TRANSACTIONS OF SOCIETY OF ACTUARIES
1965 VOL. 17 PT. 1 NO. 48

DISCUSSION OF PAPERS PRESENTED AT THE REGIONAL MEETINGS

PERIODOGRAMS OF GRADUATION OPERATORS

GEORGE H. ANDREWS AND CECIL J. NESBITT

SEE PAGE 1 OF THIS VOLUME

HUGH H. WOLFENDEN:

In this interesting and mathematically detailed paper the reader is assumed to be quite familiar with Fourier analysis—a valuable portion of mathematical theory seldom employed by actuaries. Two important instances of its application, however, are to be found in Anderson and O'Brien's 1928 paper, "Notes upon Experiments with Actuarial Functions and Fourier's Series" (*JIA*, LIX, 256), and Elphinstone's 1950 contribution on "Summation and Some Other Methods of Graduation—the Foundations of Theory" (*TFA*, XX, 15). Anderson and O'Brien explored the possibility of applying Fourier's methods of trigonometrical expansion to "achieve a quick means of arriving at a_x from ungraduated data, and possibly to get a joint-life annuity . . . [and] coefficients for N'_x and l_x"; but their conclusion was that "the results obtained were not all that they had hoped," so that the paper must be viewed, in the light of that statement and the valuable discussion which accompanied it, as an interesting though mainly discouraging investigation of the practicability of using trigonometrical instead of ordinary interpolation for the representation of actuarial functions. Elphinstone's highly original paper is, of course, the foundation on which Andrews and Nesbitt have built their further analysis of that author's ideas. It is therefore essential for the reader to be well informed concerning both the basis of Fourier's methods and Elphinstone's applications of them to the problems of graduation in the special forms in which those problems usually present themselves to actuaries.

Jean Baptiste Joseph Fourier, born in France in 1768, showed early in his life remarkable mathematical and teaching abilities; surviving the terrors of the French Revolution, and having accompanied Napoleon's expedition to Egypt, he was appointed to the chair of mathematics at the École Normale (created in 1794 through Napoleon's efforts) in Paris, and later taught at the Polytechnique. In 1807 he submitted to the French Academy of Sciences his first memoir on the conduction of heat, and in

1812 won the Academy's Grand Prize (refereed, with controversial criticisms, by Laplace, Lagrange, and Legendre) with his famous *Théorie Analytique de la Chaleur*. In that work he developed the general problem of representing "any function whatever in an infinite series of sines or cosines of multiple arcs," and established his trigonometric "Fourier series" by which an arbitrary function $f(x)$ in the interval $-\pi$ to π can be expanded as

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx);$$

where (with certain restrictions) a_0, a_n, and b_n are given by

$$\pi a_m = \int_{-\pi}^{\pi} f(y) \cos my\,dy \qquad \text{and} \qquad \pi b_m = \int_{-\pi}^{\pi} f(y) \sin my\,dy\,.$$

The important problems of determining the points x for which the series converges, and allied questions, are discussed in many mathematical texts (see, as examples, Courant's *Differential and Integral Calculus*, I, 447, and Rogosinski's *Fourier Series*).

Other forms are shown in Andrews and Nesbitt's paper. The representation of any finite sequence of n terms $u_0 \ldots u_{n-1}$ by such a Fourier sum is discussed well by Whittaker and Robinson in their chapter on practical Fourier analysis, while for a sequence of $2n+1$ terms $u_{-n} \ldots u_n$ the formulae are shown conveniently on p. 28 of Elphinstone's paper.

The Fourier expansion reproducing a sequence of given values thus represents the sum of a set of superimposed sinusoidal waves, in which the relative intensities of the different sinusoids can be determined. When a linear compounding graduation process is then applied, using a coefficient-operator G_t at t, the ratios in which the waves of different periods are retained by the graduation can be measured by the values of the "periodogram function," shown by Elphinstone, p. 30, and by Andrews and Nesbitt in their formula (9).

With regard to the desirable shape of a periodogram, Elphinstone observed truly that "enough is known to make a wise man very cautious in interpreting a periodogram" (p. 29). He remarked that the periodogram values "should remain small for short periods and rise to 1 at infinity" (p. 33)—indicating an ability to suppress short waves and retain the overall trend. He commended "a steady sweep without substantial negative values" (p. 38), and Professor Aitken similarly implied merit if "smooth behavior" is found (p. 69). Andrews and Nesbitt of course interpret the periodograms in the same general manner, adding that clearly they may show negative values which should not be large though "small values need not be alarming." Nothing specific is said about what is meant by "small" and "large," except that Elphinstone expressed several judgments with respect to waves of periods 7, 9, etc., with some of the restricted list of graduation formulae which he examined. Absurdly shaped periodograms, like those exhibited by Elphinstone for his obviously unsound cases numbered 1, 2, and 7 (pp. 32–37 and 45–46), undoubtedly indicate rejection of those processes; but it is practically impossible, on the other hand, to choose with any confidence between the periodograms of Spencer's formula (which Elphinstone approves somewhat too readily), De Forest's, Whittaker's, and Andrews and Nesbitt's new summation form. The method embodies an a priori test for rejection rather than for confident acceptance. In its attempts to appraise, a priori, the virtues of competing graduation processes, closely similar (though slightly different) periodograms of various graduation methods do not give any really trustworthy indications of the relative effectiveness of those processes in practice.

Elphinstone adopted the view, as Dr. Leon Solomon said in his penetrating discussion (*TFA*, XX, 61), that "the strictly limited object of graduation by means of such [linear compounding] operators . . . is merely to remove the short-wave components of a mortality curve; we must remove them, it seems, because we have an innate objection to them; that is all, the author concludes, that a graduation can do; any attempt to arrive at truth, to eliminate random errors, to appreciate underlying patterns—these fond hopes have been dashed . . ." (see also the remarks of J. W. Sutherland, p. 63). Solomon added the pertinent question: "I wonder whether waves are not given undue prominence in this theory simply because it uses harmonic analysis as its main tool" (p. 62). With mortality data it must be remembered that the total elimination of all waves may not be desirable when, as sometimes happens, certain waves are unquestionably inherent in the data, and that the conflict between fit and smoothness (as indicated by the orders of differences) is not always easily resolved by distorting the graduation through undue suppression of inherent waves.

The main defects of Elphinstone's stimulating paper which seemed evident to me when it appeared in 1950 were his concentration on the Schiaparelli–De Forest least-squares "fitting" R_0 formulae (which again were erroneously called by Sheppard's name), and his surprising failure even to mention De Forest's greatly superior and more logical "smoothing" R_4 formulae which minimize the mean-square error in the 4th differences when $j = 3$ (see my paper on De Forest's work, *TASA*, XXVI, 109). The results were that the R_0 formulae were held, almost as if it were a new conclusion, to be inferior—though that inferiority, except for fitting only, was well known to De Forest and to others (see my paper, pp. 103–4), because they aim to effect only a best fit without regard to smoothness; the failure to mention the much more important R_4 smoothing formulae gave a seriously misleading impression of the scope and efficacy of De Forest's methods; and the paper approved too readily Whittaker's approach which brings together criteria for fit and smoothness by merely adding in an arbitrary proportion—an addition which prompted Solomon (in his discussion, p. 62) to record, as others like myself also have felt despite the rationale offered in Whittaker's book, his "especial difficulty in understanding why roughness and distortion can be balanced simply by adding."

The paper by Andrews and Nesbitt now shows also the periodograms for De Forest's R_4 formulae of 21 and 29 terms, and of the two R_3 formulae of the same lengths (which, however, are relatively unimportant because obviously they have an inferior logical basis in minimizing the mean-square error in the 3d differences, instead of the 4th, when $j = 3$). The acceptability of those periodograms, as would be expected, is certainly indicated. Elphinstone's conclusion that only Spencer's and a Whittaker method are satisfactory is thus upset, as it should be.

These periodogram analyses apply an abstrusely theoretical approach to what is eventually an essentially practical problem. In any graduation of mortality data it may be difficult to decide how closely $j = 3$ (or even 2 or 5) may be a justifiable assumption (having regard to the number of decimal places to be retained in the graduated results), and the suppression of all waves may not always be desirable. The periodogram method of attempting a priori appraisals of linear compounding formulae, whether applied by summation, reduction of mean-square error, or difference-equation procedures, therefore cannot be expected to establish the inevitable superiority of any of those processes, just as the a priori values of R_0 and R_4 (or even R_3)—useful as they are—cannot be regarded as infallible guides. Every set of data has its own special characteristics, for which theoretical a priori assumptions must often be highly uncertain. Since every graduation method in every practical case should be selected with careful regard to the identifiable characteristics of the data, it must always be unwise to suppose that some summation formula (whether Hardy's, Spencer's, King's, Kenchington's, Vaughan's, Davidson's and Reid's, or the one now suggested by Andrews and Nesbitt), or De Forest's excellent R_4 (or R_0) formulae of some selected length, or some type of Whittaker's method, will necessarily yield superior results when those results are subjected to strict a posteriori tests of fit and smoothness. When extended comparable graduations of different sets of actuarial data are made, it will almost always be found in practice that no one type of process can have any claim to be taken as a standard against which all others should be judged (see also my book *The Fundamental Principles of Mathematical Statistics*, pp. 147–48). For these reasons the periodogram approach, with its particular emphasis on the suppression of waves, seems likely to remain more of an interesting mathematical theory than a practical working tool for selecting linear compounding methods in the graduation of actuarial data.

THOMAS N. E. GREVILLE:

Professors Andrews and Nesbitt are to be congratulated on a lucid and balanced presentation of periodogram analysis of graduation operators. I wish to thank them for their kind references to my work and for their acknowledgment of my role in bringing to their attention the important work of I. J. Schoenberg in this area.

At the Mathematics Research Center (of which Schoenberg is also a member), I have been doing some research on smoothing formulae that has some relationship to the subject of this paper. In this discussion I shall refer briefly to two problems concerning characteristic functions of graduation operators that are not yet completely solved.

It is evident that k-fold iteration (i.e., repeated application) of a graduation formula is equivalent to a single formula of wider range. It is natural to inquire whether the (suitably normalized) coefficients corresponding to k-fold iteration of a given formula approach a definite limiting curve as k increases without limit. A graduation formula is called "stable" if it has this property. The question of stability of graduation formulas was first investigated in 1878 by E. L. De Forest (see Wolfenden's paper, *TASA*, XXVI, 116–18), who showed that certain formulae are indeed stable, and that if $j = 1$ (j being the degree of polynomial reproduced by the formula), the limiting curve is the normal frequency curve. In numerical experiments for larger values of j he obtained curves having the general appearance of the normal curve in the middle portion and that of a damped sine curve in the tails.

For the case of a symmetrical graduation formula this problem was solved in 1948 by Schoenberg ("Some Analytical Aspects of the Problem of Smoothing," *Studies and Essays Presented to R. Courant on His 60th Birthday* [New York: Interscience]), who showed that such a formula is stable if and only if the characteristic function $\varphi(u)$ satisfies the condition $|\varphi(u)| < 1$ for $0 < u < 2\pi$, and gave an explicit formula for the limiting function. In the paper cited by Andrews and Nesbitt he showed also that the satisfaction of this condition insures that application of the formula will, in a certain sense, actually increase the smoothness of any sequence of data. This is therefore a reasonable requirement to impose on a graduation formula. His method of proof is easily extended to unsymmetrical formulae provided j remains odd. De Forest also investigated the case of unsymmetrical formulae with even j; his results are suggestive but inconclusive. This problem is still not fully solved (see my abstract in *SIAM Review*, VI, 92–93).

Recently, Schoenberg proposed the problem of proving rigorously the conjecture that the minimum R_0 formula is stable for all odd n (n being the number of terms) and for all odd $j < n$. This turns out to be unexpectedly difficult. The conjecture is easily established for $j = 1$ and for $j = n - 2$, and I have obtained a recursive relation from which it would follow that all minimum R_m formulae are stable (for $m > 0$) if it could only be shown that all minimum R_0 formulae are stable. If we define $R_\infty = \lim_{m\to\infty} R_m$, it can be shown that R_∞ is given in terms of the characteristic function by $R_\infty = |\varphi(\pi)|$, and a minimum R_∞ formula can be derived (for given n and j). This minimum R_∞ formula has the interesting property that its characteristic function is never negative, and in the range $(0, 2\pi)$ is zero only at $u = \pi$. Thus, as might be expected, R_∞ is actually zero for a minimum R_∞ formula.

The following table shows 7-term minimum R_m formulae for $j = 3$ and $m = 0, 4$, and ∞, together with the values of R_0, R_4, and R_∞ in each case.

m	Minimum R_m Formula	R_0	R_4	R_∞
0.	$P=1-\frac{3}{7}\delta^4-\frac{2}{21}\delta^6$	.577	.181	.238
4.	$P=1-\frac{18}{65}\delta^4-\frac{12}{221}\delta^6$	.602	.006	.044
∞.	$P=1-\frac{3}{16}\delta^4-\frac{1}{32}\delta^6$	.640	.127	0

As might be expected, the minimum R_∞ formula is a more "gentle" smoothing formula than the others. As measured by the value of R_4, it is a better smoothing formula than the minimum R_0.

HARWOOD ROSSER:

This is not the first time that Dr. Nesbitt has lent at least his name to publication in the *Transactions*, in curtailed form, of a doctoral dissertation. This is mutually advantageous to the Society and to the primary author. The latter gets a broader exposure for the results of his research than just his faculty adviser!

The members of the Society, in turn, have the results of recent research in more accessible form, in two senses. Elphinstone's 1951 paper, cited by the authors, contains much of the theory, including some graphs; but the *Transactions of the Faculty of Actuaries* has very limited circulation in this hemisphere. Like a true Scot, he mentions Whittaker, but not Henderson—an omission remedied by today's authors, as well as by one of Elphinstone's reviewers, A. C. Aitken. The latter individual's opening comments on Elphinstone express admirably my feeling about the current paper—in fact, about both:

> This paper . . . is not, I feel, a paper to be commented upon at any shorter notice than two or three months. My own perusal of it is very recent; I should really like more time to form an opinion. Some parts of it I have not fully grasped.

Having considerable interest in music, and hence in the physics of sound, I was aware that Fourier series were quite useful in the analysis of the composite effect of waves of different frequencies—in one of the simpler cases, that of a single tone plus all its overtones. Before these two papers, it would never have occurred to me that a series of points to be graduated bore any resemblance to the auditory image of a vibrating string. It now appears that a proper graduation formula will have an effect analogous to suppressing, or substantially damping, all overtones beyond, say, the second octave, whose frequency is four times that of the primary note. (I believe such things can now be done electronically. However, the principle involved does not depend, for its validity, upon actual physical realization in the realm of sound, or elsewhere.)

The Fourier analysis enables us to measure the degree of reduction of waves of various lengths. The periodogram of a graduation formula is simply a table, in the form of a graph, of such results. A value of 1 on the vertical scale would mean that the corresponding wavelength is reproduced intact by the graduation process.

For the reader who peruses the discussions before he tackles a paper seriously—a frequent practice with me—and who probably will not read every word of either, I suggest first a look at the Introduction, then one at the graphs in Appendix B, and, finally, a glance at "Interpretation of Periodograms," on page 14, particularly the first two paragraphs. A key sentence is this:

> Thus the values of the periodogram are weights and indicate to what extent waves of a given length are preserved by the graduation.

These weights are not to be confused with the multipliers of surrounding terms, when the graduation is put into linear compound form.

Generally, Part III will receive more attention than the other two parts. From the standpoint of readability, this might have been put first; but then the results would have preceded the requisite theoretical development, which is a little like popping the question at the beginning of a courtship. Andrews and Nesbitt have already, through the use of appendices, made substantial concessions to the average reader of these pages. This is part of what I meant above, when I spoke of "more accessible form." Let any dissenter turn to the original paper (No. 2 of the "References")!

As an intellectual exercise, Dr. Nesbitt and his protégé have produced a 21-term summation formula which, they suggest, may put the British to shame. This might be more evident at a glance if they had superimposed its periodogram upon that of Spencer's summation formula of the same length, shown by Elphinstone. I attempted a visual comparison to confirm their remarks, without much success. It would be interesting to see values of R_m^2, for $m = 0$, 3, and 4, for this new formula, to compare with those for Spencer's formula and for the 21-term minimum R_0^2 formula (Sheppard's), shown on page 35 of Elphinstone's paper. Even so, I would not consider it a practical formula until expressed in linear compound form. None of these values is theoretically difficult to obtain, but the computation is laborious for a 21-term formula.

Being one of the latter-day apostles of the linear compound form,* I was most happy to see recognition of Greville's excellent but neglected work on minimum R_m^2 graduation formulae (Nos. 4 and 5 of the "References"). In this he gives coefficients, for a wide range of terms, which include those shown at the bottom of page 32 of Miller's monograph on graduation. In addition, he has given a solution to the notorious end-of-series problem, which I believe the authors failed to note. With his tables, graduation by electronic computer becomes relatively simple. The chief problem is whether to select a given length, such as, say, nine terms, and use the resulting program on an all-purpose basis, or to have a choice of formula lengths. It is not necessary to be familiar with Tchebycheff polynomials in order to use Greville's results. The first four graphs of periodograms are for this type of graduation, and they are seen to be quite satisfactory.

The major value of the paper for actuaries is not a new 21-term formula but the insight it offers into existing graduation methods. For instance, the fifth graph, showing Q_5 and Q_6, shows the periodograms for the Whittaker-Henderson "A" formula with $k = 1$ and $k = 20$, respectively. The authors' formula (25) corresponds to (5.31) in Miller's monograph, if we

* Cf. *Proceedings of the Conference of Actuaries in Public Practice*, XII, 298.

replace their u_x by u_x'', their v_x by u_x, and set $g = k = 0$. It corresponds to Miller's (5.51), if their u_x is replaced by q_x'', their v_x by q_x, and $g = h = 0$. This is the case shown in the graph. The greater emphasis on smoothness in Q_6 produces a greater tendency to eliminate short waves.

Using still other values in (25), the authors go on to examine the mixed difference case, and to illustrate it in the sixth graph. Again, the periodograms help in drawing conclusions.

Before leaving difference-equation graduation, it is worth noting that this is one type where the linear compound form is not necessarily the best for actual computation. This is due to the fact that, in this form, the number of coefficients is usually not limited but actually represents an infinite series. While convergence is usually rapid, this introduces complications. Hence other approaches are usually preferable. This means that each graduated point is a function of *every* ungraduated point, rather than of a certain number of points. In practice, the effect of distant points is normally lost in the rounding off. The breaking point, however, tends to be difficult of prediction.

The last graph has already been commented upon.

In summary, this is an excellent paper, but not an easy one. Elphinstone said, on his opening page, that "a paper on graduation always seems to arouse controversy. . . ." I trust I have not taken the authors to task unduly. On the other hand, I hope that this paper will attract, if not controversy, at least some of the notice that the very substantial labor of the writers deserves. To the Education and Examination Committee, I would make a strong recommendation: that they abstract, for Part 5 Study Notes, the notion of periodogram, including some of the graphs. In so doing, perhaps they can give belated recognition to Greville's excellent and practical work, mentioned above.

(AUTHORS' REVIEW OF DISCUSSION)

GEORGE H. ANDREWS AND CECIL J. NESBITT:

First we should like to thank Mr. Wolfenden, Dr. Greville, and Mr. Rosser for their informative and illuminating discussions. Also, we were pleased to receive by letter some reactions of Professor James Hickman to our paper.

Mr. Wolfenden has given much interesting background for the paper, both in the notes concerning Fourier and the series which bears his name, and in the remarks concerning Elphinstone's paper. The remarks provide insight into the concept of periodogram and its limitations as a practical working tool for graduation purposes. His criticisms concerning the a priori nature of the periodogram appraisal of a linear compounding formula are echoed in Professor Hickman's letter. Professor Hickman prefers a statistical approach that seeks "most probable values" for a set of function values, given some observed values. But the actuary is also interested in the elusive quality of smoothness, and it may be difficult to achieve this quality by a purely statistical approach. However, interesting possibilities coming out of a statistical approach are appearing in work now going on under the direction of Professor D. A. Jones.

It may be well at this stage to qualify the statement in the first paragraph of our paper that a periodogram can be obtained in relation to each adjusted average and difference equation method of a graduation. It is not clear how to determine a periodogram for a Whittaker-Henderson Type B formula, or for other graduation processes where something more than a single symmetric linear operator is applied. As noted by Mr. Rosser, we have not discussed nonsymmetric linear operators such as might be used at the end of a series. For a nonsymmetric linear operator, the periodogram definition given by Formula (12) would not hold. However, a characteristic function in Schoenberg's sense could still be defined.

Dr. Greville indicates some of the results obtained by Schoenberg and himself in regard to the question of stability which was first investigated by E. L. De Forest. He also introduces the definition

$$R_\infty = \lim_{m \to \infty} R_m ,$$

and states that R_∞ is given in terms of the characteristic function by $R_\infty = |\phi(\pi)|$. Since, as indicated in A.10 of the Appendix to our paper, $\varphi(2\pi/\beta) = \mathcal{P}(\beta)$, then R_∞, when it exists, is equal to $|\mathcal{P}(2)|$. From the graphs in Appendix B, which begin at $\beta = 2$, it appears that R_∞, if it exists, is near zero for the given graduation operators. In particular, for Q_9 the value of $\mathcal{P}(2) = 0$, as is the case for any summation graduator with a factor $[m]$, m an even integer.

The summation graduator Q_9 was developed somewhat incidentally as an application of the periodogram concepts. Following Mr. Rosser's suggestion, we have expressed Q_9 in linear compound form,

$$\sum_t q_t G_t ,$$

where t runs over the even integers from 0 to 20, and have computed the values of R_m^2 for $m = 0, 3$, and 4 for this formula. The results are summarized in the accompanying tabulations. Although the coefficients q_t are presented to five places, the R_m^2 values were computed using additional accuracy.

COEFFICIENTS OF Q_9 IN LINEAR COMPOUND FORM, $\sum_t q_t G_t$

t	q_t
0	.16746
2	.15665
4	.13145
6	.09425
8	.05466
10	.02232
12	−.00040
14	−.01230
16	−.01339
18	−.01091
20	−.00605

VALUES OF R_m^2 FOR VARIOUS 21-TERM GRADUATORS

m	GRADUATOR				
	Minimum R_0^2	Minimum R_3^2 Q_1	Minimum R_4^2 Q_2	Spencer's 21-Term	Q_9
0	.1076	.132215	.139035	.1432	.137387
3	.0032	.000023	.000027	.000039	.000039
4	.0030	.000005	.000003	.000010	.000016

It has been noted earlier that the value of $\mathcal{P}(2)$ for the graduator Q_9 is 0, and this is the value of R_∞ if the latter exists. From the value of R_m^2, it appears that Q_9 is equivalent to Spencer's formula by the R_3^2 criterion, but by the R_4^2 criterion, Spencer's formula smooths more strongly. However, R_0^2 for Spencer's formula exceeds that for Q_2 which in turn exceeds R_0^2 for Q_9.

As an attempt to evaluate periodograms and how they fit into actuarial knowledge, we may say that they are a special case of characteristic functions where by reason of symmetry the interpretation in regard to wave modification follows. Characteristic functions provide a powerful tool in regard to probability distribution theory, and work such as that of Greville and Schoenberg indicate they are significant also for graduation and interpolation. Periodogram theory is a mathematical not a statistical approach and may, as Mr. Wolfenden suggests, provide only an a priori test for rejection of a graduation process rather than for confident acceptance. It may remain simply an interesting mathematical theory, but one never knows when such a theory may yield a practical application. It seems limited to graduation and interpolation processes based on symmetric linear operators.

For the benefit of Mr. Rosser, students, and others who may be interested in the application to interpolation, we give the subtabulator or symmetric linear operator that may be read off from the coefficients given by Boyer (*RAIA*, XXXI [1942], 341) for subtabulating at subinterval 1/5 by means of the Jenkins modified fifth-difference osculatory interpolation formula, namely:

$$\begin{aligned} S = {} & .8\dot{3}\,G_0 + .779\dot{5}\,G_2 + .643\dot{1}\,G_4 + .461\dot{3}\,G_6 + .271\dot{5}\,G_8 \\ & + .\dot{1}\,G_{10} + .008\dot{6}\,G_{12} - .041\dot{7}\,G_{14} - .054\dot{8}\,G_{16} - .045\dot{3}\,G_{18} \\ & - .02\dot{7}\,G_{20} - .014\dot{2}\,G_{22} - .006\,G_{24} - .001\dot{7}\,G_{26} - .000\dot{2}\,G_{28} \end{aligned}$$

When S is applied to a series of values U_y obtained from given values u_{-2}, $u_{-1}, u_0, u_1, u_2, u_3$, by the rule $U_y = 0$, y not a multiple of 5, and $U_{5x} = u_x$, $x = -2, -1, 0, 1, 2, 3$, it subtabulates the values v_x at $x = 0, .2, .4, .6, .8$, and 1.0 which would be given by direct application of the Jenkins formula. A periodogram for S may be obtained by applying formulae (16) and (13) of the paper. If one prefers, one may work instead with (1/5) S which is a graduator in the sense of Part III of the paper. For further information concerning subtabulators such as S, one may refer to the paper by Greville (listed as reference [6] in our paper) or to reference [2].

MINIMUM PREMIUMS PROMULGATED BY NEW YORK FOR GROUP LIFE INSURANCE ISSUED IN CANADA—ACTUARIAL NOTE

MORTON D. MILLER

SEE PAGE 28 OF THIS VOLUME

GERALD B. ANGER:

The propriety of using population mortality differences between geographic regions to approximate group insurance mortality differences is open to question.

Mr. Miller states that the advisory committee felt that such use was in order since "group insurance covers such a broad cross-section of individuals." But is the cross-section broad enough when we note that farmers, ranchers, fishermen, share-croppers, field laborers, sole proprietors, employees of small firms, professionals (such as physicians, dentists, lawyers etc.), indigents, inmates of institutions, and others, are not normally covered by group insurance and that the proportion of total population of these excluded classes is not necessarily the same from one region to another? For example, is the cross-section broad enough such that the relative proportions of nonwhites to whites covered under group insurance in the United States and Canada are the same as those given by population tables?

Perhaps the experience results (Table A) of the Canadian Association of Actuaries on 1963 Canadian Group Life Experience as compared to the 1955–59 Society of Actuaries group experience rates will tend to confirm the view that the use of population data is questionable. Table A actually suggests that there is no significant difference between United States and Canadian group mortality—especially if the mortality improvement trend has continued and the 1955–59 experience rates are projected to 1963!

Of course, if the premise that population mortality differences do reflect group insurance differences is accepted as valid (or can be substantiated from group insurance experience results), it would seem (when we note that the population of Canada is of the same order of magnitude as New York or California) that the argument for variation in minimum rates between the United States and Canada could be made with equal force to produce various minimum scales within the several major regions of the United States. Table B suggests the pattern which might obtain.

TABLE A

CANADIAN ASSOCIATION OF ACTUARIES
1963 CANADIAN GROUP LIFE EXPERIENCE—RATED AND NONRATED INDUSTRIES—POLICIES COVERING 25 LIVES OR MORE—ALL DISABILITY PROVISIONS COMBINED

CENTRAL AGE	EXPOSED TO RISK	DEATH AND DISABILITY CLAIMS*	FORCE OF DECREMENT 1963 C.A.A.	FORCE OF DECREMENT 1955–59 S. of A.†	RATIO 1963 C.A.A./1955–59 S. OF A.
18........	40,429	47.5	.00117	.00113	1.04
23........	118,931	118.5	.00100	.00107	0.93
28........	121,859	133.5	.00110	.00106	1.04
33........	134,237	173.5	.00129	.00124	1.04
38........	135,029	244.5	.00181	.00188	0.96
43........	124,389	354.0	.00285	.00316	0.90
48........	108,424	523.0	.00482	.00538	0.90
53........	92,754	738.0	.00796	.00885	0.90
58........	69,303	1076.5	.01553	.01477	1.05
18–58.....	945,355	3409.0	.00361	.00375	0.96
63........	45,792	857.0	.01872	.01994	0.94
68........	22,016	674.0	.03061	.03110	0.98
73........	11,658	595.0	.05104	.04888	1.04
78........	5,764	451.0	.07824	.07854	1.00
83........	2,266	254.0	.11209	.12219	0.92
88........	601	115.0	.19134	.18555	1.03
93........	107	31.0	.28972	.25473	1.14
98........	9	5.0	.55555	.22318	2.49
63–98.....	88,213	2982.0	.03380	.03444	0.98
Total....	1,033,568	6391.0	.00618	.00637	0.97

SOURCE: Table 11, Appendix B, December, 1964, *Report of the C.A.A. Mortality Committee (Group)*.

* The number of death and disability claims is equal to the number of death claims plus 75 per cent of the Waiver of Premium Disability claims plus 50 per cent of the Total and Permanent Disability claims.

† Derived from Report of the Committee on Group Insurance Mortality, Table 2, *1960 Reports of the Society of Actuaries*, all disability provisions combined with the number of death and disability claims computed as in * note above.

TABLE B

1961 MALE AND FEMALE* MORTALITY

AREA	20–44 YEARS		45–64 YEARS	
	Rate per 1,000	Ratio to U.S.	Rate per 1,000	Ratio to U.S.
Canada..............	1.59	0.79	9.47	0.83
New England.........	1.67	0.83	11.23	0.99
Middle Atlantic.......	1.96	0.97	12.06	1.06
East North Central....	1.87	0.93	11.16	0.98
West North Central...	1.73	0.86	9.98	0.88
South Atlantic........	2.43	1.20	12.78	1.12
East South Central....	2.45	1.21	11.88	1.04
West South Central....	2.12	1.05	10.88	0.96
Mountain............	2.09	1.03	10.24	0.90
Pacific...............	1.87	0.93	10.48	0.92
United States.........	2.02	1.00	11.39	1.00

SOURCES: Dominion Bureau of Statistics—Cat. No. 84-202, Vital Statistics, 1961. U.S. Dept. of Health, Education, and Welfare, *Vital Statistics of the United States 1961*, Vol. XI, Sec. 1.

* Male only was not available by U.S. major regions.

(AUTHOR'S REVIEW OF DISCUSSION)

MORTON D. MILLER:

I am glad that Mr. Anger has put into the *Transactions* the initial year's compilation of Canadian group life insurance experience by the Canadian Association. As mentioned in the Actuarial Note, intercompany data under Canadian policies were not available when the Advisory Committee was called upon to make its study. It was therefore necessary to look to other indicators of comparative mortality levels such as the population tables which were used.

Such Canadian experience as the companies represented on the Advisory Committee had was fragmentary. What there was tended to support a level of Canadian mortality on the order of that recommended to the Department.

Mr. Anger's material is, of course, only one year's experience, so that the significance of the Canadian studies cannot be fully assessed until additional statistics are accumulated.

BAYESIAN STATISTICS

DONALD A. JONES

SEE PAGE 33 OF THIS VOLUME

JOHN M. BOERMEESTER:

Dr. Jones states that the objective of the paper is to bring Bayesian statistics to the attention of the members of the Society of Actuaries. Confronted by such an open invitation, I decided this would be an opportunity to review the subject of statistics from the viewpoint of a member of the Society who last studied statistical theory for examination purposes more than 25 years ago.

Dr. Jones refers to a paper by Mr. A. L. Bailey in the *1950 Proceedings of the Actuarial Society*. Mr. Bailey, in effect, stated that there was no formal statistical technique available to an actuary to develop so-called "credibility factors" (Z) for experience rating purposes and that he (the actuary) had to depend on empirical methods to derive them. This paper then developed a number of illustrative formulae which may be used to calculate credibility factors under Bayesian theory. The paper also contained a reference to Mr. Ralph Keffer's paper published in *TSA*, Vol. XXX (1929).

I became curious concerning the nature of Mr. Keffer's credibility factor (Z), shown on page 113 of *Actuarial Studies No. 6*. This credibility factor had been suggested by Mr. Keffer for use in the experience rating process for group life contracts. Was Mr. Keffer's factor just another empirical device, or could it be defended by Bayesian statistical theory?

In Mr. Keffer's development of the credibility factor, he first assumed that the prior frequency distribution of loss ratios for all group cases was in the form of a gamma distribution. He then assumed that the conditional frequency distribution for deaths in connection with the experience for any one case was in the form of a Poisson distribution. As a consequence of the theory of conjugate sets, the posterior frequency distribution for the loss ratios emerged in the form of a gamma distribution. Mr. Keffer used the posterior distribution to obtain a point estimate for the mean loss ratio applicable to the actual experience of a particular group. The result was recast in the familiar formula for the credibility factor Z which is stated in the Actuarial Study.

It would appear to me that Mr. Keffer used the principles outlined by Dr. Jones. If my reasoning is correct, I wonder whether Mr. Keffer was, in fact, a Bayesian statistician. If so, can we be consoled by the Bayesian school that the stigma of empiricism should never have been given to those who have used Mr. Keffer's credibility theory for the past 35 years?

In conclusion, I wish to state that we should be most grateful to Dr. Jones for giving us the opportunity to become acquainted with some of the more recent developments in Bayesian statistical theory.

HUGH H. WOLFENDEN:

Professor Jones' paper performs a useful service in once more bringing to the attention of actuaries the famous work of the Reverend Thomas Bayes. His references to Whittaker's views on probability and graduation theory also provide a needed opportunity to re-examine Whittaker's opinions, for the latter's *TFA* paper and its accompanying discussions left the entire subject of the place of the Bayes-Laplace formulae in actuarial techniques in a state of largely unresolved confusion.

When Professor Jones notes that the courses of reading suggested in this Society for years have omitted any convenient reference to Bayes' theorem, I am bound to point out that in my book *The Fundamental Principles of Mathematical Statistics*—which was published by the Actuarial Society, and is still available, for the express purpose of providing information in an accessible form for students on this and many other matters—the basic principles are explained on pp. 7–8, 165, and 221–23, in statements which should be understandable easily to actuaries already trained to think in terms of numerical probabilities of death, q_x, based on observed values of θ_x' and E_x'. (In the subsequent portions of this discussion it will be convenient to refer to that book as "*FPMS*.")

One of the points emphasized particularly therein (pp. 221–23) is the distinction to be maintained between "Bayes' Theorem" of formula (43*b*) for the special case when the a priori existence probabilities, κ_r, are all equal, and the generalized "Bayes-Laplace Theorem" (43*a*) when the κ_r's are not all equal. Here it may also be of assistance to observe, in slightly different notation and form (cf. *TASA*, XXX, 279), that the Bayes-Laplace formula states that when p_i is the a priori probability of the existence of the *i*th cause F_i, and P_i is the a priori probability that when F_i exists the event E will happen, then the probability, a posteriori, that E actually happened from the cause F_i is $p_iP_i/\Sigma_ip_iP_i$; and that Bayes' formula is P_i/Σ_iP_i when all the p_i's are taken as equal. The Bayes-Laplace formula "represents a perfectly sound and logical argument, and leads to unexceptionable results when it can be applied rigorously"; but when the prior probabilities are supposed, in the absence of specific information, to be equal merely because their real values are unknown (in accordance with the "principle of insufficient reason"—Boole's "equal distribution of ignorance"), Bayes' formula may lead to questionable or absurd results—for "Ellis' remark must always be remembered, that 'mere ignorance is no ground for any inference whatever; *ex nihilo nihil*'" (*FPMS*, p. 222).

In this necessarily short discussion there is no space to examine in any detail the implications and possible uses of these Bayes-Laplace formulae in actuarial procedures. Their consequences, however, have interested me for so many years that recently—prior to the receipt of Professor Jones' paper—I have prepared a careful exploration of those questions (particularly in relation to the determination of probabilities from observations, and Whittaker's approach to his graduation rationale), with the intention of including the material as one of several commentaries in two volumes on the life and work of Erastus Lyman De Forest (see my paper, *TASA*, XXVI, 81), with reproductions of his writings, which I plan to publish soon with the collaboration of Dr. T. N. E. Greville.

Without attempting at this time, therefore, to summarize the character or conclusions of that projected publication, it may be useful here to remark that:

1. Throughout the development of their practical techniques, actuaries have, quite properly, instinctively preferred the "maximum likelihood" approach of estimating q by maximizing the probability, $\binom{E'}{\theta'} q^{\theta'} p^{E'-\theta'}$, that θ' deaths did actually occur by chance out of E' at risk (cf. *FPMS*, p. 239). Thereby usually they have not had to reconcile that obvious method (wherein q_x is carefully defined under specified conditions as a ratio obeying the addition and multiplication rules of probability), with the metaphysical concepts of "probability" (cf. *FPMS*, pp. 4, 7, and 179–87) which lead into many obscure difficulties in the abstract mathematical interpretations of such hypothetical approaches.

2. The highly controversial possibility of setting up some hypothesis for the prior probabilities in the Bayes-Laplace formula (43*a*) has been considered also in a long contribution by Perks, in *JIA*, LXXIII, 285—to which Professor Jones does not refer. Perks commented (p. 285; see also E. S. Pearson, p. 324) that "the paper is concerned with fundamental questions of a controversial nature, and has little, if any, immediate practical aspect, at any rate so far as applied actuarial science is concerned," though he added that "the time [1947] is more than ripe for actuaries to re-examine the fundamental bases of their processes."

3. Whittaker's use of Bayes' formula with its arbitrary assumption of an equal distribution of ignorance inevitably introduces grave difficulties into his theoretical treatment of mortality formulae.

4. The rationale, as set out by Whittaker (his book, pp. 303–6, and restated by Jones), of his idea of basing a graduation method on the addition of criteria for fit and smoothness in an arbitrary proportion again involves the "equal distribution of ignorance" assumption of Bayes' formula (see *FPMS*, p. 138), which is one of my reasons for remarking—in another discussion on the paper by Andrews and Nesbitt in this *TSA* volume—upon the difficulty of fully accepting Whittaker's reasoning.

On the wider question of the general appropriateness of employing the latest interpretations of "Bayesian methods" in actuarial work, it is of course true that the Bayes-Laplace approach is exceedingly intriguing, both philosophically and mathematically; occasionally, as already noted, it is directly applicable to practical statistical problems when the prior probabilities can be stated specifically and surely; but the assumption in Bayes' formula of equally distributed ignorance, or the invention of some artificial hypothesis of "cogent reason" for inclusion in the Bayes-Laplace expression, may introduce unavoidably an undesirable or intractable element of doubt into any apparent inference—and this is so even though it often may be found in practice that the form of such a hypothetical prior probability function may not influence the numerical results greatly when the data are large (as indeed has been emphasized by Laplace and Poisson, Sir George Hardy, and by modern writers like E. C. Molina, Sir Harold Jeffreys, and M. G. Kendall).

Professor Jones' interesting paper, nevertheless, perhaps may encourage actuaries in the future to consider more closely these efforts to incorporate a "well defined judgment factor" (to use his description) in a few of their objective technical analyses.

This discussion considers only the essential distinction, which for the sake of justice and clarity ought to be maintained, between the formulae of Bayes and Bayes-Laplace, with certain applications of their approaches to the determination and graduation of such basic actuarial functions as q_x. It is not intended to deal with their possible applications to credibility and experience rating problems which are examined in papers at the Casualty Actuarial Society included in Professor Jones' bibliography.

CHARLES GREELEY:

The author has written an excellent paper which should certainly fulfill his objective "to bring Bayesian statistics to the attention of the members of the Society of Actuaries." He chose the right audience, since Actuaries, consciously or otherwise, have been practicing Bayesian statistics for many years, as can be proven by the following syllogism.

1. Whatever actuaries do, they are expert at it.
2. Most actuaries play bridge.
3. Therefore, most actuaries play expert bridge.
4. Bridge experts can solve the following type of problem.
5. No one unfamiliar with Bayesian statistics could solve it.
6. Therefore, a bridge expert knows Bayesian statistics, and that includes most actuaries.

Suppose declarer (South) holds A 4 3 2 of Spades opposite North's K J 7 6 5. When South first plays Spades, by leading the Ace, suppose that both opponents follow suit with low cards, and when South next leads a low Spade, West again follows suit with a low Spade. Now the odds in favor of playing the King as opposed to the Jack are 12 to 11, since for East the Queen might be one of 12 other cards, but for West it can be only one out of 11.

Consider next a different holding in which North holds the Spade ten instead of the Jack. When South plays the Spade Ace, suppose West follows low, and East drops the Spade Jack. Next, when a low Spade is led and West again follows low, we must choose between playing the King or the ten. Superficially, the odds seem 12 to 11 in favor of playing the King, but, surprisingly, assuming expert opponents, it will be shown that the odds are 11 to 6 in favor of the play of the ten.

Given South's Spade holding of A 4 3 2 and North K 10 7 6 5, consider the universe of possible hands in which West holds the 9 and 8 of Spades plus eleven other cards, one of which may be the Queen of Spades, and East holds the Jack of Spades plus twelve other cards (in other words, East's Spade holding is either the singleton Jack or the doubleton Queen and Jack).

Let A represent "East *plays* the Jack," and B represent "East *has* the Queen."

Within the given universe, let us first evaluate the a priori probability $P(A)$ that East will play the Jack.

$P(A) \quad = P(A/B)\,P(B) + P(A/\bar{B})\,P(\bar{B})$

$P(B) \quad = 12/23$

$P(\bar{B}) \quad = 11/23$

$P(A/B) = 1/2$ because an expert East, holding the Queen as well as the Jack, would falsecard half the time.

$P(A/\bar{B}) = 1$ since East would have to play the Jack if it were a singleton.

$$\therefore P(A) = (\tfrac{1}{2} \times \tfrac{12}{23}) + (1 \times \tfrac{11}{23}) = \tfrac{17}{23}\,.$$

Let us next evaluate $P(B/A)$, the a posteriori probability, within the universe defined, of East having the Queen of Spades, *given the fact that he played the Jack.*

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} = (\tfrac{1}{2} \times \tfrac{12}{23})/(\tfrac{17}{23}) = \tfrac{6}{17}.$$

Conclusion: Since the odds are 11 to 6 that West has the Queen of Spades, you should definitely play the ten, not the King, and then thank Mr. Bayes when you collect your winnings more often than you used to.

Observe the typically Bayesian method of analysis used in this problem to reach a decision. In the game of bridge, the effect on probabilities of every bid and play of the opponents can best be evaluated in a Bayesian manner. Your course of action depends first on your prior evaluation of the probabilities, including your personal estimate of the probability of an opponent acting in a particular manner, and then on your computations, according to Bayes' theorem, reflecting any significant and relevant evidence obtained.

Thanks to Dr. Jones, I expect that, in the future, many more actuarial applications of Bayesian statistics will be found; they will be well worth the effort if they touch the practical interests of Actuaries as much as does the application to the game of bridge!

JAMES C. HICKMAN:

I have the feeling that Professor Jones' paper may mark a significant turning point in the relationship between actuarial science and statistics. In the past, actuaries and other business executives who must make decisions on quantitative matters have not relied on classical statistical methods to the extent that natural scientists have. This caution has, in part, been caused by a feeling that classical statistical models are somewhat divorced from business reality. In particular, there is the common complaint that past experience is given insufficient weight in these models. The warning that current statistical results must be modified by that desirable but elusive quality, common sense, pops up again and again in actuarial literature. The Bayesian approach to statistics now offers a formal framework within which past experience and current results may be explicitly introduced into a decision problem. This is all to the good. The requirement that the information which summarizes past experience must be quantified, in order to define a prior distribution, will force decision-makers to clear thinking by more sharply defining the areas and the extent of personal differences in viewing the situation surrounding a decision problem.

It is not surprising that classical statistics has developed largely as a tool of natural science, while Bayesian statistics seems to have greater applicability in the solution of social, business, political (including war) problems. Since Galileo's dramatic success in the Renaissance, it has been part of the method and folklore of natural science that the first step in seeking new truth is to discount past theory. This preliminary intellectual sterilization, and a general and uniform interest in all alternative hypotheses, are common characteristics of both natural science and classical statistics.

In the fields of business and politics, however, decision-makers have been trained by intensively studying past experience. They have been impressed with the profusion of the states of nature that they face in their decision-making, and they have not enjoyed the luxury of being able to suspend judgment on urgent questions. In addition, some consequences of possible decisions in these fields involve such large losses that the probability of making a decision that could lead to these abhorred results must be minimized. Thus it seems natural that the Bayesian approach to decision-making with its promise of a formal mathematical framework that forces the blending of past experience and current results, and with the flexibility that it permits in ordering possible consequences, will be vigorously developed in these fields.

Professor Jones has provided actuaries with a program for applying Bayesian statistics to some of their problems The two principal points in this program are the application of Bayesian approach to graduation and to credibility theory. As he points out in the paper, both these actuarial problems have long and interesting histories and both have been attacked in the past by almost-Bayesian approaches. I am particularly interested in the history of difference equation graduation as it is traced in the paper. If one follows the graduation path blazed by Bayes, King, Whittaker, and Jones, do statistical tests of the graduation, as are mentioned in the Miller monograph and developed in Seal's paper (*JIA*, Vol. LXXI), any longer have relevance? It would appear that perhaps within the framework of this theory the "best" graduation would have already been done.

Before turning our attention to another example of the application of Bayesian statistics in actuarial science, it is interesting to note a certain parallelism between objective and personalistic probability on the one hand, and risk and uncertainty on the other. Risk and uncertainty were the terms adopted by Knight,* in his pioneering work on the economic consequences of random losses, for differentiating between possible losses that are unique (uncertainty) and those where empirical results provide

* Frank H. Knight, *Risk, Uncertainty and Profit* (Boston: Houghton Mifflin, 1921).

an estimate of the probability of occurrence (risk). Situations where the probability of loss is obvious from the very nature of the situation are also called risks under this classification. These definitions have been used by many subsequent authors in this field. However, this is a classification that is somewhat difficult to maintain. The dynamic nature of the forces that determine the distribution of random losses serves to soften this distinction considerably. Similarly, statisticians have sought to build walls, both a conceptual and a notational wall, between objective and personal probabilities. This effort also seems to hold little promise of complete success, for in a real sense the personal approach contains the objective. Certainly a rational decision-maker would not overlook the results of a great many iterations of the random experiment under study. It is also interesting to note that the mental process through which a rational decision-maker assigns personal probabilities to events is very similar to that employed by an experienced underwriter who is faced with the problem of rating a variety of risks.

I am convinced that the development of Bayesian approaches to actuarial problems should be one of the principal current intellectual goals of the actuarial profession. Therefore it seems appropriate in the remainder of this discussion to turn to such an application that is mentioned at the end of Section 4 of the paper. However, it will first be necessary to introduce some additional ideas.

One aspect of recent developments in statistics has been in the area of statistical decision theory. The problem in this area is to formulate a decision rule that depends on a random variable X where the distribution of X in turn depends on a parameter θ. To each pair $[\theta, d(x)]$, we assign a nonnegative number $L[\theta, d(x)]$ which in some sense measures the loss in making decision $d(X)$ when the distribution of X is determined by θ. The expected value of this loss function, for fixed θ, $E_{X|\theta}\{L[\theta, d(X)]\}$, is called, in this theory, the risk function and is denoted by $R(\theta, d)$. One approach to the problem, which leads to a solution in only some special cases, is to select that decision rule which will minimize $R(\theta, d)$ for all θ. Another approach to selecting an optimum $d(X)$ involves considering θ as a random variable, with a distribution determined on the basis of prior information. This approach calls for determining the $d(X)$ that minimizes $E_{\theta|X}\{L[\theta, d(X)]\}$, the expected value being taken with respect to the posterior distribution. A decision rule selected by this method is called a Bayes solution.

An illustration of an application of this method may be developed by building on example E of the paper. This seems to be an especially appropriate example for actuaries, for although it is in a biostatistical setting, the change of a few words makes it into a problem of estimating the probability of no claim for an individual risk belonging to a proposed new risk class.

Our problem is to estimate t by $\hat{T}(n)$ where $\hat{T}(n)$ is a function of the $x + y = n$ outcomes of Bernoulli experiments. A natural and very tractable loss function is given by $L[t, \hat{T}(n)] = c[\hat{T}(n) - t]^2, 0 < c$. If we follow the Bayesian procedure, outlined previously, we will seek to find $\hat{T}(n)$ so that

$$E_{\theta|x}\{L[t, \hat{T}(n)]\} = \int_0^1 c[t(n) - t]^2 K(x+p, n-x+q) \times t^{x+p}(1-t)^{n-x+q} dt$$

is a minimum. It is easy to show that the required Bayes solution (estimate) is $\hat{T}(n) = (x + p + 1)/(n + p + q + 2)$ and that

$$E_{\theta|x}[T] = \hat{T}(n) = [n/(n + p + q + 2)][x/n] + [(p + q + 2)/(n + p + q + 2)][(p + 1)/(p + q + 2)] .$$

Note that as the number of observed individual risks (n) increases, the prior estimate $(p + 1)/(p + q + 2)$ is overwhelmed by the experimental data (x, n). This simple example serves to illustrate a Bayesian approach to an estimation decision. It could, just as well, have been stated in terms of a credibility problem (see Mayerson [12]).

JOHN A. MEREU:

Dr. Jones has written a very readable and interesting paper on the subject of Bayesian statistics, a subject which apparently is experiencing a revival of interest.

Many of us have had only a limited contact with the subject, and that under the name of "inverse probability." You may recall the type of problem where the poser says that either there are four white balls or two white balls and two black balls in a box and that either situation is equally likely. If a white ball is drawn you are asked to calculate that all of the balls are white. In practical applications of Bayesian statistics the prior information is never so clearly specified.

It is interesting to note that the Whittaker-Henderson methods of graduation can be justified by Bayesian statistics. The choice of the factor h to reflect the relative emphasis to be given to fit and smoothness is left for personal judgment, and no light is shed on the most probable value of h.

Perhaps it would be useful to simulate crude data on a computer by reversing the usual graduation process. One could start with a set of smooth values, say, from some mathematical function, and then cause each value to wander about its true value in some random fashion. For each set of crude data so generated one could determine the value of h which would again produce the original smooth values.

I believe that because of the controversy in the field of statistics between classical and Bayesian statistics, it would be useful to have the classical side of the arguments before making the decision to become a Bayesian. If the Bayesian approach proves itself, the Society motto might be reworded to: "The work of science is to substitute facts for appearances by blending demonstrations and impressions."

In conclusion, may I say that Dr. Jones has done the actuarial profession a real service by giving us his survey of Bayesian statistics. In it there is much food for thought.

EDWARD A. LEW:

We are greatly indebted to Dr. Jones for his highly intelligible presentation of the Bayesian approach. His paper and that of Allen Mayerson on "A Bayesian View of Credibility" show that certain actuarial problems can be handled with advantage by calculating probabilities on the basis of assumed prior knowledge and observational data, using Bayes' theorem.

When the assumed prior knowledge is derived from other statistical data, there is substantial agreement about the application of Bayes' theorem. The issue between the classical and Bayesian statisticians is joined on the question of whether prior information that is not statistical in nature can properly be included in reaching useful decisions. This turns on our willingness to accept a theory of personal probability.

If we take the position that any theory of probability is pure mathematics arising from a set of postulates, then different theories of probability can be obtained by changing postulates. In applying a particular theory of probability, the pragmatic test is whether such theory and its underlying postulates fit the real life situation in which we are interested. It is possible that under some circumstances the theory of personal probability may give more meaningful answers than the frequency theory, while in certain problems the reverse may be true; it depends on the questions we ask (e.g., inferences about hypotheses or a decision rule) and the assumptions we believe to be pertinent to a particular real life situation (e.g., a set of repetitive events or a specific event). In many situations it would be convenient if we could retain the advantages of the personalistic view without the sacrifice of objectivity which this approach entails.

The major difficulty with the Bayesian approach lies in assigning objective values to the a priori distributions in terms of personal probabilities:

Even L. J. Savage is on record that "it is usually possible to determine the personal probabilities of important events only very crudely," although he warns about drawing the "over hasty conclusion that such determinations are of little worth."

An important advantage of classical confidence intervals is that they can be calculated precisely on the frequency theory without any assumptions about prior distributions. This economy of assumptions leads to very general statements about intervals containing an unknown parameter at a preselected probability level, but such general statements are sometimes not very informative. The Bayesian approach leads to a logically more pertinent statement about an unknown parameter which is related to our prior knowledge of the parameter in terms of personal probability, but the numerical answer may be highly subjective.

This criticism is obviously valid, since personal probabilities may differ widely from one individual to another and even for the same individual from time to time, and they may ultimately rest on the variable idiosyncrasies of human beings. L. J. Savage has stated that any personal probability should "in principle be indexed with the name of the person or people whose opinion it describes." Bayesians are likely to insist that the personal probabilities of an ideal individual reflect not merely his vague introspections but rather the results of a search for prior information. It may further be argued that normal people will have had similar experiences and hence may reasonably be expected to interpret them within limits in the same way. Even when such is not the case, the use of personal probabilities might be warranted in circumstances where their combination with sufficient observational data by the application of Bayes' theorem yields similar posteriori probabilities for different individuals. The discrepancies between the posteriori probabilities of individuals with different a priori opinions tend to become smaller as observations mount, in spite of conflicting a priori judgments.

L. J. Savage rightly points out that the classical approach to inference has been pursued without reference to the existence of personal differences in judgment. Classical statistics, says Savage, "is largely devoted to exploiting similarities in judgments of certain classes of people and in seeking devices, notably relevant observation, that tend to minimize their differences."

Such devices include also the design of experiments and formalized approaches to the analysis of data which introduce some subjective elements into the application of the frequency theory. Even more sobering is the realization that after inferences based on the frequency theory have been drawn from observations, the results may still have to be considered in relation to prior knowledge. In this connection, it is interesting to note that C. B. Winsten has suggested that the term "weight" rather than "probability" be used in describing prior distributions.

When we do decide to follow a theory of personal probability, we will need a systematic approach for selecting a prior distribution that will best summarize a priori knowledge, be relevant to the problem at hand, and appraise the stability of the process under observation, at least when there are marked divergences between the prior distribution and the distribution implied by the observations.

Because Bayesian distribution theory is only about ten years old and apt to be more complicated than classical distribution theory, it must be used with great circumspection. The possible usefulness in the Bayesian approach of nonparametric methods, where the main concern is to estimate the parameters of unknown distributions, remains to be explored.

The concept of "consistent behavior" on which the theory of personal probability relies needs to encompass more explicitly the question of how relevant are the gambles representing prior knowledge to the gambles concerning anticipated events. The importance of basing judgments on relevant information is illustrated for actuaries by the early history of total and permanent disability benefits when for lack of better information the experience of fraternal orders was initially taken as the main source of prior knowledge about disability rates.

Underlying all theories of statistical inference is the principle of stability. In the frequency theory we rely on the premise that indefinite replication of observations will continue to yield essentially the same results. When the parameters of the process in which we are interested are liable to change with time or otherwise, we cannot count on stability, and in such circumstances neither classical theory nor the Bayesian approach will provide fertile grounds for induction. In his recent presidential address to the Royal Statistical Society, J. O. Irwin commented on inductive inference as follows:

> I have never concealed my personal conviction that we cannot get beyond the position, stated as early as Hume, that belief in the validity of inductive inference is, in the last resort, a matter of faith and faith only. It seems, therefore, largely a matter of taste at what logical depth we seek our primary postulates and axioms and whether we give more emphasis to objective or subjective theories of probability. This particular difficulty is not peculiar to the theory of statistics but concerns scientific inference in general.

In my opinion the Bayesian viewpoint can be very helpful if we look upon prior personal probabilities simply as useful hypotheses rather than as uniquely intelligent guides for decision. If we make several plausible a priori judgments and regard them as alternative hypotheses, we can learn a great deal about the range of posteriori probabilities produced by plausible assumptions as to prior probabilities and thus be in better position to make optimal decisions. This approach has some points of resemblance to the method of stable estimation advocated by L. J. Savage.

HARRY M. SARASON:

Professor Jones has made some of the statistical mathematical viewpoints more understandable to actuaries by use of the probability approach. He has also introduced numerical illustrations which are still more understandable. I wish to enlarge upon use of numerical aids to understanding, by use of samples for survival rates as an illustration. Suppose the survival rate is three in four. We set up the hypothesis of independent probabilities; no catastrophe or epidemic. Then (1) trial-and-error based on a four-sided log marked d on one side could be used as a demonstration. (2) The Monte Carlo technique with random numbers could be used. (3) With some loss of realism a mathematical model based on an even distribution could be used; e.g., if sequences of eight tosses were used, we would have two to the eighth power or 256 sequences of deaths and survivals illustrating all possibilities, with some repeated to indicate relative probabilities, for a grand total of four to the eighth power. (4) With additional loss of realism we could replace this by four series, as follows: (*a*) dsss, dsss; (*b*) sdss, sdss; (*c*) ssds, ssds; (*d*) sssd, sssd. From this we could assume samples entering the first sequence of four either evenly or at random. (5) With additional loss of realism, we can assume an average start for our sample (i.e., after the first two items); we then have samples of (*a*′) ss, dsss; (*b*′) ss, sdss; (*c*′) ds, ssds; (*d*′) sd, sssd. In this series we have, all told, twenty-four items of which six are deaths and eighteen (or three-fourths) are survivors. (6) If we close our samples on the first death in the second sequence of four, we have (*a*″) ss, d; (*b*″) ss, sd; (*c*″) ds, ssd; (*d*″) sd, sssd. In this there are eighteen items with six deaths and twelve survivors with a survival rate of two-thirds. The unvarying inclusion of a death as the last item in our sample would produce a bias in our sampling procedure.

Mathematical statistics is used in a judgmental process and is not an exact science. Our students need to be made well aware of the fact that somewhat different numerical answers can be almost equally useful in judgmental process (without lengthening our examination syllabus, of course). It is much more important, however, for students to learn that there are a thousand and one pitfalls which every one of us finds difficult to avoid in the development and use of statistics. In my experience, the major pitfalls have been avoided only when I was wary and was willing to examine all ideas. The pitfalls have not been avoided when I was not wary of prejudging. In forecasting disability rates among insured lives in the early thirties, for example, I never really believed that the depression would end. My forecasts were just as far off as were the earlier forecasts of those who believed that the prosperity of the twenties would continue forever.

The worst pitfall that any statistician can fall into is to underestimate the fact that being an expert in the figures does not make him an expert in the facts which underlie the figures. You can't get the facts by sitting in an armchair, and there are very few who like to get out of their armchairs and go dig up the facts; as Francis Bacon bemoaned somewhat more elegantly in his introductory essay in the *Novum Organum* when he decided that he had to do the job himself. One method of properly emphasizing reality is by giving case histories. My own disability-depression case is one example of this. The first part of the statistics textbook of the Institute of Actuaries is along the case-history method.

Another way to help the students is to emphasize the hypothetical bases of our numerical calculations so as to help them differentiate between thinking based on reality and thinking based on hypotheses. I have seen statisticians try to develop the mathematics of subgroup sampling, rather than sampling by total population, on a hypothesis of statistical homogeneity of the entire population.

Subgroup sampling would work very well in practice because, statistically, birds of a feather tend to flock together, and defective nuts or defective bolts tend to flock together. But subgroup analysis is merely devious and inefficient when used with a hypothesis of statistical homogeneity. If we emphasize the word "hypothetical," it may help our thinking. Perhaps Hall and Knight, in the second half of their sovereigns and shillings illustrative example, used the devious inverse probability solution because in real life sovereigns tend to congregate with sovereigns and shillings tend to congregate with shillings—in supermarket cash registers, for example, and, presumably, in the eighteenth-century English green grocer's cash boxes also. For the benefit of students, then, my hypothetical answer to Dr. Jones' question about tossing a coin is as follows: "If a hypothetical coin is hypothetically tossed, then every hypothetical body would be hypothetically wasting their hypothetical time if they hypothetically bet on the hypothetical outcome of the hypothetical event because both hypothetical bettors would hypothetically know that the hypothetical answer is, quite exactly, the real number, one-half." Curiously enough, we can inject a note of reality by spinning the coin instead of tossing the coin. The probability of heads for a spun coin may be far from one-half. Right here in Denver a large number of pennies were minted which, when spun, came up heads five-sixths of the time—approximately, that is. It may help our perspective to realize that this can lead to three kinds of thinking: (1) testing a single hypothesis, *either* one-half or five-sixths; (2) testing both hypotheses jointly; (3) getting more facts; from me, from the source of my information, by actual trial. The statistician always has various choices of mathematical processes and, more importantly, always has the option of digging deeper into the underlying facts and the related environment, the historical changes, and the shadow of coming events.

HARWOOD ROSSER:

Dr. Jones' "personalistic definition of probability" appeals to me intuitively much more than does the "statistical" definition. So also does his single "black box" with the extra hole. However, the Examination Committee, with the aid of a syllabus change, persuaded me years ago that statistics was not one of my strong subjects. For that reason, among others, I shall confine my remarks to what Dr. Jones has to say about difference-equation graduation.

For one who has made something of a hobby of graduation, it was a pleasant surprise to find two papers being presented simultaneously, one directly on the subject and the other with a section devoted to it. Both deal with the Whittaker-Henderson method of graduation, but from quite different viewpoints. The paper by Andrews and Nesbitt shows that certain difference-equation graduation formulae, including mixed difference types, have desirable periodograms. These they identify as Q_5 through Q_8 on their graphs.

Dr. Jones, in turn, explains *why* this family of graduation formulae gives good results. Much of this is not new, and is, as he notes, available in Miller's monograph on the subject. He goes considerably deeper, however, and gives some history as to controversial aspects with which this reviewer was not familiar.

He also makes a very trenchant capsule commentary on many actuaries, present company not excepted: "To use or not to use the difference-equation method has probably been decided on computational grounds rather than on the merits of its theoretical justification." What a gem! This is particularly true of the Whittaker-Henderson "B" formula. On the other hand, many substitutions could be made for "difference-equation method" in that sentence, and it would still be all too true. Fortunately, in the electronic-computer age, this attitude is on the wane.

Since this reviewer does not feel competent to comment on the broader aspects of the paper, it remains only to compliment Dr. Jones on a very provocative viewpoint and an excellently organized presentation of his subject. In so stating, I make the fairly safe assumption that the quality of what I didn't fully understand, at first reading, measures up to that of what I did. But I refuse to specify any confidence limits!

WILLIAM H. CROSSON:

In his paper, Dr. Jones referred to the problem that the Part II Committee should have. We have recognized for some time that there we have a problem, not merely with respect to Bayesian statistics, but with the question of how to set an examination that properly discriminates between the better and poorer students, and can be described as being "based on the material usually covered in undergraduate mathematics courses in probability and statistics," in view of the increasing diversity in fundamental approaches to statistics.

The purpose of this discussion is to record that the committee has been studying this problem for some time. We do not have the answer yet, and we solicit any suggestions that anyone may have which would help in solving this problem.

(AUTHOR'S REVIEW OF DISCUSSION)

DONALD A. JONES:

My many thanks to the discussants who assisted immeasurably in "bringing Bayesian statistics to the attention of the members of the Society of Actuaries." The number of receptive discussants, the history of "Bayesian actuaries" contributed by Mr. Boermeester and Mr. Wolfenden, and the conspicuous absence of comments by classical statisticians, all seem to me to be evidence of the affinity of Bayesian theory and actuarial practice.

Mr. Greeley and Professor Hickman have contributed excellent examples to the paper. I wish I had opened the paper with Mr. Greeley's clear and concrete application to bridge and closed with Professor Hickman's sophisticated example in decision theory. Both examples should help "bring . . . attention . . ." (though Mr. Greeley's syllogism might distract us mediocre bridge players).

I should try to clear up the misunderstanding induced by the ambiguity of the name "Bayesian statistics." In a school of thought so named one rightfully expects to find those statisticians who use Bayes' theorem:

$$P(B|A) \propto P(A|B)P(B) .$$

Among these statisticians are those who derived their statistical methods from personal probability, i.e., those called Bayesians in my paper. But there are some other Bayesians in the broad sense of using Bayes' theorem. Sir Harold Jeffreys is the chief author for this other Bayesian viewpoint which derives its statistical methods from the *necessary definition* of probability. This is a sophisticated descendant of Laplace's *equally likely definition*. Thus the personalistic Bayesian interprets the probability $P(A)$ as a characteristic of the event A and the individual in contrast to the necessary Bayesian who interprets $P(A)$ as an inherent property of the event A which may be derived on the basis of assuming symmetry at the "right point."

While necessary probability is older than statistical and personal probabilities, the search for symmetry in applications other than games of chance has not been too successful, and hence the number of necessary Bayesians is small.

The first three-fourths of Mr. Wolfenden's discussion with its history and references to the necessary viewpoint is an appreciated addition to the paper. One additional reference to the considerations of "equal distribution of ignorance" is M. T. L. Bizley's *Some Notes on Probability* in *JIASS*, X (1951), 161–203.

I regret that Mr. Wolfenden did not detail his objection (4) to Whittaker's basis for the difference equation method of graduation. As I understand Whittaker in [20], [21], and [22], he used Bayes' theorem, not formula, as Mr. Wolfenden states. As shown in [21], it was G. J. Lidstone, not Whittaker, who suggested "basing a graduation method on the addition of criteria for fit and smoothness in an arbitrary proportion." This proportion was not arbitrary in [22] but rather was the ratio of the variances of two normal distributions.

Mr. Mereu also raised questions about the proportion between the measures of fit and smoothness in the Whittaker-Henderson method. Since h is the ratio of the variances of two prior distributions, the only "light to be shed on the [most probable] value of h" is from the graduator's knowledge. If this knowledge is limited, then he needs to do research, perhaps along the line indicated by Mr. Mereu. For one graduator and graduation problem h is neither a random variable nor an unknown parameter but is a characteristic of the graduator's prior opinion about the problem. Incidentally the chore of making "Bayesian graduation" operational as Mr. Mereu's query suggests is a nontrivial exercise which is the topic of a recently completed Ph.D. thesis at the University of Michigan.

Mr. Boermeester's query about Mr. Keffer's work is related to the ambiguity of the term Bayesian. In the broad sense of using Bayes' theorem Mr. Keffer was using methods of the Bayesian School. The defense of Mr. Keffer's factor turns on one's interpretation of the six assumptions, especially (6), on page 131 of *TASA*, XXX. A classical statistician might defend the six assumptions as "prior knowledge . . . derived from other statistical data" (see Mr. Lew's discussion of this paper) and hence "apply Bayes' theorem rigorously." On the other hand, the six assumptions may be accepted for reasons of symmetry by a necessary Bayesian. But certainly Mr. Keffer may adopt the assumptions from the personalistic view if he intends to use them as a guide to consistent behavior—which was his objective as I see it.

Another major issue which is raised by Mr. Lew and Mr. Wolfenden is the difficult one of objectivity vs. subjectivity. These two words denote extremes of a very elusive concept, but as I interpret them a statistical theory based upon personal probability is more objective than a statistical theory based upon statistical probability. Professor Hickman says this in a positive tone in the last sentence of the first paragraph of his discussion.

Objectivity is sometimes characterized by the appeal, "let the data speak for themselves." L. J. Savage uses the following example in his classroom teaching, and it speaks to this point for me. Suppose that you are a consulting statistician called in on the following three problems. Case A: A musicologist claims he can distinguish between the works of Haydn and the works of Bach. Case B: A lady tea-taster claims she can distinguish between a cup with cream added to the tea and a cup with the cream poured before the tea. Case C: An inebriated person boasts he can predict whether a head or tail will turn up when a coin is tossed by a disinterested third party. Now suppose that ten trials are performed by each of the three claimants and that each obtains nine successes. My reactions are that the musicologist is slipping, the lady tea-taster may have a sensitive mouth, the inebriate was exceedingly lucky—and if objectivity admonishes me to draw the same conclusion about the abilities of all three then objectivity should "be sacrificed."

I don't know how a classical statistician would advise his three clients for there is no place in the classical framework to insert the prior opinion which distinguishes each of these cases. On the other hand, the personal probability Bayesian and his clients have no choice but to construct a prior distribution which represents the client's opinion and knowledge before the data are in. The latter is more objective (Webster's 3d ed. [1*d*]: "expressing or involving the use of facts without distortion by personal feelings or prejudices") for me than the classicist's unguided insertion of the "nondata" information. Mr. Lew sees this chink in the classicist's objective armor when he says, "Even more sobering is the realization that after inferences based on the frequency theory have been drawn from observations, the results may still have to be considered in relation to prior knowledge."

This objectivity versus subjectivity struggle may be eased some by using the terms public information versus private information to remove the emotional attachment that we have to the former terms. If in Mr. Lew's discussion of confidence intervals, we insert "private" for "subjective," "the Bayesian approach leads to a logically more pertinent statement about an unknown parameter which is related to our prior knowledge of the parameter in terms of personal probability, but the numerical answer may be highly private (subjective)," then it sounds about right to me. If the blend of prior information to data is high, then the Bayesian credible intervals will be very dependent on the individual, hence private (subjective) as they should be. On the other hand, when the blend of prior information to data is low, the credible intervals will be approximately equal for most individuals, hence public (objective). For more discussion on this issue see B. de Finetti, "Probability, Philosophy and Interpretation," *International Encyclopaedia of the Social Sciences*, 1963, L. J. Savage [23] p. 178, and [16] Chap. 4, especially Section 4.6.

I would make one more comment on Mr. Lew's discussion of the pros and cons of Bayesian statistics. In his second paragraph he says, "The issue . . . is . . . whether prior information that is not statistical in nature can properly be included in reaching useful decisions." Later he says, "for lack of better information the experience of fraternal orders was initially taken [by actuaries] as the main source of prior knowledge about disability rates." Here I think Mr. Lew has a strong case for actuaries to be on the Bayesian side of the issue, for most certainly some prior information of a nonstatistical nature accompanied the fraternals' disability experience into company use. It would also appear more objective to display this other prior information by a probability distribution than to simply make arbitrary adjustments to the fraternal experience.

Mr. Lew and Mr. Sarason both have understandings of statistics in practice that cannot be gained in the cloistered halls of the academic place. Their discussions of homogeneity and stability are in this esoteric realm of understanding. However, Mr. Sarason tried to climb into my ivory tower with his rigged sample for my survival rate example (p. 46), and I think I'll leave him up there until he "closes his sample on the first death in the second sequence of four" in the series dssd, ssss—"curiously enough, we can inject this note of realism. . . ."

In conclusion, I would assert that classical statistics has failed Mr. Lew's "pragmatic test . . . [of] fitting the real life situation in which we are interested." If the contrary were true, then we should find classical statisticians' contributions to actuarial practice in *TSA*, among these discussions, and among the discussions given before the open meeting of the Research Committee. Due to its failure of the pragmatic test, I suspect that Mr. Mereu will have a long wait to hear "the classical side of the arguments before making the decision to become a Bayesian," so I urge him and others to follow Professor Hickman and Mr. Greeley up the path of applications of personal probability and Bayes' theorem.

SOME INSTANCES OF THE SUPERIORITY OF GEOMETRIC METHODS OVER ARITHMETIC METHODS OF INTERPOLATION AND EXTRAPOLATION

CHARLES B. BAUGHMAN

SEE PAGE 159 OF THIS VOLUME

ROBERT C. TOOKEY:

Mr. Baughman has pointed out the superiority of geometric interpolation and extrapolation methods which we tested by extrapolating cash values. Frequently, the actuary will be trying to design a policy that will produce a desired cash value by the end of the twentieth year. Sometimes it is the level of cash value required that finally determines his selection of interest rate. When cash values according to a set formula are available at $2\frac{1}{2}$ per cent and 3 per cent, the corresponding values at 2 per cent and $3\frac{1}{2}$ per cent may be obtained by extrapolation.

COMPARISON OF ACTUAL AND EXTRAPOLATED MINIMUM CASH VALUES, ISSUE AGE 30

Duration	Whole Life			Endowment at 65			20-Pay Life		
	Actual	Geometric	Arithmetic	Actual	Geometric	Arithmetic	Actual	Geometric	Arithmetic
5.....	$ 29.24	$ 29.53	$ 28.73	$ 57.36	$ 57.62	$ 56.95	$ 65.75	$ 65.86	$ 63.68
10.....	96.13	96.33	95.24	160.84	161.14	160.20	180.14	179.74	176.12
15.....	170.48	170.65	169.39	279.94	280.32	279.34	312.59	311.70	316.99
20.....	251.31	251.49	250.20	416.77	417.22	416.39	466.13	464.66	459.25
25.....	336.78	336.97	335.78	575.03	575.48	574.96			
30.....	424.69	424.92	423.89	763.07	763.43	763.24			
35.....	511.90	512.16	511.32						

The table above compares $3\frac{1}{2}$ per cent minimum cash values at issue age 30 on whole life, 20-pay life and endowment at 65 plans as extrapolated from $2\frac{1}{2}$ per cent and 3 per cent values by the arithmetic and geometric methods of extrapolation with the actual values. It is somewhat surprising that both the arithmetic and geometric methods produce such close results in the case of the whole life and endowment at 65 plans. However, the superiority of the geometric method is quite apparent when extrapolating the values under the 20-pay life plan, leading one to conclude that for the most consistently accurate results, the geometric method is the most reliable in extrapolating cash values. It would follow that this would usually hold true in the extrapolation of policy reserves as well.

ROBERT F. LINK:

The Equitable faced essentially this problem recently, with the following requirements:

1. We wished to derive values of reserves at various odd rates of interest from values at *two* specified rates.
2. A high degree of accuracy was desired, both in interpolation and in extrapolation to higher interest rates.
3. Several thousand such determinations were to be made in one EDP run, with uniform pivotal interest rates but the desired rate varying from case to case.

What we required was a formula of the general type $u_x = f(u_i, u_j)$, adapted particularly to the reproduction of present value interest and annuity functions with average discount periods of from ten to twenty-five years.

COMPARISON OF GEOMETRIC EXTRAPOLATION WITH SPECIAL FORMULA EXTRAPOLATION

(Based on 3 Per Cent and 3½ Per Cent Values)

f (4 Per Cent)	Actual Value	Geometric Value	Special Formula Value	Geometric Per Cent Error	Special Formula Per Cent Error
$(1+i)^3$	1.12486	1.12494	1.12487	.01%	0
$(1+i)^{33}$	3.64838	3.65119	3.64492	.08	− .09%
$(1+i)^{100}$	50.50495	50.62295	48.94920	.23	−3.08
v^3	.88900	.88893	.88901	− .01	0
v^{10}	.67556	.67542	.67570	− .02	.02
v^{33}	.27409	.27388	.27517	− .08	.39
v^{100}	.01980	.01975	.02210	− .25	11.62
$s_{\overline{3}\rceil}$	3.12160	3.12163	3.12156	0	0
$s_{\overline{33}\rceil}$	66.20953	66.10756	66.07397	− .15	− .20
$s_{\overline{100}\rceil}$	1237.62370	1225.28223	1207.99855	−1.00	− 2.39
$a_{\overline{3}\rceil}$	2.77507	2.77492	2.77508	− .01	0
$a_{\overline{10}\rceil}$	8.11090	8.10837	8.10974	− .03	− .01
$a_{\overline{33}\rceil}$	18.14765	18.10575	18.11953	− .23	− .15
$a_{\overline{100}\rceil}$	24.50500	24.20408	24.27745	−1.23	− .93
a_{10}	21.5207	21.3835	21.4183	− .64	− .48
a_{65}	11.6923	11.6755	11.6804	− .14	− .10
a_{90}	2.2692	2.2689	2.2691	− .01	0
A_{10}	.13383	.13241	.13398	−1.06	.11
A_{65}	.51183	.51065	.51113	− .23	− .14
A_{90}	.87425	.87408	.87417	− .02	− .01
P_{10}	.00594	.00592	.00593	− .34	− .17
P_{65}	.04033	.04029	.04030	− .10	− .07
P_{90}	.26742	.26741	.26742	0	0

We first approached the question empirically. Harmonic interpolation seemed to have the right general characteristics and produced rather good results. However, we decided to derive a formula directly based on the characteristics of a present value function, and arrived at the following:

$$u_x = u_i + (u_j - u_i)\left(\frac{x-i}{j-i}\right)\left[\frac{1-\frac{1}{2}(x+i)-\frac{1}{2}(x-i)n}{1-\frac{1}{2}(j+i)-\frac{1}{2}(j-i)n}\right],$$

where

$$0 < i < j \quad \text{and} \quad n = \frac{1}{u_i}\left(\frac{u_i - u_j}{j-i}\right).$$

Our reasoning was threefold:

1. If an average value of the duration until payment, n, could be assigned in a given case, the value of n would aid in establishing the curve, u_x, based on u_i and u_j. We deduced the formula above for an approximate value of n in terms of u_i and u_j.

2. Using this value of n, the value of u_x could be expressed as

$$u_x = u_i + (u_j - u_i)\frac{(1+i/1+x)^n - 1}{(1+i/1+j)^n - 1}.$$

3. By an expansion of the above, and a judicious dropping of terms, we arrived at the formula given.

The table on page 202 illustrates the results of extrapolation in accordance with this formula, in comparison with that suggested by Mr. Baughman. Ours is more accurate in the area of present values of annuities certain, life annuities, life insurances, and net level premiums. Geometric is superior for interest accumulation and discount factors and for annuity accumulation factors.

DAVID M. GOOD:

Mr. Baughman has given a useful reminder that there are more ways than one of performing an interpolation. A particular feature of his development is the criteria supplied for choosing between linear and geometric interpolation, most of which are in a form new to me. His comments and examples illustrate the point that the choice of method depends on available knowledge of the function; this is particularly important when extrapolation is involved. An outstanding example is the extrapolation from $f(0) = 0$ and $f(90) = 1$ to $f(180) = 2$, only to find later that the function is a sine curve!

It has been said that numerical analysis is as much an art as it is a science. This is perhaps because the aim of interpolation, for example, is not to get the "best" result but to get a satisfactory result with the least effort. Here "satisfactory" depends entirely on the use to be made of the result; to achieve this sometimes calls for skill and intuition. To illustrate some other devices that may be used, I will first discuss geometric interpolation as it relates to other methods and then mention a similar useful method adapted to desk calculator computation.

1. The familiar methods of interpolation are based on developing a polynomial which is a useful model of the desired function. Geometric interpolation can be considered a generalization of this in two ways.

First, functions other than polynomials may be used. In Mr. Baughman's method, the function is represented by an exponential function such as

$$u_x = (u_1/u_0)^x \, u_0 \, .$$

The general approach of using exponential functions is of course familiar in the case of the Gompertz and Makeham formulae. However, other functions may well be used. A numerical method of representing the function by a rational function by means of "reciprocal differences" is due to Thiele and may be found in Milne-Thomson, *Calculus of Finite Differences*. This method may be indicated when the function becomes infinite at finite values of x, or approaches a finite value as x approaches infinity. Another example is the method of representing the function in terms of trigonometric functions (harmonic analysis). This is indicated when the function is apparently periodic. Discussion may be found in Whittaker and Robinson, *Calculus of Observations*.

Second, the function may be replaced by another more tractable function, from which the original function may be obtained and for which polynomial interpolation is appropriate. Mr. Baughman's method consists of replacing the function by its logarithm for interpolation purposes. This device has many forms; possibilities sometimes used are polynomial interpolation on $1/u(x)$ or on $x \cdot u(x)$. Combinations of these methods also may be useful.

2. A method similar to geometric interpolation is polynomial interpolation on the reciprocal of the function, which might be called harmonic interpolation. An example of this method is the Balducci approximation for l_x. For the two-point case, the formula is

$$u_x = \frac{u_0 u_1}{x u_0 + (1 - x) u_1} ,$$

which is relatively easy to calculate by desk calculator (geometric interpolation for fractional x may require use of a seven-place logarithm table). This method frequently gives results on the order of geometric interpolation for functions of i (Mr. Baughman's Table 1), but for functions of age or duration it may not be so satisfactory (Mr. Baughman's Table 2). If second differences are ignored and if $u(x)$ does not vary too much in the interval concerned, the respective errors in linear, geometric, and harmonic interpolation are on the order of

$$\epsilon_L = \frac{x(x-1)}{2}[u'']$$

$$\epsilon_G = \frac{x(x-1)}{2}\left[\frac{uu''-u'^2}{u}\right]$$

$$\epsilon_H = \frac{x(x-1)}{2}\left[\frac{uu''-2u'^2}{u}\right],$$

where the quantity in square brackets is to be evaluated at some undetermined point z in the interval.

These indicate that geometric is superior to linear if

$$|uu''-u'^2| < |uu''|$$

or

$$u'^2 < 2uu''.$$

Harmonic is superior to linear if

$$|uu''-2u'^2| < |uu''|$$

or

$$u'^2 < uu''.$$

Neither method is superior to linear if u'' and u are of opposite signs (the curve concave toward the x-axis). This situation is illustrated by the function $a_{\overline{n}|}$ as a function of n in Mr. Baughman's Table 2, and may commonly arise for reserves considered as a function of duration.

3. As an illustration of the use of a preparatory transformation before interpolating, consider the case mentioned above of $a_{\overline{n}|}$ as a function of n. For direct interpolation, linear interpolation is superior. If, however, the function

$$V(n) = 25 - a_{\overline{n}|}$$

is constructed, harmonic interpolation will be useful, and geometric interpolation will be exact. The device may be useful for functions for which the algebra is not so obvious, such as those involving life contingencies. In this case the transformation is related to the limiting value as n approaches infinity.

As Mr. Baughman suggests, such methods are worthy of the further consideration by actuaries. They are practical in those situations where linear interpolation is unsatisfactory; the occasion and manner of their use can only be determined by the computer's knowledge and intuition.

I would be happy to see more actuaries get their feet wet in this stream; they will find many such devices of practical use to them.

HARWOOD ROSSER:

Mr. Baughman's numerical examples are more impressive, or at least easier to understand, than his accompanying theoretical development. He mentions interpolation but confines his illustrations to extrapolation. Accordingly, I have given below a numerical comparison of the results of interpolation under the two methods. In so doing, I have recast the work in what may be a more practical form, somewhat like the usual approach. This simplifies extension to the case where more than three values are given. It also permits some judicious blending of the two methods.

My numerical example assumes that only an isolated value or so is required, or that accuracy is highly important. For interpolation in quantity, especially if quinquennial values are available, I would prefer the linear-compound approach given in my 1962 paper in the *Proceedings of the Conference of Actuaries in Public Practice*, XII, 298.

This reviewer questions the value—for actuaries, anyway—of the elaborate use of remainder terms. In actuarial work rarely does one seek to interpolate or extrapolate a mathematically expressible function. The author attempts to set up criteria to ascertain when geometric methods are preferable to arithmetic ones. These seem, however, to require that the underlying function be known.

Mr. Baughman has, by implication, motivated us to re-examine a very fundamental actuarial subject: the choice of an appropriate formula or method for approximating additional points on a curve. This is, unfortunately, an area in which there is a plethora of formulae but a dearth of tools for prediction of comparative accuracy. More on this later.

For a paper of this brevity, the author has wisely avoided the complementary question of smoothness of results. Also, in government work, the emphasis is usually on forecasting rather than on obtaining intervening figures. The reader who seeks to apply these methods in a different context should, however, be aware of these limitations. This comment on the scope of the paper should not be construed as a criticism of the author's very ingenious, and yet basically simple, approach.

One is moved to speculate as to what would be the impact on the linear compound interpolation formulae of Beers and Greville, if the tests of smoothness were based not on differences of the observed values but on those of their logarithms, as suggested by the author. In the computer age this is something less than utterly fantastic!

COMPARISON OF INTERPOLATION RESULTS

Arithmetic Methods

To obtain what Mr. Baughman calls "arithmetic" results, we use classical finite difference methods. For the value of $a_{\overline{100}|}$ when $i = .0275$, or $n = 2\frac{1}{2}$ (see Table 1), the appropriate formula is

$$U_{n+x} = (1 + \Delta)^x U_n \,. \tag{1}$$

Setting $x = \frac{1}{2}$ and expanding, we obtain

$$U_{n+1/2} = U_n + \tfrac{1}{2}\Delta U_n - 1/8\ \Delta^2 U_n + 1/16\ \Delta^3 U_n + \ \ldots . \tag{2}$$

We employ as many terms on the right-hand side as we have given values. Thus, if three values are to be used, we proceed as though third and higher differences were all zero. It will be noted that picking up an additional value, to improve the approximation, thus means correcting the previous figure by the value of another term of the series.

Substituting values, both given and derived (those employed are italicized), from Table 1 in formula (2), first using only two given values, then three, and finally all four, we obtain results as shown in Table 2. In

TABLE 1

DIFFERENCES AND RATIOS OF $a_{\overline{100}|}$

n (1)	Interest Rate i (2)	$a_{\overline{100}\|}$ (3)	Δ (4)	Δ^2 (5)	Δ^3 (6)	R_1 $(3)_{n+1}/(3)_n$ (7)	R_2 $(7)_{n+1}/(7)_n$ (8)	R_3 $(8)_{n+1}/(8)_n$ (9)
1...	.020	43.09835	−6.48424	1.46904	*−.39732*	.8495478	1.015865	*.9982714*
2...	.025	*36.61411*	*−5.01520*	*1.07172*	....	*.8630255*	*1.014109*	....
3...	.030	31.59891	−3.94348	....	....	.8752020	....	....
4...	.035	27.65543	....	....	....	....	....	....

TABLE 2

"ARITHMETIC" VALUES FOR $a_{\overline{100}|}$ AT 2.75 PER CENT

Approximation (1)	n—Subscripts of Values Used (2)	Approximate Value (3)	Correction to Preceding Figure (4)	Highest Order of Differences Involved (5)
1st........	2–3	34.10651	....	....
2d.........	2–4	33.97254	−.133965	2d
3d.........	1–4	33.94771	−.024832	3d

each case we take $n = 2$. However, in the last case, using $n = 1$ and $x = 3/2$ would give the same numerical result.

Geometric Methods

For "geometric" results, we replace addition by multiplication, use coefficients as exponents, and substitute R's for Δ's, where the former represent the ratios between two successive terms, rather than differences. Thus the counterpart of formula (2) is

$$U_{n+1/2} = U_n \frac{(R_1U_n)^{1/2}(R_3U_n)^{1/16}\ldots}{(R_2U_n)^{1/8}(R_4U_n)^{5/128}\ldots}. \qquad (3)$$

Just as higher differences are deemed to be zero, higher "ratios," or R's, are considered to be unity, and hence do not affect the result. Thus the

TABLE 3

"GEOMETRIC" VALUES FOR $a_{\overline{100}|}$ AT 2.75 PER CENT

APPROXIMATION (1)	n—SUBSCRIPTS OF VALUES USED (2)	APPROXIMATE VALUE: Pure (3)	APPROXIMATE VALUE: Blended (4)	CORRECTION FACTOR TO PRECEDING FIGURE (COL. 3): Pure (5)	CORRECTION FACTOR TO PRECEDING FIGURE (COL. 3): Blended (6)	SYMBOL FOR COL. 5 (7)
1st....	2–3	34.01420	34.10651	....	....	....
2d.....	2–4	33.95468	33.95430	$\div$1.001753	$\div$1.001764	$(R_2U_2)^{1/8}$
3d.....	1–4	33.94815	33.95101	$\times$ .9998078	$\times$ .9998920	$(R_3U_1)^{1/16}$

first approximation, based on two given values, uses only R_1U_n. Each successive approximation refines the previous one by introducing an R-term of higher order.

Substituting values from Table 1 in formula (3), using two, three, and four given values, respectively, we obtain results parallel to those in Table 2. These are shown in Table 3, in the columns labeled "Pure."

In this terminology, Mr. Baughman's formula (4) would appear as follows:

$$U_{n+3} = U_n(R_1U_n)^3\,(R_2U_n)^3\,. \qquad (4)$$

In fact, substitution therein from Table 1, with $n = 2$, will give his extrapolated value of 24.54558.

Fractional Powers of R-*Terms*

An obvious reason why "arithmetic" formulae have commonly been preferred is the labor of computing fractional powers of the R-terms, as shown in formula (3). This may be reduced by the use of logarithms, generally with some loss of accuracy. It is not usually practical, however, to employ logarithms on an electronic computer.

A method that is perfectly compatible with either a desk computer or an electronic one is Newton's iterative approximation formula:

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}. \tag{5}$$

This will be illustrated in obtaining the value of $(R_2U_2)^{1/8}$ shown in Table 3.

The basic equation is

$$F(x) = x^8 - R_2U_2 = 0 . \tag{6}$$

Differentiating this, and substituting in formula (5), gives a working formula:

$$x_{n+1} = .875x_n + .125\ R_2U_2/x_n^7 . \tag{7}$$

In worksheet form, the application of formula (7) appears in Table 4.

TABLE 4

USE OF NEWTON'S FORMULA TO OBTAIN A FRACTIONAL POWER

Trial	$(R_2U_2)^{1/8}=x_n$ (1)	x_n^7 (2)	$R_2U_2\div(2)$ 1.014109÷(2) (3)	x_{n+1} .875(1)+.125(3) (4)	Check: (1) (2) (5)
1......	1.000000	1.000000	1.014109	1.001764	....
2......	1.001764	1.012414	1.001674	1.001753	....
3......	1.001753	1.012336	1.001751	1.001753	....
4......	1.001753	1.012336	....	....	1.014111

"Blended" Methods

If for the last R-value used under the "geometric" approach we employ a simple approximation to the fractional power, the work is substantially reduced. Also, as will be seen, the accuracy is often actually improved. This approximation is

$$(R_kU_n)^{p/q} = \frac{1}{q}[(q-p) + p(R_kU_n)]. \tag{8}$$

An example is the second trial value in Table 4, where $p = 1$ and $q = 8$.

The results of using formula (8), in each case for a single R-term only, are shown in Table 3, in the columns labeled "Blended." To illustrate the saving in labor, the last figure in Column 5 required twelve trial values. Also, since the correct value of $a_{\overline{100|}}$ at 2.75 per cent is 33.95104, the "blended" values are preferable to the "pure" ones in all but the first instance (where, of course, the former coincides with the "arithmetic" result). This is more readily seen in Table 5, which compares the errors of approximation in the values in Tables 2 and 3.

POSSIBLE PREDICTION THEORY

A glance at Table 1 suggests the possibility of some sort of practical criterion to determine whether an arithmetic or a geometric method will give better results, assuming that the underlying function is not formulated. If a certain order of differences—as the fourth, say—were all zero, then any standard finite difference formula extending to third differences should give an exact result. Similarly, if all the R-terms in a certain column were unity, then a method using all R-terms of lower order should

TABLE 5

COMPARISON OF ERRORS OF APPROXIMATION TO $a_{\overline{100}|}$ AT 2.75 PER CENT

Approximation (1)	"Arithmetic" Method (2)	"Geometric" Method: Pure (3)	"Geometric" Method: Blended (4)
1st........	.15547	.06316	.15547
2d........	.02150	.00364	.00326
3d........	−.00333	−.00289	−.00003

give a precisely accurate answer. Obviously, in the first instance, an arithmetic method would be preferable, and a geometric one in the second.

In between these two extremes, the choice is less clear. A direct comparison between differences and ratios (less unity) of the same order would not be appropriate, since multiplication of all the basic values by a constant would affect the former, but not the latter. It would be more reasonable first to divide each difference by something like: (*a*) the basic value on the same line; (*b*) the average of the basic values involved in the difference; (*c*) the next lower difference; or (*d*) the mean of the two differences defining such difference. Then such quotients might be compared with the departures from unity of the corresponding ratios.

On such a basis—admittedly somewhat intuitive—the geometric approach would be favored for the figures in Table 1. Time does not permit more investigation into the theory or more extensive empirical testing.

One objection to such an attempt is that a constant addition to the basic values—pictorially, a translation—would not affect the differences but would alter all the ratios. This objection applies equally to all of

Mr. Baughman's geometric results as well; that is, they are not independent of the position of the given values on the graph. No ready remedy suggests itself.

(AUTHOR'S REVIEW OF DISCUSSION)

CHARLES B. BAUGHMAN:

That this paper has only scratched the surface is further evidenced by the discussion.

It was interesting to see Mr. Tookey's results from extrapolating to derive estimates of minimum cash values. His experience suggests that empirical testing is more useful than theoretical analysis in practical applications.

Mr. Link has demonstrated Equitable's success in using a formula of the type

$$u_x \doteqdot u_i f(i, j, n) + u_j g(i, j, n) .$$

This suggests formulae of the type

$$\log u_x \doteqdot F(i, j, n) \log u_i + G(i, j, n) \log u_j .$$

Mr. Good makes a number of helpful points which are so clearly stated as to require no comment.

The author was particularly interested in Mr. Good's polynomial interpolation on the reciprocal of the function, the process of harmonic interpolation. If $T_{n+1}(x)$ and W_x are defined so as to be consistent with the definitions of $R_{n+1}(x)$, $S_{n+1}(x)$, U_x, and V_x, we then have a formula analogous with formulae (5) and (6) in the paper,

$$T_{n+1}(x) = \frac{1}{u_x} - W_x .$$

It is then relatively easy to derive theorems involving $T_{n+1}(x)$ and W_x which are analogous to those in the paper. For example, a theorem analogous to Theorem VII is:

If

$$\left|\frac{u_x^2 T_{n+1}(x)}{1 - u_x T_{n+1}(x)}\right| < |R_{n+1}(x)| ,$$

then

$$\left|u_x - \frac{1}{W_x}\right| < |u_x - U_x| .$$

Also, if

$$\left|\frac{u_x T_{n+1}(x)}{1 - u_x T_{n+1}(x)}\right| < |1 - e^{-S_{n+1}(x)}| ,$$

then

$$\left|u_x - \frac{1}{W_x}\right| < |u_x - e^{V_x}| .$$

It can also be noted from Mr. Good's discussion that, if second differences are ignored and if $u(x)$ does not vary too much in the interval concerned, geometric is superior to harmonic if $uu'' - u'^2 < 0$ and harmonic is superior to geometric if $0 < uu'' - 2u'^2$.

It is interesting to compare values obtained by harmonic extrapolation with those in Table 1 in the paper. These values are contained in the following table, and it will be noted that harmonic extrapolation provides the best approximation for the present value of an annuity certain.

ILLUSTRATIONS OF VALUES
EXTRAPOLATED HARMONICALLY
(*Extrapolated from 3 Per Cent and 3½ Per Cent Values*)

$(1+n)^3$	1.12518
$(1+i)^{33}$	3.76422
$(1+i)^{100}$	82.73352
v^3	0.88912
v^{33}	0.27999
v^{100}	0.02317
s_3	3.12170
s_{33}	66.71559
s_{100}	1488.09522
a_3	2.77517
a_{33}	18.18546
a_{100}	24.58690
$\ddot{a}_{10}$	23.0309
$\ddot{a}_{55}$	13.2031
$\ddot{a}_{90}$	3.3563
A_{10}	0.1212
A_{55}	0.4943
A_{90}	0.8709
P_{10}	0.00515
P_{55}	0.03734
P_{90}	0.25944

Mr. Rosser's discussion is a helpful addition to the paper, inasmuch as he has given us a number of practical suggestions and has cast his development in a more familiar form. It is hoped he will develop his prediction theory in a future paper.

The author greatly appreciates the valuable contributions of the discussants and the stimulation to further exploration which they offer.