# DISTRIBUTION BY SIZE OF RISK

## A MODEL

CHARLES C. HEWITT, JR.

Distribution of risks by size is important in many lines of commercial casualty insurance, and yet there seems to be no evidence in the *Proceedings* of any attempt to provide a workable mathematical model for this distribution. This paper will indicate that there is a basic model which provides excellent fit of the raw data in many instances. Also, the paper will illustrate an application of the model to a study of certain types of expense by size of risk.

### THE MODEL—LOG-GAMMA DISTRIBUTION

The Gamma Distribution (also referred to as Pearson Type III) has been used in several recent articles in these *Proceedings* with excellent results. Dropkin[1] gives a readily understandable discussion of the gamma function, including the use of the Pearson Tables[2] in his Appendix D to his 1959 paper, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records." In the Gamma Distribution the probability density function, p. d. f., is given by:

$$T\,(x,a,p)\;dx = \frac{a^{p+1}}{\Gamma\,(p+1)}\,x^{p}\,e^{-ax}\;dx$$

where

$$[0 \leq x] \text{ and } \left.\begin{matrix} a \\ p+1 \end{matrix}\right\} > 0$$

In Dropkin's work $r\ (= p + 1)$ is used; however, $p$ is one of the entry values into the Pearson Tables and is, therefore, a little handier in going back and forth between the theory and the tables.

$$E\,(x) = \frac{p+1}{a} \text{ and } \sigma^2\,(x) = \frac{p+1}{a^2}$$

$$\text{Mode}\,(x) = \frac{p}{a}$$

$$\text{If}\quad X = \text{Risk premium,}$$

$$\text{and } x = \log_e X,$$

then $T(x,a,p)$, or a compound thereof, produces reasonably good fits to distributions by size of risk. In this instance it is appropriate to refer to

[1] L. B. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS XLVI*, p. 165.

[2] Karl Pearson, *Tables of the Incomplete Gamma Function*, Cambridge University Press (1957).

*T(x,a,p)* as the Log-Gamma Distribution. This particular model is extremely flexible in shifting back and forth between distributions of *number of risks* by size of risk and distributions of *amount of premiums* by size of risk. E.g., if *T(x,a,p)* represents a distribution of amount of premium, then it is easily seen that:

$$T(x, a+1, p)\,dx = \frac{(a+1)^{p+1}}{\Gamma(p+1)} x^p e^{-(a+1)x}\,dx$$

is the distribution by number of risks,

from which it can be shown that $E(X) = \left(\frac{a+1}{a}\right)^{p+1}$

The modal value of $X$ for the distribution of *number of risks* is:

$$e^{\frac{p}{a+1}}$$

while for the distribution of *amount of premium* the mode is:

$$e^{\frac{p}{a}}$$

## THE FIT—WORKMEN'S COMPENSATION STUDY BY SIZE OF RISK

The method of fitting the data for non-participating stock companies from the 1965 Study of Expenses by Size of Risk by the National Council on Compensation Insurance is discussed in Appendix 1. A comparison of the theoretical and actual values for number of risks and amount of premium is set forth in Table 1.

To verify that this model fits other size-of-risk data, a number of similar tests were made on workmen's compensation insurance statistics for the National Council and for some of the independent state rating bureaus. In general it was found that the basic model described above works quite well for distributions of non-participating stock company business and distributions of "All Other" risks in workmen's compensation. A compound of *T(x,a,p)* was used to fit data for mutual carriers (see Appendix 2).

## THE APPLICATION—EXPENSES BY SIZE OF RISK

Certain overhead items of expense analyzed during the course of the National Council's 1965 Study are capable of being expressed analytically by the following formula:

$$\epsilon = \alpha + \beta e^{-\gamma x}$$

where $\epsilon$ is the expense ratio for a particular premium size, and $X$, and $x$ have the same meaning as heretofore; when $x$ has its minimum value of zero, the expense ratio becomes $\alpha + \beta$; as $x$ increases, the second

## TABLE 1

### COMPARISON OF THEORETICAL AND ACTUAL* DISTRIBUTIONS BY SIZE OF RISK (EXCLUDING 3-YEAR FIXED RATE POLICIES)

### NON-PARTICIPATING STOCK CARRIERS

| | Number of Policies (000's omitted) | | Standard Premium (Excl. $10 Expense Constant) (000,000's omitted) | |
|---|---|---|---|---|
| Annual Premium Size | Theoretical | Actual | Theoretical | Actual |
| Under $100 | 381.3 | 397.2 | $ 16.8 | $ 16.1 |
| $ 100-$ 499 | 359.0 | 348.6 | 78.2 | 81.4 |
| 500- 749 | 54.0 | 55.6 | 32.7 | 34.2 |
| 750- 999 | 28.5 | 28.0 | 24.3 | 24.3 |
| Under $1,000 | 822.8 | 829.4 | 152.0 | 156.0 |
| $ 1,000-$ 4,999 | 77.0 | 71.5 | 157.6 | 152.5 |
| 5,000- 24,999 | 15.2 | 13.9 | 148.4 | 139.0 |
| 25,000- 49,999 | 1.5 | 1.5 | 50.8 | 50.6 |
| 50,000- 99,999 | 0.6 | 0.7 | 41.4 | 43.9 |
| $ 1,000-$ 99,999 | 94.3 | 87.6 | 398.2 | 386.0 |
| $100,000-$249,999 | 0.3 | 0.4 | 41.3 | 56.6 |
| $250,000 and over | 0.1 | 0.1 | 73.9 | 66.8 |
| $100,000 and over | 0.4 | 0.5 | 115.2 | 123.4 |
| TOTAL | 917.5 | 917.5 | $ 665.4 | $ 665.4 |

*Source: National Council on Compensation Insurance – Report of the Special Committee to Study Expenses by Size of Risk (May 28, 1965) – Exhibit I – Non-Participating Stock Carriers term in the above relationship approaches zero and the expense ratio tends toward $\alpha$ as a limiting value.

Since the minimum value for $X$ is unity $(x = 0)$, $\alpha + \beta$ represents the "expense constant" for those items of expense being represented by the formula.

Using $T(x,a,p)$ and the above form for expense ratios, it follows that:

$$E(\epsilon) = \alpha + \beta \left(\frac{a}{a+\gamma}\right)^{p+1}$$

or, for any premium range up to some value X'

$$E_{x'}(\epsilon) = \alpha + \frac{\beta \left(\frac{a}{a+\gamma}\right)^{p+1} I(u^*, p)}{I(u,p)}$$

where

$$I(u,p) = \int_0^{x'} T(x,a,p)\,dx, \text{ and } u = \frac{a}{\sqrt{p+1}}\,x'$$

$u$ is necessary since it is the other entry value (with $p$) in Pearson's Tables of the Incomplete Gamma Function.

$$I(u^*, p) = \int_0^{x'} T(x,a+\gamma,p)\,dx, \text{ and } u^* = \frac{a+\gamma}{\sqrt{p+1}}\,x'$$

The method of fitting the data for certain overhead expenses for non-participating stock companies from the National Council's 1965 Study is discussed in Appendix 1. The three parameters are found to be:

$$\alpha = .050; \qquad \beta = 11.69; \qquad \gamma = 0.94$$

so that

$$\epsilon = .050 + 11.69\, e^{-0.94x},$$

$$\text{or expense \$ \$ \$} = .050\,\text{X} + \$11.69\, e^{0.06x}$$

$$(\text{since X} = e^{x})$$

As pointed out, when the premium is $1, the "expense constant" for these particular items is $\alpha + \beta$ or $11.74[3]. This is perhaps the first analytical derivation of an *expense constant.*

[3] Cf Report of April 28, 1965, Meeting of (NAIC) Subcommittee to Study Expenses by Size of Workmen's Compensation Risk, which suggests a figure of $12 for an expense constant.

TABLE 2

COMPARISON OF THEORETICAL AND ACTUAL*
EXPENSE RATIOS BY SIZE OF RISK

(EXCLUDING 3-YEAR FIXED RATE POLICIES AND $10 EXPENSE CONSTANT)

NON-PARTICIPATING STOCK CARRIERS

| | Actual* Expense Ratios | | | | Theoretical Expense Ratios |
|---|---|---|---|---|---|
| Annual Premium Size | Inspection Boards and Bureaus | Payroll Audit | Other General | Total | Total |
| Under $100 | .032 | .105 | .243 | .380 | .380 |
| $ 100-$ 499 | .018 | .039 | .062 | .119 | .124 |
| 500- 749 | .018 | .027 | .042 | .087 | .079 |
| 750- 999 | .022 | .021 | .044 | .087 | .071 |
| Under $1,000 | (.020) | (.041) | (.073) | (.134) | (.134) |
| $ 1,000-$ 4,999 | .017 | .014 | .026 | .057 | .059 |
| 5,000- 24,999 | .018 | ..009 | .023 | .050 | .052 |
| 25,000- 49,999 | .021 | .007 | .026 | .054 | .051 |
| 50,000- 99,999 | .020 | .006 | .026 | .052 | .051 |
| $ 1,000-$ 99,999 | (.019) | (.010) | (.025) | (.054) | (.055) |
| $100,000-$249,999 | .020 | .005 | .023 | .048 | .050 |
| 250,000 and over | .024 | .005 | .024 | .053 | .050 |
| $100,000 and over | (.022) | (.005) | (.024) | (.051) | (.050) |
| TOTAL | (.020) | (.016) | (.036) | (.072) | (.072) |

*Source: Same as for Table 1 (Expense transfers ignored)

A comparison of the theoretical average expense ratios within each premium grouping with the actual expense ratios is set forth in Table 2. Although the fit is fairly good, it is far from perfect, but the raw data is itself rather erratic from one interval to the next. In any event this example serves to illustrate the applicability of the Log-Gamma model in determining mean expense ratios for premium size intervals and in total.

## CONCLUSION

The Log-Gamma Distribution is a flexible, easily applied model which provides relatively good fits in either the basic or a compound form to commercial risk distributions by size. When the parameters of the model have been determined, the Log-Gamma is readily applicable to analysis of factors, such as expenses, which appear to vary with risk size in a polynomial or exponential form.

## APPENDIX 1—FITTING THE DATA

While the results produced by an appropriate model and the ease with which the model may be applied are the important considerations, the technique of fitting a particular set of data is also of some interest. Size of risk distributions generally have two characteristics that produce problems in fitting, unless proper precautions are taken. The characteristics are (1) a great majority of the risks are at the lowest premium sizes, and (2) jumbo risks at the opposite end of the spectrum distort the moments of the premium distribution. The precautions are (1) make the initial fit on distributions of premium amounts rather than number of risks—the former distribution is always far less skewed than the latter, and (2) make the initial fit on the logarithm of premium size rather than the premium size itself—the distortion created by the jumbo risks is minimized. (These general comments are also appropriate for fitting distributions by size of loss.)

*Log-Gamma Fitting.* This is a two-parameter distribution and the ultimate determination of the parameters, $a$ and $p$, was by solution of the two equations for mean value:

(1) $E(x) = \dfrac{p+1}{a}$ (on distribution by amount of premium)

(2) $E(X) = \left(\dfrac{a+1}{a}\right)^{p+1}$

However, the latter equation is not easily solved without a good approximation for $p$. This approximation was obtained by using the sample mean and variance of $T(x,a,p)$; the former is given in (1) just above and the latter is, of course gixen by $\frac{p+1}{a^2}$. Any reasonable value for the "interval $E(x)$ and $E(x^2)$" in each of the premium intervals will do for this approximation, except that the values for $E(x)$ and $E(x^2)$ in the uppermost premium interval should be repaired at each successive approximation, since this is an open-end interval and even logarithm values need to be carefully selected. Once a stabilized value of $p$ (it is easiest to round to the nearest entry value in the Pearson Tables) is obtained, then equation (2) is readily solved.

*Expense distribution.* The expression for expense ratio at a particular premium size is a three-parameter exponential formula. The determination of the parameters was achieved by combining analytically the expense ratio for a particular premium size with the frequency of premium amounts at that particular premium size (as fitted to the Log-Gamma function) and producing arithmetic mean values for:

(1) the entire premium range,

(2) the first \$100 of the premium range, and

(3) the first \$1,000 of the premium range.

The latter two conditions were chosen after an examination of the source data indicated that these premium intervals were critical in obtaining a good fit of expense ratios. The three conditions produced three equations which were then solved for the three parameters on a trial-and-error basis (with a minimum of difficulty).

## APPENDIX 2—COMPOUND LOG-GAMMA

The basic Log-Gamma is not a good model for mutual carrier distributions or for "Manufacturing" risk distributions by size. However, a compound Log-Gamma of the form:

$$h\,T(x,a_1,p_1) + (1\text{-}h)\,T(x,a_2,p_2),\ (0 < h < 1)$$

does produce the results set forth in Table 3. (Subscript 1 parameters were "borrowed" from the non-participating stock carrier distribution.)

This compound distribution can then be applied to an analysis of expenses by size of risk, where the parameters in the expense ratio formula are different for the separate elements of the compound Log-Gamma function. The result of this fitting of expense data for mutual carriers (Table 4) is included for the sake of completeness.

## TABLE 3

## COMPARISON OF THEORETICAL AND ACTUAL* DISTRIBUTIONS BY SIZE OF RISK (EXCLUDING 3-YEAR FIXED RATE POLICIES)

### MUTUAL CARRIERS

| | Number of Policies (000's omitted) | | Standard Premium (Excl. $10 Expense Constant) (000,000's omitted) | |
|---|---|---|---|---|
| Annual Premium Size | Theoretical | Actual | Theoretical | Actual |
| Under $100 | 89.8 | 95.7 | $ 4.0 | $ 3.6 |
| $ 100-$ 499 | 103.2 | 100.3 | 23.8 | 23.3 |
| 500- 749 | 21.2 | 21.5 | 12.9 | 13.0 |
| 750- 999 | 12.7 | 12.5 | 10.9 | 10.8 |
| Under $1,000 | 226.9 | 230.0 | 51.6 | 50.7 |
| $ 1,000-$ 4,999 | 45.3 | 42.7 | 99.3 | 93.0 |
| 5,000- 24,999 | 13.9 | 13.3 | 141.1 | 140.5 |
| 25,000- 49,999 | 1.7 | 1.7 | 57.7 | 61.3 |
| 50,000- 99,999 | 0.7 | 0.7 | 49.6 | 50.1 |
| $ 1,000-$ 99,999 | 61.6 | 58.4 | 347.7 | 344.9 |
| $100,000-$249,999 | 0.3 | 0.4 | 50.0 | 58.9 |
| 250,000 and over | 0.1 | 0.1 | 74.3 | 69.1 |
| $100,000 and over | 0.4 | 0.5 | 124.3 | 128.0 |
| TOTAL | 288.9 | 288.9 | $ 523.6 | $ 523.6 |

*Source: National Council on Compensation Insurance – Report of the Special Committee to Study Expenses by Size of Risk (May 28, 1965) – Exhibit II – *Mutual Carriers*

## TABLE 4

## COMPARISON OF THEORETICAL AND ACTUAL* EXPENSE RATIOS BY SIZE OF RISK

### (EXCLUDING 3-YEAR FIXED RATE POLICIES AND $10 EXPENSE CONSTANT)

### MUTUAL CARRIERS

| Annual Premium Size | Actual* Expense Ratios: Inspection Boards and Bureaus | Payroll Audit | Other General | Total | Theoretical Expense Ratios: Total |
|---|---|---|---|---|---|
| Under $100 | .051 | .088 | .244 | .383 | .381 |
| $ 100-$ 499 | .024 | .036 | .062 | .122 | .141 |
| 500- 749 | .027 | .027 | .045 | .099 | .103 |
| 750- 999 | .034 | .023 | .041 | .098 | .095 |
| Under $1,000 | (.029) | (.034) | (.067) | (.130) | (.140) |
| $ 1,000-$ 4,999 | .033 | .016 | .032 | .081 | .076 |
| 5,000- 24,999 | .029 | .008 | .024 | .061 | .059 |
| 25,000- 49,999 | .027 | .005 | .021 | .053 | .053 |
| 50,000- 99,999 | .027 | .004 | .020 | .051 | .051 |
| $ 1,000-$ 99,999 | (.030) | (.009) | (.024) | (.063) | (.062) |
| $100,000-$249,999 | .027 | .004 | .019 | .050 | .049 |
| 250,000 and over | .029 | .003 | .017 | .049 | .048 |
| $100,000 and over | (.028) | (.003) | (.018) | (.049) | (.049) |
| TOTAL | (.029) | (.010) | (.027) | (.066) | (.066 |

*Source: Same as for Table 3

APPENDIX 3—PARAMETERS

| Non-Participating Stock (Subscript 1) | Parameter | Mutual (Subscript 2) |
|---|---|---|
| 1.220 | a | 2.223 |
| 10.0 | p | 21.8 |
| .050 | $\alpha$ | .046 |
| 11.69 | $\beta$ | 5.25 |
| 0.94 | $\gamma$ | 0.63 |
| h | Compound Log-Gamma | 1−h |
| .731 | Number of risks | .269 |
| .293 | Amount of premium | .707 |

DISCUSSION BY JAMES R. BERQUIST

We are, indeed, indebted to Mr. Hewitt for his continual efforts to provide us with practical applications of the theoretical techniques developed by mathematical statisticians.

In this paper Mr. Hewitt suggests a model which gives a good fit for size of risk distributions. That this technique does, in fact, fit the industry data is shown in Tables I and III.

The value of the suggested model is not limited to industry statistics, however, as its most practical application for the company actuary will be in fitting the distribution of business by size of risk of his own company to the model.

For example, the table on the following page shows the differences between the actual distribution of Employers Mutuals workmen's compensation risks by size and the theoretical distribution obtained by using a compound Log-Gamma as Mr. Hewitt suggests in Appendix 2. In this case the $a_2$ and $p_2$ were determined by using the method outlined in Appendix 1. The "h's" turned out to be .861 for the distribution of business by amount of premium, and .466 for the distribution of the number of risks.

Typical of the authors of many good mathematical textbooks, Mr. Hewitt assumes a rather high degree of mathematical proficiency on the part of his readers, and leaves the reader on his own to supply some of the missing proofs.

On page 107, for example, he says the following: "if $T(x, a, p)$ repre-