DISCUSSION BY GARY VENTER

Background

Aggregate losses are easily defined as the sum of individual claims, but the distribution of aggregate losses has not been easy to calculate. In fact, this has been a central, and perhaps *the* central, problem of collective risk theory. The mean of the aggregate loss distribution can be calculated as the product of the means of the underlying frequency and severity distributions; similarly, there are well known formulas for the higher moments of the aggregate distribution in terms of the corresponding frequency and severity moments (e.g., see [5] Appendix C). However the aggregate distribution function, and thus the all important excess pure premium ratio, has been awkward to calculate from the distribution functions of frequency and severity. It is this calculation problem that is addressed and solved in this important paper. The result is generalized somewhat to the case where the severity distribution is known only up to a scale multiplicative factor, which itself follows a specific distribution (inverse gamma). In this review the approach in the paper is abstracted somewhat in an attempt to focus on the areas where the specific assumptions come into play.

Principal Idea

The derivation of the results involves complex mathematics, but the results themselves and the ideas behind the derivation can be easily understood. It is not necessary to know what a characteristic function or a convolution or a complex number is to understand these basic ideas and to use the results. The following properties of the characteristic function are germane to this understanding.

1) It is a transformation of the distribution function.

2) It has an inverse transformation; i.e., the distribution function can be calculated from the characteristic function.

3) The characteristic function of aggregate losses can be calculated from the moment generating function of frequency and the characteristic function of severity.

The basic idea, then, is to calculate the characteristic function of severity and the moment generating function of frequency; use them to compute the characteristic function of aggregate losses; and, use that to calculate the distribution function of aggregate losses and the excess pure premium ratios.

None of the above is actually new to risk theory or even to North American casualty actuaries. What is new and is the heart of this paper's contribution centers around a snag in the above method: the characteristic function of severity is not directly calculable from the distribution function in most cases. The gamma severity is an exception and Mong presented its use to the CAS in this context in the 1980 call paper program. The authors point out that the characteristic function is also calculable when severity is piecewise linear, and the solution they present is for this case. They then assert that any severity distribution needed in property-casualty practice can be closely approximated by a piecewise linear form, which seems reasonable, and thus that this method is completely general. This summarizes the basic ideas of the derivation.

Results

The results can be expressed fairly simply without reference to complex numbers. The formulas below are essentially those derived in the paper, although generalized slightly in that they hold for any severity random variable S, not just one that is piecewise linear, and for binomial or negative binomial frequency with parameters c and λ, defined below. Mong's paper and others have also presented very similar general formulas. As usual, E denotes the expected value.

Result 1: F, the aggregate distribution function, can be expressed as

$$F(x) = \tfrac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin\,(g(t) + tx)\; dt}{t\, f(t)}$$

where $f(t) = [(1 + c\lambda - c\lambda E(\cos tS))^2 + (c\lambda E(\sin tS))^2][1/2c]$

$g(t) = (-1/c) \arctan\,[E(\sin tS)/((1/c\;\lambda) - E(\cos tS))]$.

Result 2: The expected losses excess of retention x, $EP(x)$, can be calculated as

$$EP(x) = \mu - (x/2) + (1/\pi) \int_0^\infty (1/f(t)t^2)\,(\cos\,(g(t)) - \cos\,(tx + g(t)))dt$$

where μ is the expected aggregate losses.

This is a very nice formula in that 1) aggregate excess losses can be computed without computing the loss probabilities; 2) the integral converges well before infinity because of the t^2 term; and 3) its error structure can be analyzed.

Following Mong, the authors transform the integrals by a change of variables $t \rightarrow t/\sigma$. It is not clear that this is necessary or even useful.

Note that the authors use the negative binomial in the form

$$\Pr(Y = y) = \binom{y + 1/c - 1}{y} (1 + c\lambda)^{-1/c} \left(\frac{c\lambda}{1 + c\lambda}\right)^y.$$

This has mean λ and ratio of the variance to the mean of $1 + c\lambda$. Taking $p = 1/(1 + c\lambda)$ and $\alpha = 1/c$ gives the more usual form

$$\Pr(Y = y) = \binom{\alpha + y - 1}{y} p^\alpha (1 - p)^y.$$

Formulas for $E(\cos tS)$ and $E(\sin tS)$ (denoted by the authors as $\tilde{h}(t)$ and $\tilde{k}(t)$ respectively) for piecewise linear S are found as formulas 5.12 and 5.13 of the paper. This is where the piecewise linear assumption is used. Mong's results can be obtained by substituting the corresponding formulas for the gamma severity, namely $E(\cos tS) = (\cos(r \arctan t/a))/(1 + t^2/a^2)^{r/2}$ and $E(\sin tS) = (\sin(r \arctan t/a))/(1 + t^2/a^2)^{r/2}$, where r and a are gamma parameters defined by $E(S) = r/a$ and $\mathrm{Var}(S) = r/a^2$.

It would also be possible to evaluate $E(\cos tS)$ and $E(\sin tS)$ for a discrete severity distribution function S and apply the above formula. Another possibility, which might turn out to be a useful alternative, would be to approximate the severity probability density function by a piecewise linear form, rather than doing so for the cumulative distribution function.

To develop the formulas for the needed trigonometric expectations in this case, suppose the severity density $g(s)$ between two points a_i and a_{i+1} is given by $g(s) = c_i + sd_i$, and there is a probability p of a claim of the largest size a_{n+1}. Then the following formulas can be readily derived using integration by parts.

$$E(\cos tS) = \frac{1}{t} \sum_{i=1}^{n} \left((c_i + sd_i) \sin ts + (d_i/t) \cos ts\right) \Big|_{a_i}^{a_{i+1}} + p \cos ta_{n+1}$$

$$E(\sin tS) = \frac{1}{t} \sum_{i=1}^{n} \left((c_i + sd_i) \cos ts - (d_i/t) \sin ts\right) \Big|_{a_{i+1}}^{a_i} + p \sin ta_{n+1}$$

Note also that for the probabilities to total 1.0,

$$p = 1 - \sum_{i=1}^{n} (a_{i+1} - a_i)(c_i + d_i(a_{i+1} + a_i)/2).$$

For discrete severity distributions, $E(\cos tS)$ and $E(\sin tS)$ can also be directly calculated. For most severity distributions, these expected values can be calculated numerically. In fact, approximating the severity distribution by a piecewise linear function can be regarded as a numerical approximation of $E(\cos tS)$ and $E(\sin tS)$. Other approximation methods are also possible. As this is the only use made of the piecewise linear severity assumption, it can be seen that this assumption is not an essential constraint of the method but rather a convenient numerical device.

In other words, the above formulas for $F(x)$ and $EP(x)$ hold for any severity distribution, S, not just piecewise linear. Since $E(\sin tS)$ and $E(\cos tS)$ need to be calculated for many t's in order to evaluate the integrals, a method is needed to calculate these trigonometric expectations. Any number of numerical integration techniques could be used for the purpose. The point of view of this paper is that approximating the density function of S by a step function provides a simple method for the calculation of $E(\sin tS)$ and $E(\cos tS)$ which is of sufficient accuracy for the end results.

Subsequent discussion with the authors uncovered that this has been supported by further empirical tests which began by approximating a smooth density (e.g. Weibull) by a step function, calculating $F(x)$ and $EP(x)$, and then refining the approximation. It was found that 20 to 25 approximating intervals provided a high degree of accuracy in this process. Thus the characteristic function method can be applied readily to any severity distribution.

Although the formulas above use functions that have not been commonly employed in casualty actuarial practice, their calculation is straightforward. The integrands themselves can be computed on many hand calculators. Carrying out the integration requires numerical methods. The authors adopt a brute force approach, and it gets the job done. More efficient methods may be possible, but a fair amount of expertise in numerical integration would be needed to determine if this were so.

Details of the Method

The formula for the characteristic function of aggregate losses in terms of the frequency moment generating function and the severity characteristic function is $\phi_x(t) = M_n(\ln \phi_s(t))$. This is readily derived from formula 5.11 of the paper. Formulas 5.14 to 5.16 follow directly from this result and the formulas for the moment generating functions of the binomial, Poisson, and negative binomial distributions. In fact, the proofs of those formulas given are essentially derivations of the corresponding moment generating functions.

The derivations of the above general formulas for $F(x)$ and $EP(x)$ are straightforward applications of the inversion formula to the aggregate characteristic function. The inversion formula is the standard procedure for getting the distribution function from the characteristic function and can be found in advanced statistical texts.

Also, the issue of discontinuities in the distribution function deserves further attention. This inversion formula for calculating the distribution function from the characteristic function is not exact at points of discontinuity. This is easy to miss in Kendall and Stuart, which is cited as the source of the inversion formula. Because this has not been taken into account, the above formula for $F(x)$ as well as the paper's formula are incorrect at the discontinuity points. The error is an understatement of the distribution function equal to one half of the jump at those points. This would be an important issue, for example, if a discrete severity were used with the formulas above. In that case the aggregate distribution would also be discrete, and thus its distribution function would be a step function. To evaluate this function at a discontinuity point, then, it would suffice to evaluate it just above the discontinuity, in fact at any point before the next discontinuity.

These errors can also be computed from the underlying distributions. In the case the authors treat most often, namely a severity distribution with a censorship point (e.g., per occurrence limit), the aggregate distribution function is discontinuous, with jumps at n times the censorship point ($n = 0,1,2,\ldots$) equal to the probability of having exactly n claims all of which are total losses (i.e., equal the censorship point). These probabilities can be computed from the frequency and severity distribution function and then the aggregate can be adjusted by half the jump at those points. As an alternative, evaluating at slightly above the discontinuity should give a reasonable approximation. The example in Table 9.2 of the paper illustrates this at $x = 1.00$, where the error is 25%.

In examples given in Exhibits II–VIII, these adjustments would probably not be significant. If, however, the expected number of claims is small (e.g., 5,1,.02) and/or the probability at the censorship point is large, the error at the discontinuity may be significant. In excess insurance/reinsurance applications both these conditions often hold. However, as discussed below under recursive computation, the characteristic function method may not be the most efficient in such applications in any case.

Parameter Uncertainty

The parameter uncertainty issue is an important one and is well considered in the paper. For large individual risks or for insurance companies, this uncertainty can far outweigh the variation that can occur from randomness within known frequency and severity distributions. For example, parameter uncertainty can arise from severity trend and development. Although these may also affect the shape of the severity distribution, they have definite effects on its scale. The authors treat the situation in which the severity distribution is known up to a scale multiplier which is itself inverse gamma distributed. (Actually, they present this as a divisor which is gamma distributed.) The gamma is selected because it leads to tractable results. Note that applying a scale multiplier to severity is equivalent to applying the same multiplier to aggregate losses. This is not true for frequency, as increasing the number of claims changes the shape of the aggregate distribution. This is reflected in the standard formulas for the coefficients of variation and skewness of aggregate losses (e.g., [5], Appendix C).

The derivation in Appendix A of the paper shows that the gamma assumption for a scale is not absolutely required. What is required is a method of calculating the characteristic function of this divisor. This characteristic function can then be plugged into the formulas A1 and A2 to yield expressions for the aggregate distribution function and the excess pure premium, respectively. In fact, the derivations labelled "case 1" and "case 2" do exactly that for the degenerate and gamma divisors, respectively.

Estimating the parameters for the mixing distribution is a problem. The mean can be selected to give the proper severity mean. The variance is more difficult to arrive at. A study of historical errors in trend and development projections could be useful in this regard. The variance of accident year or policy year loss ratios for a large segment of the industry, where process variance can be assumed minimal, should also be a viable approach. The authors seem to suggest comparing the observed variance in loss ratios with the theoretical variance that would occur without parameter risk in order to estimate the degree of parameter risk. This also seems to be a potentially useful approach.

The inverse gamma distribution, i.e., the distribution of X where $1/X$ is gamma distributed, has density $f(x) = \beta\, e^{-1/x\beta} \div \Gamma(r)\, (\beta x)^{1+r}$. This is a fairly dangerous probability distribution, more so than the gamma, in that only finitely many moments exist. In fact $E(X^n) = \Gamma(r - n) \div \beta^n \Gamma(r)$ exists if and only if $n < r$. It is an open question whether or not this will prove appropriate for a mixing distribution.

Besides trend and development factors, parameter uncertainty also arises from risk classification. For computing the aggregate loss distribution of a large and diverse portfolio of risks, this may not be an important factor. However, for a single risk or a carrier specializing in a few classes, this could be an essential consideration. If the risk is not typical of the classification or the class rate is based on insufficient data, the dispersion of possible results will be greater than frequency and severity considerations might suggest. Historical errors in trend and development will also understate the parameter risk in this case.

The parameter uncertainty approach discussed by Bühlmann [2] and developed further by Patrik and John [4] can also be used with the characteristic function method. Bühlmann allowed all parameters of the distributions to have uncertainty and introduced a probability function, called the structure function, to describe the relative weights given to different parameter sets. If the structure function is approximated by a finite number of points, the distribution function of aggregate losses can be calculated for each parameter set by the authors' method and then weighted together by the structure function. This gives a quite general method of dealing with parameter uncertainty.

Recursive Computation of Aggregate Functions

Another method of computing the aggregate distribution function was recently developed by Panjer [3] generalizing Adelson [1]. It is interesting to compare this to the current paper.

Panjer's method involves a recursive formula for $F(x)$ based on discrete severity distributions. For his formula the severity probability function must be given at every multiple of some unit value up to the largest possible loss size, for example $g(1) = .5$, $g(2) = .3$, $g(3) = .1$, $g(4) = .05$, $g(5) = .05$, where g is the severity probability function, 10,000 is the unit, and 50,000 is thus the largest possible loss. In this case the aggregate losses will also come in multiples of the unit. If we now let f denote the aggregate probability, Panjer's formula is

$$f(x) = \sum_{i=1}^{x} (a + b\, i/x)\, g(i)\, f(x - i),$$

where a and b come from the frequency distribution. This formula is valid for binomial, negative binomial, and Poisson frequencies. For the negative binomial

$$\Pr(Y = y) = \binom{\alpha + y - 1}{y} p^{\alpha} (1 - p)^{y},$$

$a = 1 - p$ and $b = (\alpha - 1)(1 - p)$. For the Poisson $a = 0$, $b = \lambda$, and for the binomial $\Pr(Y = y) = \binom{m}{y} P^{y} (1 - p)^{m-y}$, $a = p/p - 1$ and $b = (m + 1)p/1 - p$.

As an example take the above g in units of 10,000 with Poisson $\lambda = 1$. Then $f(x) = \Sigma_{i=1}^{x} i\, g(i)\, f(x - 1)/x$.

Now $f(0) = \Pr\,(N = 0) = e^{-1}$. Thus $f(1) = .5e^{-1}$, $f(2) = .5\, f(1)/2 + .3\, f(0) = .425e^{-1}$, $f(3) = .5\, f(2)/3 + .2\, f(1) + .1\, f(0) = .8125\, e^{-1}/3$, etc.

Thus the aggregate distribution function can be built up by quite simple arithmetic operations using this method.

The excess pure premium can be derived from the aggregate probabilities. The definition in discrete terms is $EP(x) = \Sigma_{i=x}^{\infty} (i = x)\, f(i)$. Calculating this requires $f(i)$ for the largest possible i's whereas the recursive procedure builds up from the smallest. But since $\mu = \Sigma_{i=0}^{\infty} i\, f(i)$ is known from frequency and severity, if it were possible to calculate $\mu - EP(x)$ then $EP(x)$ would fall out.

$$\text{Now } \mu - EP(x = \sum_{i=0}^{\infty} i\, f(i) - \sum_{i=x}^{\infty} i\, f(i) + x \sum_{i=x}^{\infty} f(i)$$

$$= \sum_{i=0}^{x-1} i\, f(i) + x(1 - \sum_{i=0}^{x-1} f(i)).$$

$$\text{Thus let } v(x) = \sum_{i=0}^{x-1} i\, f(i),\ v(0) = 0, \text{ and } w(x) = 1 - \sum_{i=0}^{x-1} f(i),\ w(0) = 1.$$

Then the excess pure premium can be calculated by

$$EP(x) = \mu - v(x) - x\, w(x)$$

where v and w can be calculated recursively by

$$v(x + 1) = v(x) + xf(x) \text{ and } w(x + 1) = w(x) - f(x).$$

By approximating the severity distribution with discrete probabilities the aggregate distribution and excess pure premium functions can thus be estimated recursively. Exhibits 1 and 2 compare this with the characteristic function method. Exhibit 1 shows the piecewise linear severity assumed and the approximating discrete probabilities. A unit of 500 was taken. The largest possible claim is taken as 250,000. The discrete approximation was constructed by matching cumulative probabilities and average severities at $250 + 500\, i$ points, to the extent possible.

Exhibit 2 shows the cumulative probabilities and excess ratios for the two methods. (The excess ratio at x is $EP(x) \div \mu$.) The excess ratio columns are practically identical, suggesting that very little is lost by the discrete approximation. The cumulative probabilities are also rather close. In fact, since the characteristic function method does not provide error estimates for cumulative probabilities, it is not clear which method is closer to the exact probabilities for the piecewise linear severity.

Although the recursive formulas are simpler than those of the characteristic function method, they do not always take less computation, especially when only one or two limits are to be evaluated. On a ground up coverage with a high occurrence limit, a large number of points would be needed to approximate the severity distribution because a small unit would be needed to represent small claims. If, in addition, there are a large number of expected claims, the recursive method can be time consuming. If, on the other hand, an aggregate distribution is being estimated for an excess occurrence layer where there are few expected claims and a large unit can be chosen, this method may be quite efficient.

The recursive method does not provide a mathematically elegant way of accounting for the crucial element of parameter risk. However, this can be handled by enumerating a list of possible scenarios (frequency and severity functions), calculating the aggregate distribution function for each scenario, and then weighting these aggregate functions together by the relative probability attached to each scenario. As discussed above, this is more general than a gamma distributed divisor approach in that it allows for more types of parameter variation.

In conclusion, the authors have produced a practical, efficient method for calculating aggregate probabilities and excess pure premiums. This is not an obscure exercise in complex mathematics but a powerful competitive tool for those who will use it.

Ackowledgment

The reviewer must acknowledge the very fine assistance provided by Farrokh Guiahi and Linh Nguyen in unravelling the complex mathematics of characteristic functions.

REFERENCES

[1] R. M. Adelson, "Compound Poisson Distributions," Operations Research Quarterly, 17, 73–75, 1966.

[2] H. Bühlmann, *Mathematical Methods in Risk Theory,* Springer-Verlag, 1970.

[3] H. H. Panjer, "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin,* Vol. 12, No. 1, 1981.

[4] G. S. Patrik and R. T. John, "Pricing Excess-of-Loss Casualty Working Cover Reinsurance Treaties," *Pricing Property and Casualty Insurance Products,* Casualty Actuarial Society Discussion Paper Program, 1980.

[5] G. G. Venter, "Transformed Beta and Gamma Distributions and Aggregate Losses," *Pricing, Underwriting and Managing the Large Risk,* Casualty Actuarial Society Discussion Paper Program, 1982.

EXHIBIT 1

AGGREGATE LOSS DISTRIBUTIONS
COMPARATIVE ASSUMPTIONS

Frequency: Poisson λ = 13.7376

Piecewise Linear CDF

Limit (000):	1	5	6	7	8	9
Cumulative Prob.:	.38935	.77870	.78438	.78981	.79498	.79993

10	12.5	15	17.5	20	25	35	50
.80466	.81564	.82553	.83449	.84264	.85690	.87927	.90280

75	100	125	150	175	200	225	250
.92739	.94256	.95277	.96009	.96556	.96979	.97316	.97590

Discrete PDF

Amount:	500	1000	1500 to 4000
Probability:	.38326640625	.03041796875	.04866875 each 500

4500	5000	5500 to 249,000 at each N = 500k
.054731628	.019691497	Piecewise linear probability from N − 250 to N + 250

249,500	250,000
.0000685	.0241137

Moments

	Mean	Coefficient of Variation	Coefficient of Skewness
Severity	18,198	2.6600	3.6746
Aggregate	250,000	.7667	1.0744

EXHIBIT 2

Aggregate Loss Distributions
Comparative Summary

Aggregate Loss (000)	Characteristic Function Method		Recursive Method	
	Cum. Prob.	Excess Ratio	Cum. Prob.	Excess Ratio
25	.0508	.9016	.0516	.9016
50	.1291	.8107	.1298	.8107
75	.2009	.7273	.2015	.7272
100	.2676	.6507	.2683	.6507
125	.3289	.5806	.3295	.5806
150	.3843	.5163	.3848	.5163
175	.4341	.4573	.4346	.4573
200	.4788	.4030	.4793	.4029
225	.5189	.3529	.5193	.3529
250	.5548	.3066	.5552	.3066
275	.6034	.2642	.6040	.2642
300	.6556	.2273	.6561	.2273
325	.7008	.1951	.7013	.1951
350	.7405	.1672	.7408	.1672
375	.7749	.1431	.7752	.1431
400	.8047	.1221	.8049	.1221
425	.8303	.1039	.8305	.1039
450	.8524	.0880	.8526	.0880
475	.8714	.0742	.8716	.0742
500	.8878	.0622	.8879	.0622
525	.9045	.0518	.9047	.0518
550	.9201	.0430	.9203	.0430
575	.9332	.0357	.9333	.0357
600	.9442	.0296	.9443	.0296
625	.9534	.0245	.9535	.0245
650	.9611	.0202	.9611	.0202
675	.9675	.0167	.9675	.0167
700	.9728	.0137	.9729	.0137
725	.9773	.0112	.9773	.0112
750	.9810	.0091	.9810	.0091
775	.9844	.0074	.9844	.0074
800	.9873	.0060	.9873	.0060
825	.9897	.0048	.9897	.0048
850	.9916	.0039	.9916	.0039