GRADUATION OF EXCESS RATIO DISTRIBUTIONS BY THE METHOD OF MOMENTS

BY

LEWIS H. ROBERTS

The Place of Excess Ratio Distributions in Casualty Insurance Rate Making

A risk who wishes to be self-insured to a degree, and whose size as measured by expected losses is sufficient to make it practicable, may elect to have his premium based in part on actual losses up to a specified limit. The balance of his premium would consist of charges by the insurance carrier for claim service and other carrier expenses plus a charge for the expected or average value, based on the experience of many risks, of losses in excess of the specified limit. Where the specified limit is a stated percentage of total expected losses, the ratio of expected losses in excess of that percentage to the total expected losses is called an "excess pure premium ratio" or more briefly, an "excess ratio" or "charge." Likewise the risk may elect to forego the full reduction in premium that would otherwise result in event a very low actual loss ratio should be incurred, in which case his premium would be adjusted up to a specified minimum percentage of the standard premium to reflect a saving to the carrier equal to the expected value of losses in excess of actual losses. The difference between the charge and the saving for selected maximum and minimum loss ratios is the net insurance charge. The standard premium is the premium that would be paid in the absence of any plan for basing premium on the actual losses of the risk. A rating plan which bases premium on actual losses is called a "retrospective" rating plan.

In order for such a plan to be equitable it is necessary for the carriers to calculate from a large body of experience the expected ratios to total losses of losses in excess of any specified loss ratio for risks of every size. From these calculations a table of charges and savings can be prepared for rating any risk under a retrospective rating plan. The table used currently for this purpose by the principal carriers is named Table M.

Previous Treatment of the Subject

In his paper entitled "On Graduating Excess Pure Premium Ratios", (P.C.A.S. Vol. XXVIII) Mr. Paul Dorweiler showed how indicated excess ratios calculated directly from actual data could be graduated for varying specified loss ratios for a given amount of expected losses and how they could be graduated for varying expected loss sizes for a given specified loss ratio. It was on the basis of his work that Table M was prepared from the 1934–37 experience of New York State Workmen's Compensation Risks.

In "Sampling Theory in Casualty Insurance", (P.C.A.S. Vol. XXX P. 56)

Mr. Arthur L. Bailey stated the linear relationship that exists between the sum of the charges in Table M and the variance of the loss ratios of risks with corresponding expected losses. (See page 10, infra.)

This convenient mathematical relationship permitted adjustment of Table M in 1954 to reflect increases in the variance of loss ratios for risks of a given expected loss size, due in large measure to increased claim costs over the average claim cost of the 1934–37 period and the consequent decrease in the number of claims required to produce a given amount of losses.

For this purpose it was necessary to find a formula for estimating the variance of the probability distribution* of loss ratios for a risk of average size from the experience of a group of risks with varying expected losses. The problem of a formula to use for the purpose arose because grouping of risks by size necessarily involves some spread in the size of risks included in any group. A straightforward calculation of the variance of their loss ratios according to elementary formulas would produce an upward bias in the estimate of the variance for a risk of average size owing to the hyperbolic relationship between expected losses and the expected values of the squares of differences between loss ratios and their expected values. The mathematical details of the relationship are covered in the Appendix, Notes 1 and 1a.

On the basis of Mr. Bailey's studies variances corresponding to various expected losses were calculated from the countrywide experience of Policy Year 1950. Table M was accordingly revised to match the calculated variances based on 1950 experience with the variances underlying the columns of insurance charges in Table M as previously developed.

Advantages of the Method of Moments

It is apparent that the so-called "Method of Moments" has already been of great use in studies of Table M through providing, by means of variance calculations, a simple check on the correctness of the totals of the insurance charges. This check, which tests the graduation of charges by size of expected losses, is sufficient where the charges in each column are believed to stand in the proper proportions to one another.

For a more complete check on the table it is necessary to study the manner in which insurance charges are graded from low loss ratios to high as well as from small risks to large risks. Since the direct computation of a table of excess ratios and their subsequent graduation is quite a laborious undertaking without, in the writer's opinion, a very satisfactory solution from either the practical or the theoretical standpoint, it should be worth while to try to extend the method of moments to cover the grading of charges. This method, which has found wide application in many fields of statistics as a tool for describing probability distributions, should make it possible by calculation of a few parameters to produce a graded table of insurance charges from a listing of individual risk experience. It has the further advantage that the economy of parameters required reduces the sampling error in the finished table. With an electronic calculator the labor would be reduced to very little.

*The probability distribution of loss ratios for a risk of given size is mathematically the same as the theoretical distribution by loss ratio of an infinite population of risks with equal expected losses.

EXPLANATION OF SYMBOLS

Quantitative Symbols	
a	accident cost, except in Eq. (9), where it is a constant in the graduating equation for V_R^2.
b_2	coefficient of u in Eq. (9).
b_3	coefficient of u in Eq. (10).
b_4	coefficient of u in Eq. (12).
c	coefficient of u^2 in Eq. (9).
e	base of natural logarithms.
f	expected ratio of losses to permissible losses; estimated value of ER.
g	dummy constant in Eqs. (2.3) et seq.
h	dummy constant in Eqs. (3.2) et seq.
i	ordinal subscript retaining same value as a quantity in Eqs. (29.1), (29.2), (29.3).
m	number of risk size-groups, except in Eqs. (1.5) to (1.7) in which it is the expected value of n.
n	number of risks in a size group, except: (1) in Eqs. (29.1), (29.2), (29.3) where values of the argument (x) are numbered from 0 to n and (2) in Note 1a, where n denotes the number of cases in a sample, and (3) in Eq. (14) where it is an exponent.
n_a	number of accidents.
u	reciprocal of EL, except in Notes 1 and 1a, where it is a dummy variable used for illustration.
v	dummy variable used for illustration.
w	weighting coefficient used in normal equations; equals $f\Sigma X$.
β_1	measure of skewness; equals μ_3^2/μ_2^3.
β_2	measure of kurtosis; equals μ_4/μ_2^2.
μ_n	n^{th} moment of a variable; equals $E(v - Ev)^n$ if v is the variable.
φ	coefficient of correlation; defined in Note 1, Appendix.
σ	standard deviation; equals square root of μ_2.
B	spacing interval for given values of R_o.
K	upper limit of the range of a probability distribution; specifically, the lowest value of R for which charges would be shown in Table M as zero.
L	losses of a risk.
M	number of possible values of n in Note 1a.
N	number of risks in all size-groups of risks.

P	premium of a risk.
R	ratio of losses to permissible losses (loss ratio) for a risk; equals L/X.
S	charge in Table M.
S_0	charge in Table M corresponding to R value of R_0.
S_R	charge in Table M corresponding to R.
V	coefficient of variation, equals the standard deviation divided by the expected value.
X	permissible losses.
Subscripts	a subscript adjoined to any symbol denotes that the value of the symbol associated with the subscript is to be used.
Accents	
$'$	shown over a symbol; denotes the value indicated by experience, without graduation.
$\wedge$	shown over a symbol; denotes graduated values derived from experience.
Operators	
$-$	bar over a symbol; denotes its average value indicated by a sample. In connection with study of experience by size-groups of risks it denotes the average value based on one group.
$=$	double bar over a symbol; denotes the average value based on all size-groups of risks combined.
av.	average value; equivalent to bar over the symbol.
E	expected value; theoretically true average.
Σ	summation. In connection with size-groups it refers to summation over a single group. Subscripts and superscripts denote limits between which summation is to be taken.
$\Sigma\Sigma$	Used here only to denote summation over all size-groups.
$\underline{\overline{\int\ldots\int}}{}_a^b$	where the number of $\int$ strokes is n, denotes the n-fold integral evaluated at b minus the same evaluated at a.

Formulas for Estimating Moments of the Probability Distribution of Loss Ratios for a Risk of Average Size from Experience of a Group of Risks of Varying Size

To permit combination of the experience of risks with different permissible* loss ratios, actual loss ratios will be expressed as ratios of actual losses to permissible losses, which is the basis on which Table M is constructed. The mean or first moment of loss ratios associated with any risk is therefore assumed under Table M to be unity.

For any group of risks the average loss ratio is $\bar{R} = \Sigma L/\Sigma X$ where L is actual losses, X is permissible losses, and R is L/X for individual risks. EL, the expected losses for any risk, is fX where f is an abbreviation for ER, the expected ratio of L to X. If we have only one group of risks with which to deal and the group is sufficiently large to make $\bar{R}$ statistically significant, $\bar{R}$ can be used as an estimator of f. Use of $\bar{R}$ adjusts for excess or deficiency in rate level on the basis of the experience of the group. Possible alternatives to $\bar{R}$ as an estimator of f are unity and $\bar{\bar{R}}$, where $\bar{\bar{R}} = \Sigma\Sigma L/\Sigma\Sigma X$, the double Σ signs indicating summation over a number of groups.ϕ

Variance

The variance or dispersion of the probability distribution of loss ratios for a given risk is defined:

$$(1) \qquad \sigma_R^2 = E(R - f)^2$$

where E denotes expected value.

V_R, the coefficient of variation of R, is defined as σ_R/f. Since the coefficient of variation of any variable is invariant for all multiples of the variable, and L is a multiple of R equal to XR, we have

$$(2) \quad V_R^2 = \frac{E(L - EL)^2}{(EL)^2}$$

Because Table M assumes ER equals unity, the variance of R underlying Table M is the same as V_R^2 in this case. We shall find it most convenient to calculate V_L^2, knowing that $V_L^2 = V_R^2$ ($= \sigma_R^2$ for $ER = 1.000$).

Since EL equals fX we can substitute

$$(3) \qquad V_R^2 = \frac{E(L - fX)^2}{(EL)^2}$$

Eq. (3) follows from the definition in Eq. (1) of the variance of R *for a given risk.* For a risk with expected losses equal to the average loss for the group we can define EL as $\frac{\Sigma L}{n}$. The numerator, however, will require close analysis. It must be estimated on the basis of the experience of all risks in a given group. Only if we can show that the value $\frac{1}{n}\Sigma(L - fX)^2$ based on the group

*The word "permissible," though superseded in current insurance usage by "expected," will be used here to avoid confusion with "expected" in the statistical sense of "average value." In this paper "expected" will be used only in the statistical sense.

ϕUnity and $\bar{\bar{R}}$ are incorrect to use for small risks, however, because the ratio of average losses to permissible losses rises sharply for small risks when experience of risks with no losses is excluded, as is done in developing a table of excess ratios.

is a proper estimate of $E(\mathrm{L} - f\mathrm{X})^2$ for a risk with expected losses equal to the group average is Eq. (2) a valid estimator of $\mathrm{V_R}^2$ for that risk.

Since the proof of this is rather lengthy it is omitted here and given in Note 1 of the appendix.

On the basis of this proof we can use group experience in Eq. (3) to estimate $\mathrm{V_R}^2$ for a risk of average expected losses.

Our estimate of $\mathrm{V_R}^2$ is therefore

$$(4) \qquad \acute{\mathrm{V}}_{\mathrm{R}}^2 = \frac{av.(\mathrm{L} - f\mathrm{X})^2}{f^2(av.\ \mathrm{X})^2} = \frac{av.\ \mathrm{L}^2 - f^2(av.\ \mathrm{X})^2}{f^2(av.\ \mathrm{X})^2} \qquad (^*)\ \phi$$

The operator "*av.*" denotes an estimate of the expected value and is equivalent to the operator Σ/n.

Where f is estimated by the ratio $\overline{\mathrm{R}}$, adjustment must be made for the loss of a degree of freedom by use of the so-called "finite multiplier" $n/(n-1)$.

$$(4a) \qquad \acute{\mathrm{V}}_{\mathrm{R}}^2 = \frac{av.\ \mathrm{L}^2 - (av.\ \mathrm{L})^2}{(av.\ \mathrm{L})^2}\ \frac{n}{n-1}$$

If f is estimated by the ratio $\overline{\overline{\mathrm{R}}}$, less than a whole degree of freedom has been lost in any group. The finite multiplier in that case uses the total number of risks in all groups and

$$(4b) \qquad \acute{\mathrm{V}}_{\mathrm{R}}^2 = \frac{av.\ \mathrm{L}^2 - \overline{\overline{\mathrm{R}}}^2\ (av.\ \mathrm{X})^2}{\overline{\overline{\mathrm{R}}}^2\ (av.\ \mathrm{X})^2}\ \frac{\mathrm{N}}{\mathrm{N} - 1}$$

If f is estimated to be unity, no degrees of freedom are lost so no finite multiplier is required. Then

$$(4c) \qquad \acute{\mathrm{V}}_{\mathrm{R}}^2 = \frac{av.\ \mathrm{L}^2 - (av.\ \mathrm{X})^2}{(av.\ \mathrm{X})^2}$$

Skewness

The skewness of a probability distribution is measured by the statistic β_1, which is invariant with respect to the origin or unit in terms of which a variable is expressed. Because of this invariance, β_1 is the same for losses as for corresponding loss ratios. We shall compute β_1 for losses and use it for loss ratios. β_1 is defined as the square of the third moment divided by the cube of the second moment, or μ_3^2/σ^6. For losses or loss ratios of a risk with average expected losses:

$$(5) \qquad \beta_1 = \frac{[E(\mathrm{L} - E\mathrm{L})^3]^2}{\sigma_{\mathrm{L}}^6}$$

$$(6) \qquad \acute{\beta}_1 = \frac{[av.\ (\mathrm{L} - f\mathrm{X})^3]^2}{[av.\ (\mathrm{L} - f\mathrm{X})^2]^3} \qquad (^{**})$$

The derivation of Eq. (6) is given in Note 2 of the appendix.

**See footnote * in Note 2 of the Appendix regarding the applicability of this equation to small risks.

*An accent over a statistic will denote an estimate of its value based on observed values, without graduation.

ϕSee Note 1a of the appendix regarding the accuracy of this equation for small risks.

Kurtosis

The peakedness (or to be more precise, the lack of peakedness) of a probability distribution is measured by the statistic β_2, which like β_1 is invariant with respect to origin or unit of measurement. We shall compute β_2 for losses and use it for loss ratios. β_2 is defined as the fourth moment divided by the square of the second moment, or μ_4/σ^4.

$$(7) \qquad \beta_2 = \frac{E(\mathrm{L} - E\mathrm{L})^4}{\sigma_{\mathrm{L}}{}^4}$$

For losses or loss ratios of a risk with average expected losses:

$$(8) \qquad \dot{\beta}_2 = \frac{av.\ (\mathrm{L} - f\mathrm{X})^4}{[av.\ (\mathrm{L} - f\mathrm{X})^2]^2} - 3\mathrm{V}_{\mathrm{X}}^2 \qquad (*)$$

where $\mathrm{V}_{\mathrm{X}}^2$ is the squared coefficient of variation of permissible losses within the group. The derivation of Eq. (8) is given in Note 3 of the Appendix.

Graduation of Indicated Moments

Variance

The relationship between $\mathrm{V_R}^2$ and the reciprocal of $E\mathrm{L}$ is practically linear for large risks. For small risks the curve is concave upward due to the u^2 term in the equation:

$$(9) \qquad \hat{\mathrm{V}}_{\mathrm{R}}{}^2 = a + b_2 u + cu^2 \text{ where } u = 1/E\mathrm{L}$$

See Note 1, Appendix, for derivation of this equation.

The constant term, a, is included because only with as yet unattained perfect rating procedures that precisely estimated in advance the expected losses of each risk, would $\mathrm{V_R}^2$ approach zero.

The weights to be applied to calculated values of u and $\mathrm{V_R}^2$ in fitting Eq. (9) should, according to the Theorem on Observation Weights#, be inversely proportional to the sampling variances** of the respective observations. Since the principal element of sampling variance is, like $\mathrm{V_R}^2$, inversely proportional to the expected number of claims underlying the expected losses used, hence to the total expected losses, the weight to be given to each pair of $\mathrm{V_R}^2$ and u values is $f\Sigma\mathrm{X}$ for the group from which $\mathrm{V_R}^2$ and u were calculated. Letting $w = f\Sigma\mathrm{X}$, the normalϕ equations for determining a, b_2 and c are:

$$(9a) \qquad \Sigma w\acute{\mathrm{V}}_{\mathrm{R}}{}^2 = a\Sigma w + b_2\Sigma wu + c\Sigma wu^2$$

$$(9b) \qquad \Sigma wu\acute{\mathrm{V}}_{\mathrm{R}}{}^2 = a\Sigma wu + b_2\Sigma wu^2 + c\Sigma wu^3$$

$$(9c) \qquad \Sigma wu^2\acute{\mathrm{V}}_{\mathrm{R}}{}^2 = a\Sigma wu^2 + b_2\Sigma wu^3 + c\Sigma wu^4$$

Eq. (9) provides $\hat{\mathrm{V}}_{\mathrm{R}}{}^2$, the graduated value of $\mathrm{V_R}^2$, for any given value of $E\mathrm{L}$ when the values given to a, b_2 and c are derived from these normal equations.

---

*See footnotes in Notes 2 and 3 of the Appendix regarding the applicability of this equation to small risks.

ϕEquations from which values of constants are derived according to the least squares method.

**Mean square error of observed values from their expected values.

#Harold Jeffreys, *Theory of Probability*, Clarendon Press, Oxford, 1948, p. 124, and other authors.

Skewness

The procedure for graduating β_1 values is the same as for V_R^2 values except that no constant term is needed since the distribution of R approaches the normal distribution for very large risks. Graduated values of β_1 are given by the equation

(10) $\hat{\beta}_1 = b_3 u$

where b_3 is determined from the normal equation

(11) $b_3 = \dfrac{\Sigma w u \beta'_1}{\Sigma w u^2}$ (*)

Kurtosis

β_2 values are graduated in the same way on the basis of the equation:

(12) $\hat{\beta}_2 = 3 + b_4 u$

with b_4 determined from the normal equation

(13) $b_4 = \dfrac{\Sigma w(\beta'_2 - 3)u}{\Sigma w u^2}$

The constant, 3, included above represents the kurtosis of a normal distribution, with the term $b_4 u$ measuring the excess of kurtosis in the observed distribution over the normal.

Relationship Between Charges in Table M and the Moments of the Underlying Probability Distribution

As pointed out by Mr. Nels M. Valerius in "Risk Distributions Underlying Insurance Charges" (P.C.A.S. Vol. XXIX), the second differences in a table of charges yield the theoretical frequency distribution of risks by size of entry ratio (ratio of actual losses to expected losses). A double integration of the risk distribution, therefore, provides a table of charges.

The mathematics of this relationship are very interesting and are readily extended to include higher moments. We use the reduction formula:

(14) $$\int x^n f(x)dx = x^n \int f(x)dx - n x^{n-1} \iint f(x)dxdx + n(n-1)\, x^{n-2} \iiint f(x)dxdxdx - \cdots + \cdots$$

For $n = 2$ we have

(15) $$\int x^2 f(x)dx = x^2 \int f(x)dx - 2x \iint f(x)dxdx + 2 \iiint f(x)dxdxdx$$

The charge in Table M for a selected loss ratio of R_0 is defined mathematically by the equation

(16) $$S_0 = \int_{R_0}^{\infty} Rf(R)dR - R_0 \int_{R_0}^{\infty} f(R)dR$$

(17) $$S_0 = 1 - \int_0^{R_0} Rf(R)dR - R_0\left[1 - \int_0^{R_0} f(R)dR\right] \quad (\text{since } ER = 1)$$

---

*Note that the normal equation is *not* $\Sigma w\beta'_1/\Sigma wu$ as might be supposed by simple averaging. Eq. (11) is derived by minimizing the quantity: $\Sigma w(\beta'_1 - b_3 u)^2$ according to the principle of least squares. The same principle applies in connection with the normal equation for any ratio estimate (see β_2, following).

On application of Eq. (14) this reduces to

$$(18) \qquad S_0 = 1 + \overline{\iint}_0^{R_0} f(R)dRdR - R_0$$

The charge for a selected loss ratio of R is therefore

$$(19) \qquad S_R = 1 + \iint f(R)dRdR - R$$

Integration and doubling gives

$$(20) \qquad 2\int S_R dR = 2R + 2\iiint f(R)dRdRdR - R^2$$

the constant of integration being zero.

If $f(R)$ is continuous over the finite range $0 \leq R \leq K$ and $\int_0^K f(R)dR = 1$, the following equations hold. The second specification is met to as close a degree of precision as required by choosing K sufficiently large.

$$(21) \qquad 2\int_0^K S_R dR = 2K - K^2 + 2\overline{\iiint}_0^K f(R)dRdRdR$$

Since for Table M $\int_0^K f(R)dR = 1$, $\int_0^K Rf(R)dR = 1$,

$$\text{and } \sigma_R^2 = \int_0^K R^2 f(R)dR \Big/ \int_0^K f(R)dR - 1,$$

we have from Eq. (15) on taking the definite integral:

$$(22) \qquad \sigma_R^2 = K^2 - 2K\overline{\iint}_0^K f(R)dRdR + 2\overline{\iiint}_0^K f(R)dRdRdR - 1$$

Again from Eq. (15) and taking the definite integral:

$$(23) \qquad \int_0^K Rf(R)dR = K\int_0^K f(R)dR - \overline{\iint}_0^K f(R)dRdR$$

$$(24) \qquad \overline{\iint}_0^K f(R)dRdR = K\int_0^K f(R)dR - \int_0^K Rf(R)dR = K - 1$$

$$(25) \qquad \sigma_R^2 = 2K - K^2 + 2\overline{\iiint}_0^K f(R)dRdRdR - 1 = 2\int_0^K S_R dR - 1$$

$$= 2\int_0^\infty S_R dR - 1$$

For values of S_R spaced at intervals of 0.1 or more for R the value of $\int_0^K S_R dR$ should be estimated by Simpson's one-third rule or other non-linear quadrature formulas, but for spacing at intervals of .01 the trapezoidal rule is sufficient. The latter rule gives for a spacing interval of B:

$$(26) \qquad \int_0^\infty S_R dR = 2B\,(\Sigma S_R - \tfrac{1}{2}) - 1$$

or if the charge (unity) for R = 0 is omitted, as in Table M,

$$(27) \qquad \int_0^\infty S_R dR = 2B \sum_{R=B}^{\infty} S_R + B - 1$$

as stated by Mr. Arthur Bailey in the paper previously mentioned.*

The principles used in deriving Eq. (25) when extended to higher moments of R give:

$$(28) \qquad \mu_3 = 6[(K - 1)\int_0^K S_R dR - \overline{\underset{0}{\overset{K}{\iint}}} S_R dR dR] + 2$$

$$(29) \qquad \mu_4 = 12[(K^2 - 2K + 1)\int_0^K S_R dR + 2(1 - K)\overline{\underset{0}{\overset{K}{\iint}}} S_R dR dR + 2\overline{\underset{0}{\overset{K}{\iiint}}} S_R dR dR dR] - 3$$

Equations (28) and (29) have the disadvantage, for purposes of practical computation, that because the values of μ_3 and μ_4 are derived as differences, accurate calculation of small values of these statistics is subject to considerable relative error unless precise values of the several definite integrals of S_R can be calculated.

In evaluating multiple integrals by single quadrature formulas it is necessary to use the calculated values of the $(n - 1)$th integral at the selected values of the argument when applying the quadrature formula to estimate the n'th integral.

---

*Mr. Bailey (page 56) showed the summation as $\sum_{0}^{\infty}$ rather than $\sum_{R=B}^{\infty}$ but this is apparently an error if the positive sign is given to B. If $\sum_{0}^{\infty}$ is used the sign of B must be negative. In its memorandum dated November 12, 1952, in which the method used in the 1953 studies of Table M is described, the National Council on Compensation Insurance indicated the summation $\sum_{1}^{\infty}$ and showed sums of charges for various expected loss sizes. The figure shown for \$300,000 expected losses, the only one checked by the writer, reflected summation correctly from R = .01.

The formulas shown below are equivalent to repeated application of the trapezoidal rule in accordance with the preceding paragraph. For completeness, the well known rule for single quadrature is shown first:

$$\text{(29.1)} \qquad \int_{x_0}^{x_n} y\mathrm{d}x \doteq h\left[\frac{y_0 + y_n}{2} + \sum_{1}^{n-1} y_i\right]$$

$$\text{(29.2)} \qquad \overline{\underline{\iint}}_{x_0}^{x_n} y\mathrm{d}x\mathrm{d}x \doteq h^2\left[(.5n - .25)y_0 + \sum_{1}^{n-1} (n - i)y_i + .25y_n\right]$$

$$\text{(29.3)} \qquad \overline{\underline{\iiint}}_{x_0}^{x_n} y\mathrm{d}x\mathrm{d}x\mathrm{d}x \doteq \frac{h^3}{2}\left\{\left[\frac{2n - 1}{4} + \frac{(n - 1)^2}{2}\right]y_0 + \sum_{1}^{n-1} [(n - i)^2 + .5]y_i + .25\, y_n\right\}$$

Conversion of Graduated Moments into a Table of Charges and Savings

Because of the relationship:

Saving = Entry ratio + Charge − Unity, it is sufficient to calculate a table of charges, from which savings are derived by use of this equation.*

Three principal types of frequency functions are available for calculating the probability distribution from the graduated moments, namely Pearson's system of curves, the Gram-Charlier Series and the Edgeworth Series. Pearson's system is recommended here. Elderton's investigationsϕ indicate that Pearson's curves are best adapted to representation of extremely skewed distributions (characteristic of loss ratios for small risks) and approach the normal distribution for such variates as the loss ratios of very large risks. Pearson's curves have the further advantage that they do not develop negative frequencies (as the other series tend to do near the tails of the distribution).

Because the procedure for fitting these curves is published elsewhereϕ, there is no need to repeat it here.

---

*The saving is defined mathematically by

$$\text{Saving} = R_0\int_0^{R_0} f(R)\mathrm{d}R - \int_0^{R_0} Rf(R)\mathrm{d}R$$

By application of Eq. (14) this reduces to

$$\text{Saving} = \overline{\underline{\iint}}_{0}^{R_0} f(R)\mathrm{d}R\mathrm{d}R$$

which is the charge (Eq. 18) minus unity plus the entry ratio, R_0.

ϕElderton, Sir W. P., *Frequency Curves and Correlation*, 3rd Edition, Cambridge University Press, explains the procedure in great detail. Many examples of fitting these curves appear in *Biometrika*.

In performing the double integrations of the risk distributions it is essential to add -1, the constant of integration, to $\int f(R)dR$, and $+1$ to $\int\int f(R)dRdR$ as is seen by differentiating Eq. (19):

(29a) $dS_R/dR = \int f(R)dR - 1$

(29b) $d^2S_R/dR^2 = f(R)$

Use of the trapezoidal quadrature rule for integration with R spaced at intervals of .01 produces the finished table for selected sizes of expected losses. Charges for intermediate expected losses should be calculated by interpolation.

The Problem of Sampling Error

The question as to whether a given volume of experience is sufficient for derivation of a usable table requires that an estimate be made of the sampling error in the final results. The best way to accomplish this in theory is to divide the available experience into a number of parts or sub-samples selected on a random basis so that a given risk has equal probabilities of being included in any of the several parts, and compute the standard deviations from the several sets of values derived from the sub-samples. Where the values of interest are the end products of a long chain of arithmetical operations, however, this procedure is prohibitive in cost unless electronic calculating equipment is available.

A short cut is to compute the sampling errors of certain key statistics. For this purpose we can best choose $\hat{V}_R^2$ since V_R^2 for given expected losses is a linear function of the sum of the charges as noted earlier. The coefficient of variation of $\hat{V}_R^2$ is therefore the relative sampling error in the charges.

The simplest method of calculating sampling errors of the $\hat{V}_R^2$ values is to compute the values of $(\acute{V}_R^2 - \hat{V}_R^2)^2 = s^2$ for each group, which is to say for each value of u used, and fit a curve to plotted values of s^2 and u. Representing this curve by $f(u)$, the coefficient of sampling error of $\hat{V}_R^2$ for a given value of u is estimated by:

(30) $V_{\hat{V}_R^2} = \sqrt{f(u)/[(m - 3)\, g(u)\hat{V}_R^2]}$

where $g(u)$ is the experience-density function described below, m the number of size-groups and 3 the number of constants in Eq. (9).

Values of s^2 may tend to be larger for large values of u, but this is not necessarily the case. It depends on the numbers of risks in the various size-groups. If all size-groups have equal total expected losses, $f(u)$ should be a straight line with zero slope. The distribution of risks by size, however, will ordinarily prevent use of such size groups without introducing excessive ranges of size within certain groups. If the total expected losses in each group increases in proportion to average expected losses (number of risks in each group constant) the curve should be a straight line with positive slope. This procedure leads to wider individual deviations of $\acute{V}_R^2$ from $\hat{V}_R^2$ for large u. The reliability of $\hat{V}_R^2$ in a given region of u values, however, depends on the total weight given to $\acute{V}_R^2$ values of that region in the derivation of Eq. (9), that is, on the total expected losses of the region. The grouping of risks by size, therefore, should be done in a regular way so that the total expected losses corresponding to a given value of u, hence to given average expected losses per risk, will be a smooth function of u, not necessarily expressed algebraically. It can be expressed merely by a graph of the total expected losses for each size-group when plotted against corresponding u values. Denote the function represented by this graph as $g'(u)$. It is also necessary to reflect the spacing of u values. This is done by plotting, against the means $(u_{i+.5})$ of successive u values, the values of $u_{i+1} - u_i$. Denote the function represented by a graph of these points as $\Delta(u)$. The product of the ratio of $g'(u)$ to the average expected losses per group times the ratio of the average separation of u values to $\Delta(u)$ gives the experience density function of u:

$$(31) \qquad g(u) = \frac{mg'(u) \cdot [u_{\text{max.}} - u_{\text{min.}}]}{[\Sigma\Sigma fX] \cdot [m\Delta(u)]} = \frac{g'(u)[u_{\text{max.}} - u_{\text{min.}}]}{\Delta(u)\Sigma\Sigma fX}$$

Choice of size-group ranges will affect $f(u)$ but this effect will be cancelled by $g(u)$ which works in the opposite direction. Narrow groups in a region of u produce unreliable $\acute{V}_R^2$ values, hence large values of $f(u)$, but there will be more values of $\acute{V}_R^2$ in that region so the reliability of $\hat{V}_R^2$ is not reduced. The presence of $g(u)$ in the denominator of the radical of Eq. (30) expresses this by dividing the $f(u)$ values by a proportionately small number.

Values of $\acute{V}_R^2$ will not have an approximately normal probability distribution for size-groups with average expected losses as low as \$1,000 unless a good many—say 100 or more—risks are included in the group. The probability distribution of $\hat{V}_R^2$, however, can be considered normal since it is a kind of average $\acute{V}_R^2$ based on all groups, only the smallest of which need be as low as \$1,000 under the present form of Table M. We are therefore justified in using the normal curve to interpret $V_{\hat{V}_R^2}$ values in terms of the probability of stated percentages of sampling error.

APPENDIX

Note 1

Derivation of formula for V_R^2*, the squared coefficient of variation of loss ratios for risks with average expected losses, estimated from experience of a group of risks of varying size.*

We define

$$(1.1) \qquad \sigma_L^2 = E(L - EL)^2 = E[L - 2LEL + (EL)^2]$$

$$(1.11) \qquad = EL^2 - (EL)^2$$

Since $L = n_a\bar{a}$ and $E\bar{a} = Ea$

$$(1.2) \qquad \sigma_L^2 = En_a^2\bar{a}^2 - (En_a\bar{a})^2$$

We use the coefficient of correlation, $\varphi_{u,v} = \dfrac{Euv - EuEv}{\sigma_u \sigma_v}$, in Eq. (1.2) to give

$$(1.3) \qquad \sigma_L^2 = En_a^2E\bar{a}^2 - (En_aEa)^2 + (\text{terms involving } \varphi)\dagger.$$

Assuming a Poisson probability distribution for n_a, $\sigma^2 n_a = En_a$ but for any

†These terms, dropping the subscript on n, are

$$\varphi_{n^2,\bar{a}^2}\,\sigma_{n^2}\,\sigma_{\bar{a}^2} - 2EnEa\varphi_{n,\bar{a}}\,\sigma_n,\,\sigma_{\bar{a}} - \varphi_{n,\bar{a}}^2\,\sigma_n^2\,\sigma_{\bar{a}}^2$$

Where for a given risk average accident cost (severity) is statistically independent of accident frequency, φ is equal to zero so these terms can be dropped. The assumption of a zero variable, z, $Ez^2 = (Ez)^2 + \sigma_z^2$, so $En_a^2 = (En_a)^2 + En_a$. Also, for any sample of n with mean $\bar{x}$ drawn from a universe with variance σ_x^2, $\sigma_{\bar{x}}^2 = \sigma_x^2/n$. Eq. (1.3) may therefore be written, if we neglect the terms involving φ:

$$(1.4) \qquad \sigma_L^2 \doteq [(En_a)^2 + En_a]\left[(Ea)^2 + \frac{\sigma_a^2}{En_a}*\right] - (En_a Ea)^2$$

Division by $(EL)^2 \doteq (Ea)^2(En_a)^2$ gives, letting $m = En_a$

$$(1.5) \qquad V_L^2 \doteq \frac{1 + V_a^2}{m} + \frac{V_a^2}{m^2}$$

$$(1.51) \qquad V_L^2 \doteq \frac{(Ea)(1 + V_a^2)}{EL} + \frac{(Ea)^2 V_a^2}{(EL)^2}$$

The second term of Eq. (1.5) is negligible for large risks but not for risks with only a few thousand dollars expected losses since with present average claim costs of about \$700 for Workmen's Compensation, m^2 in such cases is not a large number.

The Poisson assumption regarding the probability distribution of the number of accidents was investigated by Mr. John Carleton (P.C.A.S. Vol. XXXII, p. 26). He stated "concern over the application of the Poisson distribution to casualty insurance accidents can be confined to special situations in which accidents are definitely known to be other than independent." We therefore assume the Poisson distribution ordinarily is valid for use in these equations.

To continue, we define V_R^2 as σ_R^2/f^2, $f = EL/EX$ so $EL = fEX$, hence $\sigma_L = \sigma_R\, EX$. Then since $V_L^2 = \sigma_L^2/(EL)^2$, $V_L^2 = [\sigma_R^2/f^2(EX)^2](EX)^2 = \sigma_R^2/f^2 = V_R^2$.

V_R^2 is therefore given by dividing Eq. (1.1) by $(EL)^2$:

For a risk with expected losses of X

$$(1.6) \qquad V_R^2 = E(L - EL)^2/(EL)^2 = [EL^2 - (EL)^2]/(EL)^2,$$

$$(1.61) \qquad = E(L - fX)^2/f^2X^2$$

We substitute for V_R^2 the value in Eq. (1.51) and multiply by $f^2X^2 = (EL)^2$, with $m = EL/Ea$

$$(1.7) \qquad f^2X^2\left[\frac{1 + V_a^2}{m} + \frac{V_a^2}{m^2}\right] = E\,(L - fX)^2$$

Since $EL = mEa = fX$:

$$(1.8) \qquad (Ea)(fX)(1 + V_a^2) + (Ea)^2\,V_a^2 = E(L - fX)^2$$

Eq. (1.8) applies to individual risks. The value of $(L - fX)^2$ for each risk is

value for φ for lines of insurance subject to retrospective rating is believed to be justified as a practical and necessary approximation. Although a risk's adoption of a new process may change the nature of the hazard and temporarily produce a correlation between severity and accident frequency (as by increasing the number of small accidents), there appears to be no *a priori* reason to expect a correlation between severity and frequency in the normal fluctuations of experience.

*The exact value of this term is $\sigma_a^2 En_a^{-1}$, rather than σ_a^2/En_a as shown. With a Poisson probability distribution of n, however, En^{-1} for non-zero values of n is closely approximated by $1/En$. The case of small values of En is discussed in Note 1a.

used as an estimate of its own expected value. Summing over the group of n risks:

$$(1.9) \qquad f(Ea)(\Sigma X)(1 + V_a^2) + n(Ea)^2 V_a^2 = \Sigma(L - fX)^2$$

$$(1.91) \qquad f(Ea)(av.X)(1 + V_a^2) + (Ea)^2 V_a^2 = av.(L - fX)^2$$

Dividing by $[f\ (av.\ X)]^2$ we get

$$(2.0) \qquad \frac{(Ea)(1 + V_a^2)}{f\ av.\ X} + \frac{(Ea)^2 V_a^2}{(f\ av.\ X)^2} = \frac{av.(L - fX)^2}{(f\ av.\ X)^2}$$

The left member of Eq. (2.0) is recognized as V_R^2 from Eq. (1.51) with EL represented by $f(av.\ X)$. The required formula is therefore

$$(2.1) \qquad \acute{V}_R^2 = \frac{av.(L - fX)^2}{(f\ av.\ X)^2} = \frac{av.\ L^2 - f^2(av.\ X)^2}{f^2(av.\ X)^2} \qquad \text{Q.E.D.}$$

As noted in connection with Eq. (4a), (4b) and (4c), finite multipliers are necessary if f is estimated from the same experience as used to compute $av.(L - fX)^2$ or from a larger body of experience which includes it.

The formula used in the 1953 studies of Table M to calculate the variance indicated by the experience of a given group of risks for a risk with average expected losses has not, to the writer's knowledge, previously been published. The worksheets for those calculations were based on the formula:

$$(2.2) \qquad \acute{V}_R^2 = \frac{av.\ P\ av.\ L^2/P - (av.\ L)^2}{(av.\ L)^2} \quad \frac{n}{n - 1}$$

It will be noted that this formula differs from Eq. (4a) in that (1) no recognition is given in this formula for variation between risks in the expected loss ratio, and (2) in the presence of P in the numerator. This writer has been unable to find the theoretical basis for Eq. (2.2) because the expected value of the first term of the numerator is not EL^2, which is needed in Eq. (1.6) above, but a complex expression involving the coefficients of correlation between $av.$ P and $av.$ L^2/P and between L^2 and $1/P$.

Note 1a

Calculation of V_R^2 *for small risks.*

It is elementary that for fixed n, $E\bar{u} = Eu$ and $\sigma_{\bar{u}}^2 = \dfrac{\sigma_u^2}{n}$

With n variable, we achieve sufficient generality by considering n free to take M possible positive integral values, the highest of which is K, not all values of n being necessarily unequal. Then the average of all possible sample means is

$$(2.21) \qquad E\bar{u} = \frac{1}{M}\left[u_{11} + \frac{u_{21} + u_{22}}{2} + \cdots + \frac{u_{K1} + u_{K2} + \cdots + u_{KK}}{K}\right]$$

$$(2.22) \qquad = \frac{1}{M}\left[Eu + \frac{2Eu}{2} + \cdots + \frac{KEu}{K}\right] \text{ with M terms in [ ].}$$

$$(2.23) \qquad = \frac{1}{M}[MEu] = Eu.$$

The expected value of a sample mean is therefore equal to the expected value of the variate, regardless of the probability distribution of the number of cases that comprise the sample.

The average of all possible squared deviations of sample means from the population mean for $n = 1, 2, 3 \ldots . \mathrm{K}$ is

$$E(\bar{u}-Eu)^2=\frac{1}{\mathrm{M}}E\left[(\bar{u}_1-Eu)^2+(\bar{u}_2-Eu)^2+\cdots+(\bar{u}_{\mathrm{K}}-Eu)^2\right] \tag{2.24}$$

$$=\frac{1}{\mathrm{M}}\left[\sigma_u^2+\sigma_u^2/2+\cdots+\sigma_u^2/\mathrm{K}\right] \text{ with M terms in [ ]} \tag{2.25}$$

$$=\sigma_u^2 En^{-1} \tag{2.26}$$

The expected value of the mean square deviation of a sample mean from the population mean is therefore the variance of the variate multiplied by the expected value of the reciprocal of the number of cases in the sample.

For Table M we are concerned with the variance of non-zero losses. Consequently, the expression En^{-1} refers to the expected number of accidents *provided at least one occurs*, which restriction is essential if the expression is to have a finite value. Likewise the value of En must reflect the same restriction. For large risks, the probability of zero accidents is negligible, hence for them the restriction against non-zero values is insignificant; but it is important for small risks where zero losses have considerable probability.

We see, therefore, that for non-zero losses the probability distribution of n is not the complete Poisson distribution, but only that portion of it for values of n equal to or greater than one. This considerably complicates the mathematics for small risks. The mean of such a distribution is $m/(1-e^{-m})$ where m is the mean of the complete Poisson distribution, and the variance is $\frac{m^2+m}{1-e^{-m}} - \frac{m^2}{(1-e^{-m})^2}$, as compared with m for both the mean and variance of the complete Poisson distribution.

Because of these mathematical complications in the way of accurate calculation of $\acute{\mathrm{V}}_{\mathrm{R}}^2$ for small risks when size groups contain a wide variation in size of risk, the most practical solution is to use Eq. (4a), which is considerably more accurate than the simple *av.* $(\mathrm{R} - \bar{\mathrm{R}})^2$ but still only an approximation, and keep the error down by making size-groups for small risks as narrow as the volume of experience and computing facilities will permit. The resulting scatter of $\acute{\mathrm{V}}_{\mathrm{R}}^2$ values will be ironed out in the graduated values.

Note 2

Derivation of formula for β_1 of loss ratios of risks with average expected losses, estimated from experience of a group of risks of varying size.

Rather than go through detailed calculations similar to those used for $\mathrm{V}_{\mathrm{R}}^2$ in Note 1, which were given at length because the formula advanced in Eq. (2.1) differs from the one used in the past, making it desirable to show its derivation from first principles, we shall simplify the derivation of β_1 by making use of the known inverse relationship between β_1 of the average of a sample and the number of cases on which the average is based.*

The experience of a risk may be regarded as that of a sum of short term exposures, each with expected losses of one dollar. The number of exposures for a risk is therefore equal to its expected losses. The loss ratio for a risk is therefore the average loss per exposure. Then for each risk

$$\beta_1 = \frac{[E(\mathrm{R} - E\mathrm{R})^3]^2}{[E(\mathrm{R} - E\mathrm{R})^2]^3} = \frac{b_3}{f\mathrm{X}} \tag{2.3}$$

Since β_1 is invariant with respect to units of measurement:

$$\beta_1 = \frac{[E(\mathrm{L} - f\mathrm{X})^3]^2}{[E(\mathrm{L} - f\mathrm{X})^2]^3} = \frac{b_3}{f\mathrm{X}} \tag{2.4}$$

We have shown in Eq. (1.8) that, except for the relatively small term† $(Ea)^2 \cdot \mathrm{V}_a^2$, $E(\mathrm{L} - f\mathrm{X})^2$ is proportional to $f\mathrm{X}$ so that approximately

$$\frac{[E(\mathrm{L} - f\mathrm{X})^3]^2}{g^3(f\mathrm{X})^3} = \frac{b_3}{f\mathrm{X}} \tag{2.5}$$

where g is a constant

$$E(\mathrm{L} - f\mathrm{X})^3 = f\mathrm{X}\sqrt{b_3 g^3} \tag{2.6}$$

Summing over all risks in the group and dividing by the number of risks, with $(\mathrm{L} - f\mathrm{X})^3$ treated as an estimate of its own expected value for each risk and then squaring:

$$[av.\ (\mathrm{L} - f\mathrm{X})^3]^2 = [f\ av.\ \mathrm{X}]^2 b_3 g^3 \tag{2.7}$$

Division by $(f\ av.\ \mathrm{X})^3 g^3$ gives

$$\frac{[av.\ (\mathrm{L} - f\mathrm{X})^3]^2}{g^3(f\ av.\ \mathrm{X})^3} = \frac{b_3}{f\ av.\ \mathrm{X}} \tag{2.8}$$

The denominator of the left member of Eq. (2.8) is equivalent to $[E(\mathrm{L} - f\mathrm{X})^2]^3$ estimated by $[av.\ (\mathrm{L} - f\mathrm{X})^2]^3$, hence for the risk with average expected losses

$$\beta_1' = \frac{[av.\ (\mathrm{L} - f\mathrm{X})^3]^2}{[av.\ (\mathrm{L} - f\mathrm{X})^2]^3} = \frac{b_3}{f\ av.\ \mathrm{X}} \tag{2.9}$$

Q.E.D.

Note 3

Derivation of formula for β_2 of loss ratios of risks with average expected losses, estimated from experience of a group of risks of varying size.

*See Kendall, *The Advanced Theory of Statistics*, Vol. 1, page 284 (Chas. Griffin & Sons, Ltd. 1948). The expected values of μ_3 and μ_4 for sample means as given there for sampling from a finite population of N individuals, reduce to β_1/n and $\beta_2/n + 3$ on taking the limit as $N \to \infty$ and dividing by μ_2^3 and μ_2^2 respectively.

†It should be realized that this term becomes important for risks with small expected losses. With average accident costs of \$700, expected losses of \$700 give the first and second terms the same order of magnitude.

For this reason Eqs. (6) and (3.2) are rather rough approximations for risks with only \$1,000 or so of expected losses when the range of sizes is wide.

The error is minimized by keeping the range of sizes in groups of small risks as narrow as possible. This will reduce the reliability of individual β values but not the reliability of graduated values, since there will be more β values underlying the graduating lines given by Eqs. (10) and (12). The exact formulas for calculating β_1 or β_2 from experience of small risks of varying size for a risk of average size are too complicated to make their use practicable.

$(\beta_2 - 3)$ of sample averages is, like β_1, inversely proportional to the number of cases.ϕ The derivation here is similar to that in Note 2. For each risk

$$(3.1) \qquad \beta_2 - 3 = \frac{E(\mathrm{L} - f\mathrm{X})^4}{[E(\mathrm{L} - f\mathrm{X})^2]^2} - 3 = \frac{b_4}{f\mathrm{X}}$$

Since $E(\mathrm{L} - f\mathrm{X})^2$ is proportional* to $f\mathrm{X}$ by Eq. (1.8)

$$(3.2) \qquad \frac{E(\mathrm{L} - f\mathrm{X})^4}{(hf\mathrm{X})^2} - 3 = \frac{b_4}{f\mathrm{X}}$$

where h is a constant

$$(3.3) \qquad E(\mathrm{L} - f\mathrm{X})^4 = 3(hf\mathrm{X})^2 + b_4 h^2 f\mathrm{X}$$

Treating the value of $(\mathrm{L} - f\mathrm{X})^4$ for each risk as an estimate of its own expected value, summing for all risks in the group and dividing by the number of risks we get

$$(3.4) \qquad av.\ (\mathrm{L} - f\mathrm{X})^4 = 3h^2f^2\ av.\ \mathrm{X}^2 + b_4 h^2 f\ av.\ \mathrm{X}$$

Dividing by $h^2f^2(av.\ \mathrm{X})^2$ we have

$$(3.5) \qquad \frac{av.\ (\mathrm{L} - f\mathrm{X})^4}{h^2f^2(av.\ \mathrm{X})^2} - \frac{3\ av.\ \mathrm{X}^2}{(av.\ \mathrm{X})^2} = \frac{b_4}{f\ av.\ \mathrm{X}} = \acute{\beta}_2 - 3$$

Since $av.\ \mathrm{X}^2 = (av.\ \mathrm{X})^2 + \sigma_{\mathrm{X}}^2$

$$(3.6) \qquad \acute{\beta}_2 = \frac{av.(\mathrm{L} - f\mathrm{X})^4}{[av.\ (\mathrm{L} - f\mathrm{X})^2]^2} - 3\mathrm{V}_{\mathrm{X}}^2$$

Q.E.D.

ϕKendall, loc. cit.

*Remarks in footnote † of Note 2 apply here as well.