CSCI 102 Special Topic: Bayesian Modeling

Conventional Models (Statistics) Are Failing Us

America's New Year's Eve: Overhyped and Embarrassing

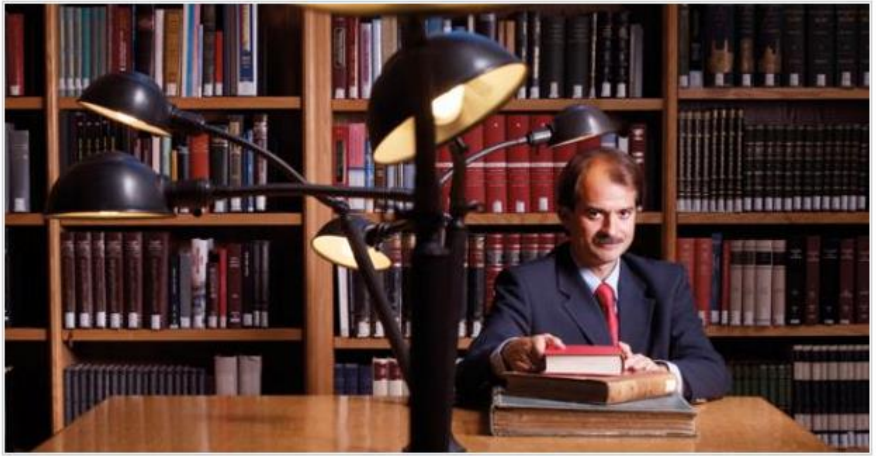
What It ... to Live

BRAVE THINKERS | NOVEMBER 2010 ATLANTIC MAGAZINE

Lies, Damned Lies, and Medical Science

Much of what medical researchers conclude in their studies is misleading, exaggerated, or flat-out wrong. So why are doctors—to a striking extent—still drawing upon misinformation in their everyday practice? Dr. John Ioannidis has spent his career challenging his peers by exposing their bad science.

By DAVID H. FREEDMAN



Robyn Twomey/Redux

PERSONAL HISTORY

SHAKEN

A mother's conviction. A son's doubts.

BY VICTOR ZAPANA

On the morning of January 25, 2007, a frail eight-year-old boy was brought into a Queens courtroom in a wheelchair. He had been the victim of a violent crime, in which three-quarters of his brain was destroyed. He wore a bandanna around his neck, to catch his drool. He would never be able to use his right arm. His feet were bound to the chair, because he could not control his legs. He was almost blind.

The boy's presence was controversial. The defense lawyer feared that it would compromise the jury's objectivity. The judge disagreed. "There's nothing grotesque or inflammatory about it," he said. "It's simply a little boy sitting in a wheelchair." At the time, the boy's father, Sung Hak An, told a reporter, "I just want to see justice. He lost his life."

The boy sat before the jury as a neurologist described his injuries: he suffered from seizures, could not respond to his name, and would never be able to eat or use the bathroom by himself. Afterward, the judge warned the jurors not to surrender to sympathy. Yet the effect on the courtroom was clear. The boy's brief appearance was enough to bring at least one person in court to tears—the defendant, my mother.

The jury later found that in March, 1999, eight years earlier, Yoon Zapana had injured eight-month-old Rew An so brutally that nerve fibres and blood vessels in his skull snapped, resulting in a condition known as shaken-baby syndrome. On February 7, 2007, the jury unanimously convicted my mother, and the police took her into custody and transported her to a holding cell in the Rose M. Singer Center, a detainment facility on Rikers Island.

I was a senior at Stuyvesant High School, and that night, when I returned to our two-story red brick house in Astoria, I heard whispering in the living room. I'd been hanging out with friends in Manhattan when, an hour earlier, Papá called and asked that I immediately come home to Queens. The request terrified me. It reminded me of a call he'd made four years earlier, shortly before he told me that he was going to be deployed to Iraq. I walked up the steps to our living room, on the second floor, and saw my grandmother, my aunts, and Papá sitting in a tight row on the leather couch, their heads bowed.

I asked what had happened, but no one looked up. Then I noticed that my mother wasn't there.

Papá said, "We didn't want to tell you, because we thought we were going to win, and then you wouldn't need to know."

Know what?

"Mom has lost a criminal case," he said. "She's going to jail."

What criminal case?

"Mom didn't want to make you worried," Papá said. "She wanted to protect you. Everything is going to be all right."

The verdict made no sense, Papá continued. She had told him she didn't do it. He knew she didn't do it. Calling collect from Rikers a few days later, my mother told me, sobbing, that she was innocent. Feigning composure, I told her that I loved her and hoped to see her soon. I couldn't bear to say that I didn't believe her. The question of her guilt was bound up for me in a larger betrayal: the very fact that the trial was taking place had been kept from me. Maybe she'd wanted to protect me, but it felt like an act of deception, a family conspiracy. How *could* I believe her?

The author, with his mother, in their Mott Street apartment, in 1990.

32 THE NEW YORKER, NOVEMBER 26, 2012

Online commentators cited my mother as an example of why no parent should hire a nanny. (In fact, parents and other family members are responsible for nearly eighty percent of cases involving shaken-baby syndrome.)

Bayesian Reasoning

Opinion

What to believe: Bayesian methods for data analysis

John K. Kruschke

Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St., Bloomington IN 47405-7007, USA

Although Bayesian models of mind have attracted great interest from cognitive scientists, Bayesian methods for data analysis have not. This article reviews several advantages of Bayesian data analysis over traditional null-hypothesis significance testing. Bayesian methods provide tremendous flexibility for data analytic models and yield rich information about parameters that can be used cumulatively across progressive experiments. Because Bayesian statistical methods can be applied to any data, regardless of the type of cognitive model (Bayesian or otherwise) that motivated the data collection, Bayesian methods for data analysis will continue to be appropriate even if Bayesian models of mind lose their appeal.

Cognitive science should be Bayesian even if cognitive scientists are not

An entire issue of Trends in Cognitive Sciences was devoted to the topic of Bayesian models of cognition [1] and there has been a surge of interest in Bayesian models of perception, learning and reasoning [2–6]. The essential premise of the Bayesian approach is that the rational, normative way to adjust knowledge when new data are observed is to apply Bayes' rule (i.e. the mathematically correct formula) to whatever representational structures are available to the reasoner. The promise that spontaneous human behavior might be normatively Bayesian on some to-be-discovered representation has driven a surge in theoretical and empirical research.

Ironically, the behavior of researchers themselves has often not been Bayesian. There are many examples of a researcher positing a Bayesian model of how people perform a cognitive task, collecting new data to test the predictions of the Bayesian model, and then using non-Bayesian methods to make inferences from the data. These researchers are usually aware of Bayesian methods for data analysis, but the mortmain of 20th century methods compels adherence to traditional norms of behavior.

Traditional data analysis has many well-documented problems that make it a feeble foundation for science, especially now that Bayesian methods are readily accessible [7–9]. Chief among the problems is that the basis for declaring a result to be 'statistically significant' is ill defined: the so-called p value has no unique value for any set of data. Another problem with traditional analyses is that they produce impoverished estimates of parameter values, with no indication of trade-offs among parameters and with confidence intervals that are ill defined because they are based on p values. Traditional methods also often impose many computational constraints and assumptions into which data must be inappropriately squeezed.

The death grip of traditional methods can be broken. Bayesian methods for data analysis are now accessible to all, thanks to advances in computer software and hardware. Bayesian analysis solves the problems of traditional methods and provides many advantages. There are no p values in Bayesian analysis, inferences provide rich and complete information regarding all the parameters, and models can be readily customized for different types of data. Bayesian methods also coherently estimate the probability that an experiment will achieve its goal (i.e. the statistical power or replication probability).

It is important to understand that Bayesian methods for data analysis are distinct from Bayesian models of mind. In Bayesian data analysis, any useful descriptive model of the data has parameters estimated by normative, rational methods. The descriptive models have no necessary relation or commitment to particular theories of the natural mechanisms that actually generated the data. Thus, every cognitive scientist, regardless of his or her preferred model of cognition, should use Bayesian methods for data analysis. Even if Bayesian models of mind lose favor, Bayesian data analysis remains appropriate.

Null hypothesis significance testing (NHST)

In NHST, after collecting data, a researcher computes the value of a summary statistic such as t or F or χ^2, and then determines the probability that so extreme a value could have been obtained by chance alone from a population with

Glossary

Analysis of variance (ANOVA): when metric data (e.g. response times) are measured in each of several groups, traditional ANOVA decomposes the variance among all data into two parts: the variance between group means and the variance among data within groups. The underlying descriptive model can be used in Bayesian data analysis.

Bayes' rule: a simple mathematical relationship between conditional probabilities that relates the posterior probability of parameter values, on the one hand, to the probability of the data given the parameter values, and the prior probability of the parameter values, on the other hand. The formula is named after Thomas Bayes (1702–1761), an English minister and mathematician.

Chi-square (χ^2): the Pearson χ^2 value is a measure of the discrepancy between the frequencies observed for nominal values and what would be expected according to a (null) hypothesis. In NHST, the sampling distribution of the Pearson χ^2 distribution is approximated by the continuous χ^2 distribution when the expected frequencies are not too small.

Descriptive versus explanatory model: descriptive models summarize relations between variables without ascribing mechanistic meaning to the functional form or to the parameters, whereas explanatory models do make such semantic ascriptions. Bayesian inference for descriptive models of data is desirable regardless of whether Bayesian explanatory models account for cognition.

***p* value:** in NHST, the p value is the probability of obtaining the observed value of a sample statistic (such as t, F, χ^2) or a more extreme value if the data were generated from a null-hypothesis population and sampled according to the intention of the experimenter, where the intention could be to stop at a pre-specified sample size or after a pre-specified sampling duration, or to check after every 10 observations and stop either when significance is achieved or at the end of the week, whichever is sooner.

Corresponding author: Kruschke, J.K. (kruschke@indiana.edu).

[C]onsider a situation in which painstaking survey work has previously established that in the general population only 1% of subjects abuse a certain dangerous drug. Suppose that a person is randomly selected from [the] population for a drug test and the test yields a positive result. Suppose that the test has a 99% hit rate and a 5% false alarm rate. [How certain are we that the person is abusing the drug?]

Basic Probability

- Probability *Distribution*: All possible values of a given variable, and their probabilities (sum = 1):
- cavity=0.8; gingivitis = 0.1; abscess = 0.05; ? = 0.05
- *Joint* probability: How likely is it that two things occur (*are observed*) together?
- rainy & cloudy = 0.3; cloudy & cool = 0.4

Bayes' Rule

- From the Product Rule:
  - $P(A|B) = P(A \& B) / P(B)$
  - $P(A \& B) = P(A|B) * P(B)$
    $= P(B|A) * P(A)$

Rev. Thomas Bayes
(1702-1761)

- We derive Bayes' Rule by substitution:
  - $P(A|B) = P(A \& B) / P(B)$
    $= P(B|A) * P(A) / P(B)$

Marginalization: Revisiting the Drug Test Example

Suppose that the test has a 99% hit rate and a 5% false alarm rate. If we ignore the prior knowledge, we would conclude that there is at least a 95% chance that the tested person abuses the drug. But if we take into account the strong prior knowledge, then we conclude that there is only a 17% chance that the person abuses the drug.

Marginalization:
Revisiting the Drug Test Example

Suppose that the test has a 99% hit rate and a 5% false alarm rate. If we ignore the prior knowledge, we would conclude that there is at least a 95% chance that the tested person abuses the drug. But if we take into account the strong prior knowledge, then we conclude that there is only a 17% chance that the person abuses the drug.

$$P(A|B) = P(B|A) * P(A) / P(B)$$

A = Abusing

B = Bad test result

Marginalization: Revisiting the Drug Test Example

Suppose that the test has a 99% hit rate and a 5% false alarm rate. If we ignore the prior knowledge, we would conclude that there is at least a 95% chance that the tested person abuses the drug. But if we take into account the strong prior knowledge, then we conclude that there is only a 17% chance that the person abuses the drug.

$$P(A|B) = P(B|A) * P(A) / P(B)$$

$$0.17 = 0.99 * 0.01 / P(B)$$

Marginalization: Revisiting the Drug Test Example

Suppose that the test has a 99% hit rate and a 5% false alarm rate. If we ignore the prior knowledge, we would conclude that there is at least a 95% chance that the tested person abuses the drug. But if we take into account the strong prior knowledge, then we conclude that there is only a 17% chance that the person abuses the drug.

$$P(A|B) = P(B|A) * P(A) / P(B)$$

$$0.17 = 0.99 * 0.01 / P(B)$$

$P(B) = .058$ Why?

Marginalization:
Revisiting the Drug Test Example

$P(B=T \mid A=T) = 0.99$	*hit*
$P(B=F \mid A=T) = 0.01$	*miss*
$P(B=T \mid A=F) = 0.05$	*false positive*
$P(B=F \mid A=F) = 0.95$	*ordinary situation*

Marginalization: Revisiting the Drug Test Example

P(B=T | A=T) = 0.99	*hit*
P(B=F | A=T) = 0.01	*miss*
P(B=T | A=F) = 0.05	*false positive*
P(B=F | A=F) = 0.95	*ordinary situation*

A	B	P	
T	T	(.01)(.99)	*P(B=T) = .99 when A=T*
T	F	(.01)(.01)	*P(B=F) = .01 when A=T*
F	T	(.99)(.05)	*P(B=T) = .05 when A=F*
F	F	(.99)(.95)	*P(B=T) = .95 when A=F*

Marginalization: Revisiting the Drug Test Example

$$P(B=T) = (.01)(.99) + (.99)(.05) = .0594$$

A	B	P
T	T	(.01)(.99)
T	F	(.01)(.01)
F	T	(.99)(.05)
F	F	(.99)(.95)

Maximum Likelihood Estimation: Determining the Likeliest Hypothesis

$P(A|B) = P(B|A) * P(A) / P(B)$

Once we've made our observation B (e.g., patient has a toothache), we can treat / P(B) as a constant:

$P(A|B) = P(B|A) * P(A) * k$

Now choose the value of A that maximizes P(A|B) ...

Maximum Likelihood Estimation:

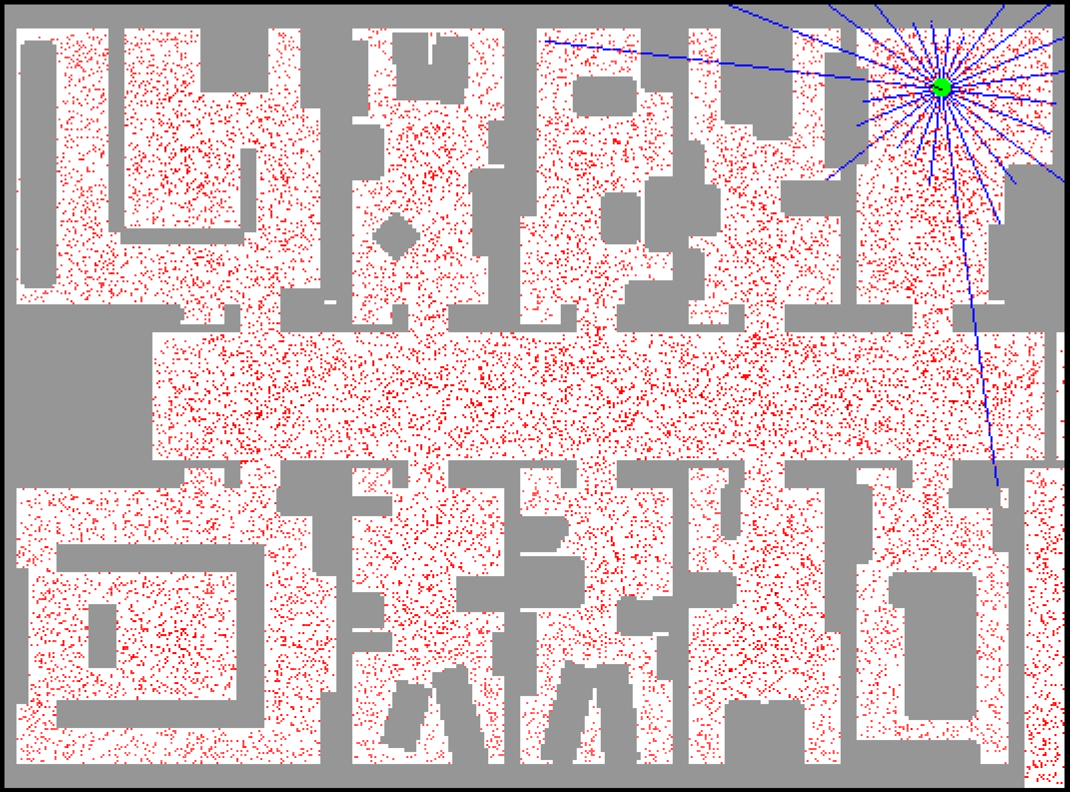
P(A=cavity|B=toothache) = 1.7 k

P(A=abscess|B=toothache) = 0.32 k

P(A=gingivitis|B=toothache) = 0.87 k

etc.

Maximum Likelihood Estimation: Robot Localization Problem



http://home.wlu.edu/~levys/courses/csci102w2015/lectures/global-floor.gif

Selection Bias and Berkson's Paradox

- Consider a College that admits students who are either Sporty, Brainy, or both:

$P(C|S) = 1.0$ $P(C|B) = 1.0$

- Assume Sporty, Brainy are independently distributed traits (knowing that someone is Brainy tells you nothing about whether she's Sporty, and vice-versa):

$P(B) = 0.5$ $P(S) = 0.5$ $P(B\&S) = 0.25$

Selection Bias and Berkson's Paradox

- If a student was admitted and she's brainy, can you say anything about whether she's sporty?

$P(B|C) = 0.67$

$P(B|C\&S) = ???$

Selection Bias and Berkson's Paradox

• If a student was admitted and she's brainy, can you say anything about whether she's sporty?

$$P(B|C) = 0.67$$

$$P(B|C\&S) = 0.5$$

• Braininess and Sportiness "compete" as explanations of getting into College.

• By looking only within the College population, we've biased our selection away from the general population