equivalent to *1.0,* the requirement of any p.d.f. Then, since the amount of premium is e^x, the frequency function for the number of risks is $x^p e^{-(a+1)x}$ upon dividing by the amount of premium. By integrating this frequency function and fulfilling the requirement that the integral equals *1.0,* we do "easily" obtain $T(x, a+1, p)$.

Mr. Hewitt's fine narrative on "fitting the data" in Appendix 1 would have been enhanced, at least for the average reader, if he had seen fit to include some of his worksheets used in obtaining the tables in the paper.

This paper is a valuable addition to our *Proceedings* despite the minor points just raised. We hope that Mr. Hewitt, and others, will continue to share their research with us.

DISCUSSION BY ROBERT L. HURLEY

While this paper, so suggestive of an austere scholarship, may seem directed to those of the avant-garde who delight in frolicking among the outer reaches of actuarial theory, Mr. Hewitt presents both a challenge and a promise to those members whose interests, like this reviewer's, may gravitate more towards the application of actuarial principles to current underwriting and rating problems.

This paper shows that the distribution by size of both the workmen's compensation standard premium and the number of policies* may be fairly described by a Log Gamma equation. It also suggests that certain workmen's compensation expenses may vary by size of risk according to a similar pattern. There is the intimation (which particularly interests this reviewer) that loss distributions may follow the same law, using the latter term in its least restrictive sense.

A quick check on Mr. Hewitt's findings by premium size (c.f. Table I) reveals a close fit of the actual to theoretical values, according to the Pearson Chi-Square or even the possibly more critical Kolmogorov-Smirnov test. While references were afforded the reader on the Gamma function, the author was understandably more interested in the potential significance of his findings to actuarial theory than in detailing the mathematics, some of which is available in the standard literature. This "Hoc age" (up and at it) approach which is not infrequently so characteristic of the scholar can be oftentimes bewildering and even exasperating to the less specialized reader.

* As given in Exhibit I of the National Council on Compensation Insurance's Report of the Special Committee to Study Expenses by Size of Risk.

It might, therefore, be not inappropriate for the reviewer to fill in with certain details which he has been able to find in the literature and to add some comment on the problem of graduation methods versus risk theory, drawing chiefly on the work of others with which, in some instances, he has had only the most casual relationship, and in others, no personal contact at all.

As Mr. Hewitt noted, the Gamma Distribution is sometimes referred to as the Pearson Type III Curve. It may be recalled that at the turn of the century Karl Pearson suggested that most of the familiar uni-modal frequency distributions could be generated by varying the numerical coefficients of a differential equation whose numerator was a linear and whose denominator was a quadratic expression in x.

The basic equation is of the general form $\frac{dy}{dx} = \frac{y\,(m-x)}{a+bx+cx^2}$

When the coefficient c equals zero, $\frac{dy}{y} = \frac{(m-x)\,dx}{a+bx}$

or $\frac{dy}{y} = -k_1\,dx + \frac{k_2}{(x-r)}dx$; And integrating: $y = Ae^{-k_1 x}(x-r)^{k_2}$

Or letting $k_1\,(x-r) = w$

$y = Be^{-w}w^{k_2}$, which is the general form of the Gamma equation given in Mr. Hewitt's paper.

The Gamma function is commonly represented as a skew shaped curve where y has its peak value at the lower end of the x scale and drops off towards zero as the x value approaches infinity. It will be sensed intuitively that the contour of such a curve might well fit the type of data, policies and premiums by size groups, with which Mr. Hewitt was working.

Now lest it be thought that the Pearson system is solely a fabrication out of sheer fancy with no foundation in reality, it should be noted that the basic differential equation cited above can be developed out of those quite practical problems as figuring the chance of getting a full house in a poker game. And the familiar Normal Curve $y = k_1e^{-k_2x^2}$ results from assigning zero values to the b and c coefficients of the x values in the denominator of Pearson's differential equation.

In many actuarial problems, reasonably satisfactory predicative statistics can be developed by recasting the original data so that tables of the probability integral (i.e. the normal curve) may be used. On occasions it is found that while, for example, the number of losses y by dollar size group x will not fit the normal curve, a reasonably good fit can be obtained by transforming to *log* x or the logarithm of the dollar loss size.

The substitution of the *log* x scale may tend to reduce both the variance and the skewness of the distribution. It is demonstrated in mathematical texts that while *log* x approaches infinity as x increases, it does so more slowly than any polynomial in $ax^{n_1} + bx^{n_2} + \ldots$ no matter how small a positive fraction n may become. Consequently, the substitution of log functions sometimes renders the data more tractible to mathematical analysis, and this seemingly was a consideration in Mr. Hewitt's decision to use the log of the gamma function.

Mr. Hewitt's paper, it is believed, represents another significant advance by the proponents of the mathematical theory of risk school in the search for a constantly more precise analytic expression for the actuarial principles underlying the casualty and property insurance business.

It may be recalled that in his review of the paper on Table M in Volume LII of the *Proceedings,* Mr. Hewett stressed the need (in support of Mr. Simon's conclusion) for determining the basic nature of underlying loss patterns rather than perpetuating the customary practice of collecting a series of observations and by some subtle ingenuity, but more commonly through the mere drudgery of actuarial sweat, devising an equation that would fit tolerably well. In this regard, it may be helpful to take just one business problem commonplace to many company actuaries, trace some intermediate solutions, and see it emerge as one of the basic situations demanding the attention of those who are interested in the possible applications of the mathematical theory of risk.

Many years ago, now, a company about to file an individual risk rating plan for fire insurance was induced to research the possibility of incorporating an optional deductible (i.e. up to \$5000) feature as "natural" for large accounts with 25 or more locations. A number of the then actuarial students were set to scurrying about the statistics to see what \$1000 to \$5000 deductibles were worth by line size.

The Loss Elimination Ratios (LERs) were computed for each deductible line size and an attempt was made to fit the observations to a rectangular hyperbola with the axes rotated minus 45°, or a curve of the general form $xy = k$. The fit was so unsatisfactory at the upper reaches of the insurable values that it was decided to draw a curve that would best fit the observed points solely on an eye control.

Somewhat later, when another company came out with a considerably less modest deductible program, additional data were taken off to check the comparative rate credits. Combining the latter data with the statistics from the earlier study, one of the investigators found that the observed LERs could be made to fit more closely a theoretical curve by changing the equation from $y = \frac{k}{x}$ to $y = \frac{k}{a + bx}$

Within the last few years many of the fire rating bureaus have filed deductable rating plans wherein the observed LERs by line size for various fire insurance deductibles from \$500 to \$75,000 will be found to fit reasonably well (at least within the range of values for which readings were available) a graduation equation of the general form $y = \frac{k}{(\log x)^n}$

And in Volume LII of the *CAS Proceedings* Mr. Simon's very readable exposition of the mathematical research underlying the 1965 revision of Table M relates that after testing some 25 different equations, it was found that the insurance charges were best described by an equation of the form $\phi(r) = 1 / (1 + r + b_2 r^2 \ldots b_7 r^7)$ where r equals the adjusted ratio of actual to expected losses.

Now, these previous references, covering different samples, different times, different coverages, all tend to describe insurance loss distribution by size as a pattern which might be generalized into an equation of:

$$y\, f(x) = k$$

Mr. Hewitt's use of the log gamma might conceivably be viewed as a further generalization on this equation with the substitution of a second variable in x for the constant k—so that the revised equation becomes $c\, y\, f_1(x) = f_2(x)$; *with* $f_2(x) < f_1(x)$ as $x \rightarrow \infty$. With the following equivalences to Mr. Hewitt $y = T$; $c = \frac{\Gamma(P+1)}{a^{P+1}}$,

$f_1(x) = e^{ax}$; $f_2(x) = x^P$ in Mr. Hewitt's first equation

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

On occasions all of us are probably bothered by the mathematical creations sometimes erected to explain situations that on the surface, at least, appear quite simple. As a case in point, we might take the basic equation $y\, f(x) = k$, discussed above.

When $f(x)$ equals x this equation reduces to an expression which is equally applicable to Boyle's Law of Gases, or to the area of a rectangle, or to any situation explaining the variation of two factors whose product tends to be constant. On the face of it, such a situation can be thought analogous to the distribution of expected losses or excess loss ratios by size since as the size increases, the expectancy decreases—but not in a straight line down to zero for some fixed x less than infinity.

And yet, on testing, we sometimes discover that the easy explanation just does not fit the facts. Consequently, additional elements must be sought to account for the underlying phenomena at play.

But on occasions this attempt to fit the mathematics to the observed facts, even with the additional data, does not work out too successfully. In such a situation the attack on the problem must be redirected, and our mathematical horizons widened.

This, as I understand it, is the goal of the Mathematical Theory of Risk school, and Mr. Hewitt's paper might be regarded as a particular approach, of some promise, to the insurance industry's possible needs area of the distribution of risk and maybe losses by size.