

MATHEMATICAL STATISTICS IN THE EARLY STATES¹

BY STEPHEN M. STIGLER

University of Wisconsin, Madison

The history of mathematical statistics in the United States prior to 1885 is reviewed, with emphasis upon the works of Robert Adrain, Benjamin and Charles Peirce, Simon Newcomb, and Erastus De Forest. While the period before 1850 produced little of substance, the years from 1850 to 1885 saw such innovations as an outlier rejection procedure, randomized design of experiments, elicitation of personal probabilities, kernel estimation of density functions, an anticipation of sufficiency, a runs test for fit, a Monte Carlo study, optimal linear smoothing, and the fitting of gamma distributions by the method of moments. Reasons for the rapid acceleration in the growth of the field are explored.

1. Introduction. In 1799 Thomas Jefferson received a letter from a young man asking which branches of mathematics it would be most useful for him to study. Jefferson's reply praised Euclid and Archimedes as useful sources, and stated that trigonometry "... is most valuable to every man. There is scarcely a day in which he will not resort to it for some of the purposes of common life; the science of calculation also is indispensable as far as the extraction of the square and cube roots, Algebra as far as the quadratic equation and the use of logarithms are often of value in ordinary cases: but all beyond these is but a luxury; a delicious luxury indeed; but not to be indulged in by one who is to have a profession to follow for his subsistence" (Smith and Ginsburg, 1934, page 62). Jefferson listed "Algebraical operations beyond the 2nd dimension" and calculus as among these luxuries, and doubtless would have added probability if he had been asked.

What follows is a report on an investigation into the question: Did Jefferson's contemporaries and their 19th century descendants follow his advice? How, by whom, and why was the study of the "delicious luxury" of probability and mathematical statistics pursued in the United States, in the first century of its existence? There were good reasons for expecting that little, if any, American work would be found. American science generally did not reach advanced stages of development before the latter half of the nineteenth century, as American capital and genius had found other pursuits more rewarding. (See de Tocqueville,

Received November 1976; revised June 1977.

¹ This research was supported by the National Science Foundation under Grant No. SOC 75-02922, and was presented as a Special Invited Paper at the Annual Meeting of the Institute of Mathematical Statistics, New Haven, Conn., on August 18, 1976.

AMS 1970 subject classifications. Primary 62-03; Secondary 01A55.

Key words and phrases. History of statistics, randomization, density estimation, sufficiency, runs test, Monte Carlo, gamma distribution, smoothing.

Reprinted from *The Annals of Statistics*, Vol. 6, No. 2, pp. 239-265, with permission of the Institute of Mathematical Statistics.

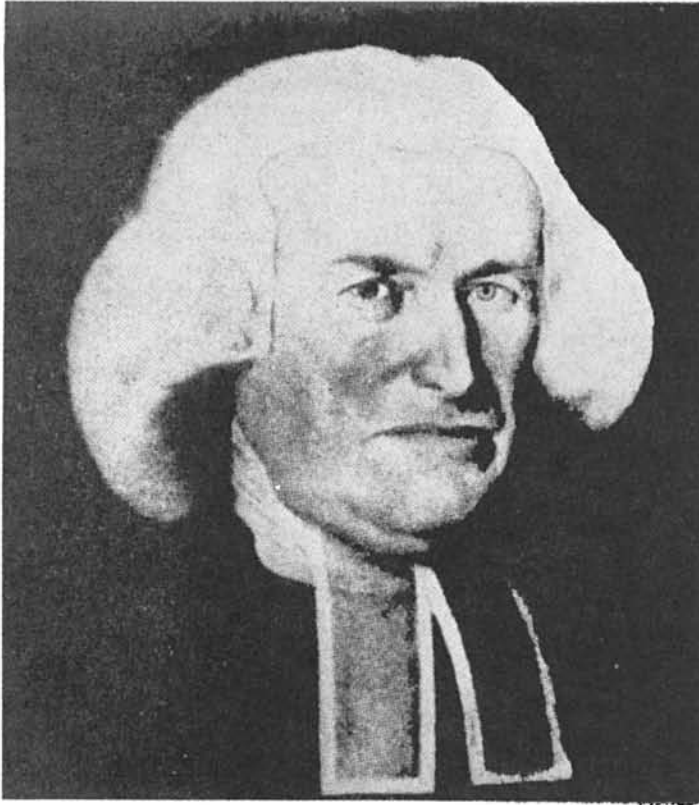
1840, and, for a modern assessment and recent references, Reingold, 1972.) Nonetheless, this investigation was undertaken with a cautious optimism. Although standard histories of statistics have little to say about American work of this period, the years from 1770 to 1850 had been ones of great interest in this subject in Europe: perhaps some Americans would have been inspired to join in its study, and contribute to its development. But, at least in the years before 1800 this was not the case. Upon closer examination it appears that the founding fathers' most significant contribution to mathematical statistics was their decision to leave the Federalist Papers unsigned! (Mosteller and Wallace, 1964.)

Fortunately as the 19th century progressed some signs of American interest in this field appeared. In what follows we shall review the early development of mathematical statistics in America, from its beginnings until 1885. The choice of 1885 as a cutoff date had been made for several reasons. First, by 1885 the field (if we may be so anachronistic as to call nineteenth century mathematical statistics a "field") has achieved a relative maturity, both in Europe and America. Those institutions most responsible for the early development of statistical techniques, the geodetic surveys, the observatories, and the insurance companies, all had reached advanced stages of growth, although the spread of these techniques to other areas of application was still only tentative. Second, 1885 marks the beginning of the major statistical works of Galton and Edgeworth, works that led directly to that of Karl Pearson, R. A. Fisher, and the twentieth century explosion of interest in the field. And third, 1885 marks the culmination of the work of one of the major American figures to contribute to the early states of statistics, Erastus De Forest.

In Section 2 the situation before 1850 will be surveyed, and the few oases in this statistical desert discussed. Section 3 will consider the Peirces, a family closely associated with the emergence of mathematical research in America, and Simon Newcomb. Section 4 will discuss early work at Yale, and the remarkable achievements of Erastus De Forest. Finally, Section 5 will consider the reasons for the late development of mathematical statistics in America, and attempt an assessment of these early efforts.

2. Before 1850. Prior to about 1850, the level of attainment in mathematical statistics in America—indeed the level of attainment in American mathematics generally—was quite primitive. In fact, the most important American mathematical publication to appear before 1850 was Nathaniel Bowditch's 1829–1834 translation of Laplace's *Mécanique Céleste*, and the only widely circulated American work on probability I have found to appear before 1850 was an anonymous book review! (Anonymous, 1832.)

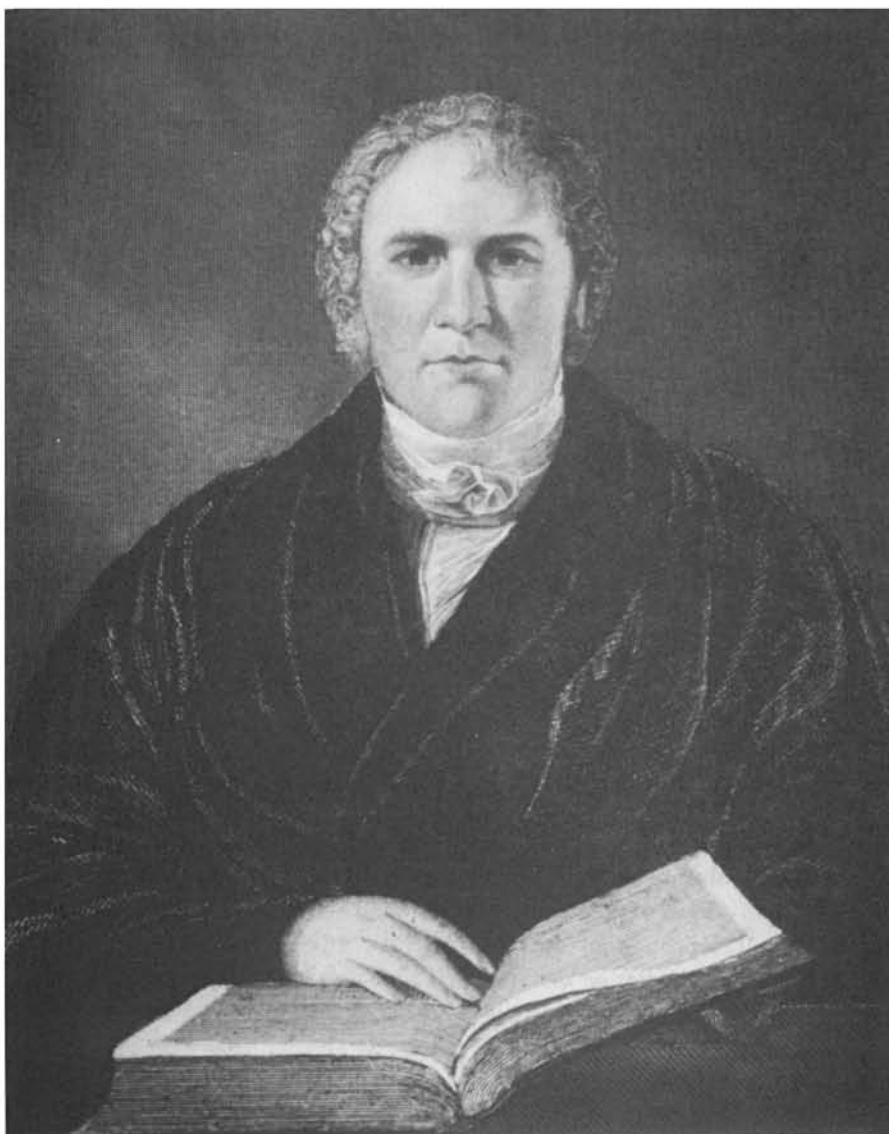
At first this dismal assessment of early American work may seem incredible. After all, Harvard could boast of a professorship of mathematics as early as 1727, the year of Newton's death. But upon closer inspection it appears that this early professor was no fit companion of Newton: his only publication was



Edward Wigglesworth (1732-1794)

an arithmetic, and, we are told, he was removed from his chair in 1738 as “guilty of many acts of gross intemperance, to the dishonor of God and the great hurt and reproach of society” (Cajori, 1890, page 24). It is true that Thomas Jefferson’s *Notes on the State of Virginia* (1785) and Benjamin Franklin’s *Observations Concerning the Increase of Mankind* (1755) made important contributions to non-mathematical statistics, but the closest I have found to a contribution to our subject is the publication in 1789 of the first American life table, by a Harvard professor of divinity, Edward Wigglesworth (1732-1794) (O’Donnell, 1936, page 371). His publication contained no mathematics, and his dour expression is adequate commentary on the fortunes of statistics in America before 1800.

Nor did the beginning of the 19th century bring early relief from this drought of mathematical and statistical research. We may recall that the first quarter of the 19th century had produced exciting work in Europe: Gauss, Laplace, and a host of lesser workers had written books on probability and statistics. The greatest of these was Laplace’s *Théorie analytique des probabilités*, first published in 1812. What, we might ask, was happening in American mathematics in 1812? You may complain that preoccupation with the war of 1812 may make such a comparison unfair, but recall that 1812 also marked the climax of the Napoleonic



Robert Adrain (1775-1843)

wars in Europe. Well, in 1812 twenty-five books on mathematics were published in the United States (Karpinski, 1940). Twenty-two these (including all those written by Americans) were books of tables or elementary texts on arithmetic or geometry. Two of the 25 were reprintings of English works on surveying, and one was a reprinting of an English work on fluxions (incidentally, this was the *first* text on calculus to be printed in the new world).

This situation persisted until about 1850. There was, however, one minor but interesting exception to this assessment, one brief but early spark, that hinted

at the latent Yankee ingenuity that would erupt in the latter half of the century. That spark was the Irish-American Robert Adrain (1775–1843). For in 1809, Adrain published two derivations of the normal probability distribution, derivations that were published independently of and nearly simultaneously with Gauss's *Theoria Motus*.

Adrain had been born in Ireland in 1775, been well-trained in mathematics there, and emigrated to America after being badly wounded while an officer with the insurgent forces in the Irish uprising of 1798. The move was a wise one in many respects; he not only escaped the English gallows, he also switched from being an Irish mathematician, of which there were many, to being an American mathematician, of which there were few. He had taught mathematics in Ireland and continued to do so in America: by his death in 1843 he had taught at several Academies, Rutgers, Columbia, and the University of Pennsylvania (Coolidge, 1926; Babb, 1926).

His sole contribution to our field (and his sole original contribution to mathematics) appeared in a mathematical magazine he started in 1808. The magazine, called the *Analyst*, was one of several dedicated to problems and recreations that appeared in the first half of the 19th century. Adrain's paper, "Research concerning the probabilities of the errors which happen in making observations," was presented as a solution to a problem in surveying that had been posed as a Prize Question in the second number of the magazine. In the fourth number Adrain began his solution by presenting two derivations of the normal distribution, both of which were more wishful thinking than proofs. Both have been analyzed in modern notation by Coolidge (1926); the first was reprinted in the original notation by Abbe (1871) and the second by Merriman (1877, page 140). Essentially the first began by supposing probabilities of errors of observation would be proportional to the quantity measured, and by an obscure argument arrived at a differential equation of which the "simplest" solution was the normal density. The second derivation argued that a bivariate error distribution should be symmetrically distributed with respect to either axis, and the chance of an error should decrease in all directions from the origin, and must have continuous contours. The contour curve "must be the simplest possible having all the preceding conditions, and must consequently be the circumference of a circle" (Adrain, 1809, page 97); with independence of coordinates this leads to the bivariate normal distribution. Having derived the normal distribution of errors, Adrain went on to deduce the "most probable" solutions to several estimation problems by maximizing the likelihood function; that is, he found the least squares solutions. The problems he considered were those of determining the most probable position of a point in space (i.e., estimating a mean), correcting the dead reckoning at sea (i.e., reconciling an observed latitude with the recorded times and directions sailed), and correcting a survey (i.e., reconciling a system of inconsistent survey measurements).

Adrain's work remained nearly totally obscure until it was rediscovered by

Abbe in 1871 (Abbe, 1871). Only two references by other writers seem to exist. His rule for correcting dead reckoning was incorporated, with a reference to Adrain, in the third edition of Bowditch's *New American Practical Navigator* (1811, page 208), and his method of correcting a survey was cited by Gummere in his treatise on that subject (1817, page 116), but neither author mentions the connection with probability or with the general method of least squares.

Adrain, in his original paper (1809) and in two later applications (1818), does not mention Legendre's earlier work on least squares, and many writers have concluded that his discovery of the method was independent of Legendre's. However, Babb (1926) tells us that Adrain had an original copy of Legendre's 1805 work in his library, and Coolidge (1926) documents an instance where Adrain borrowed from a contemporary without citation.

Also, Adrain's formulae deriving the method of least squares are quite similar to Legendre's, and his groping toward the normal distribution would be more easily explained if Adrain had Legendre's work in hand and was working toward the method of least squares. On the other hand, there is no reason to doubt that his derivation of the normal distribution was done in ignorance of Gauss's work. The publication of the number of Adrain's magazine containing his paper was evidently delayed; notwithstanding the nominal year of publication 1808, internal evidence (such as a May 1809 date in a problem on page 110) suggests a spring 1809 publication as more likely. But Adrain's *manuscript* was dated 1808 (Abbe, 1871), and Gauss's book *Theoria Motus* did not reach even Paris until May 1809 (Plackett 1972, page 243), the preface being dated March 28, 1809. These facts and the total dissimilarity of their derivations of the normal density make it clear that Adrain must be counted an independent discoverer of this density, although his work had no apparent impact on the development of statistics.

3. The Peirces and Simon Newcomb. The emergence of American mathematical research was closely linked to Benjamin Peirce (1809–1880). Peirce was born in 1809 and graduated from Harvard in 1828, a classmate of Oliver Wendell Holmes. His most influential teacher was the self-taught sea captain and translator of Laplace, Nathaniel Bowditch (1773–1838). Peirce revised and proof-read Bowditch's translation, and wrote several elementary books before being appointed to Harvard's Perkins professorship of Astronomy and Mathematics in 1842, a chair he held until his death in 1880. Incidentally, his first published work was the solution to a problem in one of Adrain's mathematical magazines in 1825 (Archibald, 1925).

Benjamin Peirce's main contributions to our field were three. The first is of a general nature, as a teacher and researcher in mathematics. He is generally regarded as the first American professor of mathematics for whom research was more than a hobby. With Peirce the character of mathematics in American universities changed from that of a service department to that of a major field



Benjamin Peirce (1809–1880)

of research. Although only his *Linear Associative Algebra* (1870) is considered today as a genuinely important piece of research, his texts on all areas of mathematics published from 1835 on and his many papers on mathematical astronomy marked a change in the level of work in American mathematics. Peirce is also reported to have been an inspiring and stimulating teacher, although descriptions of his technique make one wonder. One of his students was Charles Eliot, later President of Harvard, who wrote: "His method was that of the lecture or monologue, his students never being invited to become active themselves in the lecture room. He would stand on a platform raised two steps above the floor of the room, and chalk in hand cover the slates which filled the whole side of the room with figures, as he slowly passed along the platform; but his scanty talk was hardly addressed to the students who sat below trying to take notes. . . . No question ever went out to the class, the majority of whom apprehended imperfectly what Professor Peirce was saying" (Eliot, 1925). Another student, later a famous mathematician himself, wrote: "Although we could rarely follow him, we certainly sat up and took notice. I can see him now at the blackboard,

chalk in one hand and rubber in the other, writing rapidly and erasing recklessly, pausing every few minutes to face the class and comment earnestly, perhaps on the results of an elaborate calculation, perhaps on the greatness of the Creator” (Byerly, 1925).

Peirce’s second contribution to statistics was specific. In 1852 he published the first significance test designed to tell an investigator whether an outlier should be rejected (Peirce, 1852, 1878). The test, based on a likelihood ratio type of argument, had the distinction of producing an international debate on the wisdom of such actions (Anscombe, 1960, Rider, 1933, Stigler, 1973a). While the debate was never satisfactorily resolved, Peirce was the victor in one sense: From 1852 to 1867 he served as director of the longitude determinations of the U. S. Coast Survey, and from 1867 to 1874 as superintendent of the Survey. During these years his test was consistently employed by all the clerks of this, the most active and mathematically inclined statistical organization of the era. Few statisticians have such an opportunity to put their test into routine use!

Benjamin Peirce’s third major contribution to our field was his son, Charles Sanders Peirce (1839–1914). It is to his son and other workers of the period after 1860 that I shall shortly turn.

Now, during the first half of the 19th century America was largely preoccupied with territorial expansion and was primarily an agricultural nation. The English divine and wit Sidney Smith spoke more in truth than jest when he wrote early in the century: “Why should the Americans write books, when a six weeks’ passage brings them, in their own tongue, our sense, science and genius, in bales and hogsheads? Prairies, steamboats, grist-mills, are their natural objects for centuries to come. Then, when they have got to the Pacific Ocean—epic poems, plays, pleasures of memory, and all the elegant gratifications of an ancient people who have tamed the wild earth, and set down to amuse themselves” (Smith, 1818). Sidney Smith’s forecast was accurate, but his timing was off. With respect to science and mathematics, the period after mid-century saw rapid advancement. For mathematical statisticians, the three most important evidences of this growth were the rapid expansion of the U. S. Coast (later Coast and Geodetic) Survey, which was charged with measuring and mapping the new land; the founding and staffing of new observatories; and the growth of higher education, with its increasing emphasis on research. Charles Sanders Peirce was a product of this changing, intellectually charged atmosphere.

C. S. Peirce, son of Benjamin Peirce, was born in 1839 and like many young men of the era, went into the family business after he finished his schooling. Only in his case, “schooling” meant Harvard University and the “family business” was the newly expanded U. S. Coast Survey. Charles Peirce is best known today as a philosopher and logician—in fact there is today a C. S. Peirce Society which publishes a journal largely devoted to his thought. But for nearly thirty years he was an assistant at the Coast Survey, and a major portion of his life’s work was tied to physical science and mathematics (Weiss, 1934; Eisele, 1974).



C. S. Peirce (1839-1914)

While his time with the Coast Survey included the years his father was superintendent, no charge of nepotism seems to have been leveled; in fact the son is generally conceded to have been the intellectual superior of the two.

Charles Peirce's interests covered an enormous range, and probability and statistics formed an integral part of both his philosophical views and his scientific method. Probability was basic to his view of scientific logic, and in one passage he defined "induction" to be "reasoning from a sample taken at random to the whole lot sampled" (Peirce, 1957, page 217). Indeed, Peirce's work contains one of the earliest explicit endorsements of mathematical randomization as a basis for inference of which I am aware (Peirce, 1957, pages 216-219).

What was perhaps Peirce's most influential statistical work came in the field of psychophysics, or experimental psychology. In 1884 Peirce and a student, Joseph Jastrow, performed an experiment to test the existence of a least perceptible difference in sensations. Gustav Fechner, in an important 1860 book (*Elemente der Psychophysik*) had argued that for each sense there was a nonzero threshold, such that if two sensations differed by less than the threshold they could not be distinguished. Peirce and Jastrow performed a large scale experiment

involving the sensation of pressure, with themselves as subjects, and they effectively refuted the existence of such a threshold. Their methodology is of particular interest.

Peirce and Jastrow's (1885) report would be considered as a good example of a well-planned and well-documented experiment today; as a nineteenth century experiment it was unexcelled. (Incredibly, Peirce later described the precautions he took as "more careful and studied and elaborate than the memoir states" (Eisele, 1957).) They sought to refute the notion of a least perceptible difference by performing what we would now call a quantal response experiment using probit analysis, and showing that the results were inconsistent with Fechner's theory. Two slightly different known weights would be presented sequentially to the subjects, and they would state (or guess) in which of two possible orders they had been presented. In addition, the subject would estimate the confidence he had in his judgment on a scale from 0 to 3. The frequency of correct guesses for differing combinations of weights was then used to fit a probit response curve. Experiments of this general type were familiar to psychophysicists; this was called the method of "right and wrong cases." The authors gave few details on the method of fitting the probit response curve. They took "dosage" to be the ratio of the two weights used, so it could be assumed that a dosage of 1.0 led to a 0.5 probability of a correct guess, and there remained only one parameter, the scale parameter, to be determined. They apparently determined this separately for each dosage level d (by $\hat{\sigma}_d = (d - 1)/\Phi^{-1}(\hat{p}_d)$) and averaged the results. But they were painstaking in their description of the experimental procedures followed, and two aspects of these were strikingly original.

The first novel point was the way in which the estimates of confidence were used. Peirce and Jastrow used them to fit a relationship of the form $m = c \log(p/(1 - p))$ for each subject where m was the estimate of confidence, c an "index of confidence" peculiar to each subject, and p the "true" probability of a correct guess, estimated by the observed relative frequency. Peirce's conception of probability was that of an objective frequentist, but his work here shows he was also one of the first individuals (perhaps the very first) to experimentally elicit subjective or personal probabilities, determining that these probabilities varied approximately linearly with the log odds.

The second departure from tradition was the manner in which the order of presentation of the weights was determined. The Peirce-Jastrow experiment is the first of which I am aware where the experimentation was performed according to a precise, mathematically sound randomization scheme! The assignment was done in blocks of twenty-five (to achieve balance) by using, alternately, two well-shuffled packs of 25 cards, one with 12 red and 13 black cards, and one with 13 red and 12 black cards. They wrote "A slight disadvantage in this mode of proceeding arises from the long runs of one particular kind of change, which would occasionally be produced by chance and would tend to confuse the mind of the subject. But it seems clear that this disadvantage was less than

that which would have been occasioned by his knowing that there would be no such long runs if any means had been taken to prevent them." Jastrow's later development and advocacy of this methodology had an important influence on later psychological research (Jastrow, 1888), although randomization in the design of experiments did not become part of the mainstream of statistical thought until R. A. Fisher's book on the subject appeared, a half-century later.

The breadth of Peirce's interests and the statistical turn of his mind are illustrated by a paper he read to the Philosophical Society of Washington in 1872. He, the abstract says, "called attention to the striking resemblance between the map showing the distribution of illiteracy . . . in the United States, given in the Report of the Census of 1870, and the map showing the distribution of rainfall during the three winter months published [by the Smithsonian]. Mr. Peirce suggested as a possible explanation for the resemblance, that the copious winter rains would produce agricultural plenty, which in its turn would favor indolence" (Peirce, 1872). I expect that farmers of his day would have taken offense at his informal path analysis, if they could have read his paper.

In another work the question whether or not meteorologists could successfully predict tornados led Peirce in 1884 to derive a latent structure measure of association for 2×2 tables (Peirce, 1884, Goodman and Kruskal, 1959). But his work for the Coast Survey is probably of more immediate interest to modern mathematical statisticians. In one 1879 paper on the "Economy of Research" (Peirce, 1879), he provided a rigorous mathematical analysis of the problem of optimally allocating experimental observations between competing experiments, under a model with two components of variance. This paper attacked the allocation problem from the point of view of a quite general utility theory, and contains an early and possibly independent formulation of a basic result in what economists call marginal utility theory.

Probably Peirce's best known statistical investigation was an 1873 paper "On the theory of errors of observations" (Peirce, 1873). At the close of this paper he presented the results of an extensive empirical investigation into the nature of laws of error. He hired an untrained 18 year old boy to react on a telegraph key to signals received. Five hundred measurements a day were recorded for 24 days, and Peirce sought to determine the distribution, that is, the density, of the reaction times for each day. The manner in which he estimated this density is interesting, particularly in view of recent work. He did *not* just present a histogram. Rather, "The curve has, however, not been plotted directly from the observations, but after they have been smoothed off by the addition of adjacent numbers in the table eight times over, so as to diminish the irregularities of the curve. The smoother curve on the figures is a mean curve for every day drawn by eye so as to eliminate the irregularities entirely." What he had done was to employ a form of repeated adjustment that was similar to techniques then in use for the interpolation and smoothing of mortality tables. What his technique did was to replace each ordinate of the histogram by a binomially

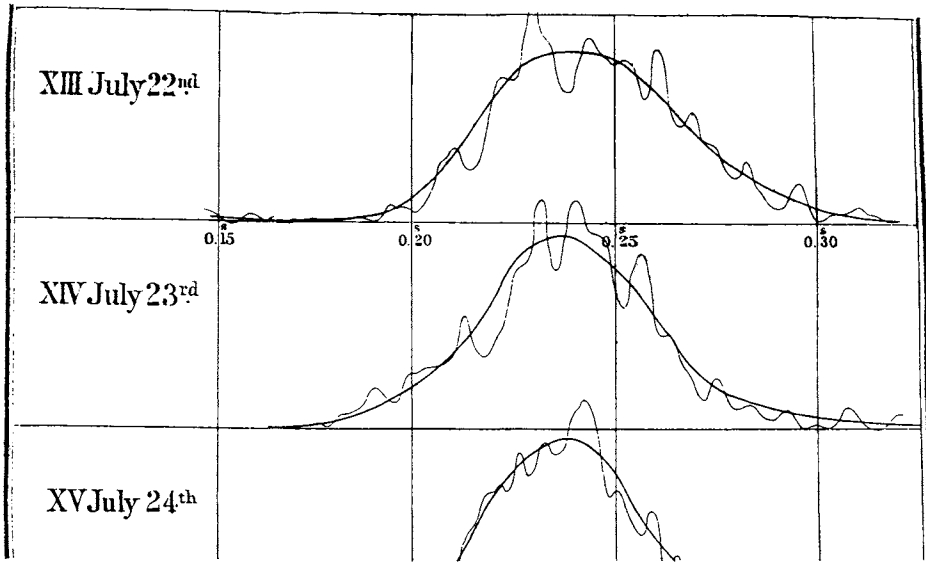


FIG. 1. C. S. Peirce's graph of his estimated probability density, by the kernel method and by a freehand sketch (smoother curves).

weighted average of nine consecutive ordinates. This is what we would now call a "kernel-type" estimate of the density using a binomial kernel that produces essentially the same effect as would a normal density kernel, although this kernel estimate produces a curve slightly out of phase with Peirce's. Fifty-five years later E. B. Wilson and Margaret Hilferty (1929) reanalyzed these data and concluded that Peirce's qualitative conclusion that the distribution differed little from the normal was not supported under closer scrutiny. In particular, data set 14 (Figure 1) was found to have a skewness of $\mu_3/\sigma^3 = 5.74$ and kurtosis of $\beta_2 - 3 = 63.6$. But Peirce had provided experimental evidence that human reaction times exhibited a regularity and distribution at least qualitatively similar to the normal curve's bell shape, and he had contributed substantially to American work on a major line of development in statistical thought, the concept of a distribution (Stigler, 1975).

Two years after Peirce's paper appeared, the British Astronomer Royal G. B. Airy added an appendix to the second edition of his text on the theory of errors (Airy, 1875) presenting an essentially similar example, apparently inspired by Peirce (although no citation was given). Unlike Peirce, however, Airy omitted the raw data and published only the smoothed frequency counts, and in this form Peirce's innovation was later to draw Karl Pearson's scorn in his famous paper on chi-square: "...that Appendix really tells us *absolutely nothing* as to the goodness of fit of his 636 observations ... to a normal curve. [We] find that he has *thrice smoothed* his observation frequency distribution before he allows us to examine it. It is accordingly impossible to say whether it really does or does not represent a random set of deviations from a normal frequency curve" (Pearson, 1900).



Simon Newcomb (1835-1909)

In his 1879 paper Charles Peirce had argued that in scientific research, the expected marginal utility of further investigation decreases as experimentation continues. He based this argument on the fact that the probable errors of estimated quantities are convex functions of sample size, and felt that the principle held more generally. "All the sciences exhibit the same phenomenon, and so does the course of life. At first we learn very easily, and the interest of experience is very great; but it becomes harder and harder, and less and less worth while, until we are glad to sleep in death" (Peirce, 1879). As if to confirm this as prophecy, Peirce became increasingly withdrawn in later life and died in isolation on a farm, in 1914.

Another, purely intellectual, descendent of Benjamin Peirce was the astronomer-statistician-economist Simon Newcomb (1835-1909). Newcomb was a student of Peirce's at Harvard's Lawrence Scientific School, where he graduated in 1858. But that cold statement of fact fails to capture the flavor of Newcomb's youth, which was like that of a plot from a Horatio Alger novel. He was born in Nova Scotia where his father, a teacher, lived a nomadic life. At 16 Simon Newcomb began an apprenticeship to a doctor that was to last 5 years. But his

autobiography (Newcomb, 1903) tells a dramatic tale of how after 2 years his prospects for a career in medicine faded as the doctor proved to be a total fraud, a quack herbalist, and he was forced to flee in the middle of the night from what had become onerous servitude, to seek his way in the world with little more than his wits to support him. At 18 he made a living as a teacher, relying on what he had taught himself in odd hours. At 21 he obtained employment he Cambridge with the Nautical Almanac office. At 23 he had a Harvard degree, and at 24 he read a paper to the American Association for the Advancement of Science on the untenability of the hypothesis that the asteroids had a common origin. He went on to become America's most honored scientist in the 19th century (Campbell, 1924).

Probability and statistical thinking played a major role in Simon Newcomb's lifework. His early work in robust estimation has received attention recently with the resurgence of interest in the subject (Stigler, 1973a), and he was the first probabilist to present the logarithmic distribution of leading digits (Newcomb, 1881). But the large volume of his published work—at least 541 notes, papers, and books (Archibald, 1924), much of this at least tangential to statistics—makes it impossible to do him justice in a short space. Rather, I will only briefly mention one minor, but interesting passage he published in some “Notes on probability” at the age of 25.

This paper appeared in the *Mathematical Monthly*, one of the better of the numerous magazines which fanned the growing interest in mathematics in America before the *Analyst* (1874–1883) approached, and the *American Journal of Mathematics* (cofounded by Newcomb in 1878) finally achieved international respectability. Among other topics, Newcomb considered the problem of estimating the number of serially numbered tickets in a bag, based on the numbers observed on s tickets drawn, with replacement. That is, based on a random sample of size s from a discrete distribution, uniform from 1 to n , estimate n . Before he went on to present a sound and well-explained Bayesian analysis of this problem, he gave the clearest statement of the idea of sufficiency I have encountered before Fisher. (See also Stigler 1973b.) Simon Newcomb wrote, “Let i be the largest number drawn in the s drawings. The number of Tickets, then, cannot be less than i . We need not know any of the drawn numbers except the largest, because after we know this, every combination of smaller numbers will be equally probable on every admissible hypothesis, and will therefore be of no assistance in judging of these hypotheses” (Newcomb, 1860–1861). Of course Newcomb did not abstract the concept, but at the least, this early statement is evidence of a clear mind capable of quickly reaching to the essentials of a problem, a signal of the brilliant career that was to come.

4. Work at Yale: E. L. De Forest. The major figures I have discussed so far—the two Peirces, Simon Newcomb, even Wigglesworth—all attended Harvard and continued their careers at Harvard, in Washington, or both. This

hints at a Boston-Washington scientific axis that did indeed exist. For example, Admiral Charles Henry Davis was instrumental in obtaining government appropriations needed to start the Nautical Almanac in 1849, and he played a key role in locating its headquarters at Harvard with Benjamin Peirce in charge. Admiral Davis was also Peirce's brother-in-law. But it would be a mistake to suppose that the *only* scientific activity in America was localized in these two centers.

Another center of learning where early and important contributions to mathematical statistics were made was Yale University. It was there in the Sheffield Scientific School that the first doctorate in America was awarded for a thesis in mathematical statistics. The thesis was on the method of least squares, and it was awarded to Mansfield Merriman in 1876, the nation's centennial. Merriman went on to become the first American statistician to capture the market for elementary statistics textbooks, with his *A Textbook on the Method of Least Squares* (First Edition, 1884; Eighth Edition, 1907), and his extensive "List of writings relating to the method of least squares" remains the best bibliography of this subject (Merriman, 1877a).

From 1871 until his death in 1903, the major intellectual force in science at Yale was J. Willard Gibbs (Wheeler, 1951). Shortly after Gibbs died, Lord Kelvin visited Yale and forecast that "by the year 2000 Yale will be best known to the world for having produced J. Willard Gibbs" (Fisher, 1930), and while some modern Yale professors might question that prediction, no one of them who is familiar with Gibbs' work could take it as an insult. Gibbs himself published nothing in statistics, although he taught least squares (Wilson, 1931) and his work on statistical mechanics relied heavily on probabilistic concepts. Gibbs, through his development of vector analysis and of statistical mechanics, may indirectly have had a more profound influence on 20th century work in mathematical statistics than any other man I have mentioned, but his more obvious impact was as a teacher. Two of his students, Irving Fisher and E. B. Wilson, served as presidents of the American Statistical Association, and a third, E. L. Dodd, had a significant impact on mathematical statistics, partly as a teacher of Sam Wilks.

But the Yale man I most wish to discuss here was not a student of Gibbs, although he was Gibbs' contemporary and his work shows some Gibbsian influences. I wish to turn to the remarkable work of Erastus Lyman De Forest. De Forest's name is not widely recognized today, but his name was well known to Edgeworth and Karl Pearson, who respected and cited his work. Between 1870 and 1885 De Forest published a series of over 20 papers which cover such topics as a runs test for the residuals from a regression; a Monte Carlo determination of the variance of a statistic; the gamma distribution for one, two, and three dimensions; a measure of skewness; and an analysis of the bivariate normal distribution.

De Forest was born in 1834, the son of a Yale graduate, and he received two degrees himself at Yale, a B.A. in 1854 and Ph. B. in 1856. I suspect he was



De Forest (1834-1888)

viewed by the Yale administration as an ideal student: he did well in his studies (Gibbs is said to have called him "one of the most brilliant and promising" of Yale's students (Wolfenden, 1968)) and he was both independently wealthy and generous. Yale's Erastus De Forest Professorship of Mathematics was endowed by him in 1888 (Anderson, 1896; Wolfenden 1925, 1968).

Shortly after he graduated, De Forest surprised his family (and possibly himself) by vanishing from sight while on a visit to New York, leaving his suitcase behind and no forwarding address. His family feared the worst, and the search concentrated on New York's East River, but two years later he turned up in Australia, teaching in Melbourne. Nowadays we would probably say he had been "finding himself." The reports of his trip are contradictory, but he must have eventually decided that he preferred bulldogs to kangaroos, for he returned to New Haven and after 1865 seldom ventured further than New York.

The direction of most of De Forest's work seems to have been determined by a project he undertook in 1867-1868 for his uncle, the president of Knickerbocker

Life Insurance Company in New York. In the process of determining the company's policy liabilities, De Forest encountered the problem of smoothing mortality or life tables.

Let u_1, u_2, \dots, u_n be a sequence of numbers; the problem is to "adjust" or smooth the sequence in the hope that a better estimate of an underlying functional relation is thus obtained. We have already encountered one example in C. S. Peirce's density estimate; in the primary application that motivated most early work on the subject, the u_i 's would be empirically determined estimates of the probabilities that an individual of age i in the class under study would die in the next year. A plot of u_i vs. i would be an empirically determined "mortality curve" that would give the chance of death in the next year for individuals of all ages. However, if the u_i 's are simple crude death rates or relative frequencies of deaths in a sample population, as may well be the case, the plot of u_i vs. i will show marked irregularities, in contradiction to the smooth relation believed to hold.

Long before De Forest, actuaries had grappled with this problem, employing a variety of parametric models and averaging schemes. De Forest's early work centered on that species of averaging which Sheppard later named "linear compounding." That is, replace each u_i by a symmetric linear function of surrounding values, say

$$v_i = l_0 u_i + l_1(u_{i+1} + u_{i-1}) + \dots + l_m(u_{i+m} + u_{i-m}).$$

Of course a different rule would be needed near the extremes of the series, and an asymmetric rule could be used, too. Many schemes equivalent to ones of this type, such as the one Peirce applied, had been considered before De Forest, but they were (with one exception) ad hoc in nature. De Forest's first innovation was the introduction, in two papers in the *Smithsonian Reports* for 1871 and 1873, of optimality criteria into this problem (De Forest, 1873, 1874).

De Forest supposed that the observed u_i differed from underlying values U_i by small errors "of an accidental nature" which he supposed independent, with equal variances (we will use σ^2 for the variance; De Forest used ε for the "probable error": $\varepsilon = .6745\sigma$). He then assumed that the U_i sequence was "smooth" in the sense that any $2m + 1$ U_i 's differed little from a polynomial of degree j in i ; that is, given $2m + 1$ U_i 's, a polynomial $g(x)$ of degree j could be found such that $U_i = g(i)$, approximately. In his 1873 paper (De Forest, 1874) he carried out much of his investigation for the case $m = 2$, $j = 3$ and so we too shall specialize to this case. Thus he supposed that any 5 consecutive U_i 's could be represented "very nearly" by a cubic in i . By making his assumption of smoothness a local one and relying on local weights, a great deal of flexibility was retained over assuming U_i cubic for *all* i , or assuming a particular parametric model.

One approach De Forest considered was to determine l_0, l_1, l_2 by least squares: if a cubic function of the index or year is fit to $u_{i-2}, u_{i-1}, u_i, u_{i+1}, u_{i+2}$ by least

squares, ignoring all other u_k 's, the ordinate of the fitted cubic at i will be the required $v_i = l_0 u_i + l_1(u_{i-1} + u_{i+1}) + l_2(u_{i-2} + u_{i+2})$ and will give the minimum mean square estimate of U_i under the cubic assumption and the restriction to estimates linear in the local u_{i-2}, \dots, u_{i+2} . An alternative (and equivalent) formulation is, since $\text{Var}(v_i) = (l_0^2 + 2(l_1^2 + l_2^2))\sigma^2$, to minimize $l_0^2 + 2(l_1^2 + l_2^2)$ subject to the condition

$$U_i = l_0 U_i + l_1(U_{i-1} + U_{i+1}) + l_2(U_{i-2} + U_{i+2})$$

for every cubic U_k in k . De Forest solved this problem and found the l 's (which of course do not depend on the u 's), but as he noted (De Forest, 1873, page 335) he was thus far anticipated by 1867 work of the Italian astronomer Schiaparelli, of which he had at first been unaware. But De Forest continued, and broke new ground when he noticed that the minimum mean square error criterion applied to each five u_i 's separately need not produce a very smooth relation globally, notwithstanding the assumption the function was cubic locally. As an alternative to the criterion "minimize $l_0^2 + 2(l_1^2 + l_2^2)$ subject to the constraint $U_i = l_0 U_i + l_1(U_{i-1} + U_{i+1}) + l_2(U_{i-2} + U_{i+2})$ for all cubics U_i " he proposed a criterion based on smoothness: minimize the probable error of the fourth difference of the smoothed series v_i , or equivalently, minimize $E(\Delta^4 v_i)^2$, subject to the same constraint. He solved this problem for several different cases.

As a contribution to nineteenth century work on smoothing or adjustment, De Forest's introduction of this measure of smoothness as an optimality criterion was well ahead of its time, and his work was not generally appreciated until Wolfenden (1925) rediscovered it in the 1920's. By then, others had come upon his main techniques independently. Variants of De Forest's and others' criteria are currently enjoying great popularity in the related field of spline interpolation.

While De Forest's introduction of optimality criteria into interpolation and smoothing problems was a major, if unappreciated, advance at the time, modern statisticians are liable to be more interested in his evaluations of the fit of the smoothed to the observed series. De Forest was acutely aware of the contradictory combination of the desires for a close fit to the observed series and for smoothness, and he devised several goodness-of-fit tests to determine whether or not the series had been over or under-smoothed.

The first tests he discussed (De Forest, 1874, 1876, 1877) were of the nature of large sample significance tests based on the magnitude of the residuals. In the first place, if independent (possibly theoretical) estimates of the variances of the errors were available, then these could be compared with the residuals. For example, if the u_i were relative frequencies based on given numbers of trials, then a binomial model using the fitted values to estimate the probabilities would provide estimates of variances to compare with the corresponding residuals. To actually perform this test he dropped the "equal variances" assumption and took, for each year, the ratio of the squared residual over the estimate of variance for that year, and averaged these ratios over all years. He then compared the

difference between this average and 1.0, with the calculated sample variance of the ratios, a sort of large sample two-tailed t -test. In suggesting that a difference of $\frac{3}{2}$ or 2 times the estimated probable error be considered large (De Forest, 1876, page 12), he seems to have had a significance level of 0.31 or 0.18 in mind. De Forest noted that this test was similar to one proposed in 1871 by Thiele, although he criticized Thiele's choice of $n-m$ as a divisor in calculating the average ratio.

If no separate or theoretically based estimate of variance was available, De Forest suggested that the residuals be compared with the fourth differences of the original series. In a privately printed pamphlet, he proposed as a statistic, the average (over the series) of $\log(r_i/d_i)$ (our notation), where r_i and d_i are the absolute values of the i th residual $u_i - v_i$ and a constant multiple of the corresponding fourth difference of the u_i 's. The multiple was chosen so that $E(d_i^2) = \sigma^2$. The basic idea was that if the U_i 's were locally cubic, fourth differencing would eliminate their effect leaving only variation due to random errors; then by choosing the constant multiplier so that $E(d_i^2) = \sigma^2$, an estimate of σ^2 not based on the residuals could be obtained. Similar procedures were later rediscovered in the ballistics literature; see Von Neumann et al. (1941).

This was an interesting and novel idea, although its execution was flawed by his neglecting the autocorrelation of the d_i 's, among other types of correlation. But the manner in which he sought to determine the asymptotic variance of this second statistic may be of more interest than the statistic itself. He began by making a quick determination of the standard deviation of $\log(r/d)$, using a first order differential approximation, based on the supposition that r and d are independent absolute values of normal random variables. Today we might recognize this as half the logarithm of a random variable with an F -distribution with 1 and 1 degrees of freedom, but De Forest's work shows little feeling for exact sampling distributions.

De Forest did not, however, have full confidence in his derivation. He wrote: "The demonstration [of the formula for the asymptotic standard deviation is] not a strictly rigorous one, and it has been thought desirable to test the accuracy of the formula by trials made on a sufficiently large scale, in the following manner" (De Forest, 1876, page 23). De Forest's "following manner" was a Monte Carlo study! From a table of the normal distribution, he found 100 percentiles for the absolute value of a normal deviate ranging from the 0.005th to the 0.995th, in steps of 0.01. These numbers he "inscribed upon 100 bits of card-board of equal size, which were shaken up in a box and all drawn out one by one, and entered in a column in the order in which they came" (De Forest, 1876, page 23). He then repeated this procedure to get four columns in all, and, considering them in pairs, took ratios, then logs. These he squared and averaged. He found close agreement with his formula, which he then adopted as "trustworthy."

Of course, we can now suggest more efficient methods of proceeding, and the correlation structures of both De Forest's Monte Carlo experiment and his

analysis were not the same as that of his intended application. Nonetheless, his appeal to a Monte Carlo experiment for verification of his analysis was a remarkable innovation in the study of sampling distributions.

Another of De Forest's innovations was first mentioned in this same privately printed pamphlet, and more fully developed in several papers in the *Analyst* (De Forest 1876, 1877, 1878a, 1878b), a journal with international circulation and impact printed in Des Moines, Iowa. This is the idea of testing fit by analyzing the grouping of signs of the residuals. Step by step he was led to a runs test.

Unknown to De Forest, Quetelet had employed one type of runs test in 1852, with a different aim (Stigler, 1975). Quetelet had examined the distribution of lengths of runs of days of rainfall to test independence against the alternative of Markov dependence; De Forest sought to examine the signs of successive residuals to determine whether or not the small number of runs would provide evidence of too much smoothing. Actually, he began by considering the distribution of the number of runs of each of several given lengths, and comparing the numbers of runs actually observed with the numbers expected under the hypothesis that the fitted curve was the actual curve (plus or minus a probable error) (De Forest, 1876, pages 29 ff). In a later paper (De Forest, 1878b), though, he approached the modern version of the test when he proposed counting the number of "permanences" (non-changes) of signs, which equals the number of terms in the series less the number of runs.

In suggesting a test of fit based on the number of permanences of signs in the residuals, De Forest provided no exact distribution theory; he did not attempt a combinatorial theory of runs. Rather he provided approximations to the mean and probable error of his statistic, based on an unproved assumption that asymptotically the number of permanences among the residual's sign behaved as would a like statistic based on tosses of a fair coin. A little reflection shows that with the mode of fitting he employed this is *not* the case. Since the fitting is accomplished by local averaging, the signs of consecutive residuals will show a strong negative association rather than be approximately independent. A test based on the latter assumption would be severely biased, as the number of runs will tend to greatly exceed what would be expected under the null hypothesis, even with an adequate fit.

This problem did not escape De Forest's notice, and he provided an approximate rule to deal with this dependence. He reasoned that if n_1 terms were averaged in a simple arithmetic average, then one would expect the signs of the n_1 residuals, on average, to be evenly divided (De Forest, 1878a). Thus given one residual is positive, the probability the succeeding one is positive is $(\frac{1}{2}n_1 - 1)/(n_1 - 1) = (1 - 2r^2)/(2 - 2r^2) = q$, where $r = n_1^{-1/2}$, the ratio of the probable error of the mean to that of a single term. Then q is the chance two successive residuals form a permanence. As he did not use simple means but weighted means, he took r to be the corresponding ratio, $(l_0^2 + 2(l_1^2 + l_2^2))^{1/2}$ in our example, so n_1 becomes a sort of effective sample size. De Forest then, by appealing

to a binomial model with probability of success q , provided an approximate means of correcting the distribution of the number of permanences for this dependence (De Forest, 1878b). He felt this correction should be adequate when $r < \frac{1}{2}$. De Forest's model for the behavior of successive signs is equivalent to a two state Markov chain with transition matrix

$$\begin{bmatrix} q & 1 - q \\ 1 - q & q \end{bmatrix}.$$

In a later series of papers in the *Analyst* De Forest was led by a series of steps to the consideration of some families of probability densities. He began by considering "repeated adjustments"; that is, iterated smoothing of a series by the same linear smoothing scheme (De Forest, 1878c, and following papers). An iterated linear smoothing scheme is itself a linear smoothing scheme whose coefficients are derivable from the original coefficients by convolution; Peirce's density estimate is one example. De Forest employed generating functions and differential equation approximations to difference equations to determine the character of the limiting curve of coefficients. For recent contributions to this problem see Greville (1966, 1974).

One limiting curve De Forest was led to was of course the normal (De Forest, 1879), but when he considered the limiting properties of unsymmetric adjustment schemes, he was led to something new. By considering differential equation approximation to the coefficients of binomial distribution, he derived the gamma distribution, which he called the "gamma curve" (De Forest, 1882-1883, page 140). The gamma distribution had appeared as a sampling distribution earlier than 1882 (see Lancaster, 1966), but apparently not outside of that sampling context. Pearson's and Edgeworth's work on asymmetric curves was yet to come, and the English school appears not to have noticed De Forest's work before about 1895, when Edgeworth called Pearson's attention to it. Pearson's own derivation of the gamma (or Type III) curve had then appeared (Pearson, 1895a), but he graciously acknowledged De Forest's priority as respects this type (saying De Forest's deduction had "the advantage of greater generality" and praising "the excellency of his work," Pearson, 1895b). Actually, De Forest's anticipation of Pearson went beyond the simple gamma curve. De Forest also showed how this density could be fit by the method of moments, and explored the third moment as a measure of skewness (he called it "cubic mean inequality") that was useful for distinguishing between the gamma and its limiting form, the normal. This work of De Forest has been commented on by Edgeworth (1896, 1902), Pearson (1895), Hatai (1910), McEwen (1921), Walker (1929) and Wolfenden (1925, 1942).

Another area in which De Forest worked was that of multivariate densities. Starting with the problem of smoothing two and three dimensional arrays, he derived differential equations for the limiting curve of coefficients after repeated adjustments, and was led to consider multidimensional normal (De Forest,

1881a, 1881b, 1882) and gamma distributions (De Forest, 1884). In the normal case he did not restrict attention to the independent case as he did in the gamma, although he noted that a simple rotation of axes was sufficient to reduce the general case to the independent (De Forest, 1881a). In this he was preceded by Bravais, to whom he referred (Walker, 1929, page 96). He added little new to the study of the bivariate normal, although in his final paper in 1885, after fitting a general bivariate normal distribution to target data, his thoughtful check for marginal asymmetry with respect to the transformed axes was a refreshing change from European work of that period.

In 1885 De Forest's health began to fail, and he ceased mathematical work. De Forest died in 1888; his work spanned two decades and was wholly on topics in mathematical statistics. It was widely circulated and extensively abstracted in the German *Jahrbuch über die Fortschritte der Mathematik* (Garver, 1932), but its impact was diminished by his failure to develop his methods much beyond the limited class of problems in adjustment which had suggested them in the beginning.

5. Additional work and conclusions. I have surveyed a major portion of American work in mathematical statistics before 1885, but the survey has by no means been complete. I have omitted discussion of early work on the errors-in-the-variables problem by a Monmouth, Illinois attorney (Adcock, 1877–1878) and by an assistant with the U.S. Lake Survey (Kummell, 1879), published in 1877–1879. I have skipped an early use of the range as a short-cut estimate of a standard deviation (Wright, 1882), and countless computational algorithms designed to simplify the calculation of least squares estimates, including the aptly named Doolittle method (Doolittle, 1881). I have included no discussion of work on the design of experiments, including an 1885 note which suggested that an experiment designed so that the factors would be orthogonal would “secure the maximum precision with the minimum of computation,” after which the discussants allowed that they had known that all along (Woodward, 1885).

Also hidden from view are the gaffes, blunders, and absurdities that have sometimes crept into our forefathers' work. I have spared you their promiscuous use of dx and ∞ as positive real numbers, and their petty disputes over ill-posed problems in probability. But lest the picture seem totally one-sided, it may be worth noting as evidence that American understanding of concepts did have limits, that just a year after getting his degree, Mansfield Merriman, Ph. D. Yale 1876, wrote of Gauss's elegant demonstration of the “Gauss–Markov” theorem that “The proof is entirely untenable” (Merriman, 1877b). In charity to Merriman it might be added that no less a mathematician than Poincaré also misconstrued the nature of Gauss's result (Poincaré, 1912, page 188).

Despite these omissions, it should be clear that by the latter part of the last century, the United States had produced a quantity and variety of work in statistics that, while not the equal of European efforts, at least permits a

respectable comparison. It had taken Americans quite a while to show an interest in mathematical statistics. I think the major reason for this was not a lack of talent, but the fact that the United States was quite late in undertaking a systematic and large scale measurement of its land, and equally late in founding observatories and beginning extensive astronomical observation. In Europe the major impetus to the development of mathematical statistics in the eighteenth and early nineteenth centuries had come from astronomy and surveying. The concepts of linear models, least squares and similar methods, had been developed between 1750 and 1820 in Europe, primarily for the reduction and analysis of astronomical observations. These techniques had then received further refinement when applied in the major geodetical surveys, for example in the survey of Britain from 1783 on.

In both spheres of activity the U.S. lagged. When President John Quincy Adams proposed a program for the construction of observatories in 1825, his phrase "lighthouses of the sky" was derisively trumpeted in the press, and funding was denied (Shepherd, 1975, page 285). The Harvard observatory was not operational before 1839, the U.S. Naval Observatory began observation in 1845. While the U.S. Coast Survey was founded in 1807 with a Swiss in charge (he was Ferdinand Hassler, Simon Newcomb's grandfather-in-law), work was begun only in 1816, and the survey did not really get on the ground on a large scale before the middle of the century, over 50 years after the British survey had reached a similar state (Cajori, 1890, pages 286 ff).

When America finally did commit itself to astronomy and land survey, it moved rapidly and energetically, and work in statistics progressed accordingly. Of the men I have discussed, only Wigglesworth and De Forest had no direct tie with astronomy or the Coast Survey. Even the European-educated Adrain's stumbling upon the normal distribution was inspired by an attempt to apply Legendre's methods for analyzing astronomical data to a problem in surveying.

De Forest's work belongs to another, separately developing tradition, that of actuarial mathematics. Here too the British led, as the major American insurance companies only reached full development in the mid-nineteenth century.

While early American work has not received much attention from historians, it did make some international impact. Peirce's outlier technique stimulated a debate which at one point involved the British Astronomer Royal. Simon Newcomb's work on robust estimation influenced the direction of Edgeworth's work. De Forest's precursor to the chi-square test and his anticipation of the gamma distribution and the method of moments may have played a role in Karl Pearson's later work on these subjects, as may also have been true of American work on the errors-in-the-variables problem, although I know of no direct evidence on this latter point. But at the least, the burst of activity in statistics after 1850, some at a remarkably high level, signaled that the talents available in America were second to none. Statistics in the early States had remained largely a dormant field; the second century would tell a different story.

Acknowledgments. I am grateful to James C. Hickman for comments and references, to T. N. E. Greville for access to Wolfenden's unpublished manuscript on De Forest, to Librarians S. Hunchar at the University of Pennsylvania, R. J. Mulligan at Rutgers, J. A. Schiff and C. M. Hanson at Yale, and C. J. Radmacher at the Warren County Library, Monmouth, Illinois; and, for comments on the first draft, to Persi Diaconis, Churchill Eisenhart, Charles C. Gillispie, Frederick Mosteller, Robin Plackett, Oscar Sheynin, George J. Stigler, John W. Tukey, and two referees.

REFERENCES

- ABBE, C. (1871). A historical note on the method of least squares. *Amer. J. of Science and Arts* 101 (Third series v. 1), 411-415.
- ADCOCK, R. J. (1877-1878). Note on the method of least squares. *Analyst* 4 183-184 (1877); 5 21-22, 53-54 (1878).
- ADRAIN, R. (1809). Research concerning the probabilities of the errors which happen in making observations, etc. *The Analyst; or Mathematical Museum* 1 (No. 4), 93-109. Available on microfilm as part of the American Periodical Series.
- ADRAIN, R. (1818). Investigation of the figure of the earth, and of the gravity in different latitudes. *Trans. Amer. Philos. Soc.* 1 (New Series) 119-135.
- ADRAIN, R. (1818). Research concerning the mean diameter of the earth. *Trans. Amer. Philos. Soc.* 1 (New Series) 353-366.
- AIRY, G. B. (1875). *On the Algebraical and Numerical Theory of Errors of Observations*, 2nd ed. Macmillan, London.
- ANDERSON, JOSEPH (Ed.) (1896). E. L. De Forest. *The Town and City of Waterbury Connecticut* 3 895-897. Price and Lee, New Haven.
- ANONYMOUS (1832). Doctrine of Probabilities. *Amer. Quarterly Rev.* 11 473-494.
- ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics* 2 123-147.
- ARCHIBALD, R. C. (1924). Simon Newcomb 1835-1909, Bibliography of his life and work. *Memoirs of the National Academy of Science* 17 19-69.
- ARCHIBALD, R. C. (1925). Benjamin Peirce: Biographical sketch and the writings of Peirce. *Amer. Math. Monthly* 32 8-30.
- BABB, M. J. (1926). Robert Adrain—Man and Mathematician. *The General Magazine and Historical Chronicle* 28 272-284.
- BOWDITCH, N. (1811). *The New American Practical Navigator*, 3rd ed. E. M. Blunt, New York.
- BYERLY, W. E. (1925). Reminiscences. *Amer. Math. Monthly* 32 5-7.
- CAJORI, F. (1890). *The Teaching and History of Mathematics in the United States*. Bureau of Education, Circular of Information No. 3, 1890 (Whole Number 167). G.P.O., Washington.
- CAMPBELL, W. W. (1924). Simon Newcomb. *Memoirs of the National Academy of Science* 17 1-18.
- COOLIDGE, J. L. (1926). Robert Adrain, and the beginnings of American mathematics. *Amer. Math. Monthly* 33 61-76.
- DE FOREST, E. L. (1873). On some methods of interpolation applicable to the graduation of irregular series, such as tables of mortality, etc. *Annual Report of the Board of Regents of the Smithsonian Institution for 1871*, 275-339.
- DE FOREST, E. L. (1874). Additions to a memoir on methods of interpolation applicable to the graduation of irregular series. *Annual Report of the Board of Regents of the Smithsonian Institution for 1873*, 319-353.
- DE FOREST, E. L. (1876). *Interpolation and Adjustment of Series*. Tuttle, Morehouse and Taylor, New Haven.
- DE FOREST, E. L. (1877). On adjustment formulas. *Analyst* 4 79-86, 107-113.

- DE FOREST, E. L. (1878a). On the grouping of signs of residuals. *Analyst* **5** 1-6.
- DE FOREST, E. L. (1878b). On repeated adjustments, and on signs of residuals. *Analyst* **5** 65-72.
- DE FOREST, E. L. (1878c). On the limit of repeated adjustments. *Analyst* **5** 129-141.
- DE FOREST, E. L. (1879). On the development of $[p + (1 - p)]^\infty$. *Analyst* **6** 65-73.
- DE FOREST, E. L. (1879-1880). On unsymmetrical adjustments, and their limits. *Analyst* **6** 140-148, 161-170 (1879); **7** 1-9 (1880).
- DE FOREST, E. L. (1880a). On some properties of polynomials. *Analyst* **7** 39-46, 73-82, 105-115.
- DE FOREST, E. L. (1880b). On a theorem in probability. *Analyst* **7** 169-176.
- DE FOREST, E. L. (1881a). Law of facility of errors in two dimensions. *Analyst* **8** 3-9, 41-48, 73-82.
- DE FOREST, E. L. (1881b). On the elementary theory of errors. *Analyst* **8** 137-148.
- DE FOREST, E. L. (1882). Law of error in the position of a point in space. *Analyst* **9** 33-40, 65-74.
- DE FOREST, E. L. (1882-1883). On an unsymmetrical probability curve. *Analyst* **9** 135-142, 161-168 (1882); **10** 1-7, 67-74 (1883).
- DE FOREST, E. L. (1883). A method of demonstrating certain properties of polynomials. *Analyst* **10** 97-105.
- DE FOREST, E. L. (1884). On an unsymmetrical law of error in the position of a point in space. *Trans. Connecticut Academy of Arts and Sciences* **6** 122-138.
- DE FOREST, E. L. (1885). On the law of error in target-shooting. *Trans. Connecticut Academy of Arts and Sciences* **7** 1-8.
- DOOLITTLE, M. H. (1881). Paper No. 3. *Report of the Superintendent of the U.S. Coast Survey for 1878*. Appendix 8, 115-120. G. P. O. Washington.
- EDGEWORTH, F. Y. (1896). The asymmetrical probability curve. *Philosophical Magazine*, Fifth series v. **41** 90-99.
- EDGEWORTH, F. Y. (1902). Article "Error, law of." In *Encyclopedia Britannica*, Vol. 28 of 9th and 10th Editions Combined, Vol. 4 of Supplement to 9th Edition.
- EISELE, C. (1957). The Charles S. Peirce-Simon Newcomb Correspondence. *Proc. Amer. Philos. Soc.* **101** 409-433.
- EISELE, C. (1974). Charles Sanders Peirce. In *Dictionary of Scientific Biography* **10** 482-488 (C. C. Gillispie, ed.). Scribner, New York.
- ELIOT, C. W. (1925). Reminiscences of Peirce. *Amer. Math. Monthly* **32** 1-4.
- FISHER, I. (1930). The application of mathematics to the social sciences. *Bull. Amer. Math. Soc.* **36** 225-243.
- FRANKLIN, B. (1755). Observations concerning the increase of mankind, peopling of countries, etc. In *The Writings of Benjamin Franklin* **3** 63-73 (Albert Henry Smyth, ed.), 1905. Macmillan, New York.
- GARVER, R. (1932). The Analyst, 1874-1883. *Scripta Mathematica* **1** 247-251, 322-326.
- GOODMAN, L. A. and KRUSKAL, W. H. (1959). Measures of association for cross classifications. II: Further discussion and references. *J. Amer. Statist. Assoc.* **54** 123-163.
- GREVILLE, T. N. E. (1966). On stability of linear smoothing formulas. *SIAM J. Numer. Anal.* **3** 157-170.
- GREVILLE, T. N. E. (1974). On a problem of E. L. De Forest in iterated smoothing. *SIAM J. Math. Anal.* **5** 376-398.
- GUMMERE, J. (1817). *A Treatise on Surveying*, 2nd ed. (First Edition published in 1814.) Richardson, Philadelphia.
- HATAI, S. (1910). De Forest's formula for an unsymmetrical probability curve. *Anatomical Record* **4** 281-290.
- JASTROW, J. (1888). A critique of psycho-physic methods. *Amer. J. Psychology* **1** 271-309.
- JEFFERSON, T. (1785). *Notes on the State of Virginia*. Reprinted 1955 by Univ. of North Carolina Press, Chapel Hill.
- KARPINSKI, L. C. (1940). *Bibliography of Mathematical Works Printed in America through 1850*. Univ. of Michigan Press, Ann. Arbor. Supplements in *Scripta Mathematica* **8** 233-236; **11** 173-177; **20** 197-202.

- KUMMELL, C. H. (1879). Reduction of observation equations which contain more than one observed quantity. *Analyst* **6** 97-105.
- LANCASTER, H. O. (1966). Forerunners of the Pearson χ^2 . *Austral. J. Statist.* **8** 117-126.
- MCEWEN, G. F. (1921). Rapid methods of approximating to terms in a binomial expansion. *J. Amer. Statist. Assoc.* **17** 606-621.
- MERRIMAN, M. (1877a). A list of writings relating to the method of least squares, with historical and critical notes. *Trans. Conn. Acad. of Arts and Sciences* **4** 151-232.
- MERRIMAN, M. (1877b). On the history of the method of least squares. *Analyst* **4** 33-36, 140-143.
- MERRIMAN, M. (1907). *A Text-Book on the Method of Least Squares*, 8th ed. Wiley, New York. First Edition 1884. Another text, *Elements of the Method of Least Squares*, published 1877.
- MOSTELLER, F. and WALLACE, D. (1964). *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Reading, Mass.
- NEWCOMB, S. (1860-1861). Notes on probability. *The Mathematical Monthly* (ed. by J. D. Runkle) **3** 119-125.
- NEWCOMB, S. (1881). Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.* **4** 39-40.
- NEWCOMB, S. (1903). *The Reminiscences of an Astronomer*. Harper, New York.
- O'DONNELL, T. (1936). *History of Life Insurance*. American Conservation Co., Chicago.
- PEARSON, K. (1895a). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Phil. Trans. Roy. Soc. Lond. A* **186** 343-414. Reprinted in *Karl Pearson's Early Statistical Papers* (pages 41-112), Cambridge Univ. Press (1956).
- PEARSON, K. (1895b). On skew probability curves. *Nature* **52** 317 (Aug. 1, 1895).
- PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, fifth series, **50** 157-175. Reprinted in *Karl Pearson's Early Statistical Papers* (pages 339-357). Cambridge Univ. Press (1956).
- PEIRCE, B. (1852). Criterion for the rejection of doubtful observations. *Astronomical Journal* **2** No. 21, 161-163.
- PEIRCE, B. (1878). On Peirce's criterion. *Proc. Amer. Acad. Arts and Sciences*, **13** (v. 5 N.S.), May 1877-May 1878, 348-349.
- PEIRCE, C. S. (1872). On the coincidence of the geographical distribution of rainfall and of illiteracy, as shown by the statistical maps of the ninth census reports (Abstract). *Bull. Philos. Soc. of Washington* **1** 68.
- PEIRCE, C. S. (1873). On the theory of errors of observations. Appendix No. 21 (pages 200-224 and plate 27) of *Report of the Superintendent of the U.S. Coast Survey for the year ending June 1870*. G.P.O., Washington. Reprinted in *The New Elements of Mathematics* by C. S. Peirce, ed. by C. Eisele, Humanities Press, Atlantic Highlands, N.J., 1976, Vol. 3, pt. 1, pp. 639-676.
- PEIRCE, C. S. (1879). Note on the theory of the economy of research. Appendix No. 14 (pages 197-201) of *Report of the Superintendent of the U.S. Coast Survey for the year ending June 1876*. G.P.O., Washington. (Also in *Collected Papers of Charles Sanders Peirce*, vol. 7. See pages 84-88 for some "Later reflections" on this subject.) Reprinted in *Operations Res.* **15** 643-648 (1967).
- PEIRCE, C. S. (1884). The numerical measure of the success of predictions. *Science* **4** 453-454. Reprinted in *The New Elements of Mathematics* by C. S. Peirce, ed. by C. Eisele, Humanities Press, Atlantic Highlands, N.J., 1976, vol. 3, pt. 1, pp. 682-683.
- PEIRCE, C. S. and JASTROW, J. (1885). On small differences of sensation. *Memoirs of the National Academy of Sciences for 1884* **3** 75-83.
- PEIRCE, C. S. (1957). *Essays in the Philosophy of Science*. (V. Tomas, ed.) Bobbs-Merrill, Indianapolis. The material on pages 195-234, "Lessons from the History of Science," was written c. 1896, originally published in *Collected Papers of Charles Sanders Peirce* (C. Hartshorne and P. Weiss, eds.), **1** 19-49. Harvard Univ. Press (1931).

- POINCARÉ, H. (1912). *Calcul des probabilités*, 2nd ed. Gauthier-Villars, Paris.
- REINGOLD, N. (1972). American Indifference to Basic Research: A Reappraisal. *Nineteenth-Century American Science, A Reappraisal*, 38-62 (George H. Daniels, ed.). Northwestern Univ. Press.
- RIDER, P. R. (1933). Criteria for rejection of observations. *Washington University Studies-New Series, Science and Technology*, No. 8.
- SHEPHERD, J. (1975). *The Adams Chronicles*. Little, Brown, Boston.
- SMITH, D. E. and GINSBURG, J. (1934). *A History of Mathematics in America Before 1900*. Carus Mathematical Monograph No. 5.
- SMITH, SYDNEY (1818). Travellers in America. *Edinburgh Review* 31 132-150. Reprinted in *The Works of the Rev. Sydney Smith*, D. Appleton, New York (1860, pages 107-115).
- STIGLER, S. M. (1973a). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920. *J. Amer. Statist. Assoc.* 68 872-879.
- STIGLER, S. M. (1973b). Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* 60 439-445.
- STIGLER, S. M. (1975). The transition from point to distribution estimation. 40th session of the I.S.I., Warsaw, Poland.
- TOCQUEVILLE, A. DE (1840). The example of the Americans does not prove that a democratic people can have no aptitude and no taste for science, literature, or art. *Democracy in America*, Vol. 2, Book 1, Chapter 9; pages 35-40 of the 1945 edition. Knopf, New York.
- VON NEUMANN, J., KENT, R. H., BELLINSON, H. R. and HART, B. I. (1941). The mean square successive difference. *Ann. Math. Statist.* 12 153-162.
- WALKER, HELEN M. (1929). *Studies in the History of Statistical Method*. William and Wilkins, Baltimore. Reprinted Arno Press (1975), New York.
- WEISS, PAUL (1934). Charles Sanders Peirce. In *Dictionary of American Biography* 14 398-403 (D. Malone, ed.). Scribner, New York.
- WHEELER, L. P. (1951). *Josiah Willard Gibbs*. Yale Univ. Press.
- WILSON, E. B. and HILFERTY, M. M. (1929). Note on C. S. Peirce's experimental discussion of the law of errors. *Proc. National Acad. Sci.* 15 120-125.
- WILSON, E. B. (1931). Reminiscences of Gibbs by a student and colleague. *Scientific Monthly* 210-227.
- WOLFENDEN, H. H. (1925). On the development of formulae for graduation by linear compounding, with special reference to the work of Erastus L. De Forest. *Trans. Actuarial Soc. of America* 26 81-121.
- WOLFENDEN, H. H. (1942). *The Fundamental Principles of Mathematical Statistics*. Macmillan, Toronto.
- WOLFENDEN, H. H. (1968). Biography of Erastus L. De Forest. Unpublished, c. 1968.
- WOODWARD, R. S. (1885). Some practical features of a field time determination with a meridian transit (Abstract). *Bull. Philos. Soc. of Washington* 8 55-58.
- WRIGHT, T. W. (1882). On the computation of probable error. *Analyst* 9 74-78.

SUPPLEMENTARY REFERENCES²

- AIRY, G. B. (1856). Letter from Professor Airy, Astronomer Royal, to the editor. *Astronomical Journal* 4 137-138.
- NEWCOMB, S. (1859-1861). Notes on the Theory of Probabilities, *Mathematical Monthly* (Runkle, ed.), Vol. 1, pp. 136-139, 233-235, 331-335; Vol. 2, pp. 134-140, 272-275; Vol. 3, pp. 119-125, 343-349.

²Added to the original article by the author for this volume.

- NEWCOMB, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, Vol. 8, pp. 343–366.
- PLACKETT, R. L. (1972). The discovery of the method of least squares. *Biometrika* **59** 239–251. Reprinted 1977 in *Studies in the History of Statistics and Probability*, Vol. 2 (M. G. Kendall and R. L. Packett, eds.) (London: Griffen).
- WIGGLESWORTH, E. (1793). A table showing the probability of the duration, the decrement and the expectation of life, in the states of Massachusetts and New Hampshire formed from sixty-two bills of mortality... in the year 1789. *Memoirs of the American Academy*, Vol. 2, part 1, Boston, 1793.
- WINLOCK, J. (1856). On Professor Airy's Objections to Peirce's Criterion. *Astronomical Journal* **4** 145–47.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

WILLIAM FELLER AND TWENTIETH CENTURY PROBABILITY

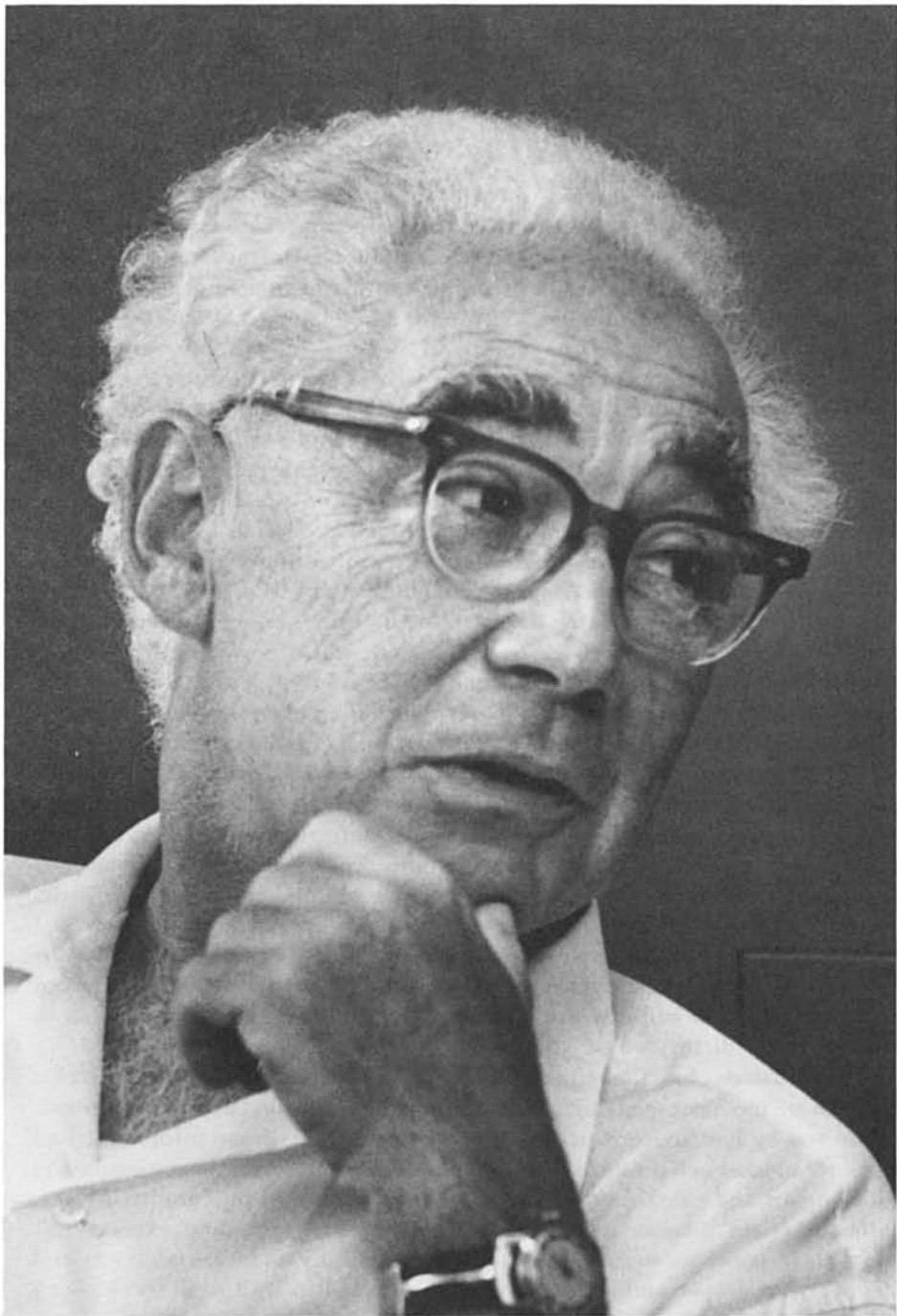
1. Twentieth Century Probability

When William Feller was born in 1906, Lebesgue measure had just been invented, and Fréchet was to introduce measure on an abstract space about ten years later. Thus, the technical basis of modern mathematical probability was developed about the time of Feller's early childhood. Since that time the subject has been transformed, by no one more than by Feller himself, into an essential part of mathematics, contributing to other parts as well as drawing from them.

In the first part of our century, few probabilists felt comfortable about the basis of their subject, either as an applied or as a purely mathematical subject. In fact, it was commonly judged that there was no specific mathematical subject "probability," but only a physical phenomenon and a collection of mathematical problems suggested by this phenomenon. A probabilist joked that probability was "a number between 0 and 1 about which nothing else is known." In the discussions of the foundations of probability, there was no clear distinction made between the mathematical and the real. For example, one influential theory was that of von Mises, based on the concept of a "collective," which was defined as a sequence of observations with certain properties. Since "observation" is not a mathematical concept and since the properties were properties which no mathematical sequence could have, the theory could survive in its original form only by an affirmation that it was not a formal mathematical theory but an attempt at a direct description of reality. Instead, the theory was restricted to remove the mathematical objection, unfortunately losing in intuitive content what it gained in mathematical significance. The fate of the theory was an inevitable result of the increasing demand of mathematicians for exact definitions and formal rigor. The present formal correctness of mathematical probability only helps indirectly in analyzing real probabilistic phenomena. It is unnecessary to stress to statisticians that the relation between mathematics and these phenomena is still obscure. Or if not obscure it is clear to many but in mutually contradictory ways.

Formalizations of mathematical probability by Steinhaus in 1923 and Fréchet in 1930 were too incomplete to have much influence. The first acceptable formalization was by Kolmogorov in his 1933 monograph. Of course before that, and in fact for at least two centuries before that, there had been mathematicians who made correct and valuable contributions to mathematical probability. Mathematicians could manipulate equations inspired by events and expectations before these concepts were formalized mathematically as measurable sets and integrals. But deeper and subtler investigations had to wait until the blessing and curse of direct physical significance had been replaced by the bleak reliability of abstract mathematics.

Reprinted from *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. xv-xx. with permission of the University of California Press.



William F. Feller

Some probabilists have scorned the measure theory, functional analysis invasion of their subject, thinking it could do no more good than the discovery by Molière's character that he had been talking prose all his life. But in fact this invasion, to which Feller contributed so much, enriched the subject enormously in bringing it into the framework of modern mathematics, providing it with the possibility of undreamed of contacts with seemingly quite different mathematical fields. The definitive acceptance of mathematical probability as mathematics was, however, quite unnecessary for a large part of probabilistic research. For example, much of the distribution theory of sums of independent random variables can be considered an analysis of the convolutions of distribution functions. Random variables need never be mentioned. But even such researches are interesting largely because of their probabilistic significance. Many would not have been thought of and many more would have not been carried out, even if thought of, without this significance. Thus, the acceptance of probability as mathematics influenced research that could have been written without the probabilistic context. Even now this acceptance is not complete. In fact, many mathematics students are unaware of the place of probability in their subject. This situation is preserved by the special flavor given to probability by its linguistic heritage. Terms like "random variable" are here to stay and to continue misleading students on the state of probability theory, although "random variable" has a purely mathematical meaning whereas other familiar terms like "inclined plane" do not.

But the situation was even more confused forty years ago. A student can hardly visualize the difficulty of working in a field without a formal basis, without any sophisticated textbooks, in which it was respectable to have a serious discussion on what "really happens" when one tosses a coin infinitely often. The first sophisticated book was Lévy's remarkable 1937 book which was not written as a textbook and which yielded its treasures only to readers willing to make extreme efforts. A distinguished statistician in the early 1930's when asked how probability was taught at his university expressed surprise at the idea of teaching probability as a separate subject—it would be a "pointless *tour de force*." At that time random variables still were so mysterious that another distinguished mathematical statistician stated in a lecture that it was not known whether two random variables which were uncorrelated had to be independent. At that time the idea that a random variable was just (mathematically) a function was still so unfamiliar that it did not occur to the speaker to consider the sine and cosine functions on $(0, 2\pi)$, with the uniform distribution on that interval, as random variables, trivially uncorrelated and not independent.

But it is true that probability has lost some of its glamour along with its mystery. Luckily, the subject still has its basic physical background to draw on, still a source of ideas and problems. A further present feature is the interplay between mathematical probability and other parts of mathematics, for example, partial differential equations and potential theory.

It was a wonderful thing to be entering the field of probability when Feller did, in the early thirties. To one with his classical background, the field was obviously full of unsolved problems. Of course, it was not obvious at the time that the field was as rich as it has turned out to be, but it was clear that the subject was new in the sense that it had been barely touched by modern techniques. For example, discrete and continuous parameter Markov processes were just beginning to be studied in a nontrivial way. The multiplicity of classical type problems suggested by probability was such that it was not surprising that many probability papers indulged in probabilistic slang only long enough to reach the safe territory of integral equations or some other respectable established topic. Thereafter both writer and reader could relax, knowing that the introductory slang was as unessential as it was unexplained.

Wiener's work on Brownian motion (1924) was an exception. Although Wiener even later never used or knew the slang or even many of the elementary results of probability theory, his Brownian motion analysis was quite rigorous. In fact, an early problem in stochastic processes was to create a general theory which would include his approach to Brownian motion!

Since the thirties, mathematical probability has exploded. Several journals are devoted to it in its pure form and some other journals, for example the *Annals of Mathematical Statistics*, are barely distinguishable from probability journals. Sophisticated text books and specialized books are appearing all the time and there is even talk that the subject has reached or passed its peak. Feller, who was one of the researchers who brought the field to its present state, liked to relax in his advanced research by playing with elementary problems, polishing their known solutions. Let us hope that the new crop of researchers will be able to continue both his research and his purifying of old results.

2. William Feller

Feller was born in Zagreb, Yugoslavia on July 6, 1906, the ninth of twelve children of the well to do owner of a chemical factory. He attended the University of Zagreb (1923–1925), where he received the equivalent of an M.S. degree, and the University of Göttingen, where he received his Ph.D. in 1926 and remained until 1928. In 1928, he left Göttingen for the University of Kiel, where he worked as Privatdozent until in 1933 he refused to sign a Nazi oath and was forced to leave. It was at Kiel that he did his first work in probability. After a year (1933–1934) in Copenhagen, he went to Stockholm where he spent the next five years at the University and (July 27, 1938) married Clara Nielsen who had been his student in Kiel.

In 1939 the Fellers emigrated to Providence, where he became associate professor at Brown University and the first executive editor of *Mathematical Reviews*, founded that year. The only current mathematics review journal was then becoming corrupted by Nazi ideas. Much of the success of *Mathematical*

Reviews has been due to the policies initiated by Feller. *Mathematical Reviews* was founded in a less frantic scientific age when it was reasonable to have critical reviews, before the age of speed and preprints. There is now some opinion that traditional reviewing is an obsolete luxury, like peaceful universities. But even if this is true, the very speed of mathematical development that has made it true is in part a tribute to the success of *Mathematical Reviews* in furthering research.

In 1945, Feller accepted a professorship at Cornell University and remained there until 1950 when he moved to Princeton University as Eugene Higgins Professor of Mathematics. He held this position until his death (January 14, 1970), but in addition was a Permanent Visiting Professor at The Rockefeller University where he spent the academic years 1965–1966 and 1967–1968. A great attraction at The Rockefeller University was the opportunity to talk to geneticists.

Feller's first probability paper (1935) was on the central limit theorem, and in fact the properties of normalized sums of independent random variables were the subject of much of his later research, both from the point of view of distribution theory and from that of asymptotic bounds of the sums. Some of his deepest analytical work was in connection with the latter, work related to the general forms of the iterated logarithm law. It was in the context of distribution functions and their convolutions, not of random variables, and thus did not need the mathematical formalization of probability provided by Kolmogorov only a few years before. The central limit theorem paper gave necessary and sufficient conditions for convergence to a Gaussian limit.

In 1906, Markov did the first work on the sequences of random variables with the property that now bears his name. Progress was slow at first and some of Markov's work was repeatedly rediscovered. Kolmogorov's 1931 paper on continuous parameter Markov processes was a turning point, the first systematic investigation of these processes including the processes of diffusion. Feller wrote his first paper on these processes in 1936, going considerably beyond Kolmogorov and proving the appropriate existence and uniqueness theorems for the integrodifferential equations governing the transition probabilities. The main interest of both authors was in these equations. The stochastic processes themselves were secondary, although they inspired the analysis, and it is not surprising in view of the general historical remarks made above in Section 1 and the state of the subject at the time that Kolmogorov defined Markov processes incorrectly and Feller added an incorrect characterization (independent increments) to Kolmogorov's definition. All they needed was the Chapman-Kolmogorov equations, and the process giving rise to them was almost irrelevant. For Feller, as distinguished, say, from Lévy, it was usually the differential or integral equations or the semigroups arising in a probability context that interested him, rather than sample properties. On the other hand he kept these properties in mind, and although he usually did not treat them specifically he had a sure feeling for them and they inspired much of his analysis.

Feller completely transformed the subject of Markov processes. Going beyond

his 1935 paper, he put the analysis into a modern framework, applying semigroup theory to the semigroups generated by these processes. He observed that the appropriate boundary conditions for the parabolic differential equations governing the transition probabilities correspond on the one hand to the specification of the domains of the infinitesimal generators of the semigroups and on the other hand to the conduct of the process trajectories at the boundaries of the process state spaces. In particular, he found a beautiful perspicuous canonical form for the infinitesimal generator of a one dimensional diffusion. In this work, he was a pioneer yet frequently obtained definitive results.

Feller is best known outside the specialists in his field for his two volume work *An Introduction to Probability Theory and Its Applications*. He never tired of revising this book and took particular pleasure in finding new approaches, new applications, new examples, to improve it. The book is extraordinary for the almost bewildering multiplicity of its points of view and applications inside and outside pure mathematics. No other book even remotely resembles it in its combination of the purest mathematics together with a dazzling virtuosity of techniques and applications, all written in a style which displays the enthusiasm of the author. This style has made the book unexpectedly popular with non-specialists, just as its elegance and breadth, not to mention its originality, has made it an inspiration for specialists. Feller had planned two more volumes, and it would have been fascinating to see if his excitement in his subject could have brightened the usual dull measure theoretic details which would inevitably have had to appear in later volumes. Perhaps his unequalled classical background could have diluted, and made more palatable with applications and examples the concentrated dosage of preliminaries other mathematicians find necessary before studying Markov processes.

Feller was never heavily involved in statistics, although he was interested in it. He was not afraid of dirtying his fingers with numbers and in fact at one time he liked to work out least squares problems on hand computers as relaxation! He was president of the Institute of Mathematical Statistics in 1946. His attitude towards applications was unusual. On the one hand, his research was almost entirely in pure mathematics. On the other hand, he had far more than an amateur's interest in and knowledge of several applied fields, including statistics and genetics. He wrote a paper on extra sensory perception, and he wrote several papers applying the sophisticated ideas of a modern probabilist to genetics. He took an excited delight in applications of pure theory and nothing pleased him more than finding new ones. On the other hand he had a low boiling point for poor thinking, and nothing made him more excited than what he considered improper scientific thinking whether he favored or opposed the conclusion. Thus, he had great contempt for those who buttressed insufficient statistics on lung cancer and cigarettes with emotionalism or those who adduced uninformed arguments against Velikovsky's theories.

Feller was a member of the U.S. National Academy of Sciences, the Royal Danish Academy of Sciences, and the Yugoslav Academy of Sciences, as well as

a member of the American Academy of Arts and Sciences and the American Philosophical Society. His wife accepted the National Medal of Science for him shortly after his death. But apart from his mathematics those who knew him personally will remember Feller most for his gusto, the pleasure with which he met life, the excitement with which he drew on his endless fund of anecdotes about life and its absurdities, particularly the absurdities involving mathematics and mathematicians. To listen to him deliver a mathematics lecture was a unique experience. No one else could generate in himself as well as in his auditors so much intense excitement. In losing him, the world of mathematics has lost one of its strongest personalities as well as one of its strongest researchers.

J. L. DOOB

Early Days in Statistics at Michigan

CECIL C. CRAIG

For me this period began in 1922 when I arrived in Ann Arbor with an M.S. degree intending to take courses in Actuarial Science. Professor J. W. Glover, who set up the actuarial program in Michigan, which still flourishes, conceived the idea in about 1910, that such a curriculum should include courses in mathematical statistics. In 1916 he brought back to Michigan a recent graduate, Harry C. Carver, to develop courses in that subject. In 1922 there were only two schools in the country, the State University of Iowa and the University of Michigan, where courses in mathematical statistics were offered. Carver's first course, Mathematics 49 and 50, each for 2 hours credit, ran throughout the year at a precalculus level. A second more mathematical course was given by Professor R. W. Barnard, who later taught pure mathematics at the University of Chicago. I took this course and learned some mathematics but not much statistics. I began teaching an advanced course after I got my doctor's degree which was a result of a year in Lund, Sweden, working under Professor S. D. Wicksell.

In those days the *Journal of the American Statistical Association* was well established, but manuscripts with any mathematical content had little chance of being published by the *Journal*. I heard Professor Carver say on more than one occasion that there ought to be a place in this country where a paper in mathematical statistics could appear. I have always thought that the trigger for the founding of the *Annals of Mathematical Statistics* was a paper of mine that was rejected by the *Journal* because it was too mathematical. Carver reacted strongly to this and shortly afterward he joined with a friend, J. W. Edwards, who was trying out a new lithoprinting process, in putting out the first issue of the *Annals of Mathematical Statistics* in 1930. Carver

Cecil C. Craig (April 14, 1898–June 16, 1985) was Emeritus Professor of Mathematics and Director of the Statistical Research Laboratory, University of Michigan. This memoir was written shortly before his death on the occasion of the 50th anniversary of the establishment of the Institute of Mathematical Statistics.

Reprinted from *Statistical Science*, Vol. 1, No. 2, pp. 292–293, with permission of the Institute of Mathematical Statistics.

assumed the financial responsibility for the new journal and with the aid of two assistants and his friend's support he served as its editor until 1935 when he turned the *Annals* over to the newly formed Institute of Mathematical Statistics. There was a sufficient supply of scholarly papers offered for publication but the supply of funds to meet the bills was not enough to avoid severe strains. At times toward the end of World War II the *Annals* came close to going broke. I don't know if Carver ever told anybody the cost in dollars of his devotion to statistics but I doubt if he knew closely. Fortunately, the publishers of the *Annals* and the officers of the Institute allowed a really large inventory of back numbers to accumulate during the second World War. Once the war was over, it turned out that there was a healthy market for those back numbers. The faithful industry of Paul Dwyer and Carl Fischer handled the sale of this merchandise. Only their friends knew how hard they worked, but enough money came in to put the *Annals* on a sound financial footing.

The remainder of the 1920s and the first of the 1930s were marked by a steady growth in this country in the number of people whose principal interest lay in mathematical statistics. By living and working in the city where the new *Annals* were edited and by regular attendance at the national meetings, it was easy for me to become widely acquainted with the members of the new group. I spent the year 1930–31 in Stanford University where Harold Hotelling was beginning a career in statistics. When I left Stanford to return to Ann Arbor, Hotelling also left to accept an appointment at Columbia University. On my way back across the country I stopped for a few days in Iowa City where Egon Pearson was lecturing. There a rather remarkable group of students was working with H. L. Rietz, who deserved to be known as the dean of American mathematical statisticians. These students were S. S. Wilks, A. T. Craig, Selby Robinson, and Carl Fischer. They all earned doctorates under Rietz, and I made friends with all of them. Only Fischer, who recently retired from Michigan, is still alive. When I left Iowa City, I went to Minneapolis where I spent five weeks listening to my first series of lectures by R. A. Fisher. Some time in the next few years I became well acquainted with B. H. Camp of Wesleyan.

In 1935, the summer meetings of the mathematics societies were held in Ann Arbor. The attendees included enough people interested in mathematical statistics to fill the reception room in the Betsy Barbour dormitory on this campus. They were convened to discuss a proposed organization of a new society devoted to mathematical statistics. I do not recall all of the thirty to forty people who were present, but I do remember Rietz, Wilks, A. T. Craig, Carl Fischer, Selby Robinson, and Paul Rider from Iowa and H. C. Carver, C. C. Craig, T. E. Raiford, and A. L. O'Toole from Michigan. Others whom I do not recall from there were Hotelling, Camp, Gertrude Cox, and W. A. Shewhart.

I know that previously Carver's idea of the proper form to be assumed by an organization of mathematical statisticians was that of the actuaries, with qualifying examinations for different grades of membership. But at the actual organization meeting this form of a society was not seriously proposed and a form very close to what we have today was adopted with only brief discussion. The Institute of Mathematical Statistics was created on September 12, 1935 in Ann Arbor, Michigan, with the following elected officers: President, H. L. Rietz; Vice President, W. A. Shewhart; Secretary-Treasurer, A. T. Craig. The five-year-old *Annals of Mathematical Statistics* was adopted as the official journal of the new society.

From his joining of the faculty of the University of Michigan until his retirement in 1960 the dominant figure in statistics in Michigan was Harry Carver. He was a native of Waterbury, Connecticut, and he took a B.S. degree from Michigan in 1915. He had a spare well-muscled figure more than 6 feet tall, a sandy complexion, and the coordination of a natural athlete. He won an "M" in track and worked out for years with the cross country team. He was good enough at pocket billiards to have made his living at that game. He could beat ordinary golfers using only a five iron. He greatly enjoyed bridge and belonged to a group which regularly met for poker. I do not know that he was exceptionally good at card games.

As a high school student, he was known for repairing and riding motorcycles. Later, as a student, he enjoyed rebuilding second-hand automobiles, making them better than new. He became known for his fast driving. He made a practice of leaving Ann Arbor at the same time as the train carrying the track team and arriving first into Chicago. Sometime later he discovered California and with his second wife he more than once drove there nonstop; one driving while the other slept in the back seat.

But soon he became interested in airplanes and became a qualified pilot. He and a friend acquired a small plane and he enjoyed taking acquaintances for rides. He quickly became aware of the problems in navigation encountered by the pilots of the long-range, high-speed planes being supplied to the Air Force. He cultivated friends among officers in the Air Force. He enrolled in and completed the training course being given to United States Air Force cadets. He applied his quick mind and mathematical ability to improving navigational methods then in use. He showed how to use small calculating machines to get numerical results in navigation problems more quickly and accurately. As he neared retirement age he spent much time in Air Force bases in Texas and California.

After retirement he made a study of climatic data for the United States and selected Santa Barbara, California, as the best place to live. He rented an apartment and moved there for several years. At age 80 he quit driving an automobile "while he was ahead" as he put it. His health deteriorated and

he moved back to Ann Arbor and ended his days at age 87 in the Michigan Union.

Carver had a very quick mind and he had a warm and sympathetic manner. Taking a course with him was an experience his students did not forget. He directed the work of ten doctoral students. He bordered on the eccentric; his diet seemed to consist largely of crackers and milk. He made a practice of offering to buy a class a dinner if it could beat him at one of five indoor sports—card games or billiards or pool—or at one of five outdoor sports such as running or putting the shot. He never lost.

Churchill Eisenhart majored in mathematical physics at Princeton University, receiving an A.B. degree in 1934. Continuing at Princeton the following year, he studied under S. S. Wilks and received an A. M. in mathematics in 1935. He then moved to University College London, where he received a Ph.D. in 1937 under the supervision of Jerzy Neyman. Later he taught at the University of Wisconsin, served as an applied mathematician during World War II, then in 1946 joined the staff of the National Bureau of Standards, where he established and headed (1947–1963) the Statistical Engineering Laboratory. He was then an NBS Senior Research Fellow until his retirement in 1983. The recipient of many honors for his work in mathematical statistics and its applications, Dr. Eisenhart is currently a Guest Scientist in the Center for Computing and Applied Mathematics of the National Institute of Standards and Technology (successor to the NBS).

S. S. Wilks' Princeton Appointment, And Statistics At Princeton Before Wilks*

CHURCHILL EISENHART

The following paragraphs provide a detailed account of the circumstances of the appointment of Samuel Stanley Wilks (1906–1964) to a position in the Department of Mathematics, Princeton University, in 1933; the state of statistics at Princeton then and in prior years; and an explanation of the three-year delay before Wilks taught his first course in statistics at Princeton.

The key figure in Wilks' appointment was my father, Luther Pfahler Eisenhart (1876–1965), who, in the spring of 1933, was not only willing, but, as Chairman of the Department of Mathematics (1928–1945), Dean of the Faculty (1925–1933), and Chairman of the University Committee on Scientific Research (1930–1945), was also able to effect Wilks' appointment to an

*Adapted from "Samuel S. Wilks and the Army Experiment Design Conference Series," in *Proceedings of the Twentieth Conference on the Design of Experiments in Army Research Development and Testing held at the U.S. Army Engineer Center, Fort Belvoir, Virginia, 23–25 October 1974* (ARO Report 75-2), Research Triangle Park, North Carolina: U.S. Army Research Office, June 1975.

instructorship in Mathematics on a more or less emergency basis over the opposition of almost every member of his department.

A few words are in order on how my father became interested in, and partial to statistics: My father's primary mathematical interest was differential geometry, and his research was exclusively in that area. Exactly when he began to take an "outside" interest in mathematical statistics I do not know. It may have been as early as 1913, when he corresponded with Edward L. Dodd on various aspects of the latter's paper entitled "The probability of the arithmetic mean compared with that of certain other functions of measurements", which was published in the *Annals of Mathematics* (Vol. 14, pp. 186–198, June 1913), of which my father was then an editor. At any rate, thereafter Dodd sent my father reprints of many of his subsequent papers on functional and statistical properties of various types of "means", which my father kept and ultimately turned over to me when I became interested in such matters in the early 1930s.

Early in 1924, "at the request of the Commission on New Types of Examination of the College Entrance Examination Board", my father "formed a committee of mathematicians to examine critically certain statistical methods used in the investigations of the Commission" (*American Mathematical Monthly*, Vol. 31, No. 4 (April 1924), p. 209). The "mathematicians" of the Committee included the economic statisticians W. Randolph Burgess and W. L. Crum (1894–1967) of the Federal Reserve System and of the Economics Department, Harvard, respectively; the mathematicians E. V. Huntington (1874–1952) and J. H. M. Wedderburn (1882–1948), of Harvard and Princeton, respectively; and the mathematical statistician, H. L. Rietz, of the University of Iowa.

The findings of this Committee, my father's continued advisory relations with the higher-ups of the College Entrance Examination Board (CEEB), and Wilks' contributions at Iowa (and under Hotelling at Columbia) to the solution of statistical problems arising in educational testing, made it possible for my father to arrange a part-time appointment with the CEEB concurrent with his initial University appointment — a relationship with the Board, and its successor, the Educational Testing Service, that continued until Wilks' death.

Harold Hotelling (1895–1973) took steps to assure that my father was kept informed of the Student-Fisher revolution in statistical theory and practice. He had gone to Princeton University as a J. S. K. Fellow in mathematics, 1921–1922, after receiving his A. B. (1919) and an M.S. (1921) from the University of Washington, in Seattle. His interest in statistics predated his going to Princeton in the fall of 1921. He had hoped to find some work in probability theory and the mathematics of statistics going on there in the Mathematics Department. Finding none, he undertook instead a program of study and research in topology (then called "analysis situs") and differential geometry, under the direction of Professor Oswald Veblen (1880–1960) and

my father, Luther Pfahler Eisenhart. He stayed on at Princeton, 1922–1924, as an Instructor in Mathematics and received his Ph.D. from Princeton University in June 1924, his doctoral dissertation being on “Three-dimensional manifolds of states in motion.” His paper “An application of analysis situs to statistics” (*Bulletin of the American Mathematical Society*, Vol. 33, (1927), pp. 467–476), had to do with topological aspects of serial and multiple correlations.

Following receipt of his Ph.D., Hotelling returned to the West Coast, to Stanford University, where he was a Junior Research Associate and then Research Associate (1925–1927), in the Food Research Institute; and finally, an Associate Professor of Mathematics (1927–1931), in the Department of Mathematics. In 1931, he was called to Columbia University, in New York City, as Professor of Economics to develop further the existing work there in mathematical economics, and to initiate a program in mathematical statistics.

Hotelling’s paper on “The distribution of correlation ratios calculated from random data”, in *Proceedings of the National Academy of Sciences*, 11, no. 10 (October 1925), pp. 657–662, made him the first person in the United States to respond *in kind* to R. A. Fisher’s signal contributions to the theory of small samples—his derivation employed the same kind of geometrical reasoning in terms of Euclidean N-dimensional space that Fisher had used so effectively. This paper carries a footnote that I’ve always considered to be very significant. I believe it affords an explanation of why so many American mathematicians had difficulty following Fisher’s geometrical proofs. Anyone who attempts to duplicate Fisher’s geometrical reasoning soon discovers that a crucial step is the correct evaluation of the relevant *element of volume*. Hotelling, at this juncture in his paper, gives a general expression for the relevant element of volume, which he numbers “(17)”, and then remarks in a footnote:

“This important expression for the volume element has been used in lectures [at Princeton University] by Professors O. Veblen and L. P. Eisenhart. I do not find it in any of the treatises on calculus, analysis or differential geometry, save for the special case in which the manifold of integration is a surface. It may readily be proved by showing first that (17) is a relative invariant under arbitrary transformations of the parameters; and second, that if the parameters of the hypersurface are orthogonal at a point, (17) becomes at this point the simple expression for the volume element in cartesian coordinates.”

During his years at Stanford, Hotelling wrote and published a stream of important original contributions to statistical theory and mathematical economics; reviews of American and English books on statistical methods, (e.g., of *Statistical Analysis* by Edmund E. Day (New York: The Macmillan Company, 1925), in *Journal of the American Statistical Association*, Vol. 21, No. 155 (Sept. 1926), pp. 360–363), in which he deplored the obsolescence of teaching and research in statistics in the United States and placed the blame

squarely on the doorsteps of departments of mathematics; and expository articles on “British statistics and statisticians today” (*Journal of the American Statistical Association*, Vol. 25, No.170 (June 1930) pp. 186–190), “Recent improvements in statistical inference” (*Journal of the American Statistical Associates*, Vol. 26, March 1931 *Supplement*, pp. 79–87; discussion, pp. 87–89) in which he did his very best to acquaint American readers with the “new look” in statistics. He regularly sent reprints of all of these to my father. When my father gave them to me in the fall of 1932, as I was reading up on “Student-Fisher statistics”, it was quite clear that my father had more than a superficial knowledge of the papers on statistical theory, and had “got the message” of Hotelling’s book reviews and expository articles.

An event that was to be instrumental in bringing both mathematical economics and modern statistical theory and methodology to the Princeton campus was the arrival of Charles F. Roos (1901–1958) as a National Research Fellow in Mathematics for the academic year 1927–1928. Roos had received his Ph.D. in theoretical economics at the Rice Institute in 1926 under Professor G. C. Evans (1887–1973), who at that time was developing a new mathematical theory of economic phenomena termed “economic dynamics”, and had spent 1926–1927 at the University of Chicago working with Professor Henry Schultz (1893–1938) who at that time was deeply engaged in his epochal research on statistical laws of demand and supply as one facet of his life’s work on the theory and measurement of demand. Roos came to Princeton primarily to broaden and sharpen his knowledge of mathematics as a basis for making further contributions to Professor Evans’ new “economic dynamics”. While there he succeeded in convincing some members of the Department of Economics and Social Institutions that the Department could not afford to continue to neglect much longer the advances in economic theory and methods pioneered by Evans and Schultz.

In 1928 my father became the chairman of the Mathematics Department. One of his early acts in this capacity was to arrange for the loan by the Bell Telephone Laboratories, Inc. of a member of its Technical Staff, Dr. Thornton C. Fry, author of *Probability and Its Engineering Uses* (D. Van Nostrand, 1928), to give a course at Princeton on “Methods of Mathematical Physics” as a Visiting Lecturer in Mathematics during the first semester 1929–1930. One result of Fry’s visit to Princeton was that a course in probability, taught by H. P. Robertson (1903–1964), Associate Professor of Mathematical Physics, using Fry’s book as the text, was offered by the Mathematics Department during the second semester of my sophomore year (1931–1932). It was this course that first interested me in probability and mathematical statistics and started me on my career.

In 1931 steps were taken that led to a course in “modern statistical theory” being offered for the first time at Princeton by the Department of Economics

and Social Institutions during the second semester of my senior year (1933–1934). What happened was this: Professor Frank D. Graham (1890–1949) of this department approached my father in his capacity as Dean of the Faculty, and suggested that one way to overcome lack of competence in his department with respect to the latest developments in mathematical and statistical methods in economics would be to send one of the young instructors in his department to study with Professor Henry Schultz at the University of Chicago. (The possibility of hiring a new staff member from the outside to this end had been considered earlier but put aside—the Depression was in full swing, and there was a freeze on new university appointments.) My father was favorable to this proposition, subject to an additional provision: that the individual concerned also study the modern theory of statistical inference with Harold Hotelling for the purpose of initiating a course in this subject on his return. The “victim” that Professor Graham had in mind was Acheson J. Duncan; and this is how it came to pass that Duncan, with financial assistance from the International Finance Section of Princeton University, spent the first half of the academic year 1931–1932 studying with Professor Henry Schultz at the University of Chicago; and the second half with Professor Hotelling at Columbia.

This assignment was very disruptive to Duncan at that time. When asked to undertake it he was already at work on his doctoral dissertation on “South African gold and international trade”; and his acceptance of it delayed until 1936 his completion of the requirements for his Ph.D. in economics. He also lost out on one of the features that “sweetened” the proposition, an opportunity to visit the West Coast—when the plans were made, Hotelling was at Stanford University, but had moved on to Columbia University before the time arrived for Duncan to study under him. This assignment was to be instrumental in changing the direction of Duncan’s subsequent career. (After serving as Assistant Professor of Economics and Statistics at Princeton until 1945, he became a Professor of Statistics at Johns Hopkins University, retiring as Professor Emeritus in 1971. Since 1960 he has been a member and Chairman (1972–1979) of the prestigious Committee E-11 (statistical methods) of the American Society of Testing and Materials.)

Duncan returned to Princeton in the fall of 1932 to resume his duties as Instructor in the Department of Economics and Social Institutions, and to begin to ready himself to teach his projected new courses, unaware—as was also my father—that before his course in “modern statistical theory” would get under way, Wilks would have joined the Princeton University faculty.

The program worked out for Duncan on his return to Princeton was this: He would participate as an assistant in the course, “Elementary Statistics”, taught by Professor James G. Smith (1897–1946) in the Department of Economics and Social Institutions, scheduled for the fall semester in 1933, serving as instructor in charge of the “laboratory” or “workshop” sessions in

which the students gained practical experience in graphical and tabular presentation, and in the computation of descriptive statistics, index numbers, moving averages, link relatives, etc. Then, as a sequel to this course, Duncan's new course on "modern statistical theory" would be offered by the same department during the second semester of the academic year 1933–1934.

I took these two courses in the fall of 1933 and spring of 1934, respectively. In Smith's course we used as text *Principles and Methods of Statistics* by Robert E. Chaddock (1879–1940), published by the Houghton Mifflin Company in 1925, but the scope, nature, and mode of presentation is more accurately reflected by Professor Smith's *Elementary Statistics. An Introduction to the Principles of Scientific Methods*, published the following year (New York: Henry Holt and Company, 1934). Some of R. A. Fisher's contributions to statistical methodology were alluded to, but only very briefly, as tips on recent developments that would warrant looking into, not as integral parts of the course. In Duncan's course, on the other hand, built as it was around Hotelling's lectures, and the then available mimeographed chapters of Hotelling's never published book, *Statistical Inference*, the contributions of "Student" (William Sealy Gosset) and R. A. Fisher occupied the center of the stage a large part of the time.

In the spring of 1933 a crisis developed of which I was totally unaware at the time, and the particulars of which I was not to learn until some years later. Wilks was at Cambridge University working with Wishart on the last lap of his two-year fellowship program and would be needing a permanent post, or at least a new source of income, by fall. He had sent résumés of his professional career to the universities in the United States known to have programs in probability and mathematical statistics, indicating that he was in need of an instructorship or other full-time position beginning with the academic year 1933–1934. The replies that he received were all negative—the United States was in the depth of the Depression, colleges and universities were having to make do with dramatically reduced income from endowment and other sources, and all, it seemed, were tightening the belt, and none were planning to take on additional personnel. With an exceptional training in mathematical statistics, with four substantial research papers, and two research notes already published, one joint research paper accepted for publication, and two research papers nearly ready for publication, he was one of the most promising young men in mathematical statistics and applied mathematics generally, yet he had no prospect of a job. Wilks' situation seemed hopeless and was rapidly becoming desperate. Here he was in England with his wife and son; his fellowship funds, which were never really adequate for married people, or couples with children, were about to run out; and no prospect of employment.

Hotelling, knowing full well of my father's desire to build up a program in probability and mathematical statistics at Princeton and of the need of

the College Entrance Examination Board for assistance from someone of Wilks' caliber on multivariate sampling distribution problems arising in educational testing, appealed directly to my father to take Wilks on at Princeton, stressing the long-term advantages to Princeton and the at-the-moment despatch of Wilks' situation. Thus it came to pass late in the spring of 1933 that my father, as Chairman of the Mathematics Department, offered Wilks an instructorship in the Department of Mathematics for the academic year 1933–1934, and advised him of a tentative arrangement that he had made with Professor Carl C. Brigham of the Department of Psychology and Associate Secretary of the College Entrance Examination Board (the central office of which had been at Princeton for some years) to work part-time also with the Board on problems arising in the scaling of achievement tests. It was not until many years later that I learned from my father that he had brought off this coup over the opposition of almost every member of his department. I have often wondered whether he would have been able to bring it off a year or even six months later because, although he continued as Chairman of the Mathematics Department until 1945, in mid-1933 he gave up his post as Dean of the Faculty to become Dean of the Graduate School.

Wilks arrived in Princeton in September 1933. As a new instructor in the Department of Mathematics, he found himself teaching the usual undergraduate courses in analytic geometry, calculus, and so forth during the academic year 1933–34. In addition to such teaching that first year, Sam continued his research, primarily in multivariate analysis; gave me helpful guidance in the preparation of my senior thesis on "The accuracy of computations involving quantities known only to a given degree of approximation"; and spent the remainder of his "spare time" on his "second job" with Professor Brigham and the College Entrance Examination Board. The following year, 1934–1935, Sam's program was much the same, except that he now guided my post-graduate reading and study in probability and statistical theory and methodology in preparation for my becoming a doctoral candidate in statistics under J. Neyman and E. S. Pearson at University College, London, 1935–1937.

Wilks taught his first statistics course at the University of Pennsylvania, in Philadelphia, during 1935–1936. (Dr. George Gailey Chambers, Professor of Mathematics, University of Pennsylvania, had died on 24 October 1935, shortly after his graduate course "Modern Theory of Statistical Analysis" had gotten under way. Sam was commissioned to complete the teaching of this course.) During the same period Sam gave an informal course—i.e., not listed in the official university course catalog—to three Princeton seniors, Walter W. Merrill, John O. Rohm, and William C. Shelton, on much the same material; and supervised Shelton's senior thesis on "Regression and analysis

of variance". (Shelton continued in statistics, rising to become Special Assistant to the Commissioner of Labor Statistics. Merrill and Rohm took up accounting and law, respectively.)

Wilks was promoted to an assistant professorship in 1936; and in 1936–1937 taught his first statistics courses at Princeton: a graduate course during the fall term and an undergraduate course during the spring term. A Princeton senior that year who took the graduate course, Irving E. Segal (later a Professor of Mathematics at Chicago and at MIT), wrote a senior thesis under Sam's supervision that was subsequently published in the *Proceedings of the Cambridge Philosophical Society* 34, pt 1 (Jan. 1938), pp. 41–47.

The publication, in the January 1973 issue of the *IMS Bulletin*, of Professor Harry C. Carver's letter of 14 April 1972 to Professor William Jackson Hall on the "beginnings of the *Annals*" prompts me to correct a mistaken conjecture contained therein on why Sam Wilks was not permitted to teach a course in mathematical statistics during his first few years as an instructor in the Mathematics Department there. Professor Carver wrote:

"... one day I asked [Wilks] how it was that he was not teaching a course in mathematical statistics at Princeton. He replied that he had tried to start such a course there, but his superiors turned down his request each time,—probably because mathematical statistics and probability had not yet rung a bell in the staid Eastern Colleges."

The fact of the matter is that mathematical statistics and probability *already had* "rung a bell" at Princeton: two years before Wilks' arrival, Acheson J. Duncan had been sent off at university expense to study with Professors Henry Schultz and Harold Hotelling for the express purpose of readying himself to initiate courses in "mathematical economics" and "modern statistical theory" on his return. It was this prior arrangement and commitment, not lack of appreciation of the importance of mathematical statistics and probability—or of Wilks' exceptional qualifications—that constituted the primary obstacle to Wilks' offering an undergraduate course in mathematical statistics during his first three years as a member of the Mathematics Department of Princeton University. Duncan's course on "modern statistical theory" had been scheduled to be offered for the first time during the spring term of 1934 before the possibility of Wilks' coming to Princeton had even been considered. In view of the expense that the university had incurred in underwriting Duncan's year of training in preparation for the offering of this course, and the sacrifice that Duncan had made in postponing work on his doctoral dissertation in order to acquire the requisite training at the university's request, it would have been very improper and cruel to have shelved

Duncan's course and let Wilks start one instead. I am sure that Wilks recognized this; and was also cognizant of the other factors that delayed his getting a course of his own in the Mathematics Department.

The three-year delay between Sam's arrival at Princeton and his first officially recognized course in statistics under the auspices of the Mathematics Department was the result of at least four factors.

First, there was the priority that circumstances had accorded to Duncan's course in the Department of Economics and Social Institutions. Furthermore, that department had taken the initiative in the matter, and was desirous of modernizing its outlook and course offerings with respect to mathematical economics and statistics. This department was aiming to improve its own offerings in statistics for economics students by integrating and updating the Smith-Duncan sequence of courses within that department. The extent to which this aim was achieved is evidenced by the two volumes *Fundamentals of the Theory of Statistics: Vol. 1, Elementary Statistics and Applications; Vol. 2, Sampling Statistics and Applications*, authored jointly by Professors Smith and Duncan and published by the McGraw-Hill Book Company, Inc., in 1944, 1945, respectively.)

Second, under the circumstances, any course on "mathematical statistics", "statistical analysis", "statistical inference", or whatever, to be offered by Wilks in the Mathematics Department would have to be an additional new course, and would require the approval of the all-powerful Course of Study Committee of the Faculty. A new course at Princeton had to be described in detail by the department proposing to offer it. Faculty approval gave the department the right to teach the described subject matter. I am not sure that this was an exclusive right, but I doubt that the Course of Study Committee would have approved teaching essentially the same material in two departments. Hence a major obstacle to Sam's teaching an undergraduate course in statistics was the historical fact that statistics had been the province of the Department of Economics and Social Institutions.

Third, until Sam was promoted to an assistant professorship in 1936, he was only an instructor; and in a department having the stature, nationally and internationally, of Princeton's Mathematics Department it was definitely not customary for an undergraduate, much less a graduate course, to be initiated by and be the sole responsibility of an individual with the rank of instructor.

A fourth, and very inhibiting factor was the unfavorable mathematical "climate" that prevailed in Fine Hall, the home of Princeton's Mathematics Department, during Sam's early years at Princeton. Geometry had occupied the center of the stage in this department, for over a quarter of a century, with algebra and analysis accorded much less exalted roles. Then, in 1932, the new Institute for Advanced Study, an institution completely distinct from Princeton University, had come into being, and the members of

its School of Mathematics were granted office space in the Mathematics Department's Fine Hall until the completion of their first building, Fuld Hall, in 1939. Albert Einstein (1879–1955) arrived to take up his post in the Institute during the Winter of 1933, and Hermann Weyl (1885–1955) arrived a few months earlier. John Von Neumann (1903–1957) was already there (Lecturer, 1930–1931, Princeton, then Professor of Mathematical Physics, 1931–1933; Professor of Mathematics, Institute for Advanced Study, 1933–1957); as were also E. U. Condon (1903–1961; Assistant Professor of Mathematical Physics, Princeton, 1928–1931; Associate Professor, 1931–1938, Professor, 1938–1947), and E. P. Wigner (Lecturer in Mathematical Physics, Princeton, 1930; Professor, 1930–1936; 1938–1971). With this galaxy of mathematical physicists all together in one place for the first time, the mathematical theory of relativity and quantum mechanics were definitely the fashion of the day in Fine Hall—a difficult “climate” in which to initiate a program in mathematical statistics.

By 1936–1937, the division of territory between the Department of Mathematics and the Department of Economics and Social Institutions had been resolved. The latter would be restricted to instruction in statistical theory and methods pertinent to the economic and social sciences; and the basic general undergraduate course (s) in statistical theory and methodology, and the graduate courses in advanced mathematical statistics would be the province of the Mathematics Department. As we have already said, Wilks taught his first statistics course at Princeton in the fall of 1936, the graduate course leading to his lithographed lecture notes on *Statistical Inference*—(1937); and in the spring of 1937, a sophomore course with calculus as prerequisite, quite possibly the first carefully formulated college underclass course in mathematical statistics at this level. It was offered thereafter for a number of years to students in all fields in the second half of the sophomore year. The material presented in this course, extended and polished, became generally available a decade later in his “blue book”, *Elementary Statistical Analysis* (1948). A third course, also one semester in length, was added in 1939–1940. It was an upperclass course for students who wanted to specialize in statistics, and consisted of a rather thorough mathematical treatment of statistical theory in the classroom plus a laboratory section devoted to applications and computations. This course was taken also by beginning graduate students. Wilks' first doctoral student, Joseph F. Daly received his Ph.D. in 1939. George W. Brown and Alexander M. Mood followed in 1940. World War II demolished his plans for sabbatical leave to lecture in South America and accept an offered exchange professorship for one semester at the National University in Santiago, Chile. As World War II progressed, Sam became ever more deeply involved in war research and in due course was released from academic duties entirely. Helped by two of his graduate students, T. W. Anderson and

D. F. Votaw, Jr., and Henry Scheffé, he succeeded in seeing through to litho-printed publication the graduate level text, *Mathematical Statistics* (1943), before becoming totally involved in war work. This was the forerunner of his polished comprehensive treatment bearing the same title published as a typeset book in 1962.

In keeping with my father's policy of promotions as soon as merited without regard to leave of absence, Sam was promoted to a full Professor of Mathematics in 1944, effective on his return to academic duties; and plans were laid for a Section of Mathematical Statistics within the Department of Mathematics. Following the war there was a steady flow of able graduate students and postdoctoral research associates, some of whom, like Robert Hooke and Henry Scheffé, were changing from mathematics to statistics. By the time of Sam's death (1964), Princeton had granted Ph.D.s to approximately forty students in mathematical statistics and probability, all of whom had studied to some extent with Wilks, and the dissertations of about half had been supervised by him.

A Conversation with David Blackwell

MORRIS H. DEGROOT

David Blackwell was born on April 24, 1919, in Centralia, Illinois. He entered the University of Illinois in 1935, and received his A.B. in 1938, his A.M. in 1939, and his Ph.D. in 1941, all in mathematics. He was a member of the faculty at Howard University from 1944 to 1954, and has been a Professor of Statistics at the University of California, Berkeley, since that time. He was President of the Institute of Mathematical Statistics in 1955. He has also been Vice President of the American Statistical Association, the International Statistical Institute, and the American Mathematical Society, and President of the Bernoulli Society. He is an Honorary Fellow of the Royal Statistical Society and was awarded the von Neumann Theory Prize by the Operations Research Society of America and the Institute of Management Sciences in 1979. He has received honorary degrees from the University of Illinois, Michigan State University, Southern Illinois University, and Carnegie-Mellon University.

The following conversation took place in his office at Berkeley one morning in October 1984.

“I EXPECTED TO BE AN ELEMENTARY SCHOOL TEACHER”

DeGroot: How did you originally get interested in statistics and probability?

Blackwell: I think I have been interested in the concept of probability ever since I was an undergraduate at Illinois, although there wasn't very much probability or statistics around. Doob was there but he didn't teach probability. All the probability and statistics were taught by a very nice old gentleman named Crathorne. You probably never heard of him. But he was a very good friend of Henry Rietz and, in fact, they collaborated on a college

Reprinted from *Statistical Science*, Vol. 1, No. 1, pp. 40–53, with permission of the Institute of Mathematical Statistics.

algebra book. I think I took all the courses that Crathorne taught: two undergraduate courses and one first-year graduate course. Anyway, I have been interested in the subject for a long time, but after I got my Ph.D. I didn't expect to get professionally interested in statistics.

DeGroot: But did you always intend to go on to graduate school?

Blackwell: No. When I started out in college I expected to be an elementary school teacher. But somehow I kept postponing taking those education courses. [Laughs] So I ended up getting a master's degree and then I got a fellowship to continue my work there at Illinois.

DeGroot: So your graduate work wasn't particularly in the area of statistics or probability?

Blackwell: No, except of course that I wrote my thesis under Doob in probability.

DeGroot: What was the subject of your thesis?

Blackwell: Markov chains. There wasn't very much original in it. There was one beautiful idea, which was Doob's idea and which he gave to me. The thesis was never published as such.

DeGroot: But your first couple of papers pertained to Markov chains.

Blackwell: The first couple of papers came out of my thesis, that's right.

DeGroot: So after you got your degree...

Blackwell: After I got my degree, I sort of expected to work in probability, real variables, measure theory, and such things.

DeGroot: And you *have* done a good deal of that.

Blackwell: Yes, a fair amount. But it was Abe Girshick who got me interested in statistics.

DeGroot: In Washington?

Blackwell: Yes. I was teaching at Howard and the mathematics environment was not really very stimulating, so I had to look around beyond the university just for whatever was going on in Washington that was interesting mathematically.

DeGroot: Not just statistically, but mathematically?

Blackwell: I was just looking for anything interesting in mathematics that was going on in Washington.

DeGroot: About what year would this be?

Blackwell: I went to Howard in 1944. So this would have been during the year 1944–1945.

DeGroot: Girshick was at the Department of Agriculture?

Blackwell: That's right. And I heard him give a lecture sponsored by the Washington Chapter of the American Statistical Association. That's a pretty

lively chapter. I first met George Dantzig when he gave a lecture there around that same time. His lecture had nothing to do with linear programming, by the way. In fact, I first became acquainted with the idea of a randomized test by hearing Dantzig talk about it. I think that he was the guy who invented a test function, instead of having just a rejection region that is a subset of the sample space. At one of those meetings Abe Girshick spoke on sequential analysis. Among other things, he mentioned Wald's equation.

DeGroot: That's the equation that the expectation of a sum of random variables is $E(N)$ times the expectation of an individual variable?

Blackwell: Yes. That was just such a remarkable equation that I didn't believe it. So I went home and thought I had constructed a counterexample. I mailed it to Abe, and I'm sure that he discovered the error. But he didn't write back and tell me it was an error; he just called me up and said let's talk about it. So we met for lunch and that was the start of a long and beautiful association that I had with him.

DeGroot: Would you regard the Blackwell and Girshick book (*Theory of Games and Statistical Decisions*. New York, John Wiley & Sons, 1954) as the culmination of that association?

Blackwell: Oh, that was a natural outgrowth of the association. I learned a great deal from him.

DeGroot: Were you together at any time at Stanford?

Blackwell: Yes, I spent a year at Stanford. I think it was 1950–1951. But he and I were also together at other times. We spent several months together at Rand. So we worked together in Washington, and then at Rand, and then at Stanford.

“I WROTE 105 LETTERS OF APPLICATION”

DeGroot: Tell me a little about the years between your Ph.D. from Illinois in 1941 and your arrival at Howard in 1944. You were at a few other schools in between.

Blackwell: Yes. I spent my first postdoctoral year at the Institute for Advanced Study. Again, I continued to show my interest in statistics. I sat in on Sam Wilks' course in Princeton during that year. Henry Scheffé was also sitting in on that class. He had just completed his Ph.D. at Wisconsin. Jimmie Savage was at the Institute for that year. He was at some of Wilks' lectures, too. There were a lot of statisticians about our age around Princeton at that time. Alex Mood was there. George Brown was there. Ted Anderson was there. He was in Wilks' class that year.

DeGroot: He was a graduate student?

Blackwell: He was a graduate student, just completing his Ph.D. So that was my first postdoctoral year. Also, I had a chance to meet von Neumann that year. He was a most impressive man. Of course, everybody knows that. Let me tell you a little story about him.

When I first went to the Institute, he greeted me, and we were talking, and he invited me to come around and tell him about my thesis. Well, of course, I thought that was just his way of making a new young visitor feel at home, and I had no intention of telling him about my thesis. He was a big, busy, important man. But then a couple of months later, I saw him at tea and he said, "When are you coming around to tell me about your thesis? Go in and make an appointment with my secretary." So I did, and later I went in and started telling him about my thesis. He listened for about ten minutes and asked me a couple of questions, and then he started telling *me* about *my* thesis. What you have really done is this, and probably this is true, and you could have done it in a somewhat simpler way, and so on. He was a really remarkable man. He listened to me talk about this rather obscure subject and in ten minutes he knew more about it than I did. He was extremely quick. I think he may have wasted a certain amount of time, by the way, because he was so willing to listen to second- or third-rate people and think about their problems. I saw him do that on many occasions.

DeGroot: So, from the Institute you went where?

Blackwell: I went to Southern University in Baton Rouge, Louisiana. That's a state school and at that time it was *the* state university in Louisiana for blacks. I stayed there just one year. Then the next year, I went to Clark College in Atlanta, also a black school. I stayed there for one year. Then I went to Howard University in Washington and stayed there for ten years.

DeGroot: Was Howard at a different level intellectually from these other schools?

Blackwell: Oh yes. It was the ambition of every black scholar in those days to get a job at Howard University. That was the best job you could hope for.

DeGroot: How large was the math department there in terms of faculty?

Blackwell: Let's see. There were just four regular people in the math department. Two professors. I went there as an assistant professor. And there was one instructor. That was it.

DeGroot: Have you maintained any contact with Howard through the years?

Blackwell: Oh yes. I guess the last time I gave a lecture there was about three years ago, but I visited many times during the years.

DeGroot: Do you see much change in the place through the years?

Blackwell: Yes, the math department now is a livelier place than it was when I was there. It's much bigger and the current chairman, Jim Donaldson, is very good and very active. There are some interesting things going on there.

DeGroot: Did you feel or find that discrimination against blacks affected your education or your career after your Ph.D.?

Blackwell: It never bothered me. I'll put it that way. It surely shaped my expectations from the very beginning. It never occurred to me to think about teaching in a major university since it wasn't in my horizon at all.

DeGroot: Even in your graduate-student days at Illinois?

Blackwell: That's right. I just assumed that I would get a job teaching in one of the black colleges. There were 105 black colleges at that time, and I wrote 105 letters of application.

DeGroot: And got 105 offers, I suppose.

Blackwell: No, I eventually got three offers, but I accepted the first one that I got. From Southern University.

DeGroot: Let's move a little further back in time. You grew up in Illinois?

Blackwell: In Centralia, Illinois. Did you ever get down to Centralia or that part of Illinois when you were in Chicago?

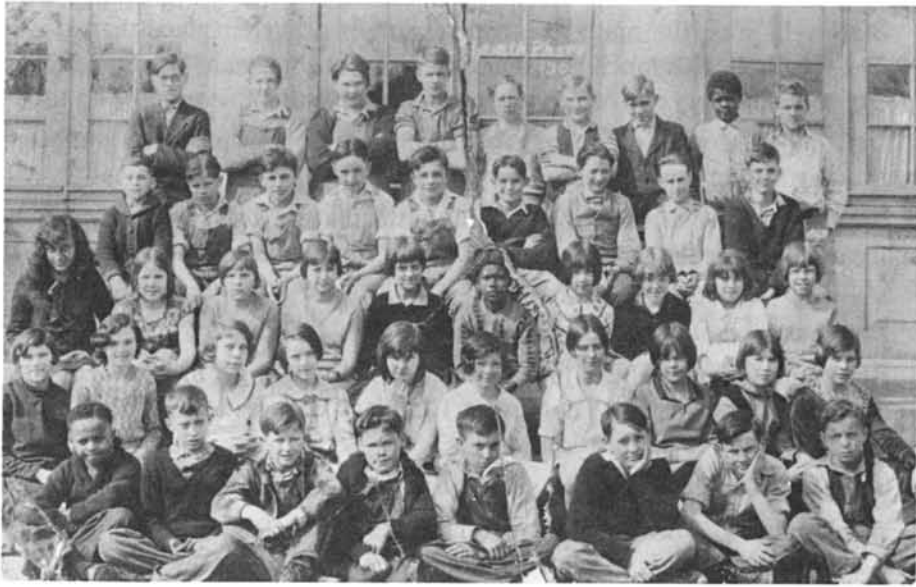
DeGroot: No, I didn't.

Blackwell: Well, it's a rather different part of the world from northern Illinois. It's quite southern. Centralia in fact was right on the border line of segregation. If you went south of Centralia to the southern tip of Illinois, the schools were completely segregated in those days. Centralia had one completely black school, one completely white school, and five "mixed" schools.

DeGroot: Well that sounds like the boundary all right. Which one did you go to?

Blackwell: I went to one of the mixed schools, because of the part of town I lived in. It's a small town. The population was about 12,000 then and it's still about 12,000. The high school had about 1,000 students. I had very good high school teachers in mathematics. One of my high school teachers organized a mathematics club and used to give us problems to work. Whenever we would come up with something that had the idea for a solution, he would write up the solution for us, and send it in our name to a journal called *School Science and Mathematics*. It was a great thrill to see your name in the magazine. I think my name got in there three times. And once my *solution* got printed. As I say, it was really Mr. Huck's write-up based on my idea. [Laughs].

DeGroot: Was your family encouraging about your education?



David Blackwell (lower left), 1930, probably sixth grade.



David Blackwell, about 1945.

Blackwell: It was just sort of assumed that I would go to college. There was no “Now be sure to study hard” or anything like that. It was just taken for granted that I was going to go to college. They were very, very supportive.

SOME FAVORITE PAPERS

DeGroot: You were quite young when you received your Ph.D. You were 21 or so?

Blackwell: 22. There wasn’t any big jump. I just sort of did everything a little faster than normal.

DeGroot: And you’ve been doing it that way ever since. You’ve published about 80 papers since that time. Do you have any favorites in that list that you particularly like or that you feel were particularly important or influential?

Blackwell: Oh, I’m sure that I do, but I’d have to look at the list and think about that. May I look?

DeGroot: Sure. This is an open-book exam.

Blackwell: Good. Let’s see... Well, my first statistical paper, called “On an equation of Wald” (*Ann. Math. Statist.* **17**, 84–87, 1946) grew out of that original conversation with Abe Girshick. That’s a paper that I am still really very proud of. It just gives me pleasant feelings every time I think about it.

DeGroot: Remind me what the main idea was.

Blackwell: For one thing it was a proof of Wald’s theorem under, I think, weaker conditions than it had been proved before, under sort of *natural* conditions. And the proof is *neat*. Let me show it to you. [Goes to blackboard.]

Suppose that X_1, X_2, \dots are i.i.d. and you have a stopping rule N , which is a random variable. You want to prove that $E(X_1 + \dots + X_N) = E(X_1)E(N)$. Well, here’s my idea. Do it over and over again. So you have stopping times N_1, N_2, \dots , and you get

$$\begin{aligned} S_1 &= X_1 + \dots + X_{N_1}, \\ S_2 &= X_{N_1+1} + \dots + X_{N_1+N_2}, \\ &\dots \end{aligned}$$

Consider $S_1 + \dots + S_k = X_1 + \dots + X_{N_1+\dots+N_k}$. We can write this equation as

$$\frac{S_1 + \dots + S_k}{k} = \left(\frac{X_1 + \dots + X_{N_1+\dots+N_k}}{N_1 + \dots + N_k} \right) \left(\frac{N_1 + \dots + N_k}{k} \right).$$

Now let $k \rightarrow \infty$. The first term on the right is a subsequence of the X averages. By the strong law of large numbers, this converges to $E(X_1)$. The second term on the right is the average of N_1, \dots, N_k . We are assuming that

they have a finite expectation, so this converges to that expectation $E(N)$. Therefore, the sequence

$$\frac{S_1 + \cdots + S_k}{k}$$

converges a.e. Then the converse of the strong law of large numbers says that the expected value of each S_i must be finite, and that

$$\frac{S_1 + \cdots + S_k}{k}$$

must converge to that expectation $E(S_1)$. Isn't that neat?

DeGroot: Beautiful, beautiful.

Blackwell: So that's the proof of Wald's equations just by invoking the strong law of large numbers and its converse. I think I like that because that was the first time that I decided that I could do something original. The papers based on my thesis were nice, but those were really Doob's ideas that I was just carrying out. But here I had a really original idea, so I was very pleased with that paper. Then I guess I like my paper with Ken Arrow and Abe Girshick, "Bayes and minimax solutions of sequential decision problems" (*Econometrica* 17, 213–244, 1949).

DeGroot: That was certainly a very influential paper.

Blackwell: That was a serious paper, yes.

DeGroot: There was some controversy about that paper, wasn't there? Wald and Wolfowitz were doing similar things at more or less the same time.

Blackwell: Yes, they had priority. There was no question about that, and I think we did give inadequate acknowledgment to them in our work. So they were very much disturbed about it, especially Wolfowitz. In fact, Wolfowitz was cool to me for more than 20 years.

DeGroot: But certainly your paper was different from theirs.

Blackwell: We had things that they didn't have, there was no doubt about that. For instance, induction backward—calculation backward—that was in our paper and I don't think there is any hint of it in their work. We did go beyond what they had done. Our paper didn't seem to bother Wald too much, but Wolfowitz was annoyed.

DeGroot: Did you know Wald very well or have much contact with him?

Blackwell: Not very well. I had just three or four conversations with him.

IMPORTANT INFLUENCES

DeGroot: I gather from what you said that Girshick was a primary influence on you in the field of statistics.

Blackwell: Oh yes.

DeGroot: Were there other people that you felt had a strong influence on you? Neyman, for example?

Blackwell: Not in my statistical thinking. Girshick was certainly *the* most important influence on me. The other person who had just one influence, but it was a very big one, was Jimmie Savage.

DeGroot: What was that one influence?

Blackwell: Well, he explained to me that the Bayes approach was the right way to do statistical inference. Let me tell you how that happened. I was at Rand, and an economist came in one day to talk to me. He said that he had a problem. They were preparing a recommendation to the Air Force on how to divide their research budget over the next five years and, in particular, they had to decide what fraction of it should be devoted to long-range research and what fraction of it should be devoted to more immediate developmental research.

“Now,” he said, “one of the things that this depends on is the probability of a major war in the next five years. If it’s large then, of course, that would shift the emphasis toward developing what we already know how to do, and if it’s small then there would be more emphasis on long-range research. I’m not going to ask you to tell me a number, but if you could give me any guide as to how I could go about finding such a number I would be grateful.” Oh, I said to him, that question just doesn’t make sense. Probability applies to a long sequence of repeatable events, and this is clearly a unique situation. The probability is either 0 or 1, but we won’t know for five years, I pontificated. [Laughs] So the economist looked at me and nodded and said, “I was afraid you were going to say that. I have spoken to several other statisticians and they have all told me the same thing. Thank you very much.” And he left.

Well, that conversation bothered me. The fellow had asked me a reasonable, serious question and I had given him a frivolous, sort of flip, answer, and I wasn’t happy. A couple of weeks later Jimmie Savage came to visit Rand, and I went in and said hello to him. I happened to mention this conversation that I had had, and then he started telling me about deFinetti and personal probability. Anyway, I walked out of his office half an hour later with a completely different view on things. I now understood what was the right way to do statistical inference.

DeGroot: What year was that?

Blackwell: About 1950, maybe 1951, somewhere around there. Looking back on it, I can see that I was emotionally and intellectually prepared for Jimmie’s message because I had been thinking in a Bayesian way about sequential analysis, hypothesis testing, and other statistical problems for some years.

DeGroot: What do you mean by thinking in a Bayesian way? In terms of prior distributions?

Blackwell: Yes.

DeGroot: Wald used them as a mathematical device.

Blackwell: That's right. It just turned out to be clearly a very natural way to think about problems and it was mathematically beautiful. I simply regretted that it didn't correspond with reality. [Laughs] But then what Jimmie was telling me was that the way that I had been thinking all the time was really the right way to think, and not to worry so much about empirical frequencies. Anyway, as I say, that was just one very big influence on me.

DeGroot: Would you say that your statistical work has mainly used the Bayesian approach since that time?

Blackwell: Yes. I simply have not worked on problems where that approach could not be used. For instance, all my work in dynamic programming just has that Bayes approach in it. That is *the* standard way of doing dynamic programming.

DeGroot: You wrote a beautiful book called *Basic Statistics* (New York, McGraw-Hill, 1970) that was really based on the Bayesian approach, but as I recall you never once mentioned the word "Bayes" in that book. Was that intentional?

Blackwell: No, it was not intentional.

DeGroot: Was it that the terminology was irrelevant to the concepts that you were trying to get across?

Blackwell: I doubt if the word "theorem" was ever mentioned in that book. That was not originally intended as a book, by the way. It was simply intended as a set of notes to give my students in connection with lectures in this elementary statistics course. But the students suggested that it should be published and a McGraw-Hill man said that he would be interested. It's just a set of notes. It's short; I think it's less than 150 pages.

DeGroot: It's beautiful. There are a lot of wonderful gems in those 150 pages.

Blackwell: Well, I enjoyed teaching the course.

DeGroot: Do you enjoy teaching from your own books?

Blackwell: No, not after a while. I think about five years after the book was published, I stopped using it. Just because I got bored with it. When you reach the point where *you're* not learning anything, then it's probably time to change something.

DeGroot: Are you working on other books at the present time?

Blackwell: No, except that I am *thinking about* writing a more elementary version of parts of your book on optimal statistical decisions because I have been using it in a course and the undergraduate students say that it's too hard.

DeGroot: Uh oh. I've been thinking of doing the same thing. [Laughs] Well, I am just thinking generally in terms of an introduction to Bayesian statistics for undergraduates.

Blackwell: Very good. I really hope you do it, Morrie. It's needed.

DeGroot: Well, I really hope you do it, too. It would be interesting. Are there courses that you particularly enjoy teaching?

Blackwell: I like the course in Bayesian statistics using your book. I like to teach game theory. I haven't taught it in some years, but I like to teach that course. I also like to teach, and I'm teaching right now, a course in information theory.

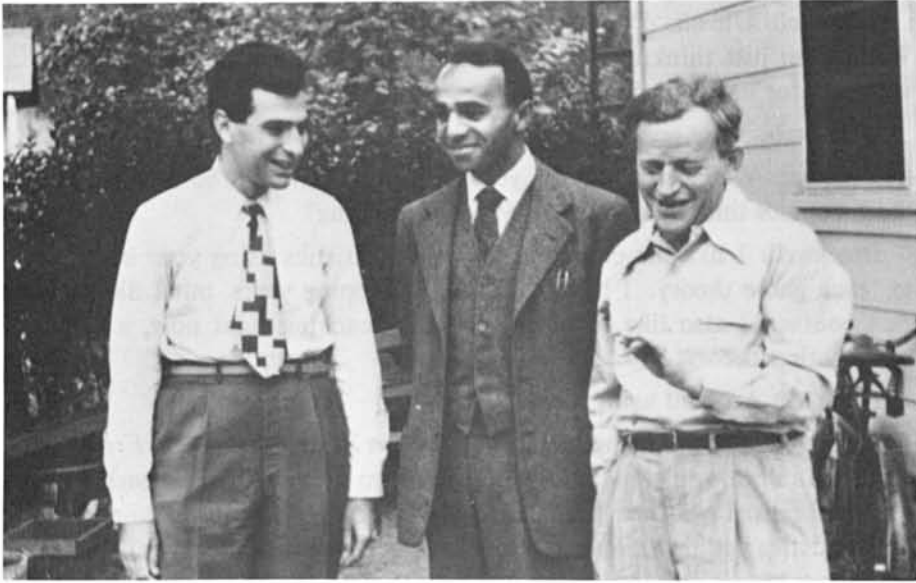
DeGroot: Are you using a text?

Blackwell: I'm not using any one book. Pat Billingsley's book *Ergodic Theory and Information* comes closest to what I'm doing. I like to teach measure theory. I regard measure theory as a kind of hobby, because to do probability and statistics you don't really need very much measure theory. But there are these fine, nit-picking points that most people ignore, and rightly so, but that I sort of like to worry about. [Laughs] I know that it is not important, but it is interesting to me to worry about regular conditional probabilities and such things. I think I'm one of only three people in our department who really takes measure theory seriously. Lester [Dubins] takes it fairly seriously, and so does Jim Pitman. But the rest of the people just sort of ignore it. [Laughs]

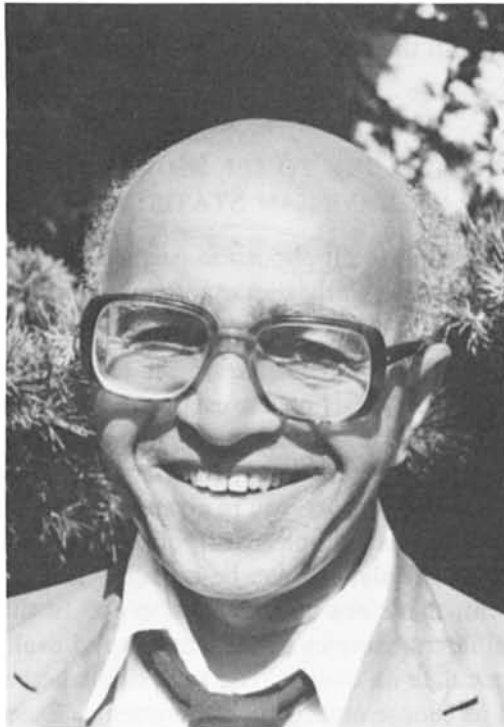
"I WOULD LIKE TO SEE MORE EMPHASIS ON BAYESIAN STATISTICS"

DeGroot: Let's talk a little bit about the current state of statistics. What areas do you think are particularly important these days? Where do you see the field going?

Blackwell: I can tell you what I'd like to see happen. First, of course, I would like to see more emphasis on Bayesian statistics. Within that area it seems to me that one promising direction which hasn't been explored at all is Bayesian experimental design. In a way, Bayesian statistics is much simpler than classical statistics in that once you're given a sample, all you have to do are calculations based on that sample. Now, of course, I say "all you have to do"—sometimes those calculations can be horrible. But if you are trying to design an experiment, that's not all you have to do. In that case, you have to look at all the different samples you might get and evaluate every one of them in order to calculate an overall risk, to decide whether the experiment is worth doing and to choose among the experiments. Except in very special situations, such as when to stop sampling, I don't think a lot of work has been done in that area.



Kenneth Arrow, David Blackwell, and M. A. Girshick,
Santa Monica, September 1948.



David Blackwell, 1984.

DeGroot: I think the reason there hasn't been very much done is because the problems are so hard. It's really hard to do explicitly the calculations that are required to find *the* optimal experiment. Do you think that perhaps the computing power that is now available would be helpful in this kind of problem?

Blackwell: That's certainly going to make a difference. Let me give you a simple example that I have never seen worked out but I am sure could be worked out. Suppose that you have two independent Bernoulli variables, say, a proportion among males and a proportion among females. They are independent, and you are interested in estimating the sum of those proportions or some linear combination of those proportions. You are going to take a sample in two stages. First of all, you can ask how large should the first sample be? And then, based on the first sample, how should you allocate proportions in the second sample?

DeGroot: Are you going to draw the first sample from the total population?

Blackwell: No. you have males and you have females, and you have a total sample effort of size N . Now you can pick some number $n \leq N$ to be your sample size. And you can allocate those n observations among males and females. Then based on how that sample comes out, you can allocate your second sample. What is the best initial allocation, and how much better is it than just doing it all in one stage? Well, I haven't done that calculation but I'm sure that it can be done. It would be an interesting kind of thing and it could be extended to more than two categories. That's an example of the sort of thing on which I would like to see a lot of work done—Bayesian experimental design.

One of the things that I worry about a little is that I don't see theoretical statisticians having as much contact with people in other areas as I would like to see. I notice here at Berkeley, for example, that the people in Operations Research seem to have much closer contact with industry than the people in our department do. I think we might find more interesting problems if we did have closer contact.

DeGroot: Do you think that the distinctions between applied and theoretical statistics are still as rigid as they were years ago or do you think that the field is blending more into a unified field of statistics in which such distinctions are not particularly meaningful? I see the emphasis on data analysis which is coming about, and the development of theory for data analysis and so on, blurring these distinctions between theoretical and applied statistics in a healthy way.

Blackwell: I guess I'm not familiar enough with data analysis and what computers have done to have any interesting comments on that. I see what

some of our people and people at Stanford are doing in looking at large-dimensional data sets and rotating them so that you can see lots of three-dimensional projections and such things, but I don't know whether that suggests interesting theoretical questions or not. Maybe that's not important, whether it suggests interesting theoretical questions. Maybe the important thing is that it helps contribute to the solution of practical problems.

INFINITE GAMES

DeGroot: What kind of things are you working on these days?

Blackwell: Right now I am working on some things in information theory, and still trying to understand some things about infinite games of perfect information.

DeGroot: What do you mean by an infinite game?

Blackwell: A game with an infinite number of moves. Here's an example. I write down a 0 or a 1, and you write down a 0 or a 1, and we keep going indefinitely. If the sequence we produce has a limiting frequency, I win. If not, you win. That's a trivial game because I can force it to have a limiting frequency just by doing the opposite of whatever you do. But that's a simple example of an infinite game.

DeGroot: Fortunately, it's one in which I'll never have to pay off to you.

Blackwell: Well, we can play it in such a way that you would have to pay off.

DeGroot: How do we do that?

Blackwell: You must specify a strategy. Let me give you an example. You know how to play chess in just one move: You prepare a complete set of instructions so that for every situation on the chess board you specify a possible response. Your one move is to prepare that complete set of instructions. If you have a complete set and I have a complete set, then we can just play the game out according to those instructions. It's just one move. So in the same way, you can specify a strategy in this infinite game. For every finite sequence that you might see up to a given time as past history, you specify your next move. So you can define this function once and for all, and I can define a function, and then we can mathematically assess those functions. I can prove that there is a specific function of mine such that no matter what function you specify, the set will have a limiting frequency.

DeGroot: So you could extract money from me in a finite amount of time.
[Laughs]

Blackwell: Right. Anyway it's been proved that all such infinite games with Borel payoffs are determined, and I've been trying to understand the

proof for several years now. I'm still working on it, hoping to understand it and simplify it.

DeGroot: Have you published papers on that topic?

Blackwell: Just one paper many years ago. Let me remind myself of the title [checking his files], "Infinite games and analytic sets" (*Proc. Natl. Acad. Sci. U.S.A.* **58**, 1836–1837, 1967). This is the only paper I've published on infinite games; and that's one of my papers that I like very much, by the way. It's an application of games to prove a theorem in topology. I sort of like the idea of connecting those two apparently not closely related fields.

DeGroot: Have you been involved in applied projects or applied problems through the years, at Rand or elsewhere, that you have found interesting and that have stimulated research of your own?

Blackwell: I guess so. My impression though is this: When I have looked at real problems, interesting theorems have sometimes come out of it. But never anything that was helpful to the person who had the problem. [Laughs]

DeGroot: But possibly to somebody else at another time.

Blackwell: Well, my work on comparison of experiments was stimulated by some work by Bohnenblust, Sherman, and Shapley. We were all at Rand. They called their original paper "Comparison of reconnaissances," and it was *classified* because it arose out of some question that somebody had asked them. I recognized a relation between what they were doing and sufficient statistics, and proved that they were the same in a special case. Anyway, that led to this development which I think is interesting theoretically, and to which you have contributed.

DeGroot: Well, I have certainly used your work in that area. And it has spread into diverse other areas. It is used in economics in comparing distributions of income, and I used it in some work on comparing probability forecasters.

Blackwell: And apparently people in accounting have made some use of these ideas. But anyway, as I say, nothing that I have done has ever helped the person who raised the question. But there is no doubt in my mind that you do get interesting problems by looking at the real world.

"I DON'T HAVE ANY DIFFICULTIES WITH RANDOMIZATION"

DeGroot: One of the interesting topics that comes out of a Bayesian view of statistics is the notion of randomization and the role that it should play in statistics. Just this little example you were talking about before with two proportions made me think about that. We just assume that we are drawing the observations at random from within each subpopulation in that example,

but perhaps basically because we don't have much choice. Do you have any thoughts about whether one should be drawing observations at random?

Blackwell: I don't have any difficulties with randomization. I think it's probably a good idea. The strict theoretical idealized Bayesian would of course never need to randomize. But randomization probably protects us against our own biases. There are just lots of ways in which people differ from the ideal Bayesian. I guess the ideal Bayesian, for example, could not think about a theorem as being probably true. For him, presumably, all true theorems have probability 1 and all false ones have probability 0. But you and I know that's not the way we think. I think of randomization as being a protection against your own imperfect thinking.

DeGroot: It is also to some extent a protection against others. Protection for you as a statistician in presenting your work to the scientific community, in the sense that they can have more belief in your conclusions if you use some randomization procedure rather than your own selection of a sample. So I see it as involved with the sociology of science in some way.

Blackwell: Yes, that's an important virtue of randomization. That reminds me of something else though. We tend to think of evidence as being valid only when it comes from random samples or samples selected in a probabilistically specified way. That's wrong, in my view. Most of what we have learned, we have learned just by observing what happens to come along, rather than from carefully controlled experiments. Sometimes statisticians have made a mistake in throwing away experiments because they were not properly controlled. That is not to say that randomization isn't a good idea, but it is to say that you should not reject data just because they have been obtained under uncontrolled conditions.

DeGroot: You were the Rouse Ball Lecturer at Cambridge in 1974. How did that come about and what did it involve?

Blackwell: Well, I was in England for two years, 1973–1975, as the director of the education-abroad program in Great Britain and Ireland for the University of California. I think that award was just either Peter Whittle's or David Kendall's idea of how to get me to come up to Cambridge to give a lecture. One of the things which delighted me was that it was named the Rouse Ball Lecture because it gave me an opportunity to say something at Cambridge that I liked—namely, that I had heard of Rouse Ball long before I had heard of Cambridge. [Laughs]

DeGroot: Well, tell me about Rouse Ball.

Blackwell: He wrote a book called *Mathematical Recreations and Essays*. You may have seen the book. I first came across it when I was a high school student. It was one of the few mathematics books in our library. I was fascinated by that book. I can still picture it. Rouse Ball was a 19th century

mathematician, I think. [Walter William Rouse Ball, 1850–1925] Anyway, this is a lectureship that they have named after him.

DeGroot: I guess there aren't too many Bayesians on the statistics faculty here at Berkeley.

Blackwell: No. I'd say, Lester and I are the only ones in our department. Of course, over in Operations Research, Dick Barlow and Bill Jewell are certainly sympathetic to the Bayesian approach.

DeGroot: Is it a topic that gets discussed much?

Blackwell: Not really, It used to be discussed here but you very soon discover that it's sort of like religion; that it has an appeal for some people and not for other people, and you're not going to change anybody's mind by discussing it. So people just go their own ways. What has happened to Bayesian statistics surprised me. I expected it either to catch on and just sweep the field or to die. And I was rather confident that it would die. Even though to me it was the right way to think, I just didn't think that it would have a chance to survive. But I thought that if it did, then it would sweep things. Of course, neither one of those things has happened. Sort of a steady 5-10% of all the work in statistical inference is done from a Bayesian point of view. Is that what you would have expected 20 years ago?

DeGroot: No, it certainly doesn't seem as though that would be a stable equilibrium. And maybe the system is still not in equilibrium. I see the Bayesian approach growing, but it certainly is not sweeping the field by any means.

Blackwell: I'm glad to hear that you see it growing.

DeGroot: Well, there seem to be more and more meetings of the Bayesians, anyway. The actuarial group that met here at Berkeley over the last couple of days to discuss credibility theory seems to be a group that just naturally accepts the Bayesian approach in their work in the real world. So there seem to be some pockets of users out there in the world, and I think maybe that's what has kept the Bayesian approach alive.

Blackwell: There's no question in my mind that if the Bayesian approach does grow in the statistical world it will not be because of the influence of other statisticians but because of the influence of actuaries, engineers, business people, and others who actually like the Bayesian approach and use it.

DeGroot: Do you get a chance to talk much to researchers outside of statistics on campus, researchers in substantive areas?

Blackwell: No, I talk mainly to people in Operations Research and Mathematics, and occasionally Electrical Engineering. But the things in Electrical Engineering are theoretical and abstract.

“THE WORD ‘SCIENCE’ IN THE TITLE BOTHERS
ME A LITTLE”

DeGroot: What do you think about the idea of this new journal, *Statistical Science*, in which this conversation will appear? I have the impression that you think the IMS is a good organization doing useful things, and there is really no need to mess with it.

Blackwell: That is the way I feel. On the other hand, I must say that I felt exactly the same way about splitting the *Annals of Mathematical Statistics* into two journals, and that split seems to be working. So I’m hoping that the new journal will add something. I guess the word “science” in the title bothers me a little. It’s not clear what the word is intended to convey there, and you sort of have the feeling that it’s there more to contribute a tone than anything else.

DeGroot: My impression is that it *is* intended to contribute a tone. To give a flavor of something broader than just what we would think of as theoretical statistics. That is, to reach out and talk about the impact of statistics on the sciences and the interrelationship of statistics with the sciences, all kinds of sciences.

Blackwell: Now I’m all in favor of that. For example, the relation of statistics to the law is to me a quite appropriate topic for articles in this journal. But somehow calling it “science” doesn’t emphasize that direction. In fact, it rather suggests that that’s *not* the direction. It sounds as though it’s tied in with things that are supported by the National Science Foundation and to me that restricts it.

DeGroot: The intention of that title was to convey a broad impression rather than a restricted one. To give a broader impression than just statistics and probability, to convey an applied flavor and to suggest links to all areas.

Blackwell: Yes. It’s analogous to computer science, I guess. I think *that* term was rather deliberately chosen. My feeling is that the IMS is just a beautiful organization. It’s about the right size. It’s been successful for a good many years. I don’t like to see us become ambitious. I like the idea of just sort of staying the way we are, an organization run essentially by amateurs.

DeGroot: Do you have the feeling that the field of statistics is moving away from the IMS in any way? That was one of the motivations for starting this journal.

Blackwell: Well, of course, statistics has always been substantially bigger than the IMS. But you’re suggesting that the IMS represents a smaller and smaller fraction of statistical activity.

DeGroot: Yes, I think that might be right.

Blackwell: You know, Morrie, I see what you're talking about happening in mathematics. It's less and less true that all mathematics is done in mathematics departments. On the Berkeley campus, I see lots of interesting mathematics being done in our department, in Operations Research, in Electrical Engineering, in Mechanical Engineering, some in Business Administration, a lot in the Economics Department by Gerard Debreu and his colleagues; a lot of really interesting, high class mathematics is being done outside mathematics departments. What you're suggesting is that statistics departments and the journals in which they publish are not necessarily the centers of statistics the way they used to be, that a lot of work is being done outside. I'm sure that's right.

DeGroot: And perhaps *should* be done outside statistics departments. That used to be an unhealthy sign in the field, and we worked hard in statistics departments to collect up the statistics that was being done around the campus. But I think now that the field has grown and matured, that it is probably a healthy thing to have some interesting statistics being done outside.

Blackwell: Yes. Consider the old problem of pattern recognition. That's a statistical problem. But to the extent that it gets solved, it's not going to be solved by people in statistics departments. It's going to be solved by people working for banks and people working for other organizations who really need to have a device that can look at a person and recognize him in lots of different configurations. That's just one example of the cases where we're somehow too narrow to work on a lot of serious statistical problems.

DeGroot: I think that's right, and yet we have something important to contribute to those problems.

Blackwell: I would say that we *are* contributing, but indirectly. That is, people who are working on the problems have studied statistics. It seems to me that a lot of the engineers I talk to are very familiar with the basic concepts of decision theory. They know about loss functions and minimizing expected risks and such things. So, we have contributed, but just indirectly.

DeGroot: You are in the National Academy of Sciences. . .

Blackwell: Yes, but I'm very inactive.

DeGroot: You haven't been involved in any of their committees or panels?

Blackwell: No, and I'm not sure that I would want to be. I guess I don't like the idea of an official committee making scientific pronouncements. I like people to form opinions about scientific matters just on the basis of listening to individual scientists. To have one group with such overwhelming prestige bothers me a little.

DeGroot: And it is precisely the prestige of the Academy that they rely on when reports get issued by these committees.

Blackwell: Yes. So I think it's just great as a purely honorific organization, so to speak. To meet just once a year, and elect people more or less at random. I think everybody that's in it has done something reasonable and even pretty good, in fact. But on the other hand, there are at least as many people *not* in it who have done good things as there are in it. It's kind of a random selection process.

DeGroot: So you think it's a good organization as long as it doesn't do anything.

Blackwell: Right. I'm proud to be in it, but I haven't been active. It's sort of like getting elected to Phi Beta Kappa—it's nice if it happens to you...

"I PLAY WITH THIS COMPUTER"

DeGroot: Do you feel any relationship between your professional work and the rest of your life, your interests outside of statistics? Is there any influence of the outside on what you do professionally, or are they just sort of separate parts of your life?

Blackwell: Separate, except my friends are also my colleagues. It's only through the people with whom I associate outside that there's any connection. It's hard to think of any other real connection.

DeGroot: It's not obvious what these connections might be for anyone. One's political views or social views seem to be pretty much independent of the technical problems we work on.

Blackwell: Yes. Although it's hard to see how it could *not* have an influence, isn't it? I guess my life seems all of a piece to me but yet it's hard to see where the connections are. [Laughs]

DeGroot: What do you see for your future?

Blackwell: Well, just gradually to wind down, gracefully I hope. I expect to get more interested in computing. I have a little computer at home, and it's a lot of fun just to play with it. In fact, I'd say that I play with this computer here in my office at least as much as I do serious work with it.

DeGroot: What do you mean by play?

Blackwell: Let me give you an example. You know the algorithm for calculating square roots. You start with a guess and then you divide the number by your guess and take the average of the two. That's your next guess. That's actually Newton's method for finding square roots, and it works very well. Sometimes doing statistical work, you want to take the square root of a positive definite matrix. It occurred to me to ask whether that algorithm works for finding the square root of a positive definite matrix. Before I got interested in computing, I would have tried to solve it theoretically. But what

did I do? I just wrote up a program and put it on the computer to see if it worked. [Goes to blackboard]

Suppose that you are given the matrix M and want to find $M^{1/2}$. Let G be your guess of $M^{1/2}$. Then your new guess is $1/2(G + MG^{-1})$. You just iterate this and see if it converges to $M^{1/2}$. Now, Morrie, I want to show you what happens. [Goes to terminal]

Let's do it for a 3×3 matrix. We're going to find the square root of a positive definite 3×3 matrix. Now, if you happen to have in mind a particular 3×3 positive definite matrix whose square root you want, you could enter it directly. I don't happen to have one in mind, but I do know a theorem: If you take any nonsingular 3×3 matrix A , then AA' is going to be positive definite. So I'm just going to enter any 3×3 nonsingular matrix [putting some numbers into the terminal] and let $M = AA'$. Now, to see how far off your guess G is at any stage, you calculate the Euclidean norm of the 3×3 matrix $M - G^2$. That's what I call the error. Let's start out with the identity matrix I as our initial guess. We get a big error, 29 million. Now let's iterate. Now the error has dropped down to 7 million. It's going to keep being divided by 4 for a long time. [Continuing the iterations for a while] Now notice, we're not bad. There's our guess, there's its square, there's what we're trying to get. It's pretty close. In fact the error is less than one. [Continuing] Now the error is really small. Look at that, isn't that beautiful? So there's just no question about it. If you enter a matrix at random and it works, then that sort of settles it.

But now wait a minute, the story isn't quite finished yet. Let me just continue these iterations... Look at that! The error got bigger, and it keeps getting bigger. [Continuing] Isn't that lovely stuff?

DeGroot: What happened?

Blackwell: Isn't that an interesting question, what happened? Well, let me tell you what happened. Now you can study it theoretically and ask, should it converge? And it turns out that it will converge if, and essentially only if, your first guess commutes with the matrix M . That's what the theory gives you. Well, my first guess was I . It commutes with everything. So the procedure theoretically converges. However, when you calculate, you get round-off errors. By the way, if your first guess commutes, then all subsequent guesses will commute. However, because of round-off errors, the matrices that you actually get don't quite commute. There are two ways to do this. We could take MG^{-1} or we could have taken $G^{-1}M$. Of course, if M commutes with G , then it commutes with G^{-1} and it doesn't matter which way you do it. But if you don't calculate G exactly at some stage, then it will not quite commute. And in fact, what I have here on the computer is a calculation at each stage of the noncommutativity norm. That shows you how different MG^{-1} is from $G^{-1}M$. I didn't point those values out to you, but they started

out as essentially 0, and then there was a 1 in the 15th place, and then a 1 in the 14th place, and so on. By this stage, the noncommutativity norm has built up to the point where it's having a sizable influence on the thing.

DeGroot: Is it going to diverge or will it come back down after some time?

Blackwell: It won't come back down. It will reach a certain size, and sometimes it will stay there and sometimes it will oscillate. That is, one G will go into a quite different G , but then that G will come back to the first one. You get periods, neither one of them near the truth. So that's what I mean by just playing, instead of sitting down like a serious mathematician and trying to prove a theorem. Just try it out on the computer and see if it works. [Laughs]

DeGroot: You can save a lot of time and trouble that way.

Blackwell: Yes. I expect to do more and more of that kind of playing. Maybe I get lazier as I get older. It's fun, and it's an interesting toy.

DeGroot: Do you find yourself growing less rigorous in your mathematical work?

Blackwell: Oh yes. I'm much more interested in the ideas, and in truth under not-completely-specified hypotheses. I think that has happened to me over the last 20 years. I can certainly notice it now. Jim MacQueen was telling me about something that he had discovered. If you take a vector and calculate the squared correlation between that vector and some permutation of itself, then the average of that squared correlation over all possible permutations is some simple number. Also, there was some extension of this result to k vectors. He has an interesting algebraic identity. He told me about it, but instead of my trying to prove it, I just selected some numbers at random and checked it on the computer. Also, I had a conjecture that some stronger result was true. I checked it for some numbers selected at random and it turned out to be true for him and *not* true for what I had said. Well, that just settles it. Because suppose you have an algebraic function $f(x_1, \dots, x_n)$ and you want to find out if it is identically 0. Well, I think it's true that any algebraic function of n variables is either identically 0 or the set of x 's for which it is 0 is a set that has measure 0. So you can just select x 's at random and evaluate f . If you get 0, it's identically 0. [Laughs]

DeGroot: You wouldn't try even a second set of x 's?

Blackwell: I did. [Laughs]

DeGroot: Getting more conservative in your old age.

Blackwell: Yes. [Laughs] I've been wondering whether in teaching statistics the typical set-up will be a lot of terminals connected to be a big central computer or a lot of small personal computers. Let me turn the interview around. Do you have any thoughts about which way that is going or which way it ought to go?

DeGroot: No, I don't know. At Carnegie-Mellon we are trying to have both worlds by having personal computers but having them networked with each other. There's a plan at Carnegie-Mellon that each student will have to have a personal computer.

Blackwell: Now when you say each student will have to have a personal computer, where will it be physically located?

DeGroot: Wherever he lives.

Blackwell: So that they would not actually use computers in class on the campus?

DeGroot: Well, this will certainly lessen the burden on the computers that are on campus, but in a class you would have to have either terminals or personal computers for them.

Blackwell: Yes. I'm pretty sure that in our department in five years we'll have several classrooms in which each seat will be a work station for a student, and in front of him will be either a personal computer or a terminal. I'm not sure which, but that's the way we're going to be in five years.

"I WOULDN'T DREAM OF TALKING ABOUT A THEOREM LIKE THAT NOW"

DeGroot: A lot of people have seen you lecture on film. I know of at least one film you made for the American Mathematical Society that I've seen a few times. That's a beautiful film, "Guessing at Random."

Blackwell: Yes. I now, of course, don't think much of those ideas. [Laughs]

DeGroot: There were some *minimax* ideas in there...

Blackwell: Yes, that's right. That was some work that I did before I became such a committed Bayesian. I wouldn't dream of talking about a theorem like that now. But it's a nice result...

DeGroot: It's a nice result and it's a beautiful film. Delivered so well.

Blackwell: Let's see... How does it go? If I were doing it now I would do a weaker and easier Bayesian form of the theorem. You were given an arbitrary sequence of 0's and 1's, and you were going to observe successive values and you had to predict the next one. I proved certain theorems about how well you could do against every possible sequence. Well, *now* I would say that you have a probability distribution on the set of all sequences. It's a general fact that if you're a Bayesian, you don't have to be clever. You just *calculate*. Suppose that somebody generates an arbitrary sequence of 0's and 1's and it's your job after seeing each finite segment to predict the next coordinate, 0 or 1, and we keep track of how well you do. Then I have to be clever and invoke the minimax theorem to devise a procedure that asymptotically does very well in a certain sense. But now if you just put a prior distribution on

the set of sequences, any Bayesian knows what to do. You just calculate the probability of the next term being a 1 given the past history. If it's more than $1/2$ you predict a 1, if it's less than $1/2$ you predict a 0. And that simple procedure has the corresponding Bayesian version of all the things that I talked about in that film. You just know what is the right thing to do.

DeGroot: But how do you know that you'll be doing well in relation to the reality of the sequence?

Blackwell: Well, the theorem of course says that you'll do well for all sequences except a set of measure zero according to your own prior distribution, and that's all a Bayesian can hope for. That is, you have to give up something, but it just makes life so much *neater*. You just know that this is the right thing to do.

I encountered the same phenomenon in information theory. There is a very good theory about how to transmit over a channel, or how to transmit over a sequence of channels. The channel may change from day to day, but if you know what it is every day, then you can transmit over it. Now suppose that the channel varies in an arbitrary way. That is, you have one of a finite set of channels, and every day you're going to be faced with one of these channels. You have to put in the input and a guy at the other end gets an output. The question is, how well can you do against all possible channel sequences?

You don't really know what the weather is out there, so you don't know what the interference is going to be. But you want to have a code that transmits well for all possible weather sequences. If you just analyze the problem crudely, it turns out that you can't do *anything* against all possible sequences. However, if you select the code in a certain random way, your overall error probability will be small for each weather sequence. So you see, it's a nice theoretical result but it's unappealing. However, you can get exactly the same result if you just put a probability distribution on the sequences. Well, the weather could be any sequence, but you expect it to be sort of this way or that. Once you put a probability distribution on the set of sequences, you no longer need random codes. And there is a deterministic code that gives you that same result that you got before. So either you must behave in a random way, or you must put a probability distribution on nature.

[Looking over a copy of his paper, BLACKWELL, D., BREIMAN, L. and THOMASIAN, A. J., "The capacities of certain channel classes under random coding," *Ann. Math. Statist.* **31**, 558–567, 1960] I don't think we did the nice easy part. We behaved the way Wald behaved. You see, the minimax theorem says that if for every prior distribution you can achieve a certain gain, then there is a random way of behaving that achieves that gain for every parameter value. You don't need the prior distribution; you can throw it away. Well, I'm afraid that in this paper, we invoked the minimax theorem. We said,

take any prior distribution on the set of channel sequences. Then you can achieve a certain rate of transmission for that prior distribution. Now you invoke the minimax theorem and say, therefore, there is a randomized way of behaving which enables you to achieve that rate against every possible sequence. I now wish that we had *stopped* at the earlier point. [Laughs] For us, the Bayesian analysis was just a preliminary which, with the aid of the minimax theorem, enabled us to reach the conclusions we were seeking. That was Wald's view and that's the view that we took in that paper. I'm sure I was already convinced that the Bayes approach was the right approach, but perhaps I deferred to my colleagues.

DeGroot: That's a very mild compromise. Going *beyond* what was necessary for a Bayesian resolution of the problem.

Blackwell: That's right. Also, I suspect that I had Wolfowitz in mind. He was a real expert in information theory, but he wouldn't have been interested in anything Bayesian.

DeGroot: What about the problem of putting prior distributions on spaces of infinite sequences or function spaces? Is that a practical problem and is there a practical solution to the problem?

Blackwell: I wouldn't say for infinite sequences, but I think it's a very important practical problem for large finite sequences and I have no idea how to solve it. For example, you could think that the pattern recognition problem that I was talking about before is like that. You see an image on a TV screen. That's just a long finite sequence of 0's and 1's. And now you can ask how likely it is that that sequence of 0's and 1's is intended to be the figure 7, say. Well, with some you're certain that it is and some you're certain that it isn't, and with others there's a certain probability that it is and a probability that it isn't. The problem of describing that probability distribution is a very important problem. And we're just not close to knowing how to describe probability distributions over long finite sequences that correspond to our opinions.

DeGroot: Is there hope for getting such descriptions?

Blackwell: I don't know. But again it's a statistical problem that is not going to be solved by professors of statistics in universities. It might be solved by people in artificial intelligence, or by researchers outside universities.

“JUST TELL ME ONE OR TWO INTERESTING THINGS”

DeGroot: There's an argument that says that under the Bayesian approach, you have to seek the optimal decision and that's often just too hard to find. Why not settle for some other approach that requires much less structure and get a reasonably good answer out of it, rather than an optimal answer?

Especially in these kinds of problems where we don't know how to find the optimal answer.

Blackwell: Oh, I think everybody would be satisfied with a reasonable answer. I don't see that there's more of an emphasis in the Bayesian approach on optimal decisions than in other approaches. I separate Bayesian inference from Bayesian decision. the inference problem is just calculating a posterior distribution, and that has nothing to do with the particular decision that you're going to make. The same posterior distribution could be used by many different people making different decisions. Even in calculating the posterior distribution, there is a lot of approximation. It just can't be done precisely in interesting and important cases. And I don't think anybody who is interested in applying Bayes method would insist on something that's precise to the fifth decimal place. That's just the conceptual framework in which you want to work, and which you want to approximate.

DeGroot: That same spirit can be carried over into the decision problem, too. If you can't find the optimum decision, you settle for an approximation to it.

Blackwell: Right.

DeGroot: In your opinion, what have been the major breakthroughs in the field of statistics or probability through the years?

Blackwell: It's hard to say... I think that theoretical statistical thinking was just completely dominated by Wald's ideas for a long time. Charles Stein's discovery that \bar{X} is inadmissible was certainly important. Herb Robbins's work on empirical Bayes was also a big step, but possibly in the wrong direction.

You know, I don't view myself as a statesman or a guy with a broad view of the field or anything like that. I just picked directions that interested me and worked in them. And I have had fun.

DeGroot: Well, despite the fact that you didn't choose the problems for their impact or because of their importance, a lot of people have gained a lot from your work.

Blackwell: I guess that's the way scholars *should* work. Don't worry about the overall importance of the problem; work on it if it looks interesting. I think there's probably a sufficient correlation between interest and importance.

DeGroot: One component of the interest is probably that others are interested in it, anyway.

Blackwell: That's a big component. You want to tell somebody about it after you've done it.

DeGroot: It has not always been clear that the published papers in our more abstract journals did succeed in telling anybody about it.

Blackwell: That's true. But if you get the fellow to give a lecture on it, he'll probably be able to tell you something about it. Especially if you try to restrict him: Look, don't tell me everything. Just tell me *one or two* interesting things.

DeGroot: You have a reputation as one of the finest lecturers in the field. Is that your style of lecturing?

Blackwell: I guess it is. I try to emphasize that with students. I notice that when students are talking about their theses or about their work, they want to tell you everything they know. So I say to them: You know much more about this topic than anybody else. We'll never understand it if you tell it all to us. Pick just one interesting thing. Maybe two.

DeGroot: Thank you, David.

A native of Canada, Cecil J. Nesbitt did his undergraduate and graduate work at the University of Toronto, where he received his Ph.D. in 1937 as a student of Richard Brauer. After a postdoctoral year at the Institute for Advanced Study, he took a position at the University of Michigan and remained there until his retirement in 1980. His early research was in algebra, but at both Toronto and Michigan his primary bent was to actuarial mathematics. With Carl H. Fischer he led a flourishing actuarial program in the Mathematics Department at Michigan, while publishing actively and serving the Society of Actuaries in various capacities. His main work has been in the areas of pension funding and social insurance.

Personal Reflections on Actuarial Science in North America from 1900

CECIL J. NESBITT

1. INTRODUCTION

At the outset, it should be made clear that this article does not pretend to be a definitive history of actuarial science developments in North America since the beginning of the century. Deadlines, and my own available time and energy, do not permit such an undertaking, worthy as it may be. Instead I shall draw on memories of almost 60 years as an actuarial student, teacher, practitioner, and researcher, to indicate actuarial highlights of that period, and also sources for further review if readers become so inclined. Such readers should turn first to *Actuarial Mathematics* (Proc. Symp. Appl. Math. 35, 1986) and peruse it alongside this article to gain detailed, introductory overviews of the diverse actuarial models that will be mentioned here. The non-exhaustive list of references at the end of the article is selected to aid such review by pointing the way to other more complete lists in regard to various topics mentioned herein. The body of ideas, known and unknown,

is infinite, and even in one special area, such as the intersection of actuarial science and mathematics, can be covered only by broad strokes.

Actuarial science has a major role in the guidance of financial security systems, developed to protect individuals and groups against a multiplicity of risks such as impairment of health, premature death, destruction of property, and extended old age. The systems may range from self-insured groups to national programs of social security. Some of these systems operate on an international basis, and more such development may lie in the future. These systems during my life have made much progress despite economic, financial, and political disturbances and disasters. The systems have been facing fast-growing environmental hazards, and military potentialities of incredible magnitude. Actuarial science has a role to play, as do all fields, in finding viable equilibria in a fast-changing world.

In the following section, there will be brief discussion of the main fields of knowledge on which actuarial science draws. Those to be mentioned are mathematics, statistics, probability, accounting, computer science, demography, economics, finance theory, law and medical science. Some of these fields were relatively undeveloped at the beginning of this century. The section on sources will be followed by one on distinctively (although not exclusively) actuarial theories. These are: estimation of mortality and other rates, life tables (now broadened to survival models), graduation theory, risk theory, credibility theory, actuarial finance theory, life insurance mathematics including growth models and stochastic models of life contingencies. The application of these theories to various fields of practice will come next, with a final summary overview.

2. SOURCES OF ACTUARIAL SCIENCE

In a broad sense, all portions of actuarial science relate to some form of mathematical theory or application. The theory may be relatively elementary, but the application may be extremely detailed and numerical. Of prime importance are the actuarial assumptions from which the mathematical model is developed. For short-term insurances, there may be a large volume of current data for statistical and probability analysis. Such current data may also be available for long-term insurances, as for example, whole life insurance, pension systems, and social security, but must be extended by projection factors to guide the future growth of the financial security system.

From the data analysis, one may estimate probabilities needed for the model of the system. The mathematical model may draw heavily on probability theory for its structure, or for the longer term it may be deterministic in character, following out the consequences of assumed rates of growth and eligibility for benefits. Statistical and probability theories, which have grown rapidly in this century, are playing an expanding role in actuarial science.

Another main source of actuarial science is the mathematics of finance. Until recent years, this has been an elementary theory, defining various discrete and continuous rates of interest, and utilizing a constant rate compound interest model to calculate present and accumulated values of series of payments. Still more recently, the turbulence of financial markets, has led to simulation studies being conducted by committees of the Society of Actuaries under the leadership of C. L. Trowbridge and D. D. Cody (Cody, 1987). About the same time there appeared Phelim Boyle's "Immunization under stochastic models of the term structure" (Boyle, 1978). Also, finance theory, with application to the pricing of options, has been advancing strongly. (For a comprehensive view of this last work, see Pedersen, Shiu and Thorlacius, forthcoming, and D'Arcy and Doherty, 1988.)

Computer science has greatly empowered actuaries in regard to: estimation of rates or probabilities, the calculation of premiums or contributions, the projection of future benefit outgo and of premium or contribution income, and the corresponding accumulation of reserves. A notable example is provided by the annual actuarial projections for old-age, survivors and disability insurance in the United States (see Andrews and Beekman, 1987).

Other bodies of knowledge or practice which impinge on actuarial practice are indicated by the examples below:

Accounting. To get a feel for some of the discussion preceding (Financial Accounting Standards No. 87, 1985), see E. L. Hicks and C. L. Trowbridge, *Employer accounting for pensions* (Hicks and Trowbridge, 1985).

Demography. A major actuarial concern here is in regard to the development of national life tables. An early reference was H. H. Wolfenden's *Population statistics and their compilation* (Wolfenden, 1925). This was followed by M. Spiegelman's *Introduction to demography* (Spiegelman, 1955). One of the current references in the actuarial education syllabus is *Demography through problems* (Keyfitz and Beekman, 1984). See also A. Wade's *Social security area population projections* (Wade, 1988) and J. Wilkins' *OASDI long-range beneficiary projection, 1987* (Wilkins, 1988).

Economics. Recently, the Office of the Actuary, Social Security Administration, has published Actuarial Study No. 101, *Economics projections for OASDHI cost and income estimates: 1987* (Goss, 1988).

Law and Regulation. Life insurance companies are supervised by the State Insurance Departments in the United States, and by

federal and provincial departments in Canada. In the United States, the Employee Retirement Income Security Act of 1974 and subsequent legislation have been a major influence on pension funds.

Medicine. This last comes to the fore in the underwriting of life insurance, in health insurance, in the projection of future mortality improvements, and in regard to current epidemics such as AIDS.

3. SOME ACTUARIAL THEORIES

This section overviews some of the theories used by actuaries in their professional practices.

3.1. ESTIMATION OF MORTALITY AND OTHER RATES

From chapter III of J. S. Elston's *Sources and characteristics of the principal mortality tables* (Elston, 1932), we have the quotation:

“United States Life Tables, 1910

United States Life Tables 1890, 1901, 1910 and 1901–1910

These tables are the first of any scientific value prepared by the U. S. Government from census returns. When the census of 1910 was taken, the Bureau of the Census called into consultation a committee of The Actuarial Society of America, and this committee gave general advice with reference to the taking of the census, the tabulation of the data, and the preparation of life tables. Although not all the committee's recommendations were followed, these tables, which were prepared under the supervision of Professor James W. Glover, mark a notable epoch in the history of mortality.”

In all, 69 life tables were prepared from the censuses, and the death statistics of the ten original registration states and the District of Columbia (Glover, 1916 and 1921). In view of the status of computing facilities in the 1910–1920 decade, the preparation and publication of these tables was a monumental task.

Another mathematician who has been a principal and innovative consultant for U. S. Life Tables for 1939–1941, and for subsequent intervals around the decennial censuses, is T. N. E. Greville (Greville, 1946). The problems he met and solved for the 1939–1941 Tables led him to many later developments in theories of interpolation, graduation, splines, generalized inverses of matrices, and life tables, as indicated in bibliographies of (Bowers et al, 1986), (London, 1985), and (Shiu, 1984 and 1986).

Turning now to insured lives mortality, we have as an early reference, *Construction of mortality tables from the records of insured lives* (Murphy and Papps, 1922). This was followed by a number of papers in the actuarial literature which plateaued in the texts (Gershenson, 1961) and (Batten, 1978). Concurrently, beginning in 1934, committees of the Society of Actuaries, or its predecessors, have published a series of annual *Reports of mortality, morbidity and other experience* based on data mainly contributed by a number of life insurance companies (Society of Actuaries, 1984).

In regard to annuitant mortality, a landmark paper by W. A. Jenkins and E. A. Lew developed the idea of scales of projection factors to allow for future mortality improvements (Jenkins and Lew, 1948).

Much statistical work has gone into the estimation of mortality rates from clinical data, and the subject has broadened to that of survival models. Simultaneously, the computer evolution has greatly facilitated the calculation of exposed to risk from the records of the individuals observed in the estimation process. These new approaches, are presented in Dick London's text, *Survival models and their estimation* (London, 1988). See also J. D. Broffitt's paper "Maximum likelihood alternatives to actuarial estimators of mortality rates" (Broffitt, 1984). It should be added that census methods used in the estimation of population mortality also have application to estimation of mortality of insured lives or of pension fund participants. There are, of course, distinctive differences in the data for the various studies.

3.2. GRADUATION THEORY

This topic is concerned with the systematic revision of estimates of series of rates, in particular, those to be used as bases for survival models. A fine survey is given by E. S. Shiu in the 1986 *Proceedings*, volume 35. His abstract is as follows: "Graduation is the process of obtaining from an irregular set of observed values, a corresponding smooth set of values consistent in a general way with the observed values. This is a survey of various methods of graduation used by actuaries."

Some early work goes back to E. L. DeForest in the 1870s which was later brought to life by H. H. Wolfenden in (Wolfenden, 1925). R. Henderson, who was prominent in the early history of both the actuarial and the mathematical professions, prepared the monograph *Mathematical theory of graduation* (Henderson, 1938). Let us pause for a moment to pay tribute to this distinguished man.

His life spanned from 1871 to 1942. He graduated from the honors mathematics program of the University of Toronto in 1891. He became a Fellow of the Actuarial Society of America and of the Casualty Actuarial Society, and was elected president of the former organization. Robert Henderson rose to become actuary of the Equitable Life Assurance Society. He served as trustee

of the Mathematical Association of America and of the American Statistical Association. In addition, "For a number of years, Mr. Henderson served as a member of the Board of Trustees of the American Mathematical Society. He felt the keenest interest in the place which the Society was taking in scientific progress and lent earnest assistance to the raising of funds in order that its work might continue unimpaired despite the economic difficulties of recent years. He also served from 1935 until shortly before his death as a Director of the Teachers Insurance and Annuity Association." (Quotation from the obituary for Robert Henderson in *Transactions of the Actuarial Society of America*, 43 (1942)).

We should also note that Robert Henderson delivered the second Gibbs Lecture on "Life insurance as a social science and as a mathematical problem" (Henderson, 1925). The entire principal of his estate was received by the American Mathematical Society in 1961 for its Endowment Fund.

Another notable author was C. A. Spoerl, a summa cum laude graduate of Harvard University. See, for instance, his paper "Whittaker-Henderson graduation formula A" (Spoerl, 1937).

For a number of years following 1950, a new monograph, *Elements of graduation*, by M. D. Miller served as education reference (Miller, 1946). Meanwhile, a succession of papers were coming from the pen of T. N. E. Greville, which may be well seen in the book, *Selected papers of T. N. E. Greville*, 1984. These have influenced the work of G. S. Kimeldorf and D. A. Jones in "Bayesian graduation" (Kimeldorf and Jones, 1967), and E. S. Shiu in "Minimum- R_z moving-weighted-average formulas" (Shiu, 1984). Another approach is exemplified by D. R. Schuette's "A linear programming approach to graduation", (Schuette, 1978).

3.3. RISK THEORY

We consider first the simpler case of short-term insurances. Here one may be concerned with the distribution of total claims in a given period for a given portfolio of insurance policies. The approach in individual risk theory is to set up a random variable

$$(3.3.1) \quad Y_j = I_j B_j \quad j = 1, 2, \dots, n$$

for each of the n insurance policies in the portfolio under consideration. Here I_j is 1 if policy j leads to a claim and is 0 otherwise; B_j is the amount of such a claim, given that it occurs. On the assumption that $I_j, B_j, j = 1, 2, \dots, n$ are mutually independent, one proceeds to approximate the distribution of aggregate claims for the period, that is, the distribution of

$$(3.3.2) \quad S_{IR} = \sum_{j=1}^n Y_j.$$

For this we need knowledge of the probability that $I_j = 1$ and of the distribution of B_j , for each j .

In the collective risk model, the basic concept is that of a random process that generates claims for a portfolio of policies. This process is in terms of a portfolio as the whole rather than in terms of the individual policies. Let N be the random number of claims for a portfolio of policies in the given period. If X_1 is the random amount of the first claim, X_2 , the random amount of the second claim, and so on, then

$$(3.3.3) \quad S_{CR} = X_1 + X_2 + \cdots + X_N$$

is the random amount of aggregate claims. The random variable N is referred to as frequency of claims and the random variables X_j measures the size of claims. In order to proceed, one makes the assumptions that:

1. X_1, X_2, \dots are identically distributed.
2. The random variables N, X_1, X_2, \dots are mutually independent.

An overview of risk theory, with emphasis on the collective theory, is given by H. Panjer's, "Models in risk theory" (Panjer, 1986). See also H. Gerber's *An introduction to mathematical risk theory* (Gerber, 1979). Both of these references provide bibliographies which indicate the historical development of risk theory. A major figure is H. L. Seal, as the bibliographies attest (see Seal, 1969). The reader interested in connecting the two approaches to risk theory for short-term insurances is referred to Section 13.5 of (Bowers et al, 1986). For information about estimating the probability distribution of the X_j 's one can refer to S. A. Klugman's "Loss Distributions" (Klugman, 1986), or to the book by R. V. Hogg and S. A. Klugman with the same title (Hogg and Klugman, 1984).

An early discussion of risk theory for individual insureds under long-term life insurance and annuity contracts was given by W. O. Menge, a later-year colleague of J. W. Glover, in the paper "A statistical treatment of actuarial functions" (Menge, 1937). An extensive development of individual risk theory for such contracts is a major theme of (Bowers et al, 1986).

3.4. CREDIBILITY THEORY

Since the early papers of F. A. Perryman (Perryman, 1937) and A. L. Bailey (Bailey, 1950), an extensive literature has grown up. Successive overviews of this literature have been presented by P. M. Kahn in 1967, 1968, 1975 and 1986. The reader is referred to this last paper, and its bibliography (Kahn, 1986).

In brief, credibility theory applies mainly to short-term insurances such as group life insurance, or those in various casualty lines, or the year-to-year risks under individual life insurances. The theory studies the revision

of premium rates in the light of current claim experience. To quote from (Kahn, 1986):

“In the classical approach the actuary must first determine the size of the experience which warrants full credibility, i.e. a credibility factor $Z(t)$ of 1, where t measures the size of the exposure or insurance experience which generated the level of claims x . The next step is to determine partial weights, or partial credibility factors for some smaller groups. Then the adjusted estimate of claims may be expressed as

$$(3.4.1) \quad Z(t)x + [1 - Z(t)]m(t)$$

where $m(t)$ is the prior estimate of expected claims.”

A. L. Mayerson’s paper “A Bayesian view of credibility” (Mayerson, 1964) was a stimulus for much further research. In 1975, J. C. Hickman drew a distinction between classical theory where the parameters of the claims process are considered as fixed constants, and the newer theories where the parameters are themselves random variables (Hickman, 1975). As with much actuarial theory, the newer concepts of credibility must undergo validation and refinement in actual insurance experience.

3.5. MATHEMATICS OF COMPOUND INTEREST

In Section 2, Sources of Actuarial Science, reference has been made already to the mathematics of finance and the direction in which it is headed. Here, and in the next section, we refer to some of the classical actuarial mathematics texts. For further information on these texts, and how they became incorporated into the education and examination processes of the profession in North America, the reader is referred to the chapter on actuarial education in E. J. Moorhead’s forthcoming 1809–1979 history of the actuarial profession, entitled *Our yesterdays* (Moorhead, forthcoming). This chapter, from a different viewpoint, gives insight about the professors and universities that have contributed to actuarial education and science.

From the University of Toronto, we have had M. A. Mackenzie’s *Interest and bond values* (Mackenzie, 1917), and N. E. Sheppard and D. C. Baillie’s *Compound interest* (Sheppard and Baillie, 1960). From the University of Michigan, there has appeared M. V. Butcher and C. J. Nesbitt’s *Mathematics of compound interest* and, as one of the leading more elementary texts, P. R. Rider and C. H. Fischer’s *Mathematics of investment* (Rider and Fischer, 1951). Since 1970, the Society of Actuaries has benefitted from S. G. Kellison’s *The Theory of Interest*. The newest text in the English language is J. J. McCutcheon and W. F. Scott’s *An introduction to the mathematics of finance* (McCutcheon and Scott, 1986). These texts treat basic finance concepts which go far back into the mists of history of civilization.

3.6. MATHEMATICS OF LIFE CONTINGENCIES

As a major part of the core of actuarial mathematics is the subject of this subsection, it is in order to discuss the principal textbooks that have appeared from time to time in the English language. If one refers to (Moorhead, forthcoming), one reads about such early works as R. Price's *Observations on reversionary payments* (1771), William Morgan's *The doctrine of annuities and assurances on lives and survivorships* (1779), Francis Bailey's *Doctrine* (1812–1813), Joshua Milne's *Treatise* (1815), and David Jones' *Value of annuity and reversionary payments* (1843). My own acquaintance goes back to G. King's *Institute of actuaries textbook, Part II* (King, 1902), and I endured through examinations on E. F. Spurgeon's *Life contingencies* (Spurgeon, 1922).

It is probably little known by now that C. H. Fischer and myself were invited in the late 1940s to undertake for the Society of Actuaries a new textbook. At that time concepts about the probability distributions of random variables were not well organized, at least in my mind, but nevertheless, it seemed to me then to be the way to proceed. The Society was not ready for what appeared to be a novel approach, and turned the project over to C. W. Jordan who by 1952 produced a book which served the profession well for over thirty years (Jordan, 1952). His book began with the notion of survival function but soon settled down to deterministic formulas. This was followed by P. F. Hooker and L. H. Longley-Cook's two-volume text, *Life and other contingencies* (Hooker and Longley-Cook, 1953, 1957). This text had brief discussion of variance around the expected values, as did also the successor book, A. Neill's *Life contingencies* (Neill, 1977). In 1978, the author team of N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbitt began work on a new textbook, entitled *Actuarial mathematics*, which emerged in final form by 1986 (Bowers et al, 1986). An enlightening overview is given by J. C. Hickman's paper "Updating life contingencies" (Hickman, 1988). This textbook goes way beyond what Fischer and I attempted forty years earlier. It intertwines individual risk theory and individual life insurance mathematics, and introduces collective risk theory, with various practical applications in group insurance and reinsurance. It ends with a chapter on "Theory of pension funding," using a mathematical deterministic model, generalizing the work of C. L. Trowbridge in "Fundamentals of pension funding" (Trowbridge, 1952). The extensive bibliography lists the many authors whose works have helped to shape the text.

For some time, a new direction in actuarial mathematics has been appearing in Europe. This is exemplified by J. Hoem's "The versatility of the Markov chain as a tool in the mathematics of life insurance" (Hoem, 1988), and by H. Wolthuis' doctoral thesis, *Savings and risk processes in life contingencies* (Wolthuis, 1988). To a considerable extent, this direction runs counter to American practice which models separately each state that an

insured may enter, for example, the state of disability, rather than use an integrated model covering all states, and transfers among them. It remains for the future to determine the usefulness of the integrated models. One indication is that the work of M. J. Cowell and W. H. Hoskins (Cowell and Hoskins, 1987), and of H. J. Panjer (Panjer, 1988) on projections regarding the AIDS epidemic, and recent work of J. Beekman in modeling decline of activity of the aged, are related thereto.

Meanwhile, actuaries like myself who are interested in the long-term guidance of pension funds and social security, are prone to use what I term mathematical deterministic (or growth) models, and to utilize a range of actuarial assumptions which are monitored regularly. This viewpoint is reflected in B. N. Berin's *The Fundamentals of pension mathematics* (Berin, 1978), and in the long-range projections for U.S. Social Security (Andrews and Beekman, 1987). This approach is also exploited in A. W. Anderson's *Pension mathematics for actuaries* (Anderson, 1985).

Another example of theory developments which have not gained much usage yet in practice is given by W. S. Bicknell's thesis "Premiums and reserves in multiple decrement theory (Bicknell and Nesbitt, 1956). This discusses three systems for premiums and reserves for the case of multiple forms of termination and benefits related thereto, as in pension plans. The second and third systems involved somewhat complex composition of the actuarial bases for the several benefits. This, we have noted, is not the American way in practice. The third system, which goes back to Alfred Loewy, has considerable possibilities, but has practical and thoretical subtleties which have been explored by (Schuette and Nesbitt, forth coming in *ARCH*).

It seems fitting to end this subsection with a tribute to H. L. Rietz who was from 1918 to 1962 influential in the development of mathematical statistics and actuarial science at the University of Iowa. He served as vice president of the American Institute of Actuaries, 1919–1920, as president of the Mathematical Association of America in 1924, as vice president of the American Statistical Association in 1925, and as vice president of the American Association for the Advancement of Science in 1929. He was the first president of the Institute of Mathematical Statistics, organized in 1935, and the 1943 volume of the *Annals of Mathematical Statistics* was dedicated to him on the occasion of his retirement. Among his doctoral students was C. H. Fischer, my long-time colleague at the University of Michigan.

4. APPLICATIONS

From 1900 through 1987, life insurance in force in the United States has grown from a little over \$7.5 billion to almost \$7.5 trillion. Some \$3 trillion of this latter amount is classified as group insurance, a form which did not exist in 1900. This period saw the development of retirement income

policies, variable life insurance and several forms of flexible life insurance. Discussions of these may be found in chapter 16 of (Bowers et al, 1986). Some initial papers were authored by E. G. Fassel (1930), J. C. Fraser, W. N. Miller and C. M. Sternhell (1969), W. L. Chapin (1976), and S. A. Chalke and M. F. Davlin (1983). (See bibliography in Bowers et al, 1986 for references.)

During the same period, the growth of pension funds is indicated by the increase from about \$20 billion of assets in 1900 to about \$2 trillion in 1986. A notable development during this period was the concept of variable annuities and the formation in 1952 of the College Retirement Equities Fund (CREF). The actuarial basis for that fund was pioneered by R. M. Duncan's "A retirement system granting unit annuities and investing in equities" (Duncan, 1952). I recall one lunchtime where Carl Fischer and I pressed Robert Duncan on the theory of dollar averaging for accumulating purchases of units by a series of regular contributions to CREF. When asked what would happen if the stock market collapsed, he thought for a moment and then with a smile said "They might not be worth very much, but you would have a lot of accumulation units."

For further information about the Teachers Insurance and Annuity Association (TIAA) and CREF, see my paper, "On the performance of pension plans" (Nesbitt, 1986). In particular, note the graded benefit annuity option which has in recent years become available from TIAA.

This section concludes with a few comments on the Old Age, Survivors and Disability Insurance (OASDI) system, popularly called Social Security but this latter also embraces the insurances under Medicare. OASDI is an extremely large system with annual benefit outgo now at the level of \$235 billion, and projected level of \$8 trillion by year 2045 under moderate growth assumptions (Annual Report, 1988). With good reason, the actuaries of the System prefer to project benefit outgo as a percent of projected taxable payroll for the System. On this basis, projected OASDI outgo in 2045 is 16.25 percent of taxable payroll. The actuarial guidance of this huge system is a major challenge for the actuarial profession.

An acknowledged leader in such guidance has been R. J. Myers who has written very extensively on Social Security (see, for instance, Myers, 1985). He set the pattern for the short-range and long-range projections, the processes for which are continuously evolving. An overview of these processes is given in (Andrews and Beekman, 1987) and (Annual Report, 1988). A recently formed National Academy of Social Insurance, with Alicia Munnell of the Federal Reserve Bank of Boston as president, will form a common ground for persons from different fields who are interested in Social Security. OASDI developments over the past fifty years have been of major importance as a foundation of benefits to be supplemented by nonfederal life insurance and pension-funding, and should remain so.

5. CONCLUDING COMMENTS

This paper has been written mainly as personal reactions to actuarial developments of the last fifty years. As I got further into it, I had more and more occasion to refer to (Bowers et al, 1986) and (*Proceedings*, 1986). For the reader interested in going beyond this paper, I recommend a perusal of the latter reference. I draw special attention to J. C. Hickman's introduction, and to his paper "Updating life contingencies" which enlarges on the concepts underlying the textbook (Bowers et al, 1986).

I have tried to make at least one reference to many, but by no means all, contributors to actuarial science in North America. Most of the omitted names may be found in the bibliographies of (Bowers et al, 1986) and (*Proc. Symp. Appl. Math.*, 1986). Some omissions relate to young men whose work is in process of recognition, and some to special areas of expertise. I have depended on the useful bibliographies in the various references cited to provide a more complete picture of developments, including much work not cited here.

Two references I wish to add here are to W. O. Menge and J. W. Glover's *An introduction to the mathematics of life insurance* (Menge and Glover, 1935), and its later revision, *Mathematics of life insurance* (Menge and Fischer, 1965). In a very real sense, these helped to clarify life insurance actuarial practice.

While risk theory and credibility theory are major elements in the actuarial mathematics of nonlife insurance, beyond these two theories and reference to (D'Arcy and Doherty, 1988) no attempt has been made to cover that field further, as it has not been part of my experience. A similar remark applies for the large field of health insurance on an individual, group or national basis.

It should be recorded that there are some actuaries who have realized that we have undergone in the last forty years the risk of incredible destruction by nuclear war. Among these was Edmund C. Berkeley who included this topic in his address "Society, computers, thinking and actuaries" to the 16th Annual Actuarial Research Conference, University of Manitoba (see Berkeley, 1982). In papers presented to the 22nd and 23rd annual actuarial research conferences, I have indicated a simple model for recognising nuclear holocaust hazard, and its pervasive effect on all longterm actuarial calculations such as those regarding average length of life, or for mortgage amortization over a term of years (Nesbitt, 1987 and 1988). This is the actuarial science that must be communicated to protect life, and to counter the weight of science that could destroy life. These tasks really fall upon teachers in all fields, but actuarial science should do its part.

Finally, this review, which was mainly retrospective but had some updating and prospective aspects, has encouraged me about this century's progress in actuarial science, and has increased my awareness of developments to come.

ADDENDUM

Here I present some information and reflections on the education and examination of actuaries. This could be a very large assignment, and I shall resolve it mainly by pointing to sources of information. An immediate problem is that there is not just one, but four, actuarial organizations directly involved in actuarial education and examination, and three others that cosponsor some or all of the examinations. The two oldest organizations, the Society of Actuaries and the Casualty Actuarial Society, have leading roles but have been supplemented by the Joint Board for the Enrollment of Actuaries (a unit of the U.S. Department of the Treasury), and the American Society of Pension Actuaries. Cosponsoring organizations are the American Institute of Actuaries, the Canadian Institute of Actuaries (CIA has a different connotation in Canada than in the U.S.), and the Conference of Actuaries in Public Practice. This may seem confusing but there is considerable coordination among the seven bodies through the Council of Presidents, and also through overlapping memberships. My own experience has been mainly with the Society of Actuaries, and I shall use the Society as my information source. The catalogs of the Society list the addresses of all seven organizations.

A second problem is that culminating in the years since 1985, there has been a restructuring of the Society's education process into a Flexible Education System (FES), and a follow-up by proposed Future Education Methods (FEM). The multi-membered Education and Examination Committee distributed two white papers, on FES in 1986, and on FEM in 1987, setting forth the proposed changes and their rationale. As Vice President for Research and Studies in 1985–1987, I witnessed the presentation of these documents and both the general support and the counter-reactions that they gathered. The new emphasis is on education that can adapt itself to our fast-changing world and that achieves a better balance with the discipline of the actuarial examinations.

FES is now in place and some steps have been taken in regard to FEM. These are reflected in the booklets, 1989 Associateship Catalog, and Spring 1989 Fellowship Catalog, where associateship is the first level and fellowship is the second level of qualification for membership in the Society. Both booklets state the following:

Principles Underlying the Education and Examination System

The Society of Actuaries administers a series of self-study courses and examinations leading to Associateship and Fellowship. The principles un-

derlying the Society's education and examination system are the following:

(1) To provide the actuary with an understanding of fundamental mathematical concepts and how they are applied, with recognition of the dynamic nature of these fundamental concepts in that they must remain up-to-date with developments in mathematics and statistics;

(2) To provide the actuary with an accurate picture of the socio-demographic, political, legal and economic environments within which financial arrangements operate, along with an understanding of the changing nature and potential future directions of these environments;

(3) To expose a broad range of techniques that the actuary can recognize and identify as to their application and as to their inherent limitations, with appropriate new techniques introduced into this range as they are developed;

(4) To expose a broad range of relevant actuarial practice, including current and potential application of mathematical concepts and techniques to the various and specialized areas of actuarial practice; and

(5) To develop the actuary's sense of inquisitiveness so as to encourage exploration into areas where traditional methods and practice do not appear to work effectively."

Under FES, a number of self-study courses are available, each providing a certain number of credits. Completion of the Series 100 requirements now satisfies the education requirements for the Associate of the Society of Actuaries (ASA) designation. A candidate must obtain 200 units of credit prior to 1995 for courses listed in Table A to satisfy the Series 100 requirements.

Table A. Course Description

| Course | Description | Credits | Type |
|---------------|----------------------------------|----------------|-------------|
| 100 | Calculus and Linear Algebra | 30 | Required |
| 110 | Probability and Statistics | 30 | Required |
| 120 | Applied Statistical Methods | 15 | Required |
| 130 | Operations Research | 15 | Elective |
| 135 | Numerical Methods | 10 | Elective |
| 140 | Mathematics of Compound Interest | 10 | Required* |
| 141 | EA-1, Segment A | 10 | Required* |
| 150 | Actuarial Mathematics | 40 | Required |
| 151 | Risk Theory | 15 | Required |
| 160 | Survival Models | 15 | Required |
| 161 | Mathematics of Demography | 10 | Elective |
| 162 | Construction of Actuarial Tables | 10 | Elective |
| 165 | Mathematics of Graduation | 10 | Elective |

Each course is designated as required or elective. A candidate must obtain 155 credits in “required” courses and 45 credits in “elective” courses to satisfy the Series 100 requirements for this catalog.

Credit for courses 140, 150, and 151 must be obtained by examinations offered by the Society of Actuaries. Credit for course 141 must be obtained by passing EA-1, Segment A of the Enrolled Actuary (EA) Examinations. Credit for all other courses must be obtained by examinations offered by the Society of Actuaries or by an alternative method which has been approved by the Board of Governors. For fall 1988 and spring 1989, credit for course 100 may be obtained by an alternative method (an appropriate score on the Graduate Record Examination Mathematics Test).”

While each 10 credits usually implies one hour of multiple-choice examination, there are exceptions. Course 140 has a one-and-one-half-hour examination, and course 150 has a four-and-one-half-hour examination split into two sessions, and including some written-answer questions.

The written-answer examinations I took years ago had algebra based on the classical Hall and Knight textbook, had analytic geometry and calculus together, and scarcely touched linear algebra. Probability was mainly combinatorics based on Whitworth’s Choice and Chance, and statistics was at a pre-calculus descriptive level. Now course 110 includes topics among those proposed for a one-year college course in probability and statistics by the Committee on the Undergraduate Program in Mathematics. Course 120 covers analysis of variance, regression analysis and time-series analysis which were largely omitted from the syllabus of my examination-writing years. Course 130 on linear programming, project scheduling, dynamic programming, relates to topics that came to the fore during World War II.

Course 135, Numerical Methods, replaces the former examination on finite differences. The finite (as opposed to the infinitesimal) calculus was one of my teaching joys. It was always a pleasure to define divided differences, proceed to the Lagrange interpolation formula with remainder, relate divided differences under prescribed conditions to derivatives at intermediate points, and pull out Newton’s divided difference interpolation formula with remainder, and as special cases obtain Taylor’s series and the various classical polynomial interpolation formulas, all with remainders. One then was set to make applications to summation, approximate integration, and difference equations. Now, the impact of computers has greatly expanded the subject to modern numerical analysis with its algorithmic approach. This is reflected in course 135 which covers iteration, interpolation, numerical integration and linear systems.

*Candidates must receive credit for either course 140 or 141 but will not receive credit for both.

Compound interest theory, the subject of course 140, while benefitting from some refinement of basic concepts, and from the enormous improvement in computing facilities, has been well established for many decades. Course 141, administered by the Joint Board for the Enrollment of Actuaries, has a two-and-one-half-hour examination covering the mathematics of compound interest and of life contingencies.

Course 150 is an extensive coverage of the mathematics of life contingencies. It is based on the new textbook (Bowers et al, 1986) which employs future lifetime as the underlying random variable. For the development of this central subject of life actuarial science over the past two centuries, and its updated setting in *Actuarial mathematics*, see Section 3.6 of my foregoing Reflections and also (Hickman, 1986).

While “Economics of insurance”, and “Individual risk models for a short term,” (chapters 1 and 2 of *Actuarial mathematics*) are recommended as background readings for course 150, these topics together with “Collective risk models for an extended period”, and “Applications of risk theory,” comprise the examination subjects for course 151. Random variables were only vaguely elaborated in the syllabus when I was a student, and much of this theory has developed since.

The subjects of courses 160, 161, 162 and 165 have been touched upon in Sections 3.1 and 3.2 of my foregoing paper. Only course 160 is required, the others are elective, but 45 credits, as of now, must be chosen from 55 available. In many cases, actuaries have very large amounts of data available (relative to insureds and deaths) in the form of policies of insurance, amounts of insurance, annuity or pension incomes, census counts, and vital statistics of births and deaths. Methods for analyzing such data may then differ in some degree from those for smaller, more detailed studies of clinical data, or of impaired lives. In any case, the actuarial profession seeks reasonable understanding of the various estimation procedures that are feasible and available.

After attaining ASA designation, many actuarial students aspire to complete the education requirements for the Fellow of the Society of Actuaries (FSA) designation. To do so, they must undertake the Series 200–500 courses. In the Fellowship Catalog, we read these Series “are divided into four groups; the common Core and three specialty tracks: the group Benefits (GB) Track, the Individual Life and Annuity (ILA) Track and the Pension (P) Track. All candidates must earn 100 credits from the core courses, 90 credits from the required courses in one of the tracks with a single national emphasis, and 60 credits from other Fellowship courses. Within a track, some courses are designed to be national in emphasis (either Canada or U.S.)”.

To give a little more insight to the nature of these requirements, I quote from the Fellowship Catalog.

COURSE 200.
INTRODUCTION TO FINANCIAL SECURITY PROGRAMS

(40 Credits) Required

The examination for this course is a four-hour multiple-choice and written-answer examination. The course covers: design, regulation and taxation of the major voluntary financial security programs involving life insurance, health insurance, property and casualty insurance, and employee benefit and pension programs; characteristics of the major social insurance programs in Canada and the U.S.; description of the providers of financial security programs; and an introduction to taxation of insurance companies in both Canada and the U.S.”

Some hardy souls, after attaining the FSA designation proceed to the ACAS and FCAS designations of the Casualty Actuarial Society. Canadian FSAs take whatever additional steps may be needed for the Fellow of the Canadian Institute of Actuaries (FCIA) designation.

Future Education Methods (FEM) are in progress. In October 1987, the Board of Governors of the Society approved implementation of five programs, namely:

- (i) a Fellowship Admissions Course, a two-and-one-half-day course focusing on professional ethics and integration of syllabus material.
- (ii) a research paper option for 30 elective Fellowship credits (details in the Fellowship Catalogs).
- (iii) credit for examinations of other actuarial organizations and complete designations of non-actuarial organizations.
- (iv) elective credit for an Intensive Seminar at the Associate level.
- (v) an experiment in allowing credit for college courses, approved by the Society of Actuaries Education and Examination Committee, covering the topics of applied statistics, operations research and numerical methods.

These programs are at various stages of implementation.

In my fellowship student days, I was required to pass three six-hour examinations, each of which had a number of subjects. Fortunately, one examination was in my special fields of interest of pensions and of social insurance. The current requirement of 250 credits may require more examination hours, although FEM programs may effect such increase.

Of 112 members of the Society of Actuaries who hold appointments in U.S. and Canadian colleges and universities, 40 are in departments of mathematics or mathematical sciences; 27, in departments of statistics and actuarial science; 22, in schools of business administration; and 10, in actuarial science and insurance programs. The remaining 13 are in miscellaneous or

unstated units. An additional 8 members of the Society are in foreign colleges and universities. As of November 1, 1988, total membership of the Society was 11,157, consisting of 6,039 Fellows and 5,118 Associates. Some 721 members reside outside Canada and the U.S.A. Additional membership statistics can be found in the 1989 Yearbook of the Society.

Much detailed information about the Society of Actuaries courses and examinations can be obtained by writing to:

Society of Actuaries
475 N. Martingale Road
Schaumburg, IL 60173
(312) 706-3500

and requesting a copy of the 1989 Associateship Catalog and of the spring 1989 Fellowship Catalog.

Education in the topics of courses 100–135 can be acquired at many colleges and universities in the United States and Canada. A list of schools which offer degree programs covering much of courses 150–165 can be obtained from the Society of Actuaries. In addition, the catalogs give information about study manuals and study groups that a student may wish to utilize.

Throughout my teaching career, and in following years, there has been a strong demand for actuarial students. These may find employment in insurance companies, consulting firms, state and federal government agencies (including insurance departments, the Internal Revenue Service and the Social Security Administration). Such organizations cover the tremendous range of financial security systems such as property-liability insurance, health insurance, life insurance, annuities and social insurance.

Actuarial education and examinations provide a rigorous but equitable process, manned by many dedicated volunteers who are complemented by an able, growing staff. The process provides a pathway to a challenging life devoted to making our financial security systems truly effective. My part therein has been a major satisfaction of my life.

REFERENCES

- [1] *Actuarial Research Clearing House (ARCH)*, C. S. Fuhrer and A. S. Shapiro, coeditors, distributed by the Society of Actuaries in 2 or 3 issues per year.
- [2] A. W. Anderson, *Pension mathematics for actuaries*, Windsor Press, P. O. Box 87, Wellesley Hills, Massachusetts 1985.
- [3] G. H. Andrews and J. A. Beekman, *Actuarial projections for the Old Age, Survivors, and Disability Insurance Program of Social Security in the United States of America*, Actuarial Education and Research Fund, 475 N. Martingale Rd., Schaumburg, IL 60173, 1987.

- [4] Annual Report, Board of Trustees, Federal Old-Age, Survivors and Disability Insurance Trust Funds, 1988.
- [5] A. L. Bailey, "Credibility procedures", *Proc. Casualty Actuar. Soc.* **37** (1950), 7–28.
- [6] R. W. Batten, *Mortality table construction*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.
- [7] J. A. Beekman, "Actuarial assumptions and models for social security projections", *Proc. Symp. Appl. Math.* **35**, 1986, 85–104.
- [8] B. N. Berin, *The fundamentals of pension mathematics*, William M. Mercer, Inc. 1978.
- [9] B. C. Berkeley, "Society, computers, thinking and actuaries", *ARCH*, New York, 1982. 1, 11–20.
- [10] W. S. Bicknell and C. J. Nesbitt, "Premiums and reserves in multiple decrement theory", *Trans. Soc. Actuar.* **8** (1956), 344–377.
- [11] P. P. Boyle, "Immunization under stochastic models of the term structure", *Jour. Inst. Actuar.* **105** (1978), 177–187. Also, *ARCH*, 1980.1, 19–29.
- [12] N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones, C. J. Nesbitt, *Actuarial Mathematics*, Society of Actuaries, 1986.
- [13] J. D. Broffitt, "Maximum likelihood alternatives to actuarial estimators of mortality rates", *Trans. Soc. Actuar.* **36** (1984), 77–142.
- [14] M. V. Butcher and C. J. Nesbitt, *Mathematics of compound interest*, Ulrich's Books, Inc. Ann Arbor, Michigan, 1971.
- [15] D. D. Cody, "Discoveries to-date on risk, valuation and surplus", *Actuarial Research Clearing House*, 1987.2, 35–44.
- [16] M. J. Cowell and W. H. Hoskins, *AIDS, HIV mortality and life insurance, parts 1 and 2*, Special Report, Society of Actuaries, 1987.
- [17] S. P. D'Arcy and N. A. Doherty, *The financial theory of pricing property-liability insurance contracts*, Richard D. Irwin, Inc., Homewood, Illinois, 1988.
- [18] R. M. Duncan, "A retirement system granting unit annuities and investing in equities", *Trans. Soc. Actuar.* **4** (1952), 317–344.
- [19] J. S. Elston, *Sources and characteristics of the principal mortality tables*, The Actuarial Society of America, 1932.
- [20.] Financial Accounting Standards Board, *Statement of financial accounting standards No. 87*, P. O. Box 3821, Stamford, CN 06905-0821, 1985.
- [21] H. Gerber, *An introduction to mathematical risk theory*, Richard D. Irwin, Inc., Homewood, Illinois, 1979.
- [22] H. Gershenson, *Measurement of mortality*, Society of Actuaries, 1961.
- [23] J. W. Glover, *United States life tables: 1910*, Bureau of the Census, 1916.
- [24] —, *United States life tables: 1890, 1901, 1910, and 1901–1910*, Bureau of the Census, 1921.
- [25] S. C. Goss, M. P. Glanz, E. Lopez, *Economic projections for OASDHI cost and income estimates: 1987*, Actuarial Study No. 101, Social Security Administration, Baltimore, MD, 1988.
- [26] T. N. E. Greville, *United States life tables and actuarial tables, 1939–1941*, Bureau of the Census, 1946.

- [27.] —, *Selected papers of T. N. E. Greville*, edited by D. S. Meek, R. G. Stanton, Winnipeg, Manitoba, 1984.
- [28] R. Henderson, "Life insurance as a social science and as a mathematical problem", *Bull. Amer. Math. Soc.* **31** (1925), 227–252.
- [29] —, *Mathematical theory of graduation*, The Actuarial Society of America, 1938.
- [30] E. L. Hicks and C. L. Trowbridge, *Employer accounting for pensions*, Pension Research Council, Wharton School, University of Pennsylvania, 1985.
- [31] J. C. Hickman, "Introduction and historical overview of credibility", in *Credibility: theory and applications*, edited by P. M. Kahn, Academic Press, New York, 1975.
- [32] —, "Updating life contingencies", *Proc. Symp. Appl. Math.* **35**, 1986, 5–15.
- [33] Jan M. Hoem, *The versatility of the Markov chain as a tool in the mathematics of life insurance transactions*, 23rd. International Congress of Actuaries, Helsinki, 1988.
- [34] R. V. Hogg and S. A. Klugman, *Loss distributions*, Wiley, New York, 1984.
- [35] D. F. Hooker and L. H. Longley-Cook, *Life and other contingencies*, Vol. I, Cambridge University Press, 1953, and *Life and other contingencies*, Vol. II, Cambridge University Press, 1957.
- [36] W. A. Jenkins and E. A. Lew, "A new mortality basis for annuities", *Trans. Soc. Actuar.* **1** (1949), 369–466.
- [37] C. W. Jordon, *Life contingencies*, Society of Actuaries, 1952 1st ed., 1967 2nd ed.
- [38] P. M. Kahn, "Overview of credibility theory", *Proc. Symp. Appl. Math.* **35**, 1986, 57–66.
- [39] S. G. Kellison, *The theory of interest*, Richard D. Irwin, Inc., Homewood, Illinois, 1970.
- [40] N. Keyfitz and J. A. Beekman, *Demography through problems*, Springer-Verlag, New York, 1984.
- [41] G. S. Kimeldorf and D. A. Jones, "Bayesian graduation", *Trans. Soc. Actuar.* **19** (1967) 66–112.
- [42] G. King, *Institute of actuaries textbook, part II*, Charles and Edwin Layton, London, 1887 1st ed., 1902 2nd ed.
- [43] S. A. Klugman, "Loss distributions", *Proc. Symp. Appl. Math.* **35**, 1986, 31–55.
- [44] R. London, *Survival models and their estimation*, ACTEX Publications, Winsted, Connecticut, 1988.
- [45] M. A. Mackenzie, *Interest and bond values*, University of Toronto Press, Toronto, 1917.
- [46] A. L. Mayerson, "A Bayesian view of credibility", *Proc. Casualty Actuar. Soc.* **51** (1964), 85–104.
- [47] J. J. McCutcheon and W. F. Scott, *An introduction to the mathematics of finance*, Heinemann, London, 1986.

- [48] W. O. Menge, "A statistical treatment of actuarial functions", *Record Amer. Inst. Actuar.* **26** (1937), 65–88.
- [49] W. O. Menge and J. W. Glover, *An introduction to the mathematics of life insurance*, MacMillan, 1935.
- [50] W. O. Menge and C. H. Fischer, *The mathematics of life insurance*, MacMillan, New York, 1965.
- [51] M. D. Miller, *Elements of graduation*, The Actuarial Society of America and American Institute of Actuaries, 1946.
- [52] E. J. Moorhead, *Our yesterdays*, a history of the actuarial profession 1809–1979 in North America, prepared for the Centennial Celebration of the actuarial profession, forthcoming. See also Actuarial History 1889–1989 in 1989 Year book, Society of Actuaries.
- [53] R. D. Murphy and P. C. H. Papps, *Construction of mortality tables from the records of insured lives*, The Actuarial Society of America, 1922.
- [54] Robert J. Myers, *Social Security*, Richard D. Irwin, Inc., Homewood, Illinois, 1985.
- [55] A. Neill, *Life contingencies*, Heinemann, London, 1977.
- [56] C. J. Nesbitt, "On the performance of pension plans", *Proc. Symp. Appl. Math.* **35**, 1986, 113–129.
- [57] —, "Exploration of actuarial mathematics with recognition of nuclear holocaust hazard", 22nd Actuarial Research Conference, University of Toronto, 1987. Forthcoming in ARCH.
- [58] —, "Further reflections on actuarial recognition of nuclear holocaust hazard", 23rd Actuarial Research Conference, University of Connecticut, 1988. To appear in *Insurance, Mathematics and Economics* and in ARCH.
- [59] H. Panjer, "Models in risk theory", *Proc. Symp. Appl. Math.* **35**, 1986, 17–30.
- [60] H. J. Panjer, *AIDS: survival analysis of persons testing HIV+*, *Trans. Soc. Actuar.* **40** (1988), forthcoming.
- [61] H. W. Pedersen, E. S. Shiu and A. E. Thorlacius, "Arbitrage-free pricing of interest-rate contingent chains", *Trans. Soc. Actuar.*, forthcoming.
- [62] F. W. Perryman, "Experience rating plan credibilities", *Proc. Casualty Actuar. Soc.* **24** (1937), 60–125.
- [63] Proceedings of Symposia in Applied Mathematics, Volume 35, *Actuarial Mathematics*, American Mathematical Society, 1986.
- [64] P. R. Rider and C. H. Fischer, *Mathematics of investment*, Rinehart, New York, 1951.
- [65] D. R. Schuette, "A linear programming approach to graduation", *Trans. Soc. Actuar.* **30** (1978), 407–431.
- [66] D. R. Schuette and C. J. Nesbitt, "Withdrawal benefit equal to reserve: non-neutrality in the discrete case", 22nd Actuarial Research conference, University of Toronto, 1987. Forthcoming in ARCH.
- [67] H. L. Seal, *Stochastic theory of a risk business*, Wiley and Sons, New York, 1969.
- [68] N. E. Sheppard and D. C. Baillie, *Compound interest*, University of Toronto Press, Toronto, 1960.

- [69] E. S. Shiu, "Minimum- R_z moving-weighted-average formulas", *Trans. Soc. Actuar.* **36** (1984), p. 489.
- [70] —, "A survey of graduation theory", *Proc. Symp. Appl. Math.* **35** (1986), 67–84.
- [71] Society of Actuaries, 1984 Reports of Mortality, Morbidity and Other Experience, 1988.
- [72] M. Spiegelman, *Introduction to demography*, Harvard University Press, Cambridge, MA., 1955 and 1968.
- [73] C. A. Spoerl, "Whittaker-Henderson graduation formula A", *Trans. Actuar. Soc. Amer.* **38** (1937), 403–462.
- [74] E. F. Spurgeon, *Life contingencies*, Cambridge University Press, 1922 1st ed., 1929 2nd ed., 1932 3rd ed.
- [75] C. L. Trowbridge, "Fundamentals of pension funding", *Trans. Soc. Actuar.* **4** (1952), 17–43.
- [76] A. Wade, *Social Security area population projections: 1988*, Actuarial Study, No. 102, Social Security Administration, Baltimore, MD, 1988.
- [77] J. C. Wilkin, OASDI, *long-range beneficiary projection: 1987*, Actuarial Study No. 100, Social Security Administration, Baltimore, MD, 1988.
- [78] H. H. Wolfenden, *Population statistics and their compilation*, The Actuarial Society of America, New York, 1925.
- [79] —, "On the development of formulae for graduation by linear compounding, with special reference to the work of Erastus L. DeForest", *Trans. Actuar. Soc. Amer.* **26** (1925), 81–121.
- [80] H. Wolthuis, *Savings and risk processes in life contingencies*, Universiteit van Amsterdam, Institut voor Actuarie en Econometrie, 1988.