Statistical Science
1986, Vol. 1, No. 2, 276–284

The Contributions of Herbert Robbins to Mathematical Statistics

Tze Leung Lai and David Siegmund

Herbert Robbins was born on January 12, 1915, in New Castle, Pennsylvania. In 1931 he entered Harvard College at the age of 16. Although his interests until then had been predominantly literary, he found himself increasingly attracted to mathematics under the influence of Marston Morse, who during many long conversations conveyed a vivid sense of the intellectual challenge of creative work in that field (cf. Page, 1984, p. 7). He received the A.B. summa cum laude in 1935, and the Ph.D. in 1938, also from Harvard. His thesis, in the field of combinatorial topology and written under the supervision of Hassler Whitney, was published in 1941 [3]. (Numbers in brackets refer to Robbins' bibliography at the end of this article.)

After graduation, Robbins worked for a year at the Institute for Advanced Study at Princeton as Marston Morse's assistant. He then spent the next three years at New York University as instructor in mathematics. He became nationally known in 1941 as the coauthor, with Richard Courant, of the classic *What Is Mathematics?* [4]. This important book has influenced generations of mathematics students here and abroad in many editions and translations. To date more than 100,000 copies have been sold.

In 1941 Robbins enlisted in the Navy. He was demobilized four years later as a lieutenant commander. His interest in probability theory and mathematical statistics began during the war and was itself something of a chance phenomenon, which arose from overhearing a conversation between two senior naval officers concerning the effect of random scatter on bomb impacts (cf. Page, 1984, pp. 8–10). Because of his lack of appropriate security clearance, he was prevented from pursuing this problem during the war. Nevertheless, his work on the naval officers' problem led to the fundamental papers [7] and [10] in the field of geometric probability.

In 1946 Harold Hotelling was setting up a department of mathematical statistics at the University of North Carolina at Chapel Hill. Having read [7] and [10], and greatly impressed by Robbins' mathematical skills, Hotelling offered him the position of associate professor to teach measure theory and probability to the graduate students in the new department. Robbins accepted the position and spent the next six years at Chapel Hill. During this relatively short period Robbins not only studied and developed an increasingly deep interest in statistics, but he also made a number of profound contributions to his new field: complete convergence [12], compound decision theory [25], stochastic approximation [26], and the sequential design of experiments [28], to name a few.

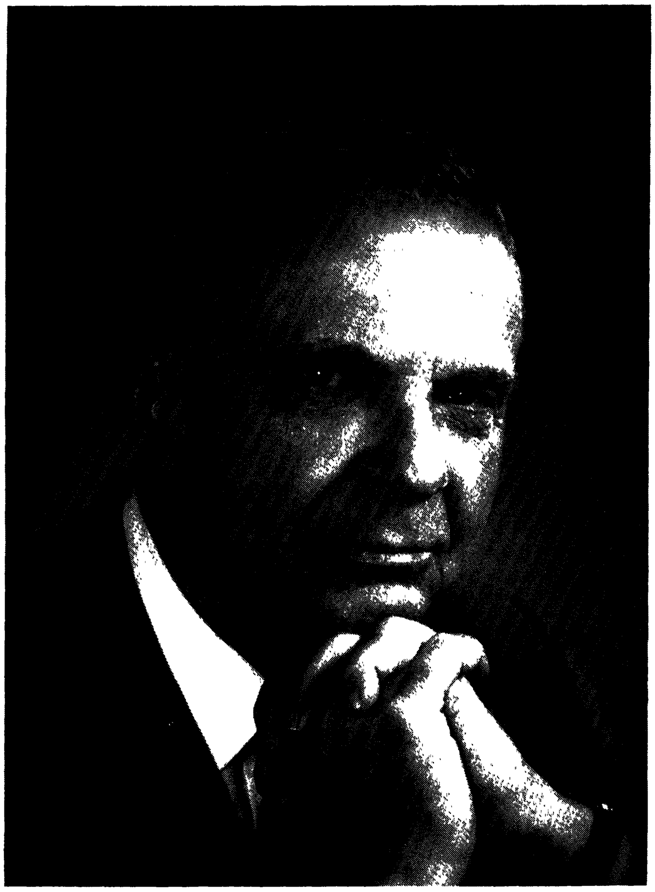
After a Guggenheim Fellowship at the Institute for Advanced Study during 1952–1953, Robbins moved from Chapel Hill to Columbia University as professor and chairman of the Department of Mathematical Statistics. Since 1953, with the exception of the three years 1965–1968 spent at Minnesota, Purdue, Berkeley, and Michigan, he has been at Columbia, where he is Higgins Professor Emeritus of Mathematical Statistics. During this period he has published over 100 papers on a variety of topics in probability and statistics. His most notable contributions include the creation of the empirical Bayes methodology, the theory of power-one tests, and the development of sequential methods for estimation, hypothesis testing, and comparative clinical trials.

Robbins was President of the Institute of Mathematical Statistics in 1965–1966, Rietz Lecturer in 1963, Wald Lecturer in 1969, and Neyman Lecturer in 1982. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences. He is widely regarded as one of the world's leading and most imaginative mathematical statisticans.

Robbins has five children and two grandchildren. Children by his first marriage to Mary Dimock are Susannah and Marcia. Children by the second marriage to Carol Hallett are Mark, David, and Emily. At the age of 71 Robbins is still young in spirit, and is as remarkably original and energetic as in the past. He continues to be a prolific contributor to the statistical literature, and much of his work continues to have a profound impact in statistics and related fields. Some of his most important contributions to mathematical statistics are discussed below.

Tze Leung Lai is Professor of Mathematical Statistics, Department of Statistics, Columbia University, New York, New York 10027. David Siegmund is Professor of Statistics, Department of Statistics, Stanford University, Stanford, California 94305.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science*.
www.jstor.org



Herbert Robbins

1. EMPIRICAL BAYES METHODOLOGY IN ESTIMATION AND PREDICTION

Robbins' pioneering paper [41] on empirical Bayes methodology and his earlier paper [25] on the related subject of compound decision theory were acclaimed by Neyman (1962) as "two breakthroughs in the theory of statistical decision making."

Consider the simple problem of estimating a real parameter θ based on the observed datum X, the probability density function of which $p(x \mid \theta)$ depends on θ. The Bayesian approach assumes the existence of a prior distribution B for θ; and the Bayes estimate minimizing squared error loss

$$E[t(X) - \theta]^2 = \int E_\theta[t(X) - \theta]^2 \, dB(\theta)$$

is given by the posterior mean

$$(1) \quad t(X) = E(\theta \mid X) = \int \theta p(X \mid \theta) \, dB(\theta)/f(X),$$

where $f(x) = \int p(x \mid \theta) \, dB(\theta)$ denotes the marginal density of X. The difficulty with implementing the Bayesian approach is specification of the prior distribution, about which it may be hard to obtain agreement in practice.

In many applications, however, one is faced with n structurally similar problems of estimating θ_i from X_i ($i = 1, 2, \cdots, n$), where X_i has probability density function $p(x \mid \theta_i)$ and the θ_i can be assumed to have the distribution B (cf. Neyman, 1962; Cover, 1968; Copas, 1972; Carter and Rolph, 1974; Efron and Morris, 1977; Hoadley, 1981; Rubin, 1980; Morris, 1983). As a concrete example suppose that the ith automobile driver in a certain sample is observed to have X_i accidents in a given year. Assuming that X_i has a Poisson distribution

$$(2) \quad p(x \mid \theta_i) = \exp(-\theta_i)\theta_i^x/x!, \qquad x = 0, 1, \cdots,$$

we want to estimate the "accident-proneness" parameter θ_i for each of the n drivers. If the distribution B of accident proneness in the population of drivers were known, the Bayes estimate of θ_i is the $t(X_i)$ given by (1), which in the present case reduces to

$$(3) \qquad t(x) = (x + 1)f(x + 1)/f(x).$$

When B is unknown, Robbins' idea is first to estimate $f(x)$ on the basis of the n observations $X_1, \cdots, X_n$ and then to use this estimate as a substitute for the unknown f in (3), leading to an empirical Bayes estimate of the form $t(x; X_1, \cdots, X_n)$. One such estimate which he proposed is

$$(4) \qquad t(x; X_1, \cdots, X_n) = (x + 1)N_{x+1}/N_x,$$

where N_x denotes the number of $X_1, \cdots, X_n$ values that are equal to x (cf. [41]). He also showed that the average squared error of estimation in using (4) is asymptotically (as $n \to \infty$) the same as that using the optimal Bayes estimate (3) (cf. [41], [59]).

The same ideas are applicable to the so-called "compound decision problems," in which the θ_i are regarded as unknown constants rather than as an unobservable random sample from a distribution B, and the overall loss function is the sum of the loss functions of the individual problems (cf. [25], [39], [52]). Robbins' asymptotically subminimax solutions of compound testing problems [25] and Stein's (1956) estimator of a multivariate normal mean are pioneering landmarks in compound decision theory.

It is worth noting the paradoxical feature of both empirical Bayes and compound decision methods that data "unrelated" to the parameter θ_i, namely the X_j for $j \neq i$, are used to form the estimate of θ_i.

Following the pioneering paper [41], Robbins made a number of major advances in the empirical Bayes methodology in [58], [59], [103], [115], [117], [118], [123], [124], [133]. In particular, in [103], [115], and [124] he extended the empirical Bayes methodology to prediction problems. Suppose in the above example that one is interested in predicting the total number S_0 of accidents in a future year incurred by those N_0

drivers who were accident-free in the observed year. Clearly 0 ("future same as past") will be an underestimate of S_0 since the good record of the n_0 drivers was partly due to luck. On the other hand, $N_0(\sum_x xN_x/N)$ will be an overestimate of S_0, since these N_0 accident-free drivers were in no sense a random sample of all the $N = \sum N_x$ drivers in the group, and should therefore perform better than the group average $\sum xN_x/N$. Under the Poisson model (2) for each driver, Robbins [103] showed that a natural empirical Bayes predictor is the number N_1 of drivers who had exactly one accident during the observed year. More generally, using (4), he showed that $(x+1)N_{x+1}$ is a good predictor of the total number S_x of accidents in a future year incurred by those N_x drivers who had x accidents in the observed year. He also obtained the asymptotic distribution of this predictor for the construction of significance tests and prediction intervals.

2. STOCHASTIC APPROXIMATION AND ADAPTIVE DESIGN

In 1951, Robbins and his student, Sutton Monro, founded the subject of stochastic approximation with the publication of their celebrated paper [26]. Consider the problem of finding the root θ (assumed unique) of an equation $g(x) = 0$. In the classical Newton-Raphson method, for example, we start with some value x_0, and if at stage n our estimate of the root is x_n, we define the next approximation by

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

Suppose that g is unknown and that at the level x we observe the output $y = g(x) + \varepsilon$, where ε represents some random error with mean 0 and variance $\sigma^2 > 0$. Suppose that in analogy with (4) we use the recursion

$$x_{n+1} = x_n - y_n/\beta, \tag{6}$$

where $\beta = g'(\theta)$ is momentarily assumed known, and $y_n = g(x_n) + \varepsilon_n$ is the observed output at the design level x_n. By (6), the convergence of x_n to θ would entail the convergence of ε_n to 0, which does not hold for typical models of random noise (e.g., independent and identically distributed ε_n).

Assuming that $g(x) > 0$ if $x > \theta$ and $g(x) < 0$ if $x < \theta$, Robbins and Monro suggested using instead of (6) the recursion

$$x_{n+1} = x_n - a_n y_n, \tag{7}$$

where a_n are positive constants such that

$$\sum a_n^2 < \infty \quad \text{and} \quad \sum a_n = \infty. \tag{8}$$

Under certain assumptions on the random errors ε_n, they showed that x_n converges to θ in L_2. Later, Blum (1954) showed that x_n also converges a.s. to θ, while Chung (1954) and Sacks (1958) showed that an asymptotically optimal choice of a_n is $a_n \sim 1/(n\beta)$, for which

$$n^{1/2}(x_n - \theta) \xrightarrow{D} N(0, \sigma^2/\beta^2). \tag{9}$$

The recent joint papers [113], [114], [119] with Lai provide asymptotically optimal stochastic approximation schemes when these recursions are of the form

$$x_{n+1} = x_n - y_n/(nb_n), \tag{10}$$

where b_n is a strongly consistent estimate of β. It is shown in [113] and [122] that such schemes not only lead to asymptotically efficient estimates of θ but also make the cumulative "cost" of the observations at the nth stage, defined as $\beta^2 \sum_1^n (x_i - \theta)^2$, to be of an asymptotically minimal order $\sigma^2 \log n$. The significance of this result in adaptive control applications is discussed in [107] and [121].

To construct strongly consistent estimates b_n of β, it is natural to try the method of least squares, thus using

$$b_n = \sum_1^n (x_i - \bar{x}_n)y_i \Big/ \sum_1^n (x_i - \bar{x}_n)^2, \tag{11}$$

at least in the linear case where $g(x) = \beta(x - \theta)$. This leads to the question of strong consistency of least squares estimates in regression models. Basic results on strong consistency are established in [105], [110], and [111] for the case where the design levels are nonrandom, and in [119] for the case where the design levels are sequentially determined random variables. In particular, it is shown in [119] that in the linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$ with independent identically distributed errors ε_i such that $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma^2$, and such that x_i is measurable with respect to the σ-field $\mathscr{F}_{i-1}$ generated by $x_1, \varepsilon_1, \cdots, \varepsilon_{i-1}$, a sufficient condition for the strong consistency of b_n is

$$\sum_1^n (x_i - \bar{x}_n)^2/\log n \to \infty \quad \text{a.s.} \tag{12}$$

This condition is shown in [119] to be minimal in some sense, and a counterexample showing inconsistency of b_n is constructed when (12) is only marginally violated. The result was subsequently extended by Lai and Wei (1982) to the multiple regression model $y_i = \theta' \varphi_i + \varepsilon_i$, where the φ_i are $\mathscr{F}_{i-1}$-measurable random vectors. In analogy with (12), a sufficient condition for the strong consistency of the least squares estimate in the multiple regression model is shown by Lai and Wei (1982) to be

$$\lambda_{\min}\left(\sum_1^n \varphi_i \varphi_i'\right) \Big/ \log \lambda_{\max}\left(\sum_1^n \varphi_i \varphi_i'\right) \to \infty \quad \text{a.s.} \tag{13}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues, respectively.

Condition (12), however, cannot be satisfied by adaptive stochastic approximation schemes (10) when $b_n \to \beta$ a.s. (cf. [113], [119]). The following sharper result is developed in [119] and is used to establish the strong consistency of b_n in adaptive stochastic approximation schemes (10): Suppose that the design levels x_n are $\mathscr{F}_{n-1}$-measurable and that there exist constants $0 < \gamma < 1/2$ and θ such that with probability 1
(a) $x_n - \theta = o(n^{-\gamma})$, and
(b) $\liminf_{n\to\infty} \sum_1^n (x_i - \theta)^2/\log n > 0$.

Then $b_n \to \beta$ a.s. on the event

$$\left\{\limsup_{n\to\infty} \sum_1^n (x_i - \theta)^2/\log n < \infty\right\} \cup \left\{\lim_{n\to\infty} \sum_1^n (x_i - \theta)^2/\log n = \infty\right\}.$$

Another contribution of Robbins to stochastic approximation is his joint paper [87] with Siegmund. The paper provides a convergence theorem (for nonnegative almost supermartingales) that can be easily applied to establish the almost sure convergence of stochastic approximation schemes and other recursive stochastic algorithms. In addition, his joint paper [88] with Goodman and Lewis proposes certain modifications of the Robbins-Monro method for quantal response models.

The Robbins-Monro stochastic approximation scheme can be readily modified to find the maximum (or minimum) of a regression function, as was first shown by Kiefer and Wolfowitz (1952). The scheme and its multivariate extensions also play a basic role in the subject of recursive estimation (cf. Sakrison, 1965; Nevel'son and Has'minksii, 1972; Fabian, 1978). Stochastic approximation has become the prototype of a broad class of control and optimization methods in various branches of engineering (cf. Tsypkin, 1973; Ljung, 1977; Kushner and Clark, 1978; Goodwin et al., 1981).

3. TESTS OF POWER ONE AND RELATED BOUNDARY CROSSING PROBABILITIES

In the series of papers [70]–[72], [77]–[83], [90]–[92], and [95], Robbins, in collaboration first with Darling and then with Siegmund, developed the theory of power one tests and made significant methodological advances in the treatment of related boundary crossing probabilities.

Let $X_1, X_2, \cdots$ be independent, identically distributed observations the common density $p(x \mid \theta)$ of which depends on an unknown parameter θ. Suppose that we want to test the hypothesis $H_0: \theta \le \theta_0$ versus the alternative $H_1: \theta > \theta_0$ with prescribed Type I error probability a. Let $U_n (= U_n(X_1, \cdots, X_n))$ be a sequence of statistics such that

$$(14) \quad P_\theta\{U_n > \theta_0 \text{ for some } n \ge 1\} \le \alpha \quad \text{for all} \quad \theta \le \theta_0,$$

and

$$(15) \quad P_\theta\left\{\lim_{n\to\infty} U_n = \theta\right\} = 1 \quad \text{for all} \quad \theta > \theta_0.$$

Consider the test which stops sampling at stage

$$N = \inf\{n \ge 1 : U_n > \theta_0\} \quad (\inf \varnothing = \infty),$$

and which rejects H_0 upon stopping. Then by (14), for $\theta \le \theta_0$,

$$P_\theta\{\text{Reject } H_0\} = P_\theta\{N < \infty\} \le a.$$

On the other hand, for $\theta > \theta_0$, it follows from (15) that

$$P_\theta\{\text{Reject } H_0\} = P_\theta\{N < \infty\} = 1.$$

Of course this procedure is not a statistical test in the usual sense of that term, because under the null hypothesis the statistican will usually continue to collect data forever and never reach a decision. To see that such a possibility is conceptually interesting, imagine that a drug is licensed for use based on a clinical trial indicating a positive treatment effect of the drug, but some concern exists that it may have deleterious side effects which will only become apparent after much more extensive use. A number of doctors agree to monitor the frequency of these side effects. If this frequency is large, some action is required—preferably as soon as possible. If the frequency of side effects is small, monitoring might continue indefinitely. Such a monitoring procedure might be realized by a test of power one. In reality, even if the null hypothesis is true, the procedure may still be terminated after some time; but the fact that this termination point need not be specified in advance and might even to some extent be data-dependent suggests fascinating theoretical and practical possibilities.

The actual construction of power one tests satisfying the basic constraints (14) and (15) and making $E_\theta(N)$ small in some sense whenever $\theta > \theta_0$ involves the development of a variety of techniques in the analysis of boundary crossing probabilities. One particularly beautiful inequality developed in this context is the following. Let $p(x \mid \theta) = \varphi(x - \theta)$, where φ is the standard normal density function (with distribution function Φ), and let $S_n = X_1 + \cdots + X_n$. Then for arbitrary positive m and a

$$(16) \quad \begin{aligned} &P_0\{|S_n| > [n(\log(n/m) + a^2]^{1/2} \\ &\quad \text{for some} \quad n \ge m\} < 2[1 - \Phi(a) + a\varphi(a)]. \end{aligned}$$

To prove (16) let $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$ be an increasing sequence of σ-fields, let P be a probability, and Q a σ-finite measure on $\mathscr{F}$ the restrictions to $\mathscr{F}_n$ of which, say $P^{(n)}$ and $Q^{(n)}$, are mutually absolutely continuous, and let L_n denote the likelihood ratio $dQ^{(n)}/dP^{(n)}$. Then for any stopping time $T \geq m$

$$P\{T < \infty\} = P\{T = m\} + \sum_{n=m+1}^{\infty} \int_{\{T=n\}} dP^{(n)}$$
$$= P\{T = m\} + \sum_{n=m+1}^{\infty} \int_{\{T=n\}} L_n^{-1}\, dQ,$$

which yields the identity

$$(17) \quad P\{T < \infty\} = P\{T = m\} + \int_{\{m<T<\infty\}} L_T^{-1}\, dQ.$$

Let $P = P_0$ and

$$Q = \int_{-\infty}^{\infty} P_\theta\, d\theta/(2\pi)^{1/2}.$$

A simple calculation gives

$$L_n = n^{-1/2}\exp(S_n^2/2n).$$

Let

$$T = \inf\{n : n \geq m,\ |S_n| \geq [n(\log(n/m) + a^2)]^{1/2}\}$$
$$= \inf\left\{n : n \geq m,\ L_n \geq m^{-1/2}\exp\left(\frac{1}{2}\, a^2\right)\right\},$$

and observe that $Q\{T = \infty\} = \int P_\theta\{T = \infty\}\, d\theta/(2\pi)^{1/2} = 0$, so (17) specializes to

$$(18) \quad P_0\{T < \infty\} = P_0\{|S_m| \geq am^{1/2}\} + \int_{\{|S_m|<am^{1/2}\}} T^{1/2}\exp(-S_T^2/2T)\, dQ.$$

From the definition of T it follows that

$$(19) \quad T^{1/2}\exp(-S_T^2/2T) \leq m^{1/2}\exp(-a^2/2)$$

whenever $T < \infty$. Substitution of this inequality into (18) and some simple calculation yield (16).

This method of mixtures of likelihood ratios has been described in [82] and [83], where it is used to obtain various inequalities like (16). The only source of inequality in (16) comes from (19). Lai and Siegmund (1977) have given the P_θ-limiting distribution, as a and $m \to \infty$, of the ratio of the two sides of (19), from which they derive an asymptotic approximation for the left-hand side of (16). Starting from a slight variation of (18) they also give an approximation to the significance level of a repeated significance test, viz.

$$P_0\{|S_n| \geq an^{1/2} \text{ for some } m_0 \leq n \leq m_1\}$$
$$\sim 2[1 - \Phi(a)] + a\varphi(a) \int_{am_1^{-1/2}}^{am_0^{-1/2}} [v(x)/x]\, dx,$$

where $v(x) = 2x^{-2}\exp[-2 \sum_1^\infty n^{-1}\Phi(-\frac{1}{2}xn^{1/2})]$. For further developments along these lines, see Siegmund (1977), Woodroofe (1982, 1983), Lalley (1983), and Siegmund (1985). An application of tests of power one to the problem of sequentially detecting a change point is given by Pollak (1985).

Robbins' idea in the sixties of terminating a test only when there is enough evidence against the null hypothesis was a characteristic example of his daring originality. Neyman (1971) gave a discussion of this revolutionary idea and described the theory of power one tests as "a remarkable achievement likely to influence theoretical-statistical and also substantive research in many domains of science."

4. SEQUENTIAL EXPERIMENTATION AND OPTIMAL STOPPING

The well known "multiarmed bandit problem" in the statistics and engineering literature, which is prototypical of a wide variety of adaptive control and design problems, was first formulated and studied by Robbins [28]. Let A, B denote two statistical populations with finite means μ_A, μ_B. How should we draw a sample $x_1, \cdots, x_n$ from the two populations if our objective is to achieve the greatest possible expected value of the $S_n = x_1 + \cdots + x_n$? Robbins [28] provided a rule which is asymptotically optimal in the sense that

$$n^{-1}ES_n \to \max(\mu_A, \mu_B) \quad \text{as} \quad n \to \infty.$$

His recent papers [129], [131], and [132] with Lai develop better rules that have the stronger property

$$n \max(\mu_A, \mu_B) - ES_n \sim C_{A,B}\log n \quad \text{as} \quad n \to \infty,$$

where $C_{A,B}$ is minimal over all allocation rules. The method also works for $k > 2$ populations. In [42], he formulated and studied the so-called "bandit problem with finite memory," in which the decision at every stage can depend on the information from no more than m previous stages, and which has useful applications in computer learning algorithms (cf. Lakshmivarahan, 1981).

The papers [89], [96], and [97] provide important advances in the related problem of treatment allocation in sequential clinical trials to test whether a new treatment is better than a standard treatment. A key observation in these papers is that for the stopping and terminal decision rules under consideration, the power function remains approximately the same over a broad class of allocation rules. Adaptive rules that substantially reduce the expected number of allocations of the inferior treatment as compared to pairwise sampling are developed. These results are of great interest because of the ethical considerations in clinical trials.

Anscombe (1963) proposed a decision-theoretic model to determine the stopping rule of a comparative clinical trial on paired data. The development of optimal stopping rules in Anscombe's model remained an important open problem for a long time. The papers [116] and [126] provide a class of stopping rules that have nearly optimal frequentist and Bayesian properties.

The monograph [86] with Chow and Siegmund presents the general theory of optimal stopping. The earlier papers [49], [55], [61], [62], [66], [67], and [85] represent fundamental contributions to the subject. Other major contributions to sequential analysis include sequential estimation [44], [64], [68], [73], [84], [98], [99], sequential selection [74], [120], and the theory of randomly stopped sums [15], [17], [56], [63], [65].

Although this brief survey has concentrated on Robbins' research in mathematical statistics and closely related aspects of probability theory, there should also be some mention of his work of a purely probabilisitic nature. In [12] with P. L. Hsu, he formulated the notion of complete convergence. This work motivated the classical paper of Erdös (1949), which in turn led to a long sequence of papers, reaching its culmination with Baum and Katz (1965). In the early 1950s Robbins wrote a series of important papers [30]–[33] on occupation times and ergodic behavior of random walks. One well known contribution is the Kallianpur-Robbins law for the occupation time of two-dimensional Brownian motion [30]. Another is his contribution together with T. E. Harris to the ergodic theory of Markov chains having an invariant distribution of infinite total mass [32].

Robbins' outstanding and prodigious research in the past 40 years has pushed the field of mathematical statistics to new heights and in important new directions. He has played a seminal role in stimulating a substantial proportion of current research in the field, and has created a variety of new specialties for subsequent generations of statisticians to explore. In addition to these accomplishments in research, he is also a superb teacher and gifted lecturer. His lectures are, like his papers, models of exposition that introduce the audience, in the simplest possible context, to important and profound ideas at the forefront of research in the field. He is widely admired by his colleagues and students for his extraordinary depth and clarity, his quick wit and lively humor, and above all, his creative intellect and exceptional originality. A very generous, warm, and caring person who is deeply involved in humanitarian causes and who has invested considerable time and effort to help refugee mathematicians from Eastern Europe become settled in the United States, he is held in high esteem and affection by his friends and colleagues.

REFERENCES

ANSCOMBE, F. J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58** 365–383.

BAUM, L. and KATZ, L. (1965). Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.* **120** 108–123.

BLUM, J. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 382–386.

CARTER, G. and ROLPH, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* **69** 880–885.

CHUNG, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25** 463–483.

COPAS, J. B. (1972). Empirical Bayes methods and the repeated use of a standard. *Biometrika* **59** 349–360.

COVER, T. (1968). Learning in pattern recognition. In *Methodologies of Pattern Recognition* (S. Watanabe, ed.) 111–132. Academic, New York.

EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Sci. Amer.* **236** 119–127.

ERDÖS, P. (1949). On a theorem of Hsu and Robbins. *Ann. Math. Statist.* **20** 286–291.

FABIAN, V. (1978). On asymptotically efficient recursive estimation. *Ann. Statist.* **6** 854–866.

GOODWIN, G. C., RAMADGE, P. J. and CAINES, P. E. (1981). Discrete time stochastic adaptive control. *SIAM J. Control Optimization* **19** 829–853.

HOADLEY, B. (1981). Quality management plan (QMP). *Bell System Tech. J.* **60** 215–273.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* **23** 462–466.

KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Springer, New York.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

LAKSHMIVARAHAN, S. (1981). *Learning Algorithms: Theory and Applications.* Springer, New York.

LALLEY, S. (1983). Repeated likelihood ratio tests for curved exponential families. *Z. Wahrsch. verw. Gebiete* **62** 293–321.

LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control* **AC-22** 551–575.

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–55.

NEVEL'SON, M. B. and HAS'MINSKII, R. Z. (1972). *Stochastic Approximation and Recursive Estimation.* Amer. Math. Soc. Translations, Providence.

NEYMAN, J. (1962). Two breakthroughs in the theory of statistical decision-making. *Rev. Internat. Statist. Inst.* **30** 11–27.

NEYMAN, J. (1971). Foundations of behavioristic statistics. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 1–19. Holt, Rinehart, & Winston, Toronto.

PAGE, W. (1984). An interview with Herbert Robbins. *College Math. J.* **15** 2–24.

POLLAK, M. (1985). Average lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **13** 206–227.

RUBIN, D. (1980). Using empirical Bayes techniques in the Law School Validity Studies. *J. Amer. Statist. Assoc.* **75** 801–827.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

SAKRISON, D. J. (1965). Efficient recursive estimation: Application to estimating the parameters of a covariance function. *Internat. J. Engrg. Sci.* **3** 461–483.

SIEGMUND, D. (1977). Repeated significance tests for a normal mean. *Biometrika* **64** 177–189.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206.

TSYPKIN, Ya. Z. (1973). *Foundations of the Theory of Learning Systems.* Academic, New York.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* SIAM, Philadelphia.

WOODROOFE, M. (1983). On sequential rank tests. In *Recent Advances in Statistics* (M. H. Rizvi, J. Rustagi, and D. Siegmund, eds.) 115–142. Academic, New York.

Publications of Herbert Robbins

1. On a class of recurrent sequences. *Bull. Amer. Math. Soc.* **41** 413–417, 1935.
2. A theorem on graphs. *Amer. Math. Monthly* **46** 281–283, 1939.
3. On the classification of the mappings of a 2-complex. *Trans. Amer. Math. Soc.* **49** 308–324, 1941.
4. *What is Mathematics?* (with R. Courant). Oxford Univ. Press, New York, 1941.
5. A note on the Riemann integral. *Amer. Math. Monthly* **50** 617–618, 1943.
6. Two properties of the function cos x. *Bull. Amer. Math. Soc.* **50** 750–752, 1944.
7. On the measure of a random set. *Ann. Math. Statist.* **15** 70–74, 1944; **18** 297, 1947.
8. On distribution-free tolerance limits in random sampling. *Ann. Math. Statist.* **15** 214–216, 1944.
9. On the expected values of two statistics. *Ann. Math. Statist.* **15** 321–323, 1944.
10. On the measure of a random set II. *Ann. Math. Statist.* **16** 342–347, 1945.
11. On the $(C, 1)$ summability of certain random sequences. *Bull. Amer. Math. Soc.* **52** 699–703, 1946.
12. Complete convergence and the law of large numbers (with P. L. Hsu). *Proc. Nat. Acad. Sci. U. S. A.* **33** 25–31, 1947.
13. Some remarks on the inequality of Tchebychef. In *Studies and Essays* (K. O. Friedrichs et al., eds.) 345–350. Interscience, New York, 1948.
14. Convergence of distributions. *Ann. Math. Statist.* **19** 72–76, 1948.
15. On the asymptotic distribution of the sum of a random number of random variables. *Proc. Nat. Acad. Sci. U. S. A.* **34** 162–163, 1948.
16. The central limit theorem for dependent random variables (with Wassily Hoeffding). *Duke Math. J.* **15** 773–780, 1948.
17. The asymptotic distribution of the sum of a random number of random variables. *Bull. Amer. Math. Soc.* **54** 1151–1161, 1948.
18. The distribution of a definite quadratic form. *Ann. Math. Statist.* **19** 266–270, 1948.
19. Mixture of distributions. *Ann. Math. Statist.* **19** 360–369, 1948.
20. The distribution of Student's t when the population means are unequal. *Ann. Math. Statist.* **19** 406–410, 1948.
21. Application of the method of mixtures to quadratic forms in normal variates (with E. J. G. Pitman). *Ann. Math. Statist.* **20** 552–560, 1949.
22. Competitive estimation. *Ann. Math. Statist.* **21** 311, 1950.
23. A generalization of the method of maximum likelihood: estimating a mixing distribution. *Ann. Math. Statist.* **21** 314, 1950.
24. The problem of the greater mean (with R. R. Bahadur). *Ann. Math. Statist.* **21** 469–487, 1950; **22** 310, 1951.
25. Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 131–148, 1951.
26. A stochastic approximation method (with Sutton Monro). *Ann. Math. Statist.* **22** 400–407, 1951.
27. Minimum variance estimation without regularity assumptions (with D. G. Chapman). *Ann. Math. Statist.* **22** 581–586, 1951.
28. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535, 1952.
29. A note on gambling systems and birth statistics. *Amer. Math. Monthly* **59** 685–686, 1952.
30. Ergodic property of the Brownian motion process (with G. Kallianpur). *Proc. Nat. Acad. Sci. U. S. A.* **39** 525–533, 1953.
31. On the equidistribution of sums of independent random variables. *Proc. Amer. Math. Soc.* **4** 786–799, 1953.
32. Ergodic theory of Markov chains admitting an infinite invariant measure (with T. E. Harris). *Proc. Nat. Acad. Sci. U. S. A.* **39** 860–864, 1953.
33. The sequence of sums of independent random variables (with G. Kallianpur). *Duke Math. J.* **21** 285–308, 1954.
34. Two-stage procedures for estimating the difference between means (with S. G. Ghurye). *Biometrika* **41** 146–152, 1954.
35. A note on information theory. *Proc. Inst. Radio Engrg.* **42** 1193, 1954.
36. A remark on the joint distribution of cumulative sums. *Ann. Math. Statist.* **25** 614–616, 1954.
37. A one-sided confidence interval for an unknown distribution function. *Ann. Math. Statist.* **25** 409, 1954.
38. A remark on Stirling's formula. *Amer. Math. Monthly* **62** 26–29, 1955.
39. Asymptotic solutions of the compound decision problem for two completely specified distributions (with J. F. Hannan). *Ann. Math. Statist.* **26** 37–51, 1955.
40. The strong law of large numbers when the first moment does not exist (with C. Derman). *Proc. Nat. Acad. Sci. U. S. A.* **41** 586–587, 1955.
41. An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163, 1956.
42. A sequential decision problem with a finite memory. *Proc. Nat. Acad. Sci. U. S. A.* **42** 920–923, 1956.
43. The theory of probability. In *Insights into Modern Mathematics* (F. L. Wren, ed.) 336–371. National Council of Teachers of Mathematics, Washington, D. C., 1957.
44. Sequential estimation of the mean of a normal population. In *Probability and Statistics* (U. Grenander, ed.) 235–245. Almqvist and Wicksell, Stockholm, 1959.
45. Probability. In *Spectrum* (R. Ginger, ed.) 100–114. Henry Holt, New York, 1959.
46. Comments on a paper by James Albertson. *J. Bibl. Lit.* **78** 347–350, 1959.
47. A statistical screening problem. In *Contributions to Probability and Statistics* (I. Olkin et al., eds.) 352–357. Stanford Univ. Press, Stanford, Calif., 1960.
48. On sums of independent random variables with infinite moments and 'fair' games (with Y. S. Chow). *Proc. Nat. Acad. Sci. U. S. A.* **47** 330–335, 1961.
49. A martingale system theorem and applications (with Y. S.

Chow). *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 93–104, 1961.

50. Recurrent games and the Petersburg paradox. *Ann. Math. Statist.* **32** 187–194, 1961.
51. A Bayes test of p < ½ versus p > ½ (with S. Moriguti). *Rep. Statist. Appl. Res. Un. Japan Sci. Engrs.* **9** 39–60, 1962.
52. Some numerical results on a compound decision problem. In *Recent Developments in Information Theory and Decision Processes* (R. E. Machol and P. Gray, eds.) 56–62. Macmillan, New York, 1962.
53. Testing statistical hypotheses; the compound approach (with E. Samuel). In *Recent Developments in Information Theory and Decision Processes* (R. E. Machol and P. Gray, eds.) 63–70. Macmillan, New York, 1962.
54. A new approach to a classical statistical decision problem. In *Induction: Some Current Issues* (H. E. Kyburg, Jr. and E. Nagel, eds.) 101–110. Wesleyan Univ. Press, Middletown, Conn., 1963.
55. On optimal stopping rules (with Y. S. Chow). *Z. Wahrsch. verw. Gebiete* **2** 33–49, 1963.
56. A renewal theorem for random variables which are dependent or non-identically distributed (with Y. S. Chow). *Ann. Math. Statist.* **34** 390–395, 1963.
57. Some problems of optimal sampling strategy (with C. L. Mallows). *J. Math. Anal. Appl.* **6** 1–14, 1963.
58. The empirical Bayes approach to testing statistical hypotheses. *Rev. Internat. Statist. Inst.* **31** 182–195, 1963.
59. The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20, 1964.
60. On the 'parking' problem (with A. Dvoretzky). *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **9** 209–225, 1964.
61. Optimal selection based on relative rank—the 'Secretary Problem' (with Y. S. Chow, S. Moriguti, and S. M. Samuels). *Israel J. Math.* **2** 81–90, 1964.
62. On optimal stopping rules for S_n/n (with Y. S. Chow). *Illinois J. Math.* **9** 444–454, 1965.
63. Moments of randomly stopped sums (with Y. S. Chow and H. Teicher). *Ann. Math. Statist.* **36** 789–799, 1965.
64. On the asymptotic theory of fixed-width sequential confidence intervals for the mean (with Y. S. Chow). *Ann. Math. Statist.* **36** 457–462, 1965.
65. An extension of a lemma of Wald (with E. Samuel). *J. Appl. Probability* **3** 272–273, 1966.
66. A class of optimal stopping problems (with Y. S. Chow). *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 419–426, 1967.
67. On values associated with a stochastic sequence (with Y. S. Chow). *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 427–440, 1967.
68. Finding the size of a finite population (with D. A. Darling). *Ann. Math. Statist.* **38** 1392–1398, 1967.
69. Some complements to Brouwer's fixed point theorem. *Israel J. Math.* **5** 225–226, 1967.
70. Iterated logarithm inequalities (with D. A. Darling). *Proc. Nat. Acad. Sci. U. S. A.* **57** 1188–1192, 1967.
71. Inequalities for the sequence of sample means (with D. A. Darling). *Proc. Nat. Acad. Sci. U. S. A.* **57** 1577–1580, 1967.
72. Confidence sequences for mean, variance, and median (with D. A. Darling). *Proc. Nat. Acad. Sci. U. S. A.* **58** 66–68, 1967.
73. A sequential analogue of the Behrens-Fisher problem (with G. Simons and N. Starr). *Ann. Math. Statist.* **38** 1384–1391, 1967.
74. A sequential procedure for selecting the largest of k means (with M. Sobel and N. Starr). *Ann. Math. Statist.* **39** 88–92, 1968.
75. Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Statist.* **39** 256–257, 1968.
76. The limiting distribution of the last time $s_n > n\varepsilon$ (with D. Siegmund and J. Wendel). *Proc. Nat. Sci. U. S. A.* **61** 1228–1230, 1968.
77. Some further remarks on inequalities for sample sums (with D. A. Darling). *Proc. Nat. Acad. Sci. U. S. A.* **60** 1175–1182, 1968.
78. Some nonparametric tests with power one (with D. A. Darling). *Proc. Nat. Acad. Sci. U. S. A.* **61** 804–809, 1968.
79. Iterated logarithm inequalities and related statistical procedures (with D. Siegmund). In *Mathematics of the Decision Sciences* (G. B. Dantzig and A. F. Veinott, eds.) **2** 267–279. Amer. Math. Soc., Providence, 1968.
80. Probability distributions related to the law of the iterated logarithm (with D. Siegmund). *Proc. Nat. Acad. Sci. U. S. A.* **62** 11–13, 1969.
81. Confidence sequences and interminable tests (with D. Siegmund). *Bull. Internat. Statist. Inst.* **43** 379–387, 1969.
82. Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397–1409, 1970.
83. Boundary crossing probabilities for the Wiener process and sample sums (with D. Siegmund). *Ann. Math. Statist.* **41** 1410–1429, 1970.
84. Sequential estimation of an integer mean. In *Scientists at Work* (T. Dalenius et al., eds.) 205–210. Almqvist and Wicksell, Uppsala, 1970.
85. Optimal stopping. *Amer. Math. Monthly* **77** 333–343, 1970.
86. *Great Expectations: The Theory of Optimal Stopping* (with Y. S. Chow and D. Siegmund). Houghton Mifflin, Boston, 1971.
87. A convergence theorem for non-negative almost supermartingales and some applications (with D. Siegmund). In *Optimizing Methods in Statistics* (J. Rustagi, ed.) 233–257. Academic, New York, 1971.
88. Simultaneous estimation of large numbers of extreme quantiles in simulation experiments (with A. S. Goodman and P. A. W. Lewis). *IBM Res.* RC3621. Yorktown Heights, New York, 1971.
89. Reducing the number of inferior treatments in clinical trials (with B. Flehinger, T. Louis, B. Singer). *Proc. Nat. Acad. Sci. U. S. A.* **69** 2993–2994, 1972.
90. On the law of the iterated logarithm for maxima and minima (with D. Siegmund). *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3** 51–70, 1972.
91. A class of stopping rules for testing parametric hypotheses (with D. Siegmund). *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **4** 37–41, 1972.
92. Statistical tests of power one and the integral representation of solutions of certain partial differential equations (with D. Siegmund). *Bull. Inst. Math. Acad. Sinica* **1** 93–120, 1973.
93. Mathematical probability in election challenges (with M. O. Finkelstein). *Columbia Law Rev.* **73** 241–248, 1973.
94. The statistical mode of thought. In *The Heritage of Copernicus* (J. Neyman, ed.) 419–432. M.I.T. Press, Cambridge, Mass., 1974.
95. The expected sample size of some tests of power one (with D. Siegmund). *Ann. Statist.* **2** 415–436, 1974.
96. Sequential tests involving two populations (with D. Siegmund). *J. Amer. Statist. Assoc.* **69** 132–139, 1974.
97. A sequential test for two binomial populations. *Proc. Nat. Acad. Sci. U. S. A.* **71** 4435–4436, 1974.
98. Sequential estimation of p in Bernoulli trials (with D. Siegmund). In *Studies in Probability and Statistics* (E. J. Williams, ed.) 103–107. Jerusalem Acad. Press, 1974.
99. Sequential estimation of p with squared relative error loss (with P. Cabilio). *Proc. Nat. Acad. Sci. U. S. A.* **72** 191–193, 1975.

100. *Introduction to Statistics* (with J. Van Ryzin). Science Research Assoc., Palo Alto, Calif., 1975.
101. Wither mathematical statistics? *Adv. Appl. Probability* (*Suppl.*) **7** 116–121, 1975.
102. Maximally dependent random variables (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **73** 286–288, 1976.
103. Prediction and estimation for the compound Poisson distribution. *Proc. Nat. Acad. Sci. U. S. A.* **74** 2670–2671, 1977.
104. Sequential decision about a normal mean (with T. L. Lai and D. Siegmund). In *Statistical Decision Theory and Related Topics II* (S. S. Gupta, ed.) 213–221. Academic, New York, 1977.
105. Strong consistency of least-squares estimates in regression models (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **74** 2667–2669, 1977.
106. A class of dependent random variables and their maxima (with T. L. Lai). *Z. Wahrsch. verw. Gebiete* **42** 89–111, 1978.
107. Adaptive design in regression and control (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **75** 586–587, 1978.
108. Limit theorems for weighted sums and stochastic approximation processes (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **75** 1068–1070, 1978.
109. Review of "Adventures of a Mathematician" by S. M. Ulam. *Bull. Amer. Math. Soc.* **84** 107–109, 1978.
110. Strong consistency of least squares estimates in multiple regression (with T. L. Lai and C. Z. Wei). *Proc. Nat. Acad. Sci. U. S. A.* **75** 3034–3036, 1978.
111. Strong consistency of least squares estimates in multiple regression II (with T. L. Lai and C. Z. Wei). *J. Multivariate Anal.* **9** 343–361, 1979.
112. Sequential rank and the Polya urn (with J. Whitehead). *J. Appl. Probability* **16** 213–219, 1979.
113. Adaptive design and stochastic approximation (with T. L. Lai). *Ann. Statist.* **7** 1196 -1221, 1979.
114. Local convergence theorems for adaptive stochastic approximation schemes (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **76** 3065–3067, 1979.
115. Estimation and prediction for mixtures of the exponential distribution. *Proc. Nat. Acad. Sci. U. S. A.* **77** 2382–2383, 1980.
116. Sequential medical trials (with T. L. Lai, B. Levin, and D. Siegmund). *Proc. Nat. Acad. Sci. U. S. A.* **77** 3135–3138, 1980.
117. Some estimation problems for the compound Poisson distribution. In *Asymptotic Theory of Statistical Tests and Estimation* (I. M. Chakravarti, ed.) 251–257. Academic, New York, 1980.
118. An empirical Bayes estimation problem. *Proc. Nat. Acad. Sci. U. S. A.* **77** 6988–6989, 1980.
119. Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes (with T. L. Lai). *Z. Wahrsch. verw. Gebiete* **56** 329–360, 1981.
120. Selecting the highest probability in binomial or multinomial trials (with B. Levin). *Proc. Nat. Acad. Sci. U. S. A.* **78** 4663–4666, 1981.
121. Iterated least squares in multiperiod control (with T. L. Lai). *Adv. Applied Math.* **3** 50–73, 1982.
122. Adaptive design and the multi-period control problem (with T. L. Lai). In *Statistical Decision Theory and Related Topics III* (S. S. Gupta ed.) **2** 103–120. Academic, New York, 1982.
123. Estimating many variances. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta, ed.) **2** 251–261. Academic, New York, 1982.
124. Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11** 713–723, 1983.
125. A note on the underadjustment phenomenon (with B. Levin). *Statist. Probab. Lett.* **1** 137–139, 1983.
126. Sequential design of comparative clinical trials (with T. L. Lai and D. Siegmund). In *Recent Advances in Statistics* (J. Rustagi et al., eds.) 51–68. Academic, New York, 1983.
127. Adaptive choice of mean or median in estimating the center of a symmetric distribution (with T. L. Lai and K. F. Yu). *Proc. Nat. Acad. Sci. U. S. A.* **80** 5803–5806, 1983.
128. Some breakthroughs in statistical methodology. *College Math. J.* **15** 25–29, 1984.
129. Optimal sequential sampling from two populations (with T. L. Lai). *Proc. Nat. Acad. Sci. U. S. A.* **81** 1284–1286, 1984.
130. A probabilistic approach to tracing funds in the law of restitution (with M. O. Finkelstein). *Jurimetrics* **24** 65–77, 1984.
131. Asymptotically optimal allocation of treatments in sequential experiments (with T. L. Lai). In *Design of Experiments: Ranking and Selection* (T. Santner and A. Tamhane, eds.) 127–142. Marcel-Dekker, New York, 1984.
132. Asymptotically efficient adaptive allocation rules (with T. L. Lai). *Adv. Applied Math.* **6** 4–22, 1985.
133. Linear empirical Bayes estimation of means and variances. *Proc. Nat. Acad. Sci. U. S. A.* **82** 1571–1574, 1985.