METHODS FOR APPROXIMATING ACCUMULATED CLAIM DISTRIBUTION

S. M. AQIL BURNEY*

SUMMARY: Different methods for approximating accumulated claim distribution are reviewed and compared with illustrations. Monte Carlo method is restressed when personal computers are available. Simulation results show that desired accuracy can be achieved by increasing the simulated sample size.

Key Words: Non-life insurance, claim distribution, polya distribution.

INTRODUCTION

Consider a non-life insurance operation being transacted through an interval of time (0, t), where t is operational time (7). Further assume that claims are occurring randomly and independently in each class of business. Furthermore, let us assume that the number of claims occurring in (0,t) follows the Poisson distribution and is given by

$$\Pr\{N(t) = k\} = q_k(t) = e^{-t} t^k / k!;\ k = 0, 1, 2,... \qquad (1.1)$$

where the probability distribution function of the independent time interval, between successive events being

$$\Pr\{T \le t\} = 1 - e^{-t};\ t > 0 \qquad (1.2)$$

Further assume that $Y \geq 0$, the size of an individual claim is a random variable independent of the epoch of claim occurrence and of the interval between it and the prior claim (16). The distribution function P(y) of Y is defined by

$$\Pr\{Y \le y\} = P(y);\ Y \ge 0 \qquad (1.3)$$

Following the above assumption, we define the distribution function of X(t), i.e. the claim incurred in (0,t):

$$F(x,t) = \Pr\{X(t) \le x\}$$

$$= \sum_{k=0}^{\infty} \Pr\{X(t) \le X \mid N(t) = k\} \Pr\{N(t) = k\}$$

$$F(x,t) = \sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} p^{k*}(x) \qquad (1.4)$$

where $P^{k*}(y)$ is called k-fold convolution of P(y) and is defined by

$$P^{k*}(y) = \int_0^Y P^{(k-1)*}(y-z)\, dp^{1*}(z);\ k = 1, 2, 3,...$$

and

$$P^{1*}(y) = P(y), \text{ for } y \ge 0$$

$$P^{0*}(y) = \begin{cases} 0, & \text{for } y < 0 \\ 1, & \text{for } y \ge 0 \end{cases} \qquad (1.5)$$

The distribution function, F(x,t) given by (1.4) is known as 'Accumulated Claim Distribution' function (ACD-function). When claim number process follows Poisson distribution and distribution of size of individual claim is not specified, the process will be denoted by 'Poisson/Generalized Model'. F(x,t) is infinitly divisible for details (7,12, 13,19, 21).

Probabilities for Poisson/Exponential model can be computed without much difficulty as in this case exact calculation are feasible (17). We shall be using Poisson/Exponential model frequently for comparative purposes whilst applying various techniques for computation and approximation of F(x,t). F(x,t) plays instrumental role for computing retention limit, safety margin loading and probability of ruin (4).

The objective of this article is to review the existing approximation techniques for F(x,t) and inform our experiences.

Some properties of interest (here) are discussed in **Properties of F(x,t)**. **Techniques for approximation of F(x,t) function** discusses different techniques of approximation of F(x,t). Along with illustration of Poisson/Exponential case for small, moderate and large values of t. A tech-

*From Department of Statistics, University of Karachi, Karachi-75270, Pakistan.

niques for approximating F(x,t), using Laguerre Polynomials (9), is discussed in this section briefly.

Simulation can also be used for computing F(x,t) and results are encouraging and discussed in **Monte Carlo method.**

PROPERTIES OF F(x,t)

We will discuss some of the properties of F(x,t) which are of use while computing this function.

Let us now define the moment generating function of the ACD-function, F(x,t), by the following integral

$$M(s,t) = \int_0^\infty e^{-sx}\, d_x\, F(x,t) \qquad (2.1)$$

when the claim number process is Poisson then (2.1) becomes

$$M(s,t) = \exp\{t\,(P(s) - 1\} \qquad (2.2)$$

where

$$P(s) = \int_0^\infty e^{-sx}\, d_x\, P(x) \qquad (2.3)$$

and s is an auxiliary variable.

The moments of ACD-function can be found using following expression:

$$\mu_r = (-1)^r d^r M(s,t)/ds^r \big|_{s=0} \text{ for } r = 0, 1, 2, ... \qquad (2.4)$$

The following first four moments about origin for Poisson/Generalized model can obtained from (2.4).

$$\mu'_1 = P_1 t$$
$$\mu'_2 = P_2 t + (P_1 t)^2$$
$$\mu'_3 = P_3 t + 3P_1P_2t^2$$
$$\mu'_4 = P_4 t + 3P_2^2\, t^2 + 4P_1P_3\, t^2 + 6P_1^2P_2t^3 + P_1^4\, t^4 \qquad (2.5)$$

where P_1, P_2, P_3 and P_4 are first four moments of P(y).

The moments about mean (μ'_r) in terms of moments about origin (μ_r) are given by

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} \mu_{r-j} (-1)^j \mu_1^j \qquad (2.6)$$

Using the expected value of a claim amount as a unit of measurement such that

$$P_1 = \int_0^\infty x\, dP(x) = 1$$

We have first four moments about mean

$$\mu_1 = 0$$
$$\mu_2 = P_2 t$$
$$\mu_3 = P_3 t$$
$$\mu_4 = P_4 t + 3P^2{}_2 t^2$$

In Polya/Generalized case, when $q_k(t_t)$ is

$$q_k(t) = \begin{bmatrix} h+k-1 \\ k \end{bmatrix} \left[\frac{h}{h+t}\right] \left[\frac{t}{h+t}\right]$$

and the m. g. f. of F (x,t) is given by

$$M(st) = \{1 - (P(s) - 1)t/n\}^{-n} \text{ for } n > 0 \qquad (2.8)$$

In case n tends to infinity, (2.8) straight away reduces to the Poisson/Generalized case given by (2.2).

TECHNIQUES FOR APPROXIMATION OF F(x,t) FUNCTION

Direct numerical calculations of F(x,t) often lead to very cumbersome expressions (17), which are not handy for practical purposes, therefore, one of the major problems in risk theory is search for a suitable approximations for F(x,t).

There are numerous methods for obtaining approximate expressions of F(x,t) such as Normal approximation, Edgeworth series expansion of F(x,t) Esscher approximation, NP method, inversion of characteristic function (Seal 1979), Pearson frequency curves (5), Gamma approximation (Seal 1976) Generalized Lambda (2, 3, 14), and Laguerre polynomial approximation (9).

Edgeworth series approximation

The normal approximation (1) to F(x,t) is in fact a very special case of a more generalized form of Edgeworth series expansion (Kendall and Stuart 1977) F(x,t) in Edgeworth expansion is given by

$$F(x,t) = \Phi\left(\frac{x-t}{\sqrt{p_2 t}}\right) - \left(\frac{p_3}{6p_3\sqrt{p_2 t}}\right)\Phi^{(3)}\left(\frac{x-t}{\sqrt{p_2 t}}\right)$$
$$+ \frac{p_4}{24tp_2^2}\Phi^{(4)}\left(\frac{x-t}{\sqrt{p_2 t}} + \frac{p_3}{72tp_2^3}\Phi^{(6)}\left(\frac{x-t}{\sqrt{p_2 t}}\right) + 0\left(t^{-3/2}\right)\right) \qquad (3.3)$$

where $\Phi^{(r)}(x) = \dfrac{d^{(r)}\,\Phi(x)}{dx^r}$

and $P_r = E(X^r) = \int_0^\infty x_r\, dP(x), r = 0,1,2,...$

(1). This series (3.3) is semi-convergent when $n \to \infty$ for a fixed value of t. However, by taking suitable number of terms this expansion provides reasonable results in the neighborhood of mean value of X(t) (Table 1). From risk theoretical point of view the main region of interest, whilst studying ACD-functions, is the neighborhood of mean value to the extend of two to three times the standard deviation. We, therefore, need some improvement in the Edgeworth expansion and this improved method of approximation for F(x,t) is called normal power method and is given by (1,18). The approximated value of F(x,t) is given by

$$F(x,t) = \Phi\left\{\sqrt{\left(1+\frac{9}{\beta_1}+\frac{6}{\sqrt{\beta_1}}\cdot\frac{x-t}{p_2 t}\right)}-\frac{3}{\sqrt{\beta_1}}\right\} \quad (3.9)$$

provided $x \geq t + (p_3 / 6p_2)(1\text{-}9 / \gamma^2_1)$ and p_1, p_2 and p_3 exist.

Accuracy of this approximation is demonstrated in case of Poisson/Negative Exponential case (Illustration 2). As apparent from Table 1, for small values of t, NP method gives poor results but provides satisfactory fit for middle range and gives surprisingly good results for large values of t. However, for x-t, NP method provides satisfactory results even for small values of t. This method uses only first three moments about mean and gives reasonably close results when the skewness is of moderate size i.e. $\gamma_1 \leq 2$.

Illustration 1 (Poisson/Exponential model)

$q_k(t) = e^{-t} \cdot t^k / k!,\ k = 0, 1, 2, \ldots$

$p(x) = 1 - e^{-t}\ x > 0$

$x_0 = (x-t) / \sqrt{(p_2 t)}$

Laguerre approximation

The F(x,t) can be approximated using Laguerre polynomials (Magnus *et al.* 1988) assuming the first m moments $\mu_1, \mu_2, \ldots, \mu_m$ about origin exist. Making use of the definition of Laguerre polynomials, we write

$$f_m(x,t) = \sum_{r=0}^{m} C_r L_r^{(n)}(x)\theta_n(x) \quad (3.10)$$

and

$$F_{m'}(x,t) = \int_0^x f_{m'}(y,t)\cdot dy \text{ for } m \leq m \quad (3.11)$$

where $\theta_n(x) = e^{-x} \cdot x^{n-1} / \Gamma n$ for $x \geq 0$ C_r are arbitrary coefficients and $L^{(n)}_r(x)$ are Laguerre polynomials for r = 0, 1, 2,...m. The values of n and C_r may be determined by the method of moments (9).

The gamma approximation

We have seen in the foregoing sections that various techniques can be followed to arrive at the approximate form of ACD-function. Among these methods NP approximation may be preferred because of its simplicity. Buhlmann (1974) has also recommended this technique. Seal (1976) on the other hand emphasized on a simple gamma distribution (Pearson-type III) for approximating F(x,t). We have also experienced that ACD-function in following form can be used

$$F\left(t + z\sqrt{\mu_{z'} t}\right) \cong (1/\Gamma\alpha)\int_0^{a+z\sqrt{\alpha}} e^{-y}\, y^{\alpha-1}\, dy \quad (3.10)$$

$$= G\left(\alpha, \alpha + z\sqrt{\alpha}\right)$$

where G (.,.) notation is for the incomplete gamma ratio, and α is a function of t given by

$\alpha = 4/\beta = 4\mu^3 / \mu^2$

and

$z = (\mu + \eta t) / \sqrt{(tp)_2}$

Table 1

x_0	t=10		t=100		t=1000	
	Seal 1972	NP	Exact Seal	NP	Exact Seal 1972	NP
-3	-	-	.00037	.00098	.00098	.00098
-2	.00234	.00338	.01669	.01683	.02091	.02092
-1	.15470	.15885	.15833	.15865	.15862	.15865
0	.54489	.54397	.51411	.51409	.50446	.50446
1	.84384	.84134	.84163	.84134	.84137	.84134
2	.96236	.96113	.97186	.97201	.97547	.97546
3	.99308	.99274	.99718	.99721	.99823	.99823
4	.99897	.99820	.99983	.99983	.99994	.99994
5	.99987	.99906	.99990	.99990	.99999	1.00000

where μ and μ are second and third moments about mean of F(x,t) and η is safety loading margin.

Simple programmable calculators can be used to compute incomplete gamma ratio in (3.24). Also tables of the incomplete gamma ratio prepared by Khamis and Rudert (11) can be used. Both Bhattacharjee (1970) and Chi-Leung Lau (1980) have produced computer programmes for computing the incomplete gamma ratios. An important feature of using the gamma distribution as an approximation (3.10) to F(x,t) is that only the first three moments of the distribution of the claim amount are required. Table 2 exhibits the closeness of the values of F(x,t), for Poisson/Exponential case when $t = 10$.

Values computed by Seal (17) and values calculated by using NP formula are also given for comparison purposes.

Pearson system of curves

Consider the ACD-function and let us assume that in the Poisson/Generalized case its parameters can be estimated from the moments of kF(x,t) given by (2.7) then an appropriate form of Pearson System of Curves (5), can be chosen to give an approximation of F(x,t) for known values of β_1 and β_2. Assuming that the given Curve is one of the Pearson System of Curves then table of the standardized deviates, for different values of β_1 and β_2, can be used for various practical purposes. However, utility of this method is restricted because the limited range of the tables of percentage points of Pearson System of Curves for given β_1 and β_2.

In the following Illustration 2, Poisson/Inverse Guassian model is considered.

Illustration 2: Poisson/Inverse Guassian model

$q_k(t) = e^{-t} t^k / k!, k = 0, 1, 2, ...$

and

$$P(x) = \{^{\lambda} / 2p\pi x^3\}^{1/2} \exp \{^{-\lambda (x-\mu)2} / 2\mu^2 x\}$$

$x > 0, \mu > 0, \lambda > 0$

The moments about mean of F(x,t) in this case are given by (20)

$\mu_2 = (1 + 1 / \lambda)^t$

$\mu_3 = (3 / \lambda^2 + 3 / \lambda + 1)t$

$\mu_4 = (1 + 6 / \lambda + 15 / \lambda^3)t + 3(1 + 1 / \lambda)^2 t^2$

Further assume $\mu = 1$ and $\lambda = 2.20408$ (18), which leads to $\beta_1 = 2.88810/t$ and $\beta_2 = 4.050869 + 3$.

The criterion for application of Pearson type III curve, i.e. $2\beta_2 - 3\beta_3 = 6$, to some extent holds in this case. It, therefore, follows that F(x,t) for the above values of β_1 and β_2 can be approximated by a Pearson type III with the parameter $p = (4 / \beta_1) - 1$ (5).

Monte Carlo method

Monte Carlo techniques can also be used to approximate F(x,t) when claim occurrence distribution $q_k(t)$ and claim amount distribution are known. Using this experimental mathematics (15) F(x,t) can be simulated. For the simulation of ACD-function, a randon number is generated to get random deviates from $q_k(t)$ by solving the equation

$$\zeta = \sum_{m=0}^{k} q_m(t)$$

Table 2: Values of F(x,t) in Poisson/Exponential case ($t = 10$ and $X_0 = (x\text{-}t) /\sqrt{(2t)}$).

X_0	Seal (1972)	NP Method	Gamma		
			Guassian*	Simpson**	$P(\alpha, a+z\sqrt{\alpha})$***
-2	.00234	.00338	.00371	.00371	.00371
-1	.15470	.15865	.15267	.15267	.15274
0	.54489	.54397	.54427	.54460	.54461
1	.84384	.84135	.84485	.84498	.84499
2	.96236	.96113	.96367	.96245	.96248
3	.99308	.99274	.99730	.99305	.99290
4	.99897	.99890	.99962	.99876	.99888
5	.99987	.99986	.99992	.99981	.99984

* Guassian Quadrature formula was used for integrating the Gamma function.
* * Simpson's general formula was used for integrating Gamma function (10).
* * * Khamis and Rudert (11) tables were used.

Table 3

x_0	Seal (1972) F (x_0,10)	Simulated F (x_0, 10)
-3	-	-
-2	.00234	.003
-1	.15470	.163
0	.54489	.539
1	.84384	.849
2	.96230	.962
3	.99308	.998
4	.99897	1.000
5	.99987	1.000

(The Kolmogorov-Smirnow Statistics, D_N is. 03869 and $D_N = 1.2)^{(.05)}$

where $0 \leq \zeta \leq 1$ is a random number. Say $k = k_1$ is the solution of (3.11), then k_1 random numbers $r_{11}, r_{12},....,r_{1k1}$ are generated using the equation
$r_{1i} = P(x)$

Using (3.12), we obtain $x_{11}, x_{12},...., x_{1k1}$ random deviates, and

$$x_1 = \sum_{i=1}^{k_1} x_{1i}$$

gives the first sample point of the ACD-function.

Following the aforestated technique, a number of sample points of F(x,t) can be generated to get desired accuracy. An estimate of f(x,t) can be obtained by constructing empirical distribution function $S_n(x)$ (8), given by

$$S_n(x) = \begin{cases} 0 & x < x_{(1)} \\ i/n & x \leq x_{(i)} \leq x_{(i+1)} \quad \text{for } i = 1,2,3 \quad n-1 \\ 1 & x \geq x_{(n)} \end{cases}$$

where n denotes the sample size and $x_{(i)}$ denotes the value of the order Statistic.

The empirical distribution function $S_n(x)$ gives an unbiased estimate of F(x,t) and standard error of $S_n(x)$ depends on n with upper bound $1/2\sqrt{n}$. In risk theory right hand tail area of F(x,t) is of importance and the values of the order 10^{-1} to 10^{-3} are of interest for practical purposes such as determining retention limit for reinsurance. If high precision is required then 10^6 or more sample points are needed from F(x,t). Monte Carlo technique, though gives reasonably good result, but due to the cost involved its use is not very popular. This technique can now be used frequently as personal computer are available and this simulation can be done on a personal computer. Another advantage is that Monte Carlo techniques can be used even when $q_k(t)$ is not a Poisson distribution provided that it is suitable for computer input (15).

In illustration 3 Poisson/Exponential model is simulated and results are given in Table 3. It is found that there is no significant difference at level .05, in simulated distribution and results of Seal (17) (Kolmogorov-Smirnov test was used). During this simulation experiment we have noted we can get desired accuracy by increasing the sample size at the cost of increase in CPU time. However, for practical purposes a simulated sample of size 1000 is sufficient to get accuracy up to third decimal place.

In last it is of interest to mention that ACD-function can be computed by numerical inversion of characteristic function. Seal (1977, 18) discussed in detail the methods for inversion of characteristic function and computed F(x,t) for Poisson/Exponential, Poisson/Inverse Gaussian, Waring/Exponential and Waring/Inverse Gaussian models.

Illustration 3 Poisson/Exponential model (t=10).

$$q_k(t) = e^{-t} t^k / k!, \quad k = 0, 1, 2,...$$

$$P(y) = 1 - e^{-y}, \quad y > 0$$

$$x_0 = (x\text{-}t) / \sqrt{p} t_2$$

ACKNOWLEDGEMENT

Author is thankful to Mr. Ghulam Hussain, Assistant Professor, Department of Statistics for providing the computational assistance during this work. Acknowledgements are also due to Dr. M. A. Hashmi (Manager for Pakistan, Royal Insurance, plt. Pakistan) for his valuable guidance through out this work.

REFERENCES

1. Beard RE, Pentikainen T, Pesonen E : Risk Theory: The Stochastic Basis of Insurance, Third Ed, Chapman and Hall New York, 1984.

2. Burney SM Aqil : Risk Theory and Insurance. Unpublished M. Phil. Thesis. University of Karachi, Karachi 75270, Pakistan, 1983.

3. Burney SM Aqil, Hussein G, Ejaz Ahmed : On the Claim Distribution in Insurance to be published in Karachi University Journal of Science, 1989.

4. Buhlmann H : Mathematical methods in risk theory. Springer-Verlag, 1970.

5. Elderton WP, Johnson NL : System of Frequency Curves. Cambridge University Press London, 1969.

6. Feller W : An Introduction to Probability Theory and its Applications, Third Ed, Vol I, Wiley, New York, 1968.

7. Feller W : An Introduction to Probability Theory and its Applications, Second Ed, Vol II, Wiley, New York, 1971.

8. Gibbon JD : Nonparametric Statistical Inference McGraw Hill Kogakusha Ltd, Tokyo, 1971.

9. Hashmi MA, Burney SM Aqil, Hussain G : Approximating Accumulated Claim Distributions using Laguerre Polynomials. Pakistan Journal of Statistics, Vol 6, 1990.

10. Kellison SG : Fundamental of Numerical Analysis, Home Wood, III Richard D Irwin, 1975.

11. Khamis SH, Rudert W : Tables of the Incomplete Gamma Function Ratio. Von Leibig, Darmstadt, 1965.

12. Lukacs E : Characteristic Functions, Second Ed, Griffin London, 1970.

13. Panjer AH, Willmot GE : Compound Poisson Models in Actuarial Risk Theory. J Econometrics, p 23, pp 63-76, 1983.

14. Ramberg JS, Tadikamla PR, Dudewicz EJ : A Probability Distribution and its uses in Fitting Data. Technometrics, 21:201-214, 1979.

15. Ripley BD : Stochastic Simulation. John Wiley, New York, 1987.

16. Robinson JM : A Stochastic Model for Non-life Insurance Claim Incurral, Accural, and Reporting. Proceeding of the Business and Economic Statistics section 148th Annual meeting. New Orleans, Lousisiana, August 22-25, 1988.

17. Seal H : Numerical calculation of the Probability of Ruin in the Poisson Exponential Case. Mitt Verin Schweiz Versich Mattr, pp 77-100, 1972.

18. Seal H : Survival Probabilities: The goal of risk theory. John Wiley, New York, 1978.

19. Steutel F : Infinite divisibility in theory and practice. Scand J Statistics, 6:57-64, 1979.

20. Tweedie MCK : Statistical Properties of the Inverse Guassian Distribution, 1. Ann Math Statistics, 28:362-377, 1957.

21. Van Harn K : Classifying infinitely divisible distributions by functional equations. Mathematical Centre Tracts no 103, Mathematical Centre Amsterdam, 1978.

Correspondence:
SM Aqil Burney
Department of Statistics,
University of Karachi,
Karachi 75270,
PAKISTAN.