SIAM Review
Issue 48, Vol. 1

Anthology of Statistics in Sports. *Edited by Jim Albert, Jay Bennett, and James J. Cochran*. SIAM, Philadelphia, PA, 2005. $65.00. x+322 pp., softcover. ASA-SIAM Series on Statistics and Applied Probability. Vol. 16. ISBN 0-89871-587-3.

This is a useful collection of articles on statistics in sports that were published between 1961 and 1999. Teachers of statistics who illustrate applications using sports will find the collection valuable. In fact, Chapter 2 discusses the use of sports and this anthology in teaching statistics. Scientists who enjoy sports and want easy access to some of the most influential articles in the subject will also find the anthology valuable.

All of the articles reproduced in this anthology were published in *Chance*, *The American Statistician*, or the *Journal of the American Statistical Association*. The level of statistical sophistication correlates with the levels of these three journals. I can't help wondering why such an anthology would restrict itself to three journals. Other journals with relevant articles include *Operations Research*, *IEEE Transactions on Systems, Man, and Cybernetics*, and *Management Science*.

The chapters are organized by sport (baseball, football, basketball, ice hockey, etc.), while Chapter 1 is a very useful introduction that includes suggested chapters by topic (prediction, hot hand phenomena, probability of victory, validity of selected data, rating players or teams, unusual outcomes, etc.). I will discuss below some of the interesting highlights that caught my attention.

The articles on baseball focus on performance. The nontechnical Chapter 15 is a good place to start. It illustrates both the power and pitfalls in interpreting baseball statistics. It is definitely possible to read more into the statistics than is justified by reality; this happens daily in sportscasts. The best teams don't always win their divisions or playoffs. The best pitchers don't always win 20 games a year. A .280 hitter isn't necessarily better than a .260 hitter, etc. This chapter addresses many facets of the question, Does the best team win? For example, how often does the best team in baseball win the World Series? This article also distinguishes between the way statisticians and baseball professionals see baseball.

The famous 1919 World Series is studied in Chapter 11, which concludes, among other things, that almost every statistical analysis supports the view that Shoeless Joe Jackson played to his full potential in that series, even though he admitted involvement in the Black Sox scandal. Being partial to baseball anyway, I found the longest chapter, Chapter 16, which analyzes scoring inning by inning, especially interesting. However, it was published in 1961 and is based on data from 1958 and 1959. The game has changed a lot since 1959, so a follow-up article would be of interest. Chapter 28 creates a model, a statistical time machine, that makes it possible to compare players from different eras by comparing players whose careers overlap.

The articles on football focus on field goal kickers, predictions, and gambling. The authors of Chapter 7 ask whether field goal kickers are lucky or good. Their analysis from 1989–1991 indicates that the data are consistent with the hypothesis that there are no skill differences among these players. The authors don't expect all fans, coaches, and kickers to agree. Later, the author of Chapter 4 analyzes 33 place kickers from the 1998 regular season using a geometry model. Field goal attempts are classified as successful, short, or wide, where short wide attempts are classified as short. Chapters 6 and 5 provide predictive models for game scores with an eye to applications to betting. Chapter 6, published in 1980, uses a linear-model methodology, while Chapter 5, published in 1998, takes a fully Bayesian approach using Markov chain Monte Carlo methods. I enjoyed the observation in Chapter 5, which includes several interesting anecdotes, that "the plural of anecdote is not data."

The articles on basketball focus on predictions of winners of elaborately designed tournaments and on the question of whether players have a "hot hand." The "hot hand" question is whether players' successes deviate significantly from random expectations of Bernoulli trials. A 1989 article, Chapter 21, makes the case that most perceptions of "hot hands" are based on misconceptions about chance processes, but two follow-up articles that year were critical of this study. Chapter 19 questions both the methodology and the conclusions, while Chapter 31 accepts the methodology but questions the interpretation. Related papers are the second and third references in Chapter 10 and the more recent paper [1]. A Brownian motion model is created in Chapter 34 that models the scoring in a basketball game as a function of time. Basketball is best suited to this model, because of the nearly continuous nature of the game.

Much of the focus of statistical research in ice hockey has been devoted to ranking or comparing teams. The methods used could be generalized to other sports (for example, soccer, lacrosse, and water polo) which share the same format of teams trying to get an object past the opposing team's players into a confined space that is guarded by a goalie. Chapter 25 deals with the thorny issue of handling ties in playoffs. Normal overtimes have the disadvantage that they can eat up a great deal of time, not to mention the energy of the participants. Shootouts have the disadvantage that they are a poor way to decide an important championship. Chapter 26 analyzes the effects of the "hot-goalie," the home ice advantage, and the intrinsically dominant team on the outcome of the Stanley Cup final series. It is interesting that the "hot-goalie" and dominant team effects are real, but the home ice advantage has minimal effect. It is also interesting that, unlike the "hot hand" in basketball, the existence of the "hot-goalie" does not seem to be in dispute.

Chapter 32 is a 1997 article by Frederick Mosteller that, in my opinion, should be Chapter 3, before the articles that specialize in various sports. The chapter presents a collection of sports-related analyses that Mosteller has done over his career. He clarifies the main goals of statistical analyses in sports, highlights some important statistical ideas, answers some critical questions about statistics in sports, and provides a means of bringing statistics to the general public.

Other sports are touched on: golf, horse racing, soccer, distance running, shooting darts, pool, figure skating, and tennis. In Chapter 42 it is argued that "psychological momentum" is considerable in tennis. Whatever they're called, streaks and slumps in baseball, the "hot hand" in basketball, and "psychological momentum" in tennis confirm our basic belief that "success breeds success, and failure breeds failure." A recent article, [2], contains convincing evidence that there are streaks in professional bowling.

Chapter 39, written in 1994, analyzes the system for judging figure skating, which the authors seem to feel is close to ideal. The judging system is based on median ranks and they "uniquely capture an important meaning of majority rule and provide strong protection against manipulation by a minority of judges." This is interesting in view of questionable scoring at recent international figure skating events. The last chapter makes a strong case that triathlons favor runners and cyclists and are biased against swimmers.

I have some criticisms about the organization of this anthology. The original sources are listed on pages ix–x instead of with the articles. This is frustrating, especially since it is difficult to even discern the years of publication of the articles, which vary wildly from 1961 to 1999. Within the sports, the chapters are listed alphabetically by authors. This seems artificial—either by publication date or by topic would be better. For example, Chapters 4 and 7 are about NFL field goal kickers, while Chapters 5 and 6 are not. As I hinted at earlier, Chapter 15 on baseball should come right after Chapter 9. Chapter 19, which is a critique of Chapter 21, should come after Chapter 21. Chapters 18 and 20 are about modeling basketball tournaments. Serendipitously, the chapters on ice hockey are simultaneously alphabetical by author, by publication date, and by topic!

Finally, it is amazing that an anthology published in 2005 does not contain a single article published after 1999.

REFERENCES

[1] J. Albert and P. Williamson, *Using model/data simulations to detect streakiness*, Amer. Statist., 55 (2001), pp. 41–50.

SIAM Review
Issue 48, Vol. 1

[2] R. Dorsey-Palmateer and G. Smith, *Bowlers' hot hands*, Amer. Statist., 58 (2004), pp. 38–45.

Kenneth A. Ross
University of Oregon

An Introduction to Continuous-Time Stochastic Processes, Theory, Models, and Applications to Finance, Biology, and Medicine. *By V. Capasso and D. Bakstein.* Birkhäuser Boston, Boston, MA, 2005. $79.95. xii+343 pp., hardcover. ISBN 0-8176-3234-4.

Chapter 1, pages 1–50, contains 158 definitions, theorems, corollaries, etc. Appendix A contains 63, and Appendix B contains 77. Section 1.8 offers 35 exercises and additions that cover topics such as entropy, infinitely divisible distributions, Lévy's measure, large deviations, law of the iterated logarithm, filtrations, adapted processes, martingale, predictability, Doob's decomposition, etc. This condenses one or two undergraduate and about four graduate courses in mathematics. My experience shows that students lacking this mathematical background do not survive a rigorous course on stochastic processes that uses the said material. By the time the course gets to applications, only the well-prepared students survive. It is either rigor or applications; both cannot coexist in the same course if the applications are serious.

Chapter 2, pages 51–126, on "Stochastic Processes," relies heavily on notation and concepts from measure theory. Writing abstract definitions and quoting theorems is insufficient for understanding the excessive generality of the definitions. Incantation of off-the-shelf theorems is much like intoning Psalms: They soothe the soul, but do little else. This is a sure way to kill any intuition. The readers, who encounter a problem that needs independent reasoning, are led to believe that the answer is in finding the appropriate theorem to quote in the given list, and not to use their own reasoning, the tools for which are not provided.

General Remarks on the Presentation of the Theory. The concept of separability of stochastic processes contributes precious little to the understanding of the theory. It would be useful to provide an explanation of what predictability of a process has to do with predictability and to provide meaningful examples. A useful example at this point is a proof that the first exit time of a continuous process from a given set is a stopping time. Another example that fits here is the proof that if a differentiable process is adapted, then its derivative is also adapted on a filtration in $C([0,1])$, endowed with the Wiener measure (generated by cylinders). The discussion of this property in the context of Markov processes (p. 77) does not use the Markov property, but merely the continuity of the paths. Sticking to formal definitions, without relevant realizations, obscures the concepts and kills the interest of the applications-oriented reader.

Disposing of Gaussian processes in two pages is a big leap of faith. Gaussian processes are usually a topic for a separate graduate course. I wonder what the purpose is of quoting advanced definitions and theorems without presenting proofs. On the one hand, the reader does not learn much on how to prove anything that has to do with stochastic processes if the proof of a given statement cannot be quoted from the book (or any book). In the absence of proofs, there seems to be no purpose for the apparent rigor of the presentation, and a merely intuitive exposition may suffice to get the message through (as in Wilmott's, Gardiner's, and other texts).

On the topic of Markov and diffusion processes: The names of Fokker and Planck should be mentioned together with Kolmogorov's in the context of the forward and backward equations. An illustration of the uses of the generator fits here, as well as connecting the probabilistic properties of diffusion processes with partial differential equations. The Itô integral is not needed for this.

The exposition of Brownian motion and the Wiener process is entirely synthetic. The reader gets no clue on where it comes from in any familiar system. The readers deserve a glimpse into Einstein's, Smoluchowski's, and Langevin's historical work, so they can get a feel of what it is all about. Reading Brush's *The Kind of Motion We*