Book Review

The signal and the noise

Authors	: Nate Silver
Publisher	: Nate Silver
ISBN	: 978-1-846-14816-3
Price	: USD 15.00
Year of Publication	: 2012

C. R. Sridhar

Nuvis Analytics Pvt Ltd, Mumbai, Maharashtra, India.
E-mail: sridhar.cr@gmail.com
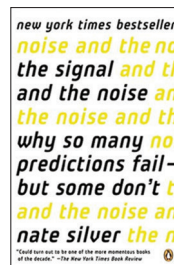
Human's obsession with predictions is legendary. We have experimented with Delphic Oracles, astrology, palmistry, and several other methods. We have also not left the birds to help us to do the job of prediction. We can still see this avian astrology in smaller towns of India with a soothsayer sitting with a brightly colored parrot picking the cards for its master. However, most of us know that predicting the future correctly all the time is impossible. The enormity of this problem is well known to pollsters and economists. The correct prediction is quite often an exception than a rule. However, the desire to completely remove uncertainty from future continues to occupy a better part of our day-to-day living. We want to know when it will rain, the prices of specific stocks, and whether your favorite cricket team will win or not, all of them with great accuracy. However, "We need to stop and admit it: We have a prediction problem. We love to predict things but we aren't very good at it." The book by Nate Silver's presents the possibilities and pitfalls in the art and science of predicting the future.

According to him, whenever we are presented with data, we fail to recognize the signal that is shrouded by noise in the data. An example he discusses in some detail to illustrate the idea of noise and signal is that of the United States Housing Bubble. As we all know, in 2008, the United States Housing Bubble started bursting. All those who watched the bubble burst never saw the signal in the financial markets, they were collapsing. Hence, none could predict that the United States Housing Bubble would carpet bomb the global economy and bring it to its heels. The economists and the investment bankers did not see the signal that indicated an economic doom because they kept their focus on the noise, the United States Housing Bubble. It is now becoming clear that "that we know and what we think we know may be widening." There lies the pitfall.

In the last two decades, we have moved from paucity to profligacy of data. Simultaneously, we have moved from fast to superfast in data crunching. With these huge technological advances in software and hardware, a total of scientists, known as Data Scientists, are ready to cut through huge amounts of data and extract relevant patterns and meaning. However, as the data get larger and faster, it becomes increasingly difficult to extract signal from the noise. They are looking for a needle in a haystack without knowing what the needle looks like. The book by Nate Silver's addresses the issue of the noise and the signal.

To illustrate the predicament of forecasters, he discusses a large number of examples in weather forecasting, stock market predictions, earth quake predictions, baseball scores, chess predictions, etc. He provides a number examples, illustrating the noise in the data and how the signals could be picked up. Politics also take a significant portion of his discussion.

The discussion on application of Bayes' Theorem is simple and quite elegant. The Bayes' approach consists of first making a judgment on probabilities of events, and then modifying the judgmental probabilities by observing the results. The method is particularly useful in analyzing the results of the drug tests where false positives need a careful analysis to identify the signal by placing the results in the correct context.

The field of statistical sciences consists of tools for forecasting the future. Most health professionals are aware that the drugs approved for treating diseases have to go through clinical trials. The clinical trials consist of using the drug on a sample of volunteers, using blind or double-blind protocols. The efficacy of the drug is then estimated using statistical methods. If the results of the new drug are statistically significant, then only

is the drug sent for approval. Subsequently, it enters the market. Failure of drugs in practice, Vioxx, for example, is not because the statistical theory is in adequate for testing but the experimental results are massaged to get an approval to enter the market.

The book is accessible to almost anyone interested in the subject of prediction and forecasting. It is written very well and does not resort to symbolic mathematics to impress the readers. Besides routine statistical approaches, it takes the reader beyond numerics. A successful predictor is a fox and not a hedgehog. We need a 360° view of a phenomenon to make a forecast. A reliance on one model, one method, and one path is a sure success on television but a pothole for prediction.

Nate Silver confirms that "All models are incorrect, but some of them work." This realization is more helpful in designing better forecasting systems. The movement in predictive modeling is from "Bias to Bayes." The book beautifully pushes that case.