FROM AGGREGATE CLAIMS DISTRIBUTION TO PROBABILITY OF RUIN

HILARY L. SEAL,
Ecole Polytechnique Fédérale de Lausanne

INTRODUCTION

When the distribution of the number of claims in an interval of time of length t is mixed Poisson and the moments of the independent distribution of individual claim amounts are known, the moments of the distribution of aggregate claims through epoch t can be calculated (O. Lundberg, 1940, ch. VI). Several approximations to the corresponding distribution function, $F(\cdot, t)$, are available (see, e.g., Seal, 1969, ch. 2) and, in particular, a simple gamma (Pearson Type III) based on the first three moments has proved definitely superior to the widely accepted "Normal Power" approximation (Seal, 1976). Briefly,

$$F(t + z\sqrt{\varkappa_2}, t) \approx \frac{1}{\Gamma(\alpha)} \int_0^{\alpha + z\sqrt{\alpha}} e^{-y} y^{\alpha - 1} \equiv P(\alpha, \alpha + z\sqrt{\alpha}) \qquad (1)$$

where the P-notation for the incomplete gamma ratio is now standard and α, a function of t, is to be found from

$$\alpha = \frac{4}{\varkappa_3^2/\varkappa_2^3} = \frac{4}{\gamma_1^2}$$

the kappas being the cumulants of $F(\cdot, t)$. An excellent table of the incomplete gamma ratio is that of Khamis (1965).

The problem that is solved in this paper is the production of an approximation to $U(w, t)$, the probability of non-ruin in an interval of time of length t, by using the above mentioned gamma approximation to $F(\cdot, t)$.

THE PROBABILITY OF NON-RUIN IN A PERIOD OF LENGTH T

In Seal (1974) it was shown that when the distribution of the number of claims in an arbitrary interval of time is generated by a stationary point process the probability of non-ruin in an interval which the insurance company enters with a risk-reserve of w and operates throughout with a risk-premium loading of η, is $U(w, t)$ given by

$$U(w,t) = F(w + \pi_1 t, t) - \pi_1 \int_0^t U(0,\tau)\, f(w + \overline{\pi_1 t - \tau, t - \tau})\, d\tau \qquad (2)$$

where π_1 is the risk-loaded pure premium rate and $f(\cdot, t)$ is the density corresponding to $F(\cdot, t)$. This is the formula which we will use for our numerical approximations.

The only stationary point processes that have been utilized by actuaries in practical applications are those that lead to ordinary or mixed Poisson distributions (O. Lundberg, *l.c.*) and in these circumstances the Prabhu-Benes-Takács formula (Seal, 1974)

$$U(0, t) = \frac{1}{\pi_1 t} \int_0^{\pi_1 t} F(y, t)\, dy \tag{3}$$

may be used to produce the first factor in the integrand of (2).

Application of Relation (1)

Considering (1) as applied to (2) we note that if the distribution of the number of claims is Poisson with mean t and the density of individual claim amounts, $b(\cdot)$, has mean

$$\mu = 1 \text{ so that } \pi_1 = 1 + \eta,\; F(w + \overline{1 + \eta} \cdot t, t) \approx P(\alpha, \alpha + z\sqrt{\alpha})$$

where

$$\alpha = \alpha(t) = \frac{4\,\varkappa_2^3}{\varkappa_3^2} = \frac{4(tp_2)^3}{(tp_3)^2} = \frac{4tp_2^3}{p_3^2}$$

p_2 and p_3 being the second and third moments about zero of the $b(\cdot)$-distribution of individual claim amounts (Seal, 1969, 2.41). In order to evaluate z we have

$$t + z\sqrt{\varkappa_2} = t + z\sqrt{(tp_2)} = w + (1 + \eta)t$$

so that

$$z = (w + \eta t)\,(tp_2)^{-1/2}$$

Further, by differentiation of (1) with respect to z,

$$f(\tau + z\sqrt{\varkappa_2}, \tau) \approx \frac{\beta}{\Gamma(\alpha)} \exp\left[-\alpha - z\sqrt{\alpha}\right] (\alpha + z\sqrt{\alpha})^{\alpha - 1} \tag{4}$$

where

$$\alpha = \alpha(\tau) = \frac{4\tau p_2^3}{p_3^2} \qquad \beta = \sqrt{(\alpha/\varkappa_2)}$$

and, when $\tau + z\sqrt{\varkappa_2} = w + (1 + \eta)\,\tau$,

$$z = (w + \eta\,\tau)\,(\tau\, p_2)^{-1/2} \quad 0 < \tau < t$$

Finally, by (3),

$$U(0, \tau) = \frac{1}{(1 + \eta)\tau} \int_0^{(1+\eta)\tau} F(y, \tau)\, dy$$

$$= \frac{\sqrt{\kappa_2}}{(1 + \eta)\tau} \int_{-\tau/\sqrt{\kappa_2}}^{\eta\tau/\sqrt{\kappa_2}} F(\tau + z\sqrt{\kappa_2}, \tau)\, dz$$

$$\approx \frac{\sqrt{\kappa_2}}{(1 + \eta)\tau} \int_{-\tau/\sqrt{\kappa_2}}^{\eta\tau/\sqrt{\kappa_2}} P(\alpha, \alpha + z\sqrt{\alpha})\, dz \qquad \text{by (1)}$$

$$= \frac{\sqrt{(\kappa_2/\alpha)}}{(1 + \eta)\tau} \int_{\alpha - \tau\sqrt{(\alpha/\kappa_2)}}^{\alpha + \eta\tau\sqrt{(\alpha/\kappa_2)}} P(\alpha, u)\, du$$

$$= \frac{1}{(1 + \eta)\tau\beta} \int_{\alpha - \tau\beta}^{\alpha + \eta\tau\beta} \frac{du}{\Gamma(\alpha)} \int_0^u x^{\alpha-1} e^{-x} dx$$

$$= \frac{1}{(1 + \eta)\, \tau\beta\Gamma(\alpha)} \left\{ \int_0^{\alpha + \eta\tau\beta} - \int_0^{\alpha - \tau\beta} \right\} du \int_0^u x^{\alpha-1} e^{-x} dx$$

$$= \frac{1}{(1 + \eta)\, \tau\beta\Gamma(\alpha)} \Big[\int_0^{\alpha + \eta\tau\beta} (\alpha + \eta\tau\beta - x)\, x^{\alpha-1} e^{-x} dx +$$

$$- \int_0^{\alpha - \tau\beta} (\alpha - \tau\beta - x)\, x^{\alpha-1} e^{-x} dx \Big]$$

$$= \frac{1}{(1 + \eta)\tau\beta} [(\alpha + \eta\tau\beta) P(\alpha, \alpha + \eta\tau\beta) - \alpha\, P(\alpha + 1, \alpha + \eta\tau\beta)$$

$$- (\alpha - \tau\beta)\, P(\alpha, \alpha - \tau\beta) + \alpha\, P(\alpha + 1, \alpha - \tau\beta)] \qquad (5)$$

where

$$\beta = \sqrt{(\alpha/\kappa_2)} \quad \text{and} \quad \alpha = \alpha(\tau).$$

A remarkable feature of the approximation (1) is that only the first three moments of the distribution of individual claim amounts are involved. If, therefore, a two-parameter distribution is successfully fitted to the observational distribution of claim amounts by means of the mean and variance it implies that the appropriateness of the chosen functional form has been determined by the approximate equivalence of the third moments of the observational and theoretical distributions of individual claims. For example, if the gamma distribution (Johnson & Kotz, 1970, ch. 17) were fitted the third central moment (or cumulant) would necessarily be twice the variance.

Now only two functional forms for $b(\cdot)$, the density function of individual claim amounts, result in explicit results for $F(x, t)$ when the distribution of the number of claims in an interval of length t is Poisson with mean t (Seal, 1969, p. 31, referring to Hadwiger, 1942). These are the gamma and the inverse Gaussian distributions and it would be convenient to use one or other of these forms for $b(.)$ so that direct checks may be made of our numerical approximations using (1).

The Inverse Gaussian Distribution

According to Seal (1969, p. 30) by far the greatest number of graduations of observed individual claim amounts have been based on the lognormal distribution, namely where the logarithm of the claim amount (the latter possibly increased or decreased by some constant) has a Normal distribution.

On the other hand the inverse Gaussian density (Tweedie, 1957)

$$b(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right] \quad x > 0,\ \mu > 0,\ \lambda > 0 \tag{6}$$

which has the distribution function

$$B(x) = \Phi\left(\sqrt{\frac{\lambda}{x}} \cdot \overline{\frac{x}{\mu} - 1}\right) + e^{2\lambda/\mu}\left\{1 - \Phi\left(\sqrt{\frac{\lambda}{x}} \cdot \overline{\frac{x}{\mu} + 1}\right)\right\} \tag{7}$$

as shown by Shuster (1968) (but misprinted in Johnson & Kotz, 1970), where $\Phi(\cdot)$ is the standardized Normal distribution function, can be made to start at the same claim amount (which we take as the origin) as the lognormal and be given the same mean μ and variance μ^3/λ. Although the Inverse Gaussian has never been used to graduate a set of individual claim amounts it may produce nearly the same γ_1-value as that possessed by the corresponding lognormal distribution and would then lead to approximately the same distribution of aggregate claims as provided by (1).

When individual claims are distributed according to the inverse Gaussian,

$$f(x, t) = e^{-t} \sum_{k=1}^{\infty} \frac{t^k}{k!} \left(\frac{\Lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\Lambda(x-M)^2}{2M^2 x}\right] \tag{8}$$

where $\Lambda = k^2\lambda$ and $M = k\mu$, and from (6) and (7)

$$F(x, t) = e^{-t} + e^{-t} \sum_{k=1}^{\infty} \frac{t^k}{k!} \left[\Phi \left(\sqrt{\frac{\Lambda}{x}} \cdot \left(\frac{x}{M} - 1 \right) \right) + e^{2\Lambda/M} \left\{ 1 - \Phi \left(\sqrt{\frac{\Lambda}{x}} \cdot \left(\frac{x}{M} + 1 \right) \right) \right\} \right] \quad (9)$$

We mention that $\beta(s)$, the Laplace transform of (6), is given by

$$ln\ \beta(s) = \frac{\lambda}{\mu} \left\{ 1 - \left(1 + \frac{2\mu^2 s}{\lambda} \right)^{1/2} \right\} \quad (10)$$

Choice of Parameters

Upwards of 50 actual individual claim distributions have been fitted by the lognormal (Seal, 1969, p. 30). The γ_1-values for 45 of these were calculated *, using the formulas provided by Johnson & Kotz (1970) applicable to the constants of the linear transform, and compared with the corresponding γ_1's calculated for (6) using the calculated mean and variance. 60% of the γ_1 pairs were approximately equal implying that the lognormal and inverse Gaussian distributions would produce nearly the same value for (1). Among the 27 distributions was Cannella's (1963) costs of 124, 279 "specialty" pharmaceutical prescriptions in the province of Rome during 1960. The two γ_1's were ·355 and ·354, respectively, but the mean and variance of the distribution were stated to be 786.4 and 280582.09 after lognormal fitting. Unfortunately this mean and variance produce γ_1's of 2.326 and 2.021, respectively, for the lognormal and inverse Gaussian indicating that, in fact, the latter distribution is not in this case a very good approximation to the lognormal. This error of Cannella was not discovered until too late and we had already chosen $\mu = 1$ and $\lambda = (786.4)^2/280582.09 = 2.20408$ for the inverse Gaussian. In order to apply this to (1) we have (Tweedie, *loc. cit.*) $p_2 = \mu^2 + \mu^3\lambda^{-1} = 1.453704$ and $p_3 = \mu^3 + 3\mu^4\lambda^{-1} + 3\mu^5\lambda^{-2} = 2.978654$ so that $\alpha(t) = 1.384993\, t$.

Results

The following Table compares the results obtained for $f(10 + t, t)$ by (4) and (8) and for $F(10 + t, t)$ by (1) and (9). In the first set of comparisons the gamma approximation is only in error by a few units in the fifth decimal place. In the second set the gamma

* It is not always easy to decide whether an author is using natural or common logarithms for his transform.

approximation is never more in error than by two units in the fourth decimal place. These are very good results.

TABLE I

Values of $f(10 + t, t)$, $F(10 + t, t)$ and $U(10, t)$

t	$f(10+t, t)$		$F(10+t, t)$		$U(10, t)$	
	(4)	(8)	(1)	(9)	(1) to (5)	method of 1974 paper
1	.00004	.00003	.99996	.99997	.9999	1.0000
2	.00019	.00016	.99978	.99983	.9997	1.0000
3	.00049	.00045	.99937	.99945	.9991	.9993
4	.00095	.00090	.99866	.99879	.9980	.9981
5	.00154	.00150	.99764	.99779	.9964	.9964
6	.00226	.00222	.9963	.9965	.9943	.9943
7	.00305	.00303	.9947	.9948	.9916	.9915
8	.00390	.00390	.9927	.9929	.9884	.9883
9	.00479	.00479	.9906	.9908	.9847	.9846
10	.00569	.00570	.9882	.9884	.9807	.9804
11	.00658	.00661	.9857	.9858	.9762	.9759
12	.00747	.00750	.9830	.9831	.9715	.9711
13	.00833	.00837	.9801	.9803	.9665	.9660
14	.00916	.00921	.9772	.9774	.9613	.9607
15	.00997	.01002	.9742	.9744	.9559	.9552
16	.01074	.01079	.9711	.9713	.9503	.9495
17	.01147	.01153	.9680	.9682	.9447	.9438
18	.01217	.01223	.9649	.9651	.9369	.9380
19	.01283	.01289	.9618	.9619	.9331	.9321
20	.01346	.01352	.9586	.9588	.9273	.9262
21	.01406	.01411	.9554	.9556	.9214	.9202
22	.01462	.01467	.9523	.9524	.9155	.9143
23	.01515	.01520	.9491	.9493	.9097	.9083
24	.01564	.01570	.9460	.9462	.9038	.9024
25	.01611	.01617	.9429	.9431	.8980	.8965

The approximate values of f, F and $U(0, t)$ (by relation (5)) were then inserted into (2) with $w = 10$ and $\eta = 0$ using repeated Simpson at unit steps in t for the value of the integral. When t was odd the last three panels were approximated by the three-eighths rule; $U(0, 1)$ was obtained by the trapezoidal. There is no "exact" result for $U(10, t)$ but the Laplace transform inversion methods described in Seal (1974) were used to produce results supposedly correct to three decimals. These, together with our new approximations appear in the last two columns of the Table. The new method appears to be producing values of $U(10, t)$ "nearly" correct to three decimals.

Conclusion

The proposed new approximation to $U(w, t)$ using the gamma approximation to $F(x, t)$ produces reasonably accurate results. Is it easy to apply? The writer confessed in his 1974 paper that steps in t at greater intervals than unity tended to harm the efficiency of the approximation to the integral in (2). For example, by using steps of five instead of unity in (2) we obtained, with the new approximations, the following values which are barely correct to two

	$U(10, t)$	
t	*Unit steps* (Table 1)	*Quinquennial steps*
5	.996	.994
10	.981	.977
15	.956	947
20	.927	918
25	.898	.887

decimals. Nevertheless this may be considered sufficient if a computer is not being used and desk calculations are the order of the day.

References

Cannella, S., 1963. "Variation de la prime d'assurance de l'assistance pharmaceutique en fonction de la participation de l'assuré au coût de l'assistance" *ASTIN Bull.* 2, 30-44

Hadwiger, H. 1942. "Wahl einer Naherungsfunktion fur Verteilungen auf Grund einer Funktionalgleichung" *Bl. Versich. Math.* 5, 345-352.

Johnson, N. L, and Kotz, S, 1970. *Distributions in Statistics. Continuous Univariate Distributions*-- 1. Houghton Mifflin, Boston.

Khamis, S H. and Rudert, W., 1965 *Tables of the Incomplete Gamma Function Ratio* Von Liebig, Darmstadt

Lundberg, O, 1940. *On Random Processes and their Application to Sickness and Accident Statistics.* Almqvist and Wiksell, Uppsala.

Seal, H. L., 1969 *Stochastic Theory of a Risk Business* John Wiley, New York.

—— 1974 "The numerical calculation of U(w, t), the probability of non-ruin in an interval (o, t)". *Scand Actu. J.* 1974, 121-139

—— 1976. "Approximations to Risk Theory's F(x, t) by means of the gamma distribution" *ASTIN Bull* 9, 213-218.

Shuster, J., 1968. "On the inverse Gaussian distribution function". *J Amer. Statist. Assn.* 63, 1514-1516.

Tweedie, M. C K, 1957. "Statistical properties of inverse Gaussian distributions" *Ann Math. Statist* 28, 362-377 and 696-705