CENTER FOR
RISK MANAGEMENT
RESEARCH

Working Paper # 2013- 05

The Signal and the Noise
by Nate Silver
reviewed by

Lisa R. Goldberg, University of California at Berkeley

September 25, 2013

University of California
Berkeley

The Signal and the Noise
by Nate Silver

reviewed by Lisa R. Goldberg*

forthcoming in *Quantitative Finance*

September 25, 2013†

Who will win the next election? What are the chances that your neighbor, Sam, will correctly predict the outcome of the next US presidential election? If Sam does not pay attention to the media, or perhaps even if he does, we might assume that his forecasts are just as good as a sequence of coin flips.[1] This means that the probability that Sam will get a single state right is one half and the chances that he'll correctly forecast the winner of the next election, say, in 49 out of 50 states, is vanishingly small:

$$\begin{aligned} \mathrm{P}(49 \text{ states right}) &= 50 \cdot .5^{49}(1-.5)^1 \\ &= 4.44 \times 10^{-14}. \end{aligned}$$

Nate Silver, formerly the lead blogger at fivethirtyeight.blogs.nytimes.com, correctly predicted the outcome of the 2008 US presidential election in 49 out of 50 states and it's hard to believe that he was just lucky. Of course, Silver's forecasting model was more complex than a sequence of coin flips. For example, predicting the election outcome was much easier in some states than in others. Nevertheless there is information in a vastly oversimplified model that likens the problem of forecasting the US presidential election to a sequence of flips of a possibly biased coin.[2] In this model, probability that Silver correctly forecasts the election in exactly 49 states is conditional on the probability, p, of forecasting a single state correctly:

$$\mathrm{P}(49 \text{ states right} \,|\, \mathrm{p}) = 50 \cdot p^{49}(1-p)^1.$$

In other words, Silver's overall success rate depends on p so the estimation of p is an *inverse probability problem*.

*Center for Risk Management Research and Department of Statistics, University of California, Berkeley, CA 94720-3880, USA, email: lrg@berkeley.edu.

†The author is grateful to Bob Anderson, Vadim Gracie, Alec Kercheval, Ola Mahmoud, Jon McAuliffe, Ken Ribet, Andrew Schein, Paul Solli and Rodney Sullivan for their contributions to this review.

[1]This is an unrealistic assumption. For example, it is plausible that most Americans would be able to forecast the outcome of a US presidential election in Texas.

[2]The model implicitly assumes that the state elections are independent, which is also unrealistic. Silver makes no such simplifying assumption in his analysis.

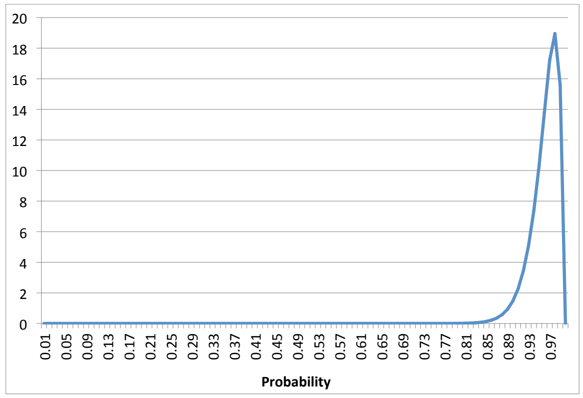

Figure 1: Posterior distribution of the probability that Nate Silver would correctly forecast the outcome of US presidential election in a single state in 2012, given that his chances were equally good in every state, that the state elections were independent and that he forecast 49 out 50 states correctly in 2008.

The method for estimating inverse probabilities is generally attributed to the eighteenth century Presbyterian minister, Reverend Thomas Bayes, and an early exposition of Bayes' rule is in Bayes and Price (1763).[3] Bayes understood the roles of variable and conditioning information could be interchanged, and Bayes' rule provides an estimate of the likelihood that Silver's predictions are based on a fair coin flip:

$$\mathrm{P}(p = .5 \,|\, 49 \text{ states right}) = \frac{\mathrm{P}(49 \text{ states right} \,|\, p = .5)}{\mathrm{P}(49 \text{ states right})} = 2.26 \times 10^{-12},$$

which is also vanishing small. The distribution of p is shown in Figure 1, and it suggests that Silver has a lot of skill: the most probable values for p are clustered near 1. Silver's election forecasts rely on Bayes' rule, and in 2012, he correctly forecast the outcome of the US presidential election in all fifty states.[4] His success was heralded as a triumph of quantitative modeling.[5]

A powerful but controversial formula. The most basic version of Bayes' rule is routinely taught in undergraduate courses in mathematics, statistics and computer science around the world. The rule relates the conditional probability of a variable X given another

[3]The Welsh moral philosopher Richard Price discovered, edited and published Bayes' mathematical notes after his death. According to Dale (1982), contemporaneous work of Pierre LaPlace has substantial overlap with the work of Thomas Bayes.

[4]This includes Florida, which Silver deemed too close to call, and which was subject to recounts and disputes in the days following the election.

[5]See, for example, Bartlett (November 7, 2012).

variable Y to the conditional probability of Y given X:

$$\mathrm{P}(X \mid Y) = \frac{\mathrm{P}(Y \mid X)\mathrm{P}(X)}{\mathrm{P}(Y)}, \tag{1}$$

In practice, the $\mathrm{P}(Y \mid X)$ is known or observed, and the left hand side of Formula (1), the *posterior distribution* of a variable X, given Y, is derived. In the assessment of Silver's forecasting ability, the variable X is, itself, a probability. In a more familiar setting, X might represent the presence of a certain illness, say a particular type of cancer, and Y might be a positive result from a cancer screen. A positive diagnostic test is disturbing but Bayes' rule provides an interpretation of the result. Suppose we are dealing with a cancer that afflicts one in one hundred individuals, $\mathrm{P}(X) = .01$, and the screen correctly identifies cancer in nine out of ten afflicted individuals, so $\mathrm{P}(Y \mid X)$ is .90. Unfortunately, the screen erroneously identifies cancer in three out of ten healthy individuals. Then the probability of a positive result is:[6]

$$\mathrm{P}(Y) = \mathrm{P}(\text{positive test}) = .90 \times .01 + .30 \times .99 = .31$$

According to Bayes rule, the probability that the illness is present, given a positive test result, is:

$$\begin{aligned}
\mathrm{P}(\text{cancer} \mid \text{positive test}) &= \frac{\mathrm{P}(\text{positive test} \mid \text{cancer})\mathrm{P}(\text{cancer})}{\mathrm{P}(\text{positive test})} \\
&= \frac{.90 \times .01}{.90 \times .01 + .30 \times .99} \\
&\approx .03.
\end{aligned}$$

This means that even in the narrow population of individuals who test positive, the presence of the illness is unlikely. It is the rarity of the illness in the overall population combined with the high rate of false positives that leads to a low rate of illness among those who test positive. Physicians are sometimes surprised by the low rate and that is troubling. Many individuals who test positive for cancer are subjected to needless worry, and perhaps unpleasant or even dangerous follow-up tests and treatments. This raises questions about whether some diagnostic tests do more harm or more good.

Forecasts that rely on Bayes' rule can be sensitive to the *base rate*, or *prior*, which is denoted $\mathrm{P}(X)$ in Formula (1). The base rate for a cancer screen is the frequency with which cancer occurs in the population, and that can be estimated from empirical data. In assessing Nate Silver's skill at predicting the outcome of elections, I did not have a reasonable estimate of the base rate, so I assumed a uniform prior distribution on the likelihood, p. A different choice of prior would have led to a different estimate. McGrayne (2011) describes heated debates about the legitimacy of Bayesian priors that raged throughout the first half of the twentieth century.

[6] Similar examples are in Chapter 8 of *The Signal and the Noise* and in Bohn and Stein (2009, Chapter 4).

The role of Bayesian evidence in courts of law is also contentious. Tribe (1971) argues that "the utility of mathematical models in the legal process have been greatly exaggerated." Schneps and Colmez (2013) describe ten legal cases where mathematical arguments were misunderstood in courts of law; the authors assert that in nine of the ten cases, the mathematical misunderstandings led to miscarriages of justice. Fienberg and Finkelstein (1996) discuss the "vigorous debate over the role of Bayesian methods ... in the presentation of such evidence and its evaluation by the triers," and they describe how arguments involving Bayes' rule can be difficult for judges and jurors to accept. The authors illustrate their point with two cases that were tried in United States courts in the early 1990s. The first case concerned proof of paternity in a case of alleged sexual assault, where a Bayesian analysis of blood tissue was ultimately excluded from consideration due to a misunderstanding about the implications of the prior. The second case concerned a closely contested election in which allegedly fraudulent absentee ballots reversed the outcome. Here again, a Bayesian analysis by a court expert was discounted as a result of confusion about priors. Fienberg and Finkelstein (1996) comment:

> We should not expect naïve jurors or judges to understand how to use Bayes' theorem, but we might learn to teach them enough about interpreting evidence to alert them to important issues...

The difficulty that judges and jurors have with Bayesian evidence may, in part, be the human tendency to neglect base rates. Gigerenzer and Hoffrage (1995) explain that this can be mitigated by presenting data as frequencies instead of as probabilities. Kahneman (2011, Chapter 30) explains that the way data are framed can have a material impact on how it is interpreted by judges and juries.

Tales of statistical forecasting. *The Signal and the Noise* is a fascinating and diverse collection of thirteen stories about statistical prediction.[7] What is most remarkable about the book is its emotional impact. Many of the chapters read like good fiction. By embedding complex and subtle mathematical ideas in accessible narratives, Silver provides a general audience with an opportunity to appreciate the role of statistics in today's world.

Silver writes, for example, about epidemiology, economics, poker, climate change, terrorism and baseball,[8] and relationships among apparently disparate topics emerge over the course of the book. We learn that financial markets may be like earthquakes in the sense that average rates of extreme events can be estimated, but their timing may be impossible to forecast. Hurricanes are different: consider how precise tracking of Hurricane Sandy facilitated an evacuation that saved countless lives. A consistent theme is the importance of consensus, or "wisdom of the crowds." Valuable information can sometimes be distilled from the views of neighbors like Sam, as long as they are sufficient in number. Even among experts, aggregate forecasts tend to be more accurate than individual forecasts.[9]

[7] A notable omission is to the role of statistics in courts of law.

[8] It is this last area in which Silver first made his mark, although he may be better known for forecasting election outcomes.

[9] *The Signal and the Noise* Chapter 11, page 335.

The Signal and the Noise begins dramatically with a negative message: some statistical forecasting models are unsuccessful. Chapter 1 is called "A Catastrophic Failure of Prediction," and the failure in question is the collapse of the market for triple-A Collateralized Debt Obligations (CDO) during the 2008–2009 financial crisis. Silver reports that the rating agency default forecast for triple-A CDOs to be .12 percent per year, while the actual default rate during the crisis was 28 percent per year, and he describes this as:

> ...about as complete a failure as it is possible to make in prediction.[10]

Indeed, the event stands in stark contrast Silver's success in forecasting elections.

While Silver's critique of the rating agencies is harsh, many of the issues he raises are compelling. For example, Silver questions the integrity of the rating agencies' business model, in which companies pay for their ratings. However, some comments seem less grounded. Arguing that the rating agencies could have foreseen the housing bubble, Silver emphasizes that capable economists such as Robert Shiller and Paul Krugman had forecast the event. But he fails to point out that those forecasts were embedded in a sea of conflicting predictions by economists who may be just as capable, despite the fact that they turned out be wrong. This is *hindsight bias*, which obscures the ability to recall the complexity and uncertainty of situations in the past.[11]

Forecasting economic variables such as inflation and gross domestic product is the subject of Chapter 6, entitled "How to Drown in Three Feet of Water." Here, Silver emphasizes some of the most basic and least popular aspects of statistical modeling. He reminds us that every statistical estimate should be bundled with a standard error indicating a range in which the value being estimated may comfortably lie. Since this extra bit of information is cumbersome, it is often neglected. The danger of this neglect is featured in a story about the 1997 flood in Grand Forks North Carolina, in which "75 percent of the city's homes were damaged or destroyed."[12] Silver argues that the damage could have been avoided if the standard error of three feet had been considered as part of the estimated height of the levee designed to protect the city against floods.

Chapter 11 is entitled "If You Can't Beat 'Em...," and it addresses predictability in financial markets.[13] Silver outlines the Efficient Market Hypothesis (EMH), which asserts (more or less) that future returns cannot be predicted, and the primary source on the EMH is Fama (1965). In an interview that appeared in *The New Yorker* in January 2010, John Cassidy asked Eugene Fama to comment on how the EMH held up during the financial crisis:[14]

> I think it did quite well in this episode. Stock prices typically decline prior to and in a state of recession. This was a particularly severe recession. Prices started to decline in advance of when people recognized that it was a recession and then continued to decline. There was nothing unusual about that. That was exactly what you would expect if markets were efficient.

---

[10] *The Signal and the Noise* Chapter 11, page 21.

[11] Hindsight bias is discussed in Kahneman (2011) and Taleb (2007).

[12] *The Signal and the Noise* Chapter 6, page 177.

[13] An application of Bayes' rule to finance is machine learning and algorithmic trading, although that subject is not treated in *The Signal and the Noise.* Chapter 11.

[14] See Cassidy (2010).

Silver comments on his own interview with Eugene Fama:

> ... in what was an otherwise friendly conversation, [Fama] recoiled when I so much as mentioned the b-word.[15]

Here, the"b-word" is "bubble," not "Bayesian." Financial bubbles are frequently discussed in the media and there is sharp disagreement about whether they do or do not conflict with EMH. Perhaps Silvers views on EMH can be gleaned from Figure 11-7, which shows scatter plots of market returns against price/earnings (P/E) ratios.[16] As the return horizon lengthens, the apparent negative relationship between the P/E ratio and market returns gets stronger. The plots are compelling and the reader may be tempted to make a ten-year investment when the market P/E next seems low. But then, past performance is no guarantee of future return.

My favorite part of *The Signal and the Noise* is Chapter 9, "The Rage of the Machines," which tells the story of chess Grand Master Gary Kasparov's defeat at the hands of Deep Blue in 1997. Deep Blue began as a science project at Carnegie Mellon University and developed into a champion under the tutelage of Feng-Hsiung Hsu and Murray Campbell at IBM. I rooted for Deep Blue throughout the chapter and my heart beat faster as Silver described the closing moments of the match. Silver is a terrific storyteller and in the words of Ben Levitt, Chief of Education at the Museum of Mathematics, he is "the toast of the town." *The Signal and the Noise* is an engrossing work that provides broad access to the obscure discipline of statistical modeling. It would be wonderful if more mathematically inclined individuals shared Silver's interest in practical problems, his gift for communication and his aspiration to connect with the outside world.

References

Tom Bartlett. The Poll Quants Won the Election. *Chronicle of Higher Education*, November 7, 2012.

Thomas Bayes and Richard Price. An Essay Towards Solving a Problem in the Doctrine of Chance. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S. *Philosophical Transactions of the Royal Society of London*, 53(0):370–418, 1763.

Jeffrey R. Bohn and Roger M. Stein. *Active Credit Portfolio Management in Practice*. Wiley Finance, 2009.

John Cassidy. Rational Irrationality: Interview with Eugene Fama. *The New Yorker*, 2010.

A.I. Dale. Bayes or Laplace? An Examination of the Origin and Early Applications of Bayes' Theorem. *Archive for History of Exact Sciences*, 27(1):23–47, 1982.

[15] *The Signal and the Noise* Chapter 11, page 347.

[16] Some financial analysts interpret a low P/E ratio as an indication of a good buy.

Eugene Fama. The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105, 1965.

Stephen E. Fienberg and Michael O. Finkelstein. Bayesian Statistics and the Law. In *Bayesian Statistics 5,* J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., pages 129–146. Oxford University Press, 1996.

Gerd Gigerenzer and Ulrich Hoffrage. How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, 102(4):684–704, 1995.

Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

Sharon Bertsch McGrayne. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, 2011.

Leila Schneps and Coralie Colmez. *Math on Trial*. Basic Books, 2013.

Nassim Nicholas Taleb. *The Black Swan*. Random House, 2007.

Laurence H. Tribe. Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review*, 84:1329–1393, 1971.