Reminiscences of the Columbia University Statistics Department in the late 1940s

Ingram Olkin
Stanford University

August 22, 2013

Every once in a while in a dinner conversation I have recalled my student days at Columbia, and have met with the suggestion that I write up these recollections. Although present-day students may recognize some of the famous names such as Hotelling, Wald, and Wolfowitz, they won't meet many faculty who were their students. The following is the result, and I hope the reader finds these reminiscences interesting. Because recollections of 60 years ago are often inaccurate, I urge readers to add to my comments.

I started City College (CCNY) in 1941 and in 1943 enlisted in the US Army Air Force meteorology program. After completion of the program I served as a meteorologist at various airports until I was discharged in 1946. I returned to CCNY and graduated in 1947, at which time I enrolled at Columbia University. As an aside, the professor at CCNY was Selby Robinson. Although not a great teacher, he somehow inspired a number of students to continue their study of statistics. Kenneth Arrow, Herman Chernoff, Milton Sobel, and Herbert Solomon are several who continued their studies at Columbia after graduating from CCNY.

After receiving a doctorate at Princeton, Harold Hotelling was at Stanford from 1924 to 1931, at the Food Research Institute and the Mathematics department. In 1927 he taught three courses at Stanford: mathematical statistics (among the very early faculty to teach a rigorous course in statistics), differential geometry, and topology (who would tackle this today?). In 1931 he moved to Columbia, where he wrote his most famous papers in economics and in statistics (principal components, canonical correlations, T^2, to mention but a few). His 1941 paper on the teaching of statistics had a phenomenal impact. Jerzy Neyman stated that it was one of the most influential papers in statistics. Faculty attempting to convince university administrators to form a department of statistics often used this paper as an argument why the teaching of statistics should be done by statisticians and not by faculty in substantive fields that use statistics. To read more about Hotelling, see Olkin and Sampson (2001a,b).

Hotelling had hired Abraham Wald, and when Hotelling left to be head of the statistics department at Chapel Hill, Wald became chair of the newly formed department at Columbia. The department was in the Division of Political Economy because the Mathematics Department objected to statistics being in the same division. The other faculty were Ted Anderson, Jack Wolfowitz, and Howard Levene; Helen Walker was in the School of Education. (Helen was one of a few well-known, influential women statisticians. One source states that she was the first woman to teach statistics.)

In the late 1940s Columbia, Chapel Hill, and Berkeley were statistical centers that attracted lots of visitors. There were other universities that had an impact in statistics such as Princeton, Iowa State, Iowa, Chicago, Stanford, and Michigan, but conferences were mostly held at the top three. The first two Berkeley Symposia were in 1946 and 1950, and these brought many visitors from around the world.

The Second Berkeley Symposium brought a galaxy of foreign statisticians to the US: Paul Lévy, Bruno de Finetti, Michel Loève, Harold Cramér, Aryeh Dvoretzky, and Robert Fortet. Domestic faculty were present as well, such as Richard Feynman, Kenneth Arrow, Jacob Marshak, Harold Kuhn, and Albert Tucker. Because some of the participants came from distant lands they visited other universities as part of the trip. I do recall that Loève visited Columbia. Because departments were small, visitors often taught special-topic courses. For example, I took a class with E.J.G. Pitman at Chapel Hill.

With the GI Bill I did not have to worry about tuition, and enrolled in two classes in the summer of 1947. The classes were crowded with post-war returnees. One class was a first course in mathematical statistics that was taught by Wolfowitz. Some of the students at Columbia during the 1947–1950 period were Raj Bahadur and Thelma Clark (later his wife), Bob Bechhofer, Allan Birnbaum, Al Bowker, Herman Chernoff (he was officially at Brown University, but worked with Wald), Herb T. David, Cyrus Derman, Sylvan Ehrenfeld, Harry Eisenpress, Peter Frank, Leon Herbach, Stanley Isaacson, Seymour Jablon, Jack Kiefer, Bill Kruskal, Gerry Lieberman, Gottfried Noether, Rosedith Sitgreaves, Milton Sobel, Herbert Solomon, Charles Stein, Henry Teicher, Lionel Weiss, and many others. Columbia statistics was an exciting place, and almost all of the students continued their career in statistics. There was a feeling that we were in on the ground floor of a new field, and in many respects we were. From 1950 to 1970 the *Annals of Mathematical Statistics* grew from 625 to 2200 pages, with many articles from the students of this era.

Some statistics classes were held at night starting at 5:40 and 7:30 so that students who worked during the day could get to class. However, math classes took place during the day. I took sequential analysis and analysis of variance from Wald, core probability from Wolfowitz, finite differences from B.O. Koopman, linear algebra from Howard Levi, differential equations from Ritt, and a computer science course at the Columbia Watson Labs.

Wald had a classic European lecture style. He started at the upper left corner of the blackboard and finished at the lower right. The lectures were smooth and the delivery was a uniform distribution. Though I had a lovely set of notes, Wald treated difficult and easy parts equally, so one did not recognize pitfalls when doing homework. The notion of an application in its current use did not exist. I don't recall the origin of the following quotation, but it is attributed to Wald: "Consider an application. Let $X_1, \ldots, X_n$ be i.i.d. random variables." In contrast to Wald's style, Wolfowitz's lectures were definitely not smooth, but he attempted to emphasize the essence of the topic. He struggled to try to explain what made the theorem "tick," a word he often used: "Let's see what makes this tick." However, as a novice in the field the gems of insight that he presented were not always appreciated. It was only years later as a researcher that they resurfaced, and were found to be illuminating.

Wolfowitz had a number of other pet phrases such as "It doesn't cut any ice," and "stripped of all baloney." It was a surprise to hear Columbia graduates years later using the same phrase. In a regression class with Wolfowitz we learned the Gauss–Seidel method. Wolfowitz was upset that the Doolittle method had a name attached to it, and he would exclaim, "Who is this Doolittle?" Many years later when Wolfowitz visited Stanford a name might arise in a conversation. If Wolfowitz did not recognize the name he would say "Jones, Jones, what theorem did he prove?"

In 1947–48 the only serious general textbooks were Cramér, Kendall, and Wilks' soft-covered notes. This was a time when drafts of books were being written. Feller's Volume I appeared in 1950, Doob's book on stochastic processes in 1953, Lehmann's notes on estimation and testing of hypotheses in 1950, Scheffé's book on analysis of variance in 1953. The graduate students at Columbia formed an organization that duplicated lecture notes, especially those of visitors. Two that I remember are Doob's lectures on stochastic processes and Loève's on probability.

The master's degree program required a thesis and mine was written with Wolfowitz. The topic

was on a sequential procedure that Leon Herbach (he was ahead of me) had worked on. Wolfowitz had very brief office hours, so there usually was a queue to see him. When I did see him in his office he asked me to explain my question at the blackboard. While talking at the blackboard Wolfowitz was multitasking (even in 1947) by reading his mail and talking on the telephone. I often think of this as an operatic trio in which each singer is on a different wavelength. This had the desired effect in that I never went back. However, I did manage to see him after class. He once said "Walk me to the subway while we are talking," so I did. We did not finish our discussion by the time we reached the subway (only a few blocks away) so I went into the subway where we continued our conversation. This was not my subway line so it cost me a nickel to talk to him. One of my students 30 years later at Stanford told me that I suggested that he walk with me while discussing a problem. There is a moral here for faculty.

Wald liked to take walks. Milton Sobel was one of Wald's students and he occasionally accompanied Wald on these walks. Later I learned that Milton took his students on walks. I wonder what is the 21st century current version of faculty-student interaction?

Appendix

The Collyer brothers became famous for their compulsive collecting. I am not in their league, but I have saved my notes from some of the courses that I took. The following is an excerpt from the course catalog.

> Mathematical Statistics 111a — Probability. 3 points Winter Session. Professor Wolfowitz.
> Tu. Th. 5:40–6:30 and 7:30–8:20 p.m. 602 Hamilton.
>
> Fundamentals. Combinatorial problems. Distribution functions in one or more dimensions. The binomial, normal, and Poisson laws. Moments and characteristic functions. Stochastic convergence and the law of large numbers. Addition of chance variables and limit theorems.
>
> This course terminates on Nov. 18. A thorough knowledge of calculus is an essential prerequisite. Students are advised to study higher algebra simultaneously to obtain a knowledge of matrix algebra for use in more advanced mathematical statistics courses.

Do note the times of this course. Milton Sobel was the teaching assistant for the 111a course; Robert Bechofer and Allan Birnbaum were also TAs. I remember that Milton Sobel sat in on Wald's class on analysis of variance. Because he was at least one year ahead of me I thought that he would have taken this course earlier. He said he did take it earlier but the course was totally different depending on who was teaching it. It was depressing to think that I would have to take every course twice! As the course progressed Wald ran out of subscripts and superscripts on the right-hand side, e.g., x_{ij}^{kl}, and subsequently added some subscripts on the left-hand side.

Wolfowitz recommended three books, and assigned homework from them:

- P.V. Uspensky, *Introduction to Probability*
- H. Cramér, *Mathematical Methods of Statistics*
- S.S. Wilks, *Mathematical Statistics*

He mentioned references to Kolmogoroff's *Foundation of Probability* and the Lévy and Roth book *Elements of Probability*.

Wolfowitz used the term "chance variables" and commented that the law of small numbers should have been called the law of small probabilities. As I look through the notes it is funny to see the old-fashioned $\lfloor n$ instead of $n!$ for factorial. As I reread my notes it seems to me that this course was a rather simplified first course in probability. Some of the topics touched upon use

independence, Markov chains, joint distributions, conditional distributions, Tchebycheff's inequality, stochastic convergence (Slutsky's theorem), law of large numbers, convolutions, characteristic functions, central limit theorem (with discussion of Lyapunov and Lindeberg conditions). I have a comment in which Wolfowitz notes an error in Cramér (p. 343): (a) if $y_1, y_2, \ldots$ is a sequence with $\sum y_i = c_i$ for all c and $\sigma^2(y_i) \to 0$ as $i \to \infty$, then $p\lim_{i\to\infty}(y_i - c_i) = 0$; (b) the converse is not true in that it may be that $\sigma^2(y_i) \to \infty$ and yet $p\lim(y_i - c_i) = 0$.

The second basic course was 111b, taught by Wald. The topics included point estimation, consistency, unbiasedness, asymptotic variance, maximum likelihood, likelihood ratio tests, efficiency. This course was more mathematical than 111a in that there was more asymptotics. In terms of mathematical background I note that he used Lagrange multipliers to show that, for $0 \leq w_i$, $\sum w_i^2/(\sum w_i)^2$ is minimized for $w_i = 1/n$, $i = i, \ldots, n$. Apparently, convexity was not discussed.

There is a derivation of the chi-square distribution that includes a discussion of orthogonal matrices. This is one of the standard proofs. Other topics include Schwarz inequality (but not used for the above minimization), and sufficiency. The second part of the course dealt with tests of hypotheses, with emphasis on the power function (Wald used the term "power curve"), acceptance sampling, and the OC curve.

References

Olkin, I. and A.R. Sampson. 2001a. "Hotelling, Harold (1895–1973)." In *International Encyclopedia of the Social & Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes. Oxford: Pergamon, 6921–6925.

Olkin, Ingram and A.R. Sampson. 2001b. "Harold Hotelling." In *Statisticians of the Centuries*, edited by C.C. Heyde, E. Seneta, P. Crepel, S.E. Fienberg, and J. Gani. New York: Springer, 454–458.