How Extensive a Payroll Exposure is Necessary to Give a Dependable Pure Premium?

ALBERT H. MOWBRAY.

The answer to this question, of fundamental importance in dealing with compensation rates, depends upon the answer to two others: (1) What are the characteristics of a dependable pure premium? (2) What factors tend to make a pure premium derived from experience undependable?

A complete answer to the first question should take into consideration a variety of factors, such as surplus, loadings, character and distribution of business, etc. It is not our present purpose to consider these factors as to the proper allowance for which wide differences of opinion will probably be found. For the purposes of the present discussion the following definition will suffice. "A dependable pure premium is one for which the probability is high (at least equal to an assigned value) that it does not differ from the absolute (true) pure premium by more than an arbitrary limit which may be selected in view of the other factors referred to."

Among the factors to be enumerated in answer to the second question are: Chance variations in the incidence of claims, number, size and character of the establishments entering into the experience, changes in legal conditions, changes in moral conditions, age of the compensation act in question, etc. It is our purpose to consider only the first of these, or what is sometimes referred to as the element of mathematical risk. It would seem that this question must be settled before the others can be appropriately taken up and by confining our attention to it we are enabled to analyze it by mathematical methods free from the confusing applications which arise when the other factors are weighed and considered.

The subject of mathematical risk was among those discussed at the Sixth International Congress of Actuaries in Vienna in 1909, but the discussion was primarily from the life insurance point of view and centered around the question of reserves to be built up as a safeguard against misfortune from this cause.

Although the point is open to some question, for most purposes, I think, we may properly consider the pure premium as made up of several elements, each having an independent probability of its own and each of which may therefore be properly considered, for purposes of discussion, alone and apart from the others and the results appropriately combined.

If it be assumed, for purposes of discussion, that in every case of fatal accident dependents are left of a certain degree of dependency, we then have a very simple case of an event which may or may not occur and when it does occur produces a certain cost. Such a case presents the problem under discussion in its simplest form. The principle having been developed under such conditions, approximations will suggest themselves for reaching such of the more complex cases as must be considered.

Let us assume that the unknown true probability of fatal accident in a given classification is q and that our experience includes n full time workers per year. Then from elementary probabilities the most probable number of fatal accidents will be the greatest integer in $(n+1)q$, which is generally the same as the nearest integer to nq. Likewise from the elementary probabilities if nq is an integer, the probability of exactly nq fatal accidents is

$$^{n}C_{nq}p^{np}q^{nq},$$

according to the usual notation p being $(1-q)$. Further the probability that the number of fatal accidents will lie within 10 per cent. of nq either way is

$$\sum_{r=0.9nq}^{r=1.1nq} {}^{n}C_{r}p^{n-r}\cdot q^{r}.$$

From this it follows that if l be the probability that the number of observed deaths arising out of a number of exposures under observation will not differ from the most probable by more than k per cent. the number of exposures under observation is given by solving for n the equation

$$\sum_{r=(1-k)nq}^{r=(1+k)nq} {}^{n}C_{r}p^{n-r}q^{r} = l. \qquad (1)$$

If it were necessary to solve this equation the determination of the problem under consideration would be little if any advanced. It can, however, be shown (e. g., see Bowley, "Elements of Statistics," p. 275 et seq.) that for a limited range and when the values of p and q are neither very small the expansion of the binomial $(p+q)^n$ for large values of n approximates very closely the normal error curve

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2} \tag{2}$$

when the origin is placed at the most probable value.

In other words, for a limited range of values on each side of the most probable, departures from that value conform closely to the "law of error." Hence we may write in place of equation (1)

$$\frac{h}{\sqrt{\pi}} \int_{-knq}^{+knq} e^{-h^2x^2} dx = l. \tag{3}$$

In equations (2) and (3) h, the measure of precision, equals $\frac{1}{\sqrt{2npq}}$. If in (3) the variable be changed to $t = hx$ the equation becomes

$$\frac{1}{\sqrt{\pi}} \int_{-khnq}^{+khnq} e^{-t^2} dt = l \quad \text{or} \quad \frac{2}{\sqrt{\pi}} \int_{0}^{khnq} e^{-t^2} dt = l. \tag{4}$$

This last integral is of sufficient importance in mathematical work of various kinds that its values have been calculated and are available in various places. For example, Bowley (opus cit., p. 281) gives a table of the values of

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{0}^{x} e^{-x^2} dx,$$

for values of x differing by .01 up to 1.50, and by .02 up to 2.00 for which $F(x) = .498$, its maximum value being .5. More extensive tables are available if needed. By use of such tables a complete solution is possible. For example, let the value of l be .90, i. e., the probability being 9 in 10 that the variation k will not be exceeded, then

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{0}^{khnq} e^{-x^2} dx = .450, \tag{5}$$

from which by the table,

$$x = khnq = 1.16, \tag{6}$$

$$\frac{knq}{\sqrt{2npq}} = 1.16, \quad n = 2\left(\frac{1.16}{k}\right)^2 \cdot \frac{1-q}{q}, \tag{7}$$

and from this q being known and k, the admissible variation having been determined upon, n can be computed.

From equations (5), (6) and (7) the following conclusions may be drawn:

1. For any particular value of q, l being fixed, n the number of exposures which must be observed, varies inversely as the square of k, the limit of admissible variation.

2. For fixed values k and l, n varies approximately inversely with q.

3. Since (x) increased more rapidly than $F(x) = l$, for fixed values of k and q, n varies directly with l in a ratio exceeding the square.

4. The values fixed for the limit of admissible variation, and for the probability of confinement of variation within such limits as necessary to make the pure premium dependable, are of much greater weight in determining the extent of data required than the probability of occurrence of the event.

Up to this point our analysis has proceeded as though q were a known quantity when, in fact q is the value we seek from experience, and is the element itself whose accuracy, as given by such experience, we are testing. For the border line cases where it is a close question whether the observed data is or is not sufficient to give a pure premium which will be dependable within the limits adopted, this creates an awkward situation and light from other sources must be sought. For most practical work, I think the value of q derived from the experience under review will be a satisfactory first approximation for the purposes of the test proposed.

Since the probabilities of the several contingencies giving rise to compensation losses are mutually independent, it is as reasonable to suppose that chance variations such as we have under consideration will tend to offset each other as to suppose they will be cumulative. In order therefore that the pure premium for the combined elements be pronounced dependable it hardly seems necessary that the probability, that the pure premium for each element does not differ from the true pure premium for that element by more than k per cent., reach the required standard, but only that there be such a probability that the pure premium for each element does not differ from the true pure premium for that element by more than a figure which is rather less than k per cent. of the aggregate pure premium.

A few numerical examples will probably make the foregoing theoretical discussion clearer, and will illustrate the way in which this error may be properly applied:

In the *Market World and Chronicle* of November 30, 1912 (new series, Vol. IV, No. 22, p. 67), is given a hypothetical table prepared by Professor Whitney for the California Industrial Accident Board, according to which the probability of accident (all kinds included) is .06, that of fatal accidents .0006, and of temporary disability .0552.

If the standard of dependable pure premium is taken to be that there is a 90 per cent. probability that variations will be limited to within 10 per cent. of the most probable, the value of k in (7) is .1, and for temporary disability q is .0552, and $(1 - q)$ is .9448. Hence, the number of employees required to be observed to find such a pure premium for temporary disability only is 4,605. If the average annual wages are $600 this means a payroll exposure of $2,763,000. Using the same standard with reference to the cost of fatal accidents above n becomes 448,264, a payroll of approximately $270,000,000.

Professor Whitney has determined the pure premium for the present California Act, on a hypothetical basis set up, as .5342 weeks' wages per man, of which .0883 weeks' wages are the fatal accident cost. A variation of 10 per cent. of the total pure premium (.05342) is about 60 per cent. of the fatal accident pure premium, hence, in (7) k may be taken as equal to .6, and n becomes 12,452, a payroll of approximately $7,471,200.

Using this value of n, taking q as .06 and solving (7) for k we find that the probability is 9 in 10 that the total number of accidents will not vary more than about 6 per cent. from the most probable. This is probably too large a variation to be coupled with a variation in the death cost of 10 per cent. of the aggregate pure premium. Hence a considerably larger payroll than $7,500,000 should be available to define a pure premium for these conditions with the limits set.

The theory here discussed may be applied in a different way to the following problem. Given a rate derived from experience with a certain number of exposures, what is the chance that the difference between such rate and the unknown true rate lies within certain limits? As an example the textile rate of $0.23 on Massachusetts Schedule Z may be taken. This is based upon a payroll of $86,339,122. The report of the Industrial Accident Board for the first year estimates the number of employees in cotton mills and woolen and worsted mills at 166,632 which at $10 per week would almost exactly give this payroll. Schedule Z combines death and dismemberment losses at $43,195 out of total incurred losses of $201,095. The Industrial Accident Board data shows 17 fatal accidents; 9 fifty week cases; 22 twenty-five week cases and 99 twelve week cases. This indicates that about $32,000 was the cost of fatal cases and about $11,000 of specified indemnities so that the total cost was about 6½ times the cost of fatal acidents. Again using the Industrial Accident Board data, 17 fatal accidents among 166,632 employees gives a probability of fatal accident of $.0001 = q$, $p = .9999$, $n = 166,632$. Hence for any assigned value of k we can compute the limit of integration in (4) and from the table find the corresponding value of l, the probability that the variation in the number of fatal accidents will not cause a change in the pure premium exceeding k.

The following table shows the probability that variation in fatal accident cost will be confined within certain limits expressed as percentages of the total pure premium

Percentage limit (k).	$khnq = (x)$.	$2F(x)$ = probability variation is confined within selected limit (k).
10.0	1.948	.994
5.0	.974	.832
2.5	.487	.509

As the weekly indemnity columns include payments and reserves on account of permanent disability as well as temporary disability, it is impossible to form such a table for any other element. It will be interesting however to form such a table for variation in the rate of accidents (all kinds included) which from the Industrial Accident Board data is approximately .06497.

Percentage limit (k).	$khnq$.	$2F(x)$ = probability variation is confined within selected limit (k).
10	7.610	Practically 1
5	3.805	Practically 1
2.5	1.903	.992
1.0	.761	.719

Even though from tests of this kind it be found that a pure premium, as for example, the textile pure premium in the Massachusetts Schedule Z, is dependable within a satisfactory definition, it should not be assumed that another experience will develop a pure premium not differing from that tested by more than the limit set or even that it will *probably* do so unless the conditions remain the same. A variation in excess of the limit would seem to be a warning that some change in underlying conditions has probably taken place, such, for example, as a gradually developing tendency to more fully claim compensation benefits.

The data in such cases should accordingly be carefully studied to see if other evidence pointing to such changed conditions can be found.

In closing it should perhaps be pointed out that according to the mathematical law of error the probability of catastrophic losses is infinitesimal. Hence pure premiums found by these tests as dependable cannot be considered to cover this risk, and the tests, unless greatly modified, are not available for dealing with such industries as coal mining, blasting, etc., where the catastrophe hazard is high.

Note.—Since this paper was written the author has found that there was a somewhat similar discussion of this problem in the paper "On the Philosophy of Statistics" by Woolhouse, *J. I. A.*, XVII, 37.