The Annals of Statistics
1978, Vol. 6, No. 2, 239-265

SPECIAL INVITED PAPER

MATHEMATICAL STATISTICS IN THE EARLY STATES[1]

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.

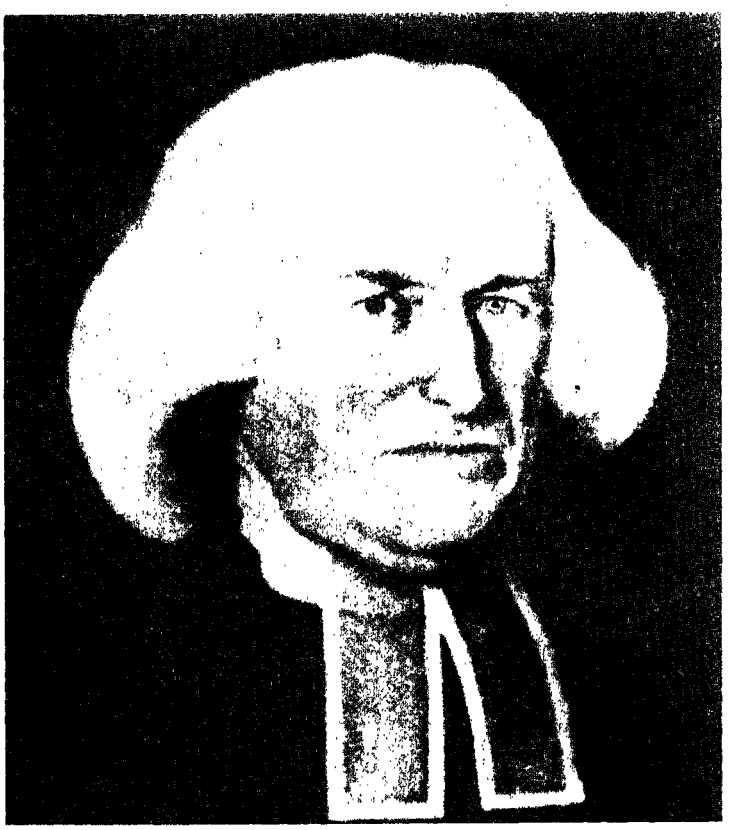
1840, and, for a modern assessment and recent references, Reingold, 1972.) Nonetheless, this investigation was undertaken with a cautious optimism. Although standard histories of statistics have little to say about American work of this period, the years from 1770 to 1850 had been ones of great interest 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 has 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.)

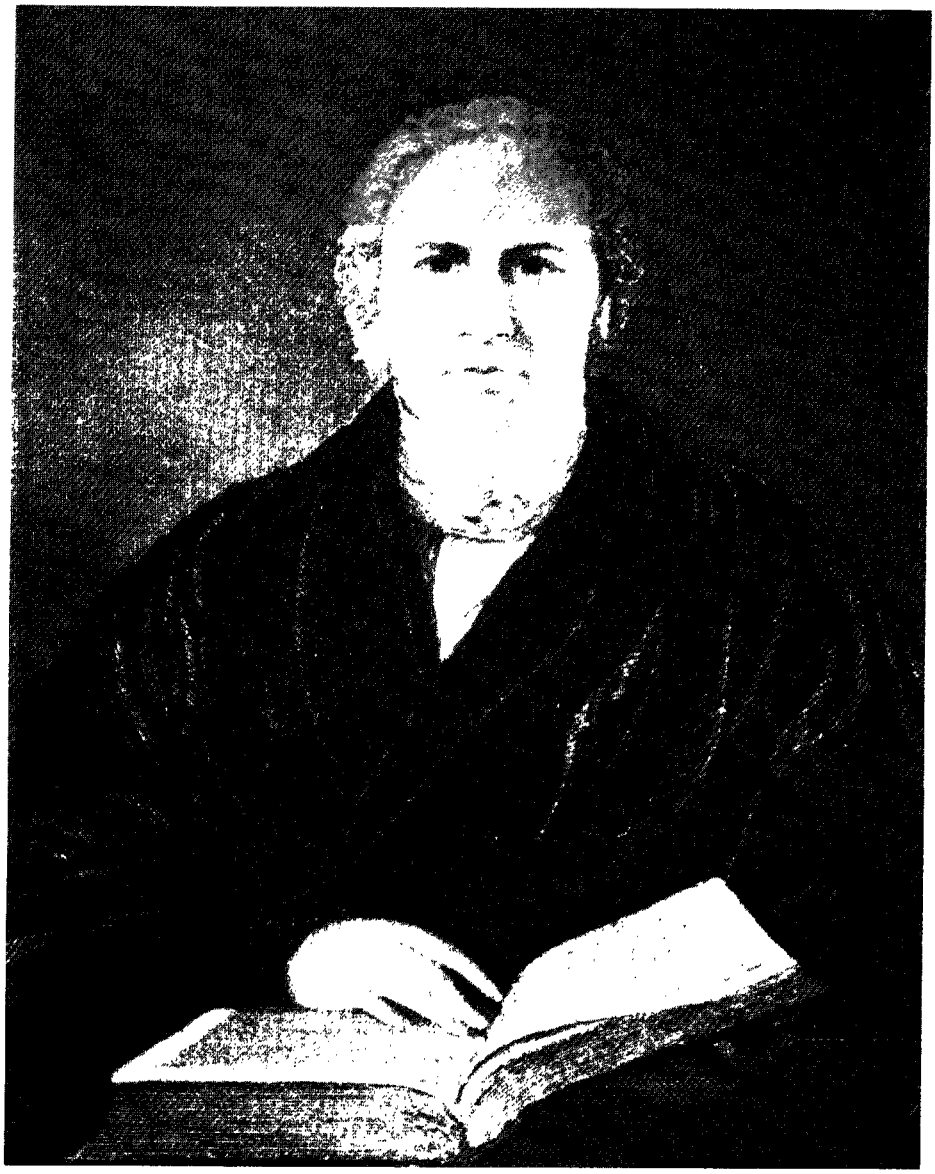
At first this dismal assessment of early American work may seem incredible. After all, Harvard could boast of a professorship of mathematics as early as 1727, the year of Newton's death. But upon closer inspection it appears that this early professor was no fit companion of Newton: his only publication 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 the 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, 1971). 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 proofread 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

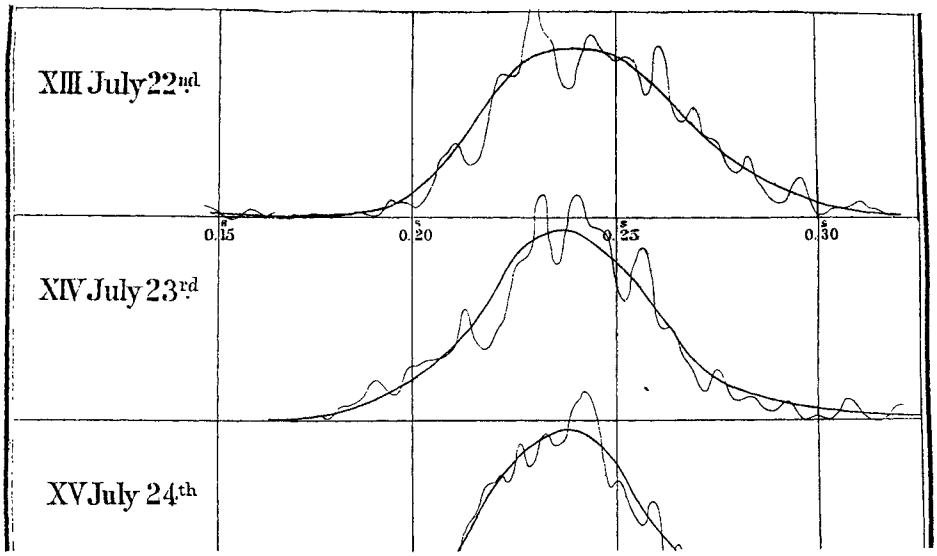


FIG. 1. C. S. Peirce's graph of his estimated probability density, by the kernel method and by a freehand sketch (smoother curves).

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 $u_1, u_2, \cdots, u_n$ 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

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

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 $v_i = l_0u_i + l_1(u_{i-1} + u_{i+1}) + l_2(u_{i-2} + u_{i+2})$ and will give the minimum mean square estimate of U_i under the cubic assumption and the restriction to estimates linear in the local $u_{i-2}, \cdots, u_{i+2}$. 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

$$U_i = l_0U_i + l_1(U_{i-1} + U_{i+1}) + l_2(U_{i-2} + U_{i+2})$$

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 $U_i = l_0U_i + l_1(U_{i-1} + U_{i+1}) + l_2(U_{i-2} + U_{i+2})$ 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 ith 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 $r = n_1^{-\frac{1}{2}}$, 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, $(l_0^2 + 2(l_1^2 + l_2^2))^{\frac{1}{2}}$ 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.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706