from *SIAM News*, Volume 35, Number 3

Playing Ball Probabilistically

Curve Ball: Baseball, Statistics and the Role of Chance in the Game. *By Jim Albert and Jay Bennett, Copernicus (An Imprint of Springer-Verlag), New York, 2001, 350 pages (alas, no index), $29.00.*

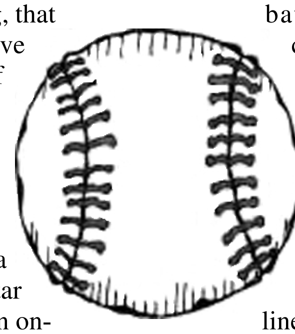
Jim Albert is a professor of mathematics and statistics at Bowling Green State University; Jay Bennett is a principal scientist with Telcordia Technologies and the editor of *Statistics in Sport.* Phil Davis, the reviewer, was an avid fan of the Red Sox when he was in elementary school, during the years when Fred Hoey broadcast the games over WNAC.

BOOK REVIEW

By Philip J. Davis

Since then I have hardly paid any attention to baseball. In view of my long inattention to the game, you may well ask why I agreed to review Albert and Bennett's book. The answer is residual nostalgia: I thought I would revisit old scenes, after which I would shake my head in amazement at how things have changed.

Inattention to baseball? One can hardly live in America and not know that players are now pulling down multimillion-dollar contracts. Or that owners are pressuring City Fathers to build them bigger and better ballparks, paid for out of public funds. Or to observe, while channel surfing, that batters now wear gloves and hard hats. That perhaps the bats and the balls and other equipment have changed. That there are "designated hitters" in the American League to relieve pitchers of the onus of batting. That there are more teams.



Many more "thats" could be listed. Perhaps two of the most significant are that the number of baseball stats seems to have grown exponentially and that the Web has added enormously to the number of peripheral baseball opportunities for managers and for Web-addicted, stats-addicted couch potatoes.

Baseball cards, caps, buttons, t-shirts? These are still around, but they bear the aura of the 1870s. We now have on-line and CD-ROM models, simulations of whole seasons, of all-star games, of World Series; determination of optimal lineups; projections of future performances; even on-line betting capabilities. All these features are based on vast databases out of which at least fifty different kinds of statistical parameters—starting with ABs (at bats) and including GOFOs (ground outs/fly outs ratio) and OFAs (outfield assists)—have been culled.

Curve Ball is targeted toward devoted baseball fans, especially those who are rabid devotees of stats, and the authors have taken time out to explain a few elementary statistical notions for the instruction of the mathematically unsophisticated.

Of all the major American sports, baseball, with its relatively simple, static, and discrete structure, is surely the most mathematized. This has been so for a very long time. Batting averages go back to 1874. Other sports, such as basketball, hockey, and soccer, have a continuous flow in time and space and are therefore more difficult to model. For some fans, stats are the name of the game—more important even than observing the play on the field. I'm sure this group will welcome the appearance of the book under review. For others, baseball will always be the sight of "DiMaggio rounding second."

We can gain an appreciation of the extent to which the lives of the players have been bugged by statisticians (employed by the team or the league), by archives, or by rank amateurs by fantasizing about a parallel mathematization of elementary school children. Suppose your kid came home from Grade 6 at the end of the term with a report card bearing the following: marks in each of eight subjects, number of times at the blackboard, ratio of average grades in science to average grades in humanities, number of absences, ratio of number of hand raises per annum to the average number of hand raises in Grade 6 over the whole of the USA, number of visits to the bathroom . . . and on and on for forty more pieces of data. Suppose also that the principal of an individual school had forty or more indices of this and that to torment her/him. Get the picture?

I wonder how the ball players can live under the tension and uncertainties that are created and magnified by the theoretical baseball analysts. Perhaps, for this reason alone, the players are entitled to every penny of their salaries, plus all they can earn from endorsements of Filboid Studge, the favorite breakfast food of the Hall-of-Famers.

Albert and Bennett's book introduces and uses elementary probability, histograms and other data displays, normal and binomial distributions, linear regression. I believe that the principal message the authors want to get across is that baseball, whatever else it is, is also a game of chance, and that chance plays a larger role in the outcomes than "pure platonic" skill of the players operating deterministically would lead us to believe. In their last chapter—titled "Did the Best Team Win?"—this is their parting message for readers:

> "Whether we like it or not, chance events have a big effect on the patterns of wins and losses that we observe."

For probability sophisticates, for whom this message is not news, baseball might be modeled as a discrete Markoff chain that has umpteen internal states and a matrix whose transitional probabilities linking the states have been estimated by passionate data collectors. This possibility goes way back, to the simple baseball board games of my youth with spinners or "select a card piles" that were divided nonuniformly to reflect a few of the probabilities.

The book discusses, among many other things, "situational effects" (e.g., home vs. away, grass vs. turf), streakiness (a player being "hot" or "cold"), measuring offensive performance, measuring clutch play (performance at moments of crisis), predictions. Albert and Bennett introduce us to a great variety of performance measures and their formulas (some of which are confidential to the teams), and to competing models. We learn of Palmer models, Lindsey models, Mills models, . . . additive models, product models. One of the more complex formulas mentioned is Bill James's Runs Created (RC) index:

$$RC = \frac{(H + BB + HBP - CS - GIDP)(TB + .26(BB - IBB + HBP) + .52(SH + SF + SB))}{AB + BB + HBP + SH + SF}$$

I have deliberately refrained from specifying the meaning of these acronymic variables so as to impress my readers with how much "alphabet soup" populates these pages, which would have profited greatly from a glossary.

The presentation might be termed one of "pure baseball stats." It is a monograph on "abstract computational baseball," employing true league data and containing numerous worked examples. Hardly anything is said about what the various stats, models, ratios, and so forth, mean *at the bottom line*. The paucity of bottom-line discussions will encourage me to seek out other books (a substantial bibliography is provided) to learn how and by whom the final computations are used. As real games are played on real grass or real turf, to what extent are the managers being advised on strategy by their models? To what extent have long years of managerial expertise been replaced by stats, models, and simulations? How are hirings, dismissals, salaries, negotiations (I just read that the Red Sox have been sold for a record $700 million) affected by the baseball statisticians?

What is presented is, without a doubt, a chapter of applied mathematics. True, the math is elementary, and much of it, I suspect, has been produced for the pure joy of the doing.

Reading Albert and Bennett led me to ask similar questions about "traditional" applied mathematics. I have a book on my desk on the econometric analysis of seasonal time series. I have another on the mathematical analysis of genomes. Will the material contained in these books ever see "real" applications at the bottom line? Come to think of it, what is a real application? What constitutes the bottom line in a piece of applied mathematics?

I doubt that there ever is a bottom line. Or perhaps there are many bottom lines, and the individuals who use the mathematics are constantly shifting them. In a way, baseball data and their analyses are paradigmatic of our increasingly mathematized civilization.

Philip J. Davis, professor emeritus of applied mathematics at Brown University, is an independent writer, scholar, and lecturer. He lives in Providence, Rhode Island, and can be reached at philip_davis@brown.edu.