. THE SPECIAL CASE OF THE PARETO DISTRIBUTION

In a lot of cases the claim distribution $s(x)$ may be represented by the Pareto distribution

$$s(x) = (\alpha - 1)\, x^{-\alpha} \qquad \text{and}$$
$$S(x) = 1 - x^{1-\alpha} \tag{11}$$

For small losses the approximation of the real frequencies by formula (11) is sometimes rather poor; for large amounts the fit is satisfactory, sometimes even very good. Moreover Benktander and Segerdahl [2] have proved that the Pareto distribution (11) is in a certain way the most stringent and cautious assumption for a claim distribution. A special analysis of the present problem when applying the Pareto distribution (11) might therefore be of interest.

The moments of the Pareto distribution (11) do not exist in general if the domain of definition is not truncated in an appropriate manner so that e.g. $A \leq x \leq M$. It is advisable to choose $A = 1$. If the parameter $\alpha > 3$ then the first and second moment exist even if there is no upper limit for the claim amounts, i.e. if M tends toward infinity. In such a case the moments for $\alpha > 3$ are:

$$S_0 = \int_1^\infty (\alpha - 1)\, x^{-\alpha}\, dx = 1$$

$$S_1 = \int_1^\infty (\alpha - 1)\, x^{1-\alpha}\, dx = \frac{\alpha - 1}{\alpha - 2} \tag{12}$$

$$S_2 = \int_1^\infty (\alpha - 1)\, x^{2-\alpha}\, dx = \frac{\alpha - 1}{\alpha - 3}.$$

If e.g. α is assumed to be 3,25
the mean is $S_1 = 1{,}8$
and the standard deviation $\sigma = 2{,}4$.

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

For the moments of the Pareto distribution truncated at the point m the following formulas hold true:

$$S_{m,1} = \frac{\alpha - 1}{\alpha - 2}(1 - m^{2-\alpha}) \qquad (12_1)$$

$$S_{m,2} = \frac{\alpha - 1}{\alpha - 3}(1 - m^{3-\alpha}). \qquad (12_2)$$

According to formula (5) the first moment of the distribution $\Phi\,(m)$ of the largest claim m is in the special case of a Pareto distribution defined as

$$M_1 = \int_1^\infty t\, m\, s(m)\, e^{-t[1-S(m)]}\, dm$$

$$= \int_1^\infty t\,(\alpha - 1)\, m^{1-\alpha}\, e^{-t\, m^{1-\alpha}}\, dm.$$

Substituting

$$u = t\, m^{1-\alpha}$$

we have for the mean value of the largest claim:

$$M_1 = t^{\frac{1}{\alpha-1}} \int_0^t e^{-u}\, u^{\frac{1}{1-\alpha}} du = t^{\frac{1}{\alpha-1}} \int_0^t e^{-u}\, u^{\frac{\alpha-2}{\alpha-1}-1}\, du = t^{\frac{1}{\alpha-1}}\, \Gamma_t\!\left(\frac{\alpha-2}{\alpha-1}\right) \qquad (13)$$

where $\Gamma_t(u)$ defines the incomplete Γ-function. Even for small values of t and, of course, for large t the incomplete Γ-function can be replaced by the complete Γ-function without noticeable lack of accuracy.

The first moment can be denoted by

$$M_1 \cong t^{\frac{1}{\alpha-1}}\, \Gamma\!\left(\frac{\alpha - 2}{\alpha - 1}\right). \qquad (13')$$

An analogous derivation gives, for the second moment, the following result:

$$M_2 \cong t^{\frac{2}{\alpha-1}} \Gamma\left(\frac{\alpha-3}{\alpha-1}\right). \tag{14}$$

It is remarkable that the gamma function appears in the formulas (13) and (14) with the reciprocal of the two first moments of the claim distribution $s(x)$ as argument.

It must be stressed that the mean value M_1 of the largest claim m increases less than linearly with the expected number of claims t. For $\alpha = 3$ e.g., the factor $t^{\frac{1}{\alpha-1}} = t^{\frac{1}{2}}$, i.e. the mean value M_1 increases proportionally to the square root of t. This means that the mean value M_1 is of the same magnitude as the standard deviation of the total loss which is defined by

$$\sigma = \sqrt{t\, S_2}.$$

For $\alpha = 3$ the second moment S_2 of $s(x)$ however does not exist; in such cases M_1 can, in a certain way, be considered as a convenient measure of dispersion for the total loss x.

The moments of the distribution of the total loss, when the largest claim is excluded, are given by equations (9) and (10). If we assume for $s(x)$ the Pareto distribution, the following expression may be determined for the mean:

$$\mu_1^{(-1)} = \int_1^\infty t^2 \frac{(\alpha-1)}{\alpha-2} m^{-\alpha} (1 - m^{1-\alpha}) e^{-tm^{1-\alpha}}\, dm =$$

$$= t^2 \frac{(\alpha-1)^2}{\alpha-2} \left[\int_1^\infty e^{-tm^{1-\alpha}} m^{-\alpha}\, dm - \int_1^\infty e^{-tm^{1-\alpha}} m^{2-2\alpha}\, dm\right].$$

With the substitution

$$u = t\, m^{1-\alpha}$$

the following equation is gained after some reductions

$$\mu_1^{(-1)} = \frac{t\,(\alpha-1)}{\alpha-2} (1 - e^{-t}) - \frac{\alpha-1}{\alpha-2} t^{\frac{1}{\alpha-1}} \Gamma_t\left(\frac{3-2\alpha}{1-\alpha}\right).$$

If the relation

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

$$\frac{\alpha-1}{\alpha-2}\Gamma\left(\frac{3-2\alpha}{1-\alpha}\right)=\Gamma\left(\frac{\alpha-2}{\alpha-1}\right)$$

is used, e^{-t} neglected and the incomplete replaced by the complete Γ-function the following expression is finally obtained:

$$\mu_1^{(-1)} \cong \frac{t(\alpha-1)}{\alpha-2} - t^{\frac{1}{\alpha-1}}\Gamma\left(\frac{\alpha-2}{\alpha-1}\right)^{*)}. \tag{15}$$

This formula ist plausible; the first term on the right side is equal to the mean value of the total loss and the second term is equal to the mean value of the largest claim according to formula (13).

In an analogous manner the expression for the second moment $\mu_2^{(-1)}$ can be obtained from formula (10). The final expression is a little more complicated:

$$\begin{aligned}\mu_2^{(-1)} = {} & \left(\frac{t(\alpha-1)}{\alpha-2}\right)^2 + \frac{t(\alpha-1)}{\alpha-3} \\ & - \Gamma\left(\frac{2\alpha-3}{\alpha-1}\right)\left[2t^{\frac{2}{\alpha-1}}\left(\frac{\alpha-1}{\alpha-2}\right)^2\right] \\ & + \Gamma\left(\frac{3\alpha-5}{\alpha-1}\right)\left[t^{\frac{2}{\alpha-1}}\left(\frac{\alpha-1}{\alpha-2}\right)^2\right] \\ & - \Gamma\left(\frac{2\alpha-4}{\alpha-1}\right)\left[t^{\frac{2}{\alpha-1}}\left(\frac{\alpha-1}{\alpha-3}\right)\right].\end{aligned} \tag{16}$$

In formula (16) the first line corresponds to the second moment of the distribution of the total loss. The other terms are, however, not equal to the second moment of the distribution of the largest claim [see formula (14)] but have a more complicated structure. This is due to the correlation between the distribution of the largest claim and the distribution of the total loss after excluding the largest claim.

It is well known that the moment of second order does not exist for a Pareto distribution with $\alpha \leqslant 3$; for $\alpha \leqslant 2$ not even the first moment exists. These properties are obvious if the moment formulas (12) are considered where the denominators $\alpha - 3$ and

*) It may even be shown that this formula is exact, if Γ is replaced by Γ_t.

$\alpha - 2$ vanish for the critical values. It is most remarkable that the existence domain for the moments of the distribution of the total loss after excluding the largest individual claim, i.e. for the moments $\mu^{(-1)}$ is broader than for the distribution of the total loss without excluding the largest claim. It may be seen that the moment $\mu_1^{(-1)}$ exists for $\alpha > 1{,}5$, whereas the moment μ_1 remains finite for $\alpha > 2$ only. For the second moment $\mu_2^{(-1)}$ finiteness is attained for $\alpha > 2$, whereas μ_2 exists only for $\alpha > 3$. This remarkable peculiarity may be seen from the following two tables where the first two moments of the three different distributions are considered for various parameters α. To simplify matters, the expected number of claims t was assumed to be 100.

Table 1

Comparison of the mean values

μ_1 of the total loss
M_1 of the largest claim
$\mu_1^{(-1)}$ of the total loss excluding the largest claim

α	μ_1 (1)	M_1 (2)	$\mu_1^{(-1)}$ (3)	decrease in % compared with (1)
1.5	∞	∞	∞	
1.75	∞	∞	585.41	
2.0	∞	∞	518.24	
2.25	500.00	182.77	317.23	36.6
2.5	300.00	57.72	242.28	19.2
2.75	233.33	28.73	204.60	12.3
3.0	200.00	17.72	182.28	8.9
3.25	180.00	12.39	167.61	6.9
3.5	166.67	9.40	157.27	5.6
3.75	157.14	7.53	149.61	4.8
4.0	150.00	6.29	143.71	4.2
5.0	133.33	3.87	129.46	2.9
10.0	112.50	1.80	110.70	1.6
∞	100.00	1.00	99.00	1.0

By excluding the largest claim, the mean is decreased to a remarkable extent for small values of the parameter α. This influence is not as big for larger values of α, but it must be borne in mind that only small values of α are of concern for practical applications.

Table 2

Comparison of the standard deviations

σ of the total loss
σ_M of the largest claim
$\sigma^{(-1)}$ of the total loss excluding the largest claim

α	σ (1)	σ_M (2)	$\sigma^{(-1)}$ (3)	decrease in % compared with (1)
1.5	∞	∞	∞	
1.75	∞	∞	∞	
2.0	∞	∞	∞	
2.25	∞	∞	86.28	
2.5	∞	∞	45.01	
2.75	∞	∞	31.22	
3.0	∞	∞	24.54	
3.25	30.00	18.90	20.92	30.3
3.5	22.36	9.72	18.77	16.1
3.75	19.15	6.12	17.22	10.1
4.0	17.32	4.26	16.19	6.5
5.0	14.14	1.66	13.78	2.5
10.0	11.34	0.27	11.32	0.2
∞	10.00	0	10.00	0

For the standard deviation the same may be said as for the mean. The decrease is considerable, especially for $\alpha < 4$.

* * *

The present paper is restricted to a few first examinations. This introduction will be continued in another paper in which not only will the properties of the critical values of the parameter α at the points $\alpha = 3$ and $\alpha = 2$ be discussed, but in which the investigations will be extended to the case where more than one largest claim is excluded.

LIST OF REFERENCES

[1] BEARD, Robert Eric: Some Notes on the Statistical Theory of Extreme Values. The ASTIN Bulletin, Vol. III, Part I, 1963.

[2] BENKTANDER, Gunnar, und SEGERDAHL, Carl-Otto: On the Analytical Representation of Claim Distributions with Special Reference to Excess

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press

of Loss Reinsurance. Transactions of the XVIth International Congress of Actuaries, Brussels 1960.

[3] FRANCKX, Edouard: Sur la fonction de distribution du sinistre le plus élevé. The ASTIN Bulletin, Vol. II, Part III, 1963.

[4] GUMBEL, E. J.: Statistics of Extremes. Columbia University Press, New York 1958.

[5] THÉPAUT, André: Le traité d'excédent du coût moyen relatif (Ecomor). Bulletin Trimestriel de l'Institut des Actuaires Français, No 192, Septembre 1950.

https://doi.org/10.1017/S0515036100010424 Published online by Cambridge University Press