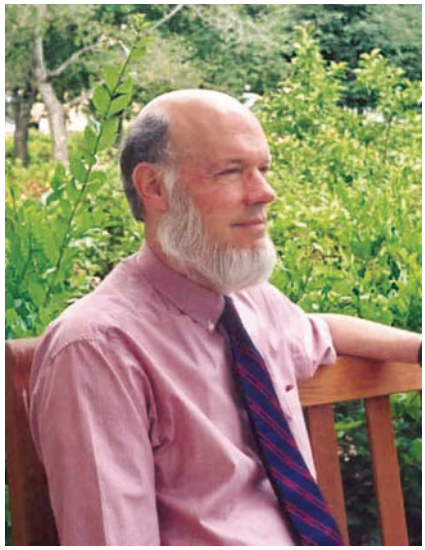
Biography of David O. Siegmund

David Siegmund, who holds the John D. and Sigrid Banks Chair at Stanford University, Stanford, CA, is a statistician who is comfortable in both the airy heights of theory and the practicalities of real-world applications. He works at the interface between probability and statistics, applying the tools he develops to topics as diverse as the design of medical clinical trials and mapping the locations of genes that are involved in specific physiological traits.

His work has earned him several awards, including a Guggenheim Fellowship in 1974, the Humboldt Prize in 1980, and membership in the American Academy of Arts and Sciences in 1994. In 2002 he was elected to the National Academy of Sciences. His Inaugural Article (1), published in this issue of PNAS, reviews recent methodological developments in quantitative trait locus mapping and addresses the problem of mapping with selected, rather than random, samples.



David O. Siegmund

Basketball and Mathematics

Siegmund grew up in Webster Groves, MO, site of a well known television documentary by Charles Kuralt titled "Sixteen in Webster Groves," in which Kuralt presented the town as the quintessential example of a suburban enclave. During childhood, Siegmund's two passions were mathematics and basketball. Like many children's environments, the popular crowd emphasized sports, not academics. "In high school, I found mathematics interesting," he said, "and surreptitiously, so as not to break the rules about the appearance of excessive mental exertion, I read parts of my father's college textbooks that I had found hidden away in a closet." A lasting influence from his high school days was 12th grade mathematics teacher George Brucker.

Unlike academics, it was acceptable to work hard at basketball, and Siegmund went to Southern Methodist University (Dallas, TX) to play basketball and study mathematics, "in that order." There he met his future wife and mother of his three children. He also learned, on the basketball court, "that hard work led to an improvement on one's natural talents, and more importantly that the same principle applied to academic work as well." Siegmund sampled several academic topics that he found interesting and was ultimately drawn to statistics, which "combined the beauty and challenge of mathematics with the possibility of continuing engagement with the natural sciences, social sciences, and even philosophy."

Siegmund received his B.A. in mathematics in 1963. Encouraged by his teachers, notably Paul Minton, and by the Woodrow Wilson and Danforth Fellowships for those interested in careers in college teaching, he went on to pursue a Ph.D. in statistics at Columbia University in New York. Columbia was, he said, "an ideal catalyst for my developing academic interests." His graduate advisor was Herbert Robbins, who was elected to the National Academy of Sciences in 1974. Together Siegmund and Robbins wrote more than a dozen papers (2–5), and Robbins "had a very large impact on my work," said Siegmund.

Bringing Theory to the Trenches

After receiving his Ph.D. in 1966, Siegmund stayed on as an assistant professor at Columbia. He relocated briefly to Stanford but returned to a full professorship at Columbia in 1971. After visiting professorships at Hebrew University (Jerusalem) and the University of Zurich, he moved back to Stanford in 1976 and attained the John D. and Sigrid Banks Professorship in 2002. He has twice been Chairman of Stanford's Department of Statistics, and from 1993 to 1996 served as Associate Dean of Stanford's School of Humanities and Sciences. Siegmund also has held visiting positions at the University of Oxford and the University of Cambridge. In 1971 he wrote his first book, with Herbert Robbins and Y. S. Chow, on the theory of optimal stopping (6). The theory deals with a class of sequential decision problems where one must choose between taking an action based on one's current state of knowledge and continuing to accumulate information in the hopes of taking a more appropriate action based on more knowledge later.

As a statistician also interested in probability theory, Siegmund's research has concentrated on statistical problems that arise in scientific applications and require novel probability theory for their resolution. Before 1985, his research concentrated on sequential analysis: the study of how data should be accumulated in an experimental situation. Siegmund's primary focus was on the design and analysis of sequential clinical trials that allow pharmaceutical workers to assess whether a new medicine is better or worse than an existing one (7–13).

"My contribution was somehow to try to bring theory a little more to bear on that approach," he said. "I tried to open up a discourse between people who were more theoretically inclined and the people who were down in the trenches actually running clinical trials." In particular, Siegmund helped answer how to decide when sequentially accumulating data are sufficient to reach a sound conclusion, and how to avoid biases in estimating a treatment effect when one stops a trial on the basis of apparently favorable (or unfavorable) outcomes.

Siegmund tried to open up a discourse between theoreticians and clinical researchers.

At the time, many scientists believed that sequential clinical trials were not useful because it was unclear how to form an estimate of the treatment effect after the trial was ended. "I guess I thought that it was a somewhat more difficult problem but not an unsolvable problem," Siegmund said. Ultimately, Siegmund was able to bridge the gap

This is a Biography of a recently elected member of the National Academy of Sciences to accompany the member's Inaugural Article on page 7845.

between theory and practice by developing a theory of sequential clinical trials and estimates of treatment effects associated with those trials. This work led to publication of his second book, in 1985, on sequential analysis (12), which he calls a "major accomplishment" in his career.

A related interest of Siegmund's is "changepoint" detection, where one looks for changes in the output of a process that signals an underlying change in the process itself. Historically the first applications involved changes that indicate a deterioration in the quality of industrial processes, but similar methods apply to monitoring changes in, for example, the frequencies of birth defects or incidence of diseases. Siegmund's interest in these applications led to several years of research on changepoint-like problems and related problems of nonlinear regression (13–15).

Mapping the Future

For the last 10 years, Siegmund has concentrated on statistical aspects of gene mapping. He notes, from the contemporary viewpoint of a "genome scan," where one uses hundreds of genetic markers at known locations throughout the genome to search for genes of interest, that gene mapping is very similar to changepoint problems, with the location of the gene as changepoint (16–18). Siegmund also has used similar techniques to analyze algorithms for pairwise sequence alignments for DNA and amino acid sequences (19).

Siegmund points to the mapping of quantitative traits in humans as an area of rapid new development. A particularly thorny issue is how to account for ascertainment bias, i.e., bias that occurs when one studies a trait based on a sample of individuals having particular phenotypes as contrasted with a random sample from the population. For example, many quantitative traits, such as blood pressure and "good" or "bad" cholesterol, are of interest because they are associated with certain diseases and are often studied by sampling pedigrees containing one or more individuals who have the associated disease. In his Inaugural Article (1), Siegmund reviews recent research based on the assumption of random sampling and demonstrates that two suggested methods to correct for ascertainment bias are asymptotically equivalent when the number of pedigrees is large.

Biology and mathematics have been in successful collaboration for approximately 20 years now, and it is an area that Siegmund finds himself coming back to again and again. "Gene mapping has changed quite a bit during the last 20 years, largely because of the different amount and kind of data that people have been able to obtain," he said. The challenge is great, but so is the reward. "There are still many things about these subjects that I don't understand—computational biology, bioinformatics, statistical genetics—the problems are sometimes extremely difficult but always very interesting," said Siegmund. "I would like to find one or two more that I care about as both a mathematical problem and as a scientific problem and to which I can contribute a piece of a solution."

David Appell, *Freelance Science Writer*

1. Peng, J. & Siegmund, D. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 7845–7850.
2. Robbins, H. & Siegmund, D. (1968) *Proc. Natl. Acad. Sci. USA* **62,** 11–13.
3. Robbins, H. & Siegmund, D. (1970) *Ann. Math. Stat.* **41,** 1410–1429.
4. Robbins, H. & Siegmund, D. (1972) in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, eds. Le Cam, L. M., Neyman, J. & Scott, E. L. (Univ. of California Press, Berkeley), Vol. IV, pp. 37–41.
5. Robbins, H. & Siegmund, D. (1974) *Ann. Stat.* **2,** 415–436.
6. Chow, Y. S., Robbins, H. & Siegmund, D. (1971) *Great Expectations: The Theory of Optimal Stopping* (Houghton Mifflin, Boston).
7. Siegmund, D. (1977) *Biometrika* **64,** 177–190.
8. Siegmund, D. (1978) *Biometrika* **65,** 341–349.
9. Lai, T. L. & Siegmund, D. (1979) *Ann. Stat.* **7,** 60–76.
10. Siegmund, D. (1980) *Biometrika* **67,** 389–402.
11. Siegmund, D. (1993) *Ann. Stat.* **21,** 464–483.
12. Siegmund, D. (1985) *Sequential Analysis: Tests and Confidence Intervals* (Springer, New York).
13. Siegmund, D. (1988) *Ann. Probab.* **16,** 487–501.
14. Venkatraman, E. S. & Siegmund, D. (1995) *Ann. Stat.* **23,** 255–271.
15. Siegmund, D. & Worsley, K. (1995) *Ann. Stat.* **23,** 608–639.
16. Feingold, E., Brown, P. & Siegmund, D. (1993) *Am. J. Hum. Genet.* **53,** 234–251.
17. Dupuis, J., Brown, P. & Siegmund, D. (1995) *Genetics* **140,** 843–856.
18. Tang, H.-K. & Siegmund, D. (2001) *Biostatistics* **2,** 147–162.
19. Siegmund, D. & Yakir, B. (2000) *Ann. Stat.* **28,** 657–680.