Notes on the Theory of Schedule Rating.

BY

ALBERT W. WHITNEY.

Mathematics is too often thought of, by the unmathematical at least, as concerned merely with numbers. In reality there are extensive demesnes in the field of mathematics in which number is of no concern. But even in those parts of mathematics which involve number the numerical result is usually of secondary importance; the fundamental interest is with the general form from which the numerical result springs as a special case. Mathematics is essentially the study of form and structure, and not only that but it is the study par excellence of form and structure.

The present paper deals with the mathematics of schedule-rating in the sense that it is a consideration of the form which a schedule should take. A schedule in the sense here used is the expression of an insurance premium-rate as a function of the elements contributory thereto. As a rate can be figured both prospectively and retrospectively it is pertinent to say that it is the prospective method which is referred to here. The rate is the premium divided by the payroll, the insurance in force or whatever the basis may be upon which the policy is written. The premium is an expectation, made up in general of the sum of other more elementary expectations. A simple expectation is the product of a contingent amount and the probability of the contingency in question.

While probabilities themselves cannot be added unless the corresponding events are mutually exclusive, this restriction does not apply to the case of expectations. This may be shown as follows: Suppose two events a and b not mutually exclusive. In terms of symbolic logic we have:

$$a \boldsymbol{+} b = a\bar{b} \boldsymbol{+} \bar{a}b \boldsymbol{+} ab. \tag{1}$$

Logical addition is here distinguished from numerical addition by a heavy symbol; logical addition is interpretable as "or"; logical multiplication is interpretable as "and"; $\bar{x}$ means not x. The right-hand side consists of mutually exclusive terms and to it can therefore be applied the principle that the probability of a logical sum is equal to the numerical sum of the probabilities. Denoting the probability of the event x by $|x|$ we have therefore:

$$|a \dotplus b| = |a\bar{b}| + |\bar{a}b| + |ab|. \tag{2}$$

Suppose that the amount payable in case the event a alone happens is α, in case the event b alone happens is β and in case they both happen is $\alpha + \beta$ and assuming that we may apply the law of addition of expectations to mutually exclusive probabilities we have: Expectation in connection with the happening of the event $a \dotplus b$

$$= |a\bar{b}|\alpha + |\bar{a}b|\beta + |ab|(\alpha + \beta), \tag{3}$$

which reduces evidently to $|a|\alpha + |b|\beta$. The law is therefore independent of whether or not the events a and b are mutually exclusive; this can of course be extended to any number of terms.

This law is strictly true for expectations of events not mutually exclusive and approximately true for the probabilities themselves provided they are small, as they usually are in the case of insurance, and provided the events are independent or approximately independent.

The form of the function expressing the hazard differs for different kinds of insurance. In fire insurance the process of damage production whose expectation is to be measured can readily be recognized as separable into three independent events or processes, viz.: ignition, combustion and damage production proper; the first two are probabilities, the last an expectation. These are related to each other dependently: unless they all concur there is no loss. There may be no ignition; there may be ignition but no combustion; there may be ignition and combustion but no damage produced. The probabilities are therefore related multiplicatively; that is, the probability of ignition is to be multiplied by the probability, if there is ignition, of combustion, and this is to be multiplied by the expectation of damage, if there are ignition and combustion. Each of these factors is expressible as a sum in view of the laws referred to above and in view of the fact that ignition, combustion and damage-production can all take place in various ways. If this analysis*

* This subject was more thoroughly discussed by the writer in Vol. 12, p. 28, of the *Transactions of the Actuarial Society of America,* and Vol. 85, p. 306, of the *Weekly Underwriter.*

were made in fire insurance schedules I am satisfied the results would be more satisfactory than under the present schedules which are found to have a very restricted region of applicability. The case of fire insurance is introduced here only as an example of the fact that each kind of insurance demands a separate analysis.

So far as I know all present rating schedules are based upon the principle of building the rate up additively item by item from a given basis rate. This can be given an interesting mathematical expression. It is an application of Taylor's theorem for several variables:

$$f(a+h, b+k, \cdots, \text{etc.}) = f(a, b, \cdots, \text{etc.}) + h\frac{\partial f}{\partial a} + k\frac{\partial f}{\partial b} + \cdots, \text{etc.} \tag{4}$$

The expression on the left represents the rate in question; $a, b, \ldots$, etc., are the values of the parameters which describe the standard condition for which the basis rate is $f(a, b, \ldots, \text{etc.})$. $h, k, \ldots$, etc., express the deviations from standard of the particular risk. The higher powers are omitted. Those who made the rating schedules doubtless did not have this formula in mind and the schedules would come far from checking up in detail and yet the formula is unquestionably the rigorous expression of the general idea which they followed.

Two general lines of procedure have been adopted with regard to the coefficients $\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}, \cdots$, etc.

In the Moore fire schedule these are for the most part assumed to be constants, that is the rate is assumed to change uniformly as the conditions change. In the Dean fire schedule, on the other hand, for the most part the coefficients are taken as proportional to the basis rate, that is the more hazardous the risk the greater the effect of a change in conditions. This latter procedure is equivalent to making the rate an exponential function of the parameters.

One very significant fact stands out at once. The Taylor expansion, particularly if the higher powers are dropped off, is limited in its application to conditions not far different from those upon which the basis rate are predicated. The schedule will break down if it is stretched to cover too wide a field. The regular fire schedules for instance cannot be applied to sprinklered risks. In fact the fire companies have found it necessary to have as many as a score of schedules adequately to cover the field.

In what follows I shall now confine myself to a discussion of rating for Workmen's Compensation insurance. It is evident that the compensation premium is susceptible of a very considerable analysis. In the first place it is a summation of the expectation of loss for each of the employees separately. In the second place for each employee it is a summation of expectations with regard to all possible injuries; for example, the compensation for loss of arm multiplied by the probability of loss of arm plus the compensation for loss of eye multiplied by the probability of loss of eye, etc., through the whole list of possible injuries. Thirdly the expectation of each employee may be analyzed on the basis of cause or hazard. There seem to be three general types of hazard, first the catastrophe hazard, second the general hazard of the industry and third the peculiar hazards to which particular employees are exposed. Very likely still further differentiations might with advantage be made.

At this point it is well to point out what is already quite evident to anyone who has given this matter any considerable attention, namely that the subject is an exceedingly difficult one and that, while it is well to have an ideal in mind as a guide, one must be prepared at almost every point to make simplifying assumptions, oftentimes violent, in order to prevent the problem from being unmanageable because of its complexity.

The catastrophe hazard is fairly simple. Certain features of the risk affect the rate only through their influence in bringing about a catastrophe. This influence in the case of standard conditions is measured by a certain part of the basic rate. In the given risk however let us suppose that there are certain sub-standard conditions. The corresponding increase in the catastrophe hazard should be independent of the basis rate. For example, a weak floor is just as great a hazard in an otherwise safe industry as in a dangerous one. This means that the coefficients in the expression for the catastrophe hazard should be constants.

The hazard of the industry however is different. There the presence of sub-standard features will be more dangerous the more dangerous the occupation. This must be taken by and large; the truth of the principle as a whole however is clear. Insufficient light where there are dangerous machines or other hazardous conditions is far more serious than where the conditions are less hazardous. It seems reasonable to assume that the coefficient expressing this shall be taken proportional to such hazard; that is in an industry where the hazard of the industry is twice that of another the charge for insufficient light should be doubled. This in effect is to say that the charge for sub-standard conditions should be a percentage of the hazard of the industry, or that the coefficients in the expression for the hazard of the industry should be proportional to that part of the rate which describes the basic hazard of the industry. As a practical matter however the hazard of the industry is not separately given in the basis rate and we shall not seriously err if we make the charge proportional to the basis rate as a whole, especially since the hazard of the industry is doubtless in general the largest part of the hazard.

The third element consists of the special hazards. If it is assumed that there are certain special hazards, exposing a limited number of employees however many the whole number of persons employed, then the premium for this part of the hazard should be the expectation for this exact number of employees and should therefore be an absolute constant independent of the payroll as a whole. Such a condition as this might arise in for instance the case of a flight of stairs which would not in the nature of things be used by more than a limited number of employees. Since in this case the premium must be an absolute constant it follows that the rate must be a constant divided by the payroll.

The coefficients then in the expression for the catastrophe hazard will be of the form l, a constant; in the expression for the hazard of the industry they will be of the form mR, where m is a constant and R is the basis rate; in the expression for the special hazards they will be of the form $\frac{n}{P}$ where n is a constant and P is the payroll on which the premium is computed. In any particular case the constants l, m, n will depend upon the particular hazards under discussion.

The general expression for the rate will therefore be: R' (the rate for sub-standard conditions) $= R$ (the basis rate) $+$ the sum of terms of the form $lh_1 +$ the sum of terms of the form $mRh_2 +$ the sum of terms of the form $\frac{nh_3}{P}$, where h_1, h_2, h_3 are the values of the departures of the parameters from the standard. Indicating the results of the summations by large letters we have:

$$R' = R + L + MR + \frac{N}{P}. \tag{5}$$

This investigation of the subject of schedule rating from a mathematical or at least quasi-mathematical point of view was undertaken without any idea of where it would lead. It is gratifying to me that it should have come out in harmony with the principles underlying the Universal Analytic Schedule. It seems to give some added weight to the schedule to have had the same results reached independently along two somewhat different lines of approach.

It should be clearly understood that the mathematical basis for schedule rating is brought forward not at all with the idea of forcing the schedule into this exact form but as a guide. Schedule rating is an art, with a scientific basis to be sure, but there must be a liberal admixture of judgment and tolerance for a considerable amount of empiricism. I should not want to insist for instance that the parameters, h, k, etc., must in practice enter into the schedule linearly; in fact in the case of height of buildings for instance I know that they should not.

Several observations may be made regarding equation (5). For one thing, when this is thrown into the form:

$$\frac{R' - R}{R} = \frac{L}{R} + M + \frac{N}{PR}, \tag{6}$$

it affords an explanation of the fact, shown in Mr. Senior's paper, that the greatest percentage reduction in rate is in connection with the low-rated risks. Secondly it suggests the possibility of determining the constants L, M and N. It would be a simple matter to determine these from the data which Mr. Senior has made use of. This determination would show how the different elements of the schedule work out in practice. To check this it would be necessary to determine them also by accident experience. This could be done by the use of a body of experience classified both as to nature of injury and cause; the nature of injury would serve to determine the cost or weight and the cause would determine in each case whether the accident were to be thrown to the determination of L, M or N.

One serious difficulty suggests itself however. L, M and N are increments. The constants determined by the causes describe the whole hazard and not merely the increments. To determine L, M and N it would be necessary therefore to make some further assumption.

I believe however that this would be in general a very fruitful line of research. In fact I believe it is practically the only way to apply statistics to a schedule. It would be out of the question to attempt to determine each of the items of a schedule separately. If we can satisfy ourselves that it is right in its larger features we must expect to put in the finer shadings by judgment.