IN DEFENSE OF MINIMUM-R_0 LINEAR COMPOUND GRADUATION, AND A SIMPLE MODIFICATION FOR ITS IMPROVEMENT

Richard L. London
University of Waterloo

I. INTRODUCTION

The purpose of this paper is to explore the rationale underlying the development of the family of linear compound graduation formulas. As the title suggests, the paper also attempts to defend the Min-R_0 member of that family against the unqualified charge that it is a "poor" graduation method. It concludes with a description of a simple modification of the Min-R_0, and an analysis of the improvement obtained by that modification.

The paper presumes that the reader has a basic understanding of this graduation method. The fundamentals can be reviewed in Miller [4] or in Greville [1]. Note that Greville prefers the term moving-weighted-average.

II. THE PURPOSE OF GRADUATION

The analysis of this graduation method will be seen to be directly related to the fundamental question of the purpose of graduation.

Miller defines graduation to be "the process of securing, from an irregular series of observed values of a continuous variable, a smooth regular series of values consistent in a general way with the observed series of values". Kimeldorf and Jones [2] observe that "graduation has traditionally been associated with smoothing". Greville, although not specifically defining graduation, clearly implies that the primary objective is smoothing.

But George King [3] is quoted as saying: "What is the real object of graduation? Many would reply, to get a smooth curve; but that is not quite correct. The reply should be, to get the most probable deaths".

As Miller points out in his chapter, a series of observed values contains random statistical fluctuations (errors) away from a true underlying series. The graduation process produces a series of graduated values which hopefully is a better representation of the true underlying values than was the series of observed values. Of course the true values are not known, but are presumed to form a smooth sequence. Thus, to approximate the true values, the graduated values should exhibit this characteristic of smoothness. Thus the objective of graduation is smoothing, goes the chain of reasoning. Indeed Miller himself is led to this conclusion as shown by his definition quoted above. But it should be stressed that the purpose is not really smoothing, per se, according to King and really Miller as well, but rather to estimate the true values, themselves believed smooth. Thus the smooth result is somewhat of a consequence, albeit a very desirable one, rather than the direct, fundament objective.

III. THE DEVELOPMENT OF LINEAR COMPOUND GRADUATION

Linear compound graduation was developed initially by E.L. De Forest in the 1870's. As reported by both Miller and Greville, his work was published in obscure places, and not brought to the attention of the actuarial community until 1925 by Wolfenden [6]. Meanwhile largely independent development of these formulas was being done by Sheppard [5].

The first attempt to develop a linear compound formula made use of the idea of "reduction of error". It is recognized that the observed values contain errors (that is, deviations away from the true values) and that the graduated values, to be a good representation of true values, should have minimum residual error. Thus the coefficients in the linear compound are derived, in part, to minimize this expected residual error. The formula so derived is called the Minimum-R_o formula.

Now it follows that if the graduation process has produced values with a minimized deviation from the true underlying values, then these values constitute our best estimate of the true values, and are thus the most "accurate" that can be produced. Both De Forest and Sheppard referred to these Min-R_o results as "accurate" in the above sense.

As stated earlier, however, graduated results should be smooth, and unfortunately the Min-R_o results generally fail to satisfy this requirement. For this reason the Min-R_z formulas, especially with $z = 2,3,$ and 4, were developed.

That a Min-R_3 formula, for example, will produce smoother results than a Min-R_o is self-evident from the criteria of smoothness traditionally used, namely the smallness of the third differences of the graduated values. The Min-R_3 coefficients were derived, in part, so as to achieve such smallness. The Min-R_o, on the other hand, has no such objective of smooth results directly incorporated in its derivation, so the resulting lack of smoothness comes as no surprise. Thus we might summarize by saying that use of the Min-R_o is consistent with an objective of "accurate" results, whereas use of the Min-R_3 (or Min-R_4) is consistent with an objective of "smooth" results. Since smoothness is such an essential result, (indeed the very purpose of graduation according to many authorities), the Min-R_o has, understandably, come to be considered a poor graduation method.

IV. TESTING FOR "ACCURACY"

Even if smoothing were the fundamental objective of a graduation process, it is interesting to further explore the concept of accuracy. Since accuracy means closeness to true underlying values, these must be "known" in order to measure the accuracy in graduated results. Perhaps the impossibility of measuring accuracy in practice contributes to the down-play of accuracy (and the corresponding up-play of smoothness) as criteria for the success of a graduation.

In a controlled, experimental environment, however, the true values are known. At a given value of x, select a rate of mortality from a standard table. This rate will now constitute what we are referring to as the true rate, and is denoted as q_x in the remainder of this paper.

Using this rate in a Monte Carlo technique, random decimal numbers are generated. Each random number is a trial; if the number is less than q_x, that trial is a death. Each trial is independent. In this way an observed number of deaths is determined, and thus an observed rate of mortality, based on a number of trials E_x.

Now the procedure is clearly defining a binomial random variable, Θ, for the number of deaths, having a mean of $E_x \cdot q_x$, and a variance of $E_x \cdot q_x \cdot (1-q_x)$. From Θ is derived the random variable $Q=\Theta/E$. Thus Q is a rate type of random variable, derived from a binomial, with mean of q_x and variance of $\dfrac{q_x(1-q_x)}{E_x}$.

From this we obtain an interpretation of our heretofore mystical "true" rate; it is the mean, or expected value, of the random variable Q, and q_x'' is one observed value of that random variable. In the graduation process, q_x'' is replaced by the graduated value, q_x^G.

Once the set of observed values q_x'', over some range of x, have been generated, and then graduated to produce the set q_x^G, the accuracy of the results can be measured by comparing the graduated set q_x^G to the true set q_x.

V. EXPERIMENTAL RESULTS

Table 1 shows an excerp of such an experiment. The exposures were taken from Miller, page 63, and the "true" rates, used in the Monte Carlo generation of the observed rates, are taken from the 1941 CSO Table. The linear compound formulas are 17-term, symmetric, reproducing a cubic.

TABLE 1

x	Exposure	True Rate	Observed Rate	Graduated Rates Min-R_0	Min-R_3
43	36007	.00715	.00733	.00761	.00743
44	40861	.00804	.00803	.00807	.00796
45	41259	.00861	.00829	.00864	.00858
46	46040	.00923	.00910	.00930	.00927
47	44534	.00991	.01053	.00994	.01002
48	48060	.01064	.01001	.01074	.01079
49	51343	.01145	.01274	.01148	.01159
50	53261	.01232	.01211	.01231	.01241
51	52689	.01327	.01330	.01324	.01327
52	54977	.01430	.01395	.01425	.01420
53	56130	.01543	.01545	.01520	.01521
54	53121	.01665	.01585	.01648	.01636
55	51909	.01798	.01822	.01761	.01764
56	51034	.01943	.01858	.01920	.01910
57	52071	.02100	.02093	.02074	.02073
Smoothness:		.00016	.03727	.00284	.00028
Accuracy:			.21108	.02651	.02934

The smoothness measure in Table 1 is the sum of absolute values of third differences. As expected, the Min-R_3 results are much more smooth than the Min-R_0 results, being, in fact, nearly as smooth as the assumed true rates.

The accuracy measure is the sum of weighted squared deviations away from true, or expected, values. It is shown for the observed values to give an indication of plausible deviation in observations, and also as a standard for comparing the accuracy measures of the graduated values.

A interesting fact revealed in Table 1 is that the accuracy measure for Min-R_o is really not much better than that for Min-R_3. This is partly a consequence of the specific data in the illustration, and partly due to assumptions underlying the derivation of the Min-R_o. A simple modification in that derivation leads to an improved accuracy measure.

VI. MODIFIED MIN-R_O

In order to develop a reasonable working formula, the traditional Min R_o makes the assumption that the variance in each random variable in the sequence is the same. Since the random variables are rates derived from a binomial, it is clear that the variance in each one is dependent inversely on the number of trials (exposure) underlying that random variable. Since the exposures are not equal for each variable, then neither should be the variances.

This leads to the development of a Modified Min R_o formula. The assumption upon which this new formula is based is that the variances in the several variables involved in one application are inversely proportional to their respective exposures.

Thus the coefficients for each application depend on the exposures for the variables involved, and thus a separate set of coefficients must be calculated for each application. This tremendous increase in amount of calculation involved may partly explain the persistence of the illogical assumption inherent in the traditional formula. The derivation of this new formula is given in the appendix.

The entire data set, from which Table 1 is excerpted, was graduated by Mod Min-R_o. The smoothness measure for the 15-age excerp was .00241, a slight improvement over traditional Min-R_o, and the accuracy measure was .01903, a more significant improvement.

Another modification is to change the constraints on the linear compound coefficients from the requirement that the formula reproduce a cubic. This requirement is traditionally imposed under the assumption that the true underlying values are closely approximated by a cubic over the limited range of one application of the formula. In this investigation, the underlying (expected) values are from the 1941 CSO, and thus follow Makeham's Law.

Several plausible observed sets of values were generated from the 1941 CSO by the Monte Carlo technique and graduated by Mod Min-R_o with the revised constraint requirements. The results generally show a better accuracy measure for Mod Min-R_o than for the other linear compound formulas.

In both cases (cubic reproduction and Makeham reproduction) certain observed data sets, being randomly generated, would have characteristics which caused more accurate results (as well as smoother results) to be produced by Min-R_3 or Min-R_4 than by Mod Min-R_o. But for the 200 total repetitions of the experiment, Mod Min-R_o produced the most accurate results more often than any other formula, as should be expected from the theory. Traditional Min-R_o (i.e., assumption of equal variance in all terms) came in second.

VII. SUMMARY

The research described herein indicates that when graduation is viewed as an attempt to approximate true underlying values, interpreted as expected values in this paper, then the Min-R_o formula, especially with the modification to recognize unequal variance in the terms, does a better job than the other linear compound formulas.

If a way could be found to improve the smoothness of these more accurate results, without reducing the accuracy level, then an ideal graduation formula would have been discovered. Further research in this area is underway.

The author would like to thank Neil L. Haynes, an Actuarial Science major at University of Waterloo, for his assistance. Mr. Haynes wrote the computer programs and guided the entire project through the Computer Center at the University.

IX. REFERENCES

1. Greville, T.N.E. "Study Note on Graduation". Chicago: Society of Actuaries, 1973.

2. Kimeldorf, G.S. and Jones, D.A. "Bayesian Graduation", TSA, XIX (1967), 66.

3. King, G. Discussion of "The Graphic Method of Adjusting Mortality Tables", by T.B. Sprague, JIA, XXVI (1887),114.

4. Miller, M.D. Elements of Graduation, New York: Actuarial Society of America and American Institute of Actuaries, 1946.

5. Sheppard, W.F. "Graduation by Reduction of Mean Square Error", JIA, XLVIII (1914), 171.

6. Wolfenden, H.H. "On the Development of Formulae for Graduation by Linear Compounding, with Special Reference to the Work of E.L. De Forest", TASA, XXVI (1925), 81.

APPENDIX

In traditional MWA graduation, the set of coefficients a_r for $r = 0,1,\ldots,8$ for a symmetric 17-term Min R_0 formula are derived by minimizing the ratio of the variance in graduated values to the variance in ungraduated values, subject to the constraints assuring that the formula reproduces a cubic.

The same set $\{a_r\}$ is used for each application of the averaging formula. In order to obtain this convenience, it was necessary to assume that each random variable in the series had the same theoretical variance.

Since variance is influenced by sample size, that is, exposure, the above assumption is quite suspect. More reasonably, the variance of a random variable should vary inversely with the exposure involved.

Thus Modified Min R_0 MWA graduation derives the coefficients for each application by considering the exposures involved in that application. As a result, a different set of coefficients is derived for each application.

Since

$$u_x = \sum_{-n}^{n} a_s u''_{x+s} \quad , \text{ where } \quad a_{-s} = a_s \ ,$$

then

$$\sigma^2 \ddot{u}_x = \sum_{-n}^{n} (\ddot{a}_s)^2 \sigma^2 \ddot{u}''_{x+s} \quad .$$

Now we define $\sigma^2 u_x'' = e^2$, and presume that the variance at other arguments relates to e^2 in proportion to exposure.

Thus

$$\sigma^2 u_x = \sum_{-n}^{n} (a_s)^2 \cdot e^2 \cdot \frac{E_x}{E_{x+s}}$$

and

$$R_0{}^2 = \frac{\sigma^2 u_x}{\sigma^2 u_x''} = E_x \sum_{-n}^{n} \frac{(a_s)^2}{E_{x+s}}$$

To minimize $R_0{}^2$, subject to the constraints that

$$\sum_{-n}^{n} a_s = 1 \quad \text{and} \quad \sum_{-n}^{n} s^2 a_s = 0, \text{ we}$$

first write $a_0 + 2\sum_{1}^{n} a_s = 1$ and $\sum_{1}^{n} s^2 a_s = 0$ to reflect the symmetric nature of the formula.

Then

$$a_1 = -\sum_{2}^{n} s^2 a_s$$

and $$a_0 = 1 - 2a_1 - 2\sum_{2}^{n} a_s = 1 + 2\sum_{2}^{n} s^2 a_s - 2\sum_{2}^{n} a_s .$$

Substituting these expressions for a_0 and a_1 into $R_0{}^2$,

$$R_0{}^2 = E_x\left[\frac{a_0{}^2}{E_x} + \frac{a_1{}^2}{E_{x+1}} + \frac{a_1{}^2}{E_{x-1}} + \frac{a_2{}^2}{E_{x+2}} + \frac{a_2{}^2}{E_{x-2}} + \ldots + \frac{a_n{}^2}{E_{x+n}} + \frac{a_n{}^2}{E_{x-n}}\right]$$

$$= E_x\left[\frac{(1+2\sum_{2}^{n} s^2 a_s - 2\sum_{2}^{n} a_s)^2}{E_x} + \frac{(\sum_{2}^{n} s^2 a_s)^2}{E_{x+1}} + \frac{(\sum_{2}^{n} s^2 a_s)^2}{E_{x-1}} + \frac{a_2{}^2}{E_{x+2}} + \frac{a_2{}^2}{E_{x-2}} + \ldots + \frac{a_n{}^2}{E_{x+n}} + \frac{a_n{}^2}{E_{x-n}}\right] .$$

Now

$$\frac{\partial R_0^{\,2}}{\partial a_r} = E_x \left\{ \frac{1}{E_x} \left[2\left(1+2\sum_2^n s^2 a_s - 2\sum_2^n a_s\right)(2r^2-2) \right] + \frac{1}{E_{x+1}}\left[2\sum_2^n s^2 a_s \cdot r^2\right] + \frac{1}{E_{x-1}}\left[2\sum_2^n s^2 a_s \cdot r^2\right] + \frac{2a_r}{E_{x+r}} + \frac{2a_r}{E_{x-r}} \right\}$$

This partial derivative of $R_0^{\,2}$ must be taken for $r=2,\ldots,n$. Each is set equal to zero, giving $n-1$ equations, each involving all $n-1$ independent unknowns.

Specifically, the equations for $n=8$, thus for $r=2,\ldots,8$, can now be derived. Since $\frac{\partial R_0^{\,2}}{\partial a_r} = 0$, it is possible to reduce the derivative equation to

$$\frac{(2r^2-2)\left(1+2\sum_2^8 s^2 a_s - 2\sum_2^8 a_s\right)}{E_x} + \frac{r^2 \sum_2^8 s^2 a_s}{E_{x+1}} + \frac{r^2 \sum_2^8 s^2 a_s}{E_{x-1}} + \frac{a_r}{E_{x+r}} + \frac{a_r}{E_{x-r}} = 0$$

or

$$\frac{4(r^2-1)\sum_2^8 (s^2-1)a_s}{E_x} + \frac{r^2 \sum_2^8 s^2 a_s}{E_{x+1}} + \frac{r^2 \sum_2^8 s^2 a_s}{E_{x-1}} + \frac{a_r}{E_{x+r}} + \frac{a_r}{E_{x-r}} = \frac{-2(r^2-1)}{E_x}$$

Now we define a diagonal matrix, D, where the elements are

$$\frac{1}{E_{x+2}} + \frac{1}{E_{x-2}}, \quad \frac{1}{E_{x+3}} + \frac{1}{E_{x-3}}, \quad \ldots, \quad \frac{1}{E_{x+8}} + \frac{1}{E_{x-8}} .$$

Also we define the matrix, C, with elements

$$c_{r,s} = \frac{r^2 s^2}{E_{x+1}} + \frac{r^2 s^2}{E_{x-1}} + \frac{4(r^2-1)(s^2-1)}{E_x}, \quad \begin{array}{l} r = 2,\ldots,8 \\ s = 2,\ldots,8 \end{array}$$

and the vector, a, of unknowns with elements a_r, $r = 2,\ldots,8$ and the vector, b, of constants with elements

$$b_r = \frac{-2(r^2-1)}{E_x} \quad \text{for } r = 2,\ldots,8.$$

Then the matrix equation $(D + C)a = b$ is solved for the unknowns $a_2, a_3,\ldots,a_8$.

Finally a_0 and a_1 are obtained by

$$a_1 = -\sum_{2}^{8} s^2 a_s$$

$$a_0 = 1+2\sum_{2}^{8} s^2 a_s - 2\sum_{2}^{8} a_s \ .$$