Teaching Innovations based on Social Science Research

Gary King
Institute for Quantitative Social Science

Talk at the Harvard Board of Overseers Meeting, 2/5/2012
The Context
What is Harvard’s Biggest Threat?
Social connections motivate
E.g., easier to get you to recycle (for the community) than to exercise
so you can live longer!
E.g., Voter turnout drives fail, unless you hear about your neighbors
People mind wander: 50% of the time; During conversation: only 25%
Teaching teaches the teacher!
(Help students teach each other)
Instant feedback improves learning
Social connections motivate

E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!

E.g., Voter turnout drives fail, unless you hear about your neighbors.

People mind wander: 50% of the time; During conversation: only 25%
1. Social connections motivate

- E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
Social connections motivate

- E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
- E.g., Voter turnout drives fail, unless you hear about your neighbors
Social connections motivate

- E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
- E.g., Voter turnout drives fail, unless you hear about your neighbors
- People mind wander: 50% of the time
Social connections motivate

- E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
- E.g., Voter turnout drives fail, unless you hear about your neighbors
- People mind wander: 50% of the time; During conversation: only 25%
Social science principles for teaching innovation

1. Social connections motivate
   - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
   - E.g., Voter turnout drives fail, unless you hear about your neighbors
   - People mind wander: 50% of the time; During conversation: only 25%

2. Teaching teaches the teacher! (Help students teach each other)
Social Science Principles for Teaching Innovation

1. Social connections motivate
   - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
   - E.g., Voter turnout drives fail, unless you hear about your neighbors
   - People mind wander: 50% of the time; During conversation: only 25%

2. Teaching teaches the teacher! (Help students teach each other)

3. Instant feedback improves learning
Collaborative Video Annotation

Let's Make a Deal

In Let's Make a Deal, Monte Hall offers what is behind one of three doors. Behind a random door is a car; behind the other two are goats. You choose one door at random. Monte peeks behind the other two doors and opens the one (or one of the two) with the goat. He asks whether you’d like to switch your door with the other door that hasn’t been opened yet. Should you switch?

```r
sims <- 1000
WinNoSwitch <- 0
WinSwitch <- 0
doors <- c(1, 2, 3)
for (i in 1:sims) {
  WinDoor <- sample(doors, 1)
  choice <- sample(doors, 1)
  if (WinDoor == choice) {  # no switch
    WinNoSwitch <- WinNoSwitch + 1
    doorsLeft <- doors[doors != choice]
    if (any(doorsLeft == WinDoor)) {  # switch
      WinSwitch <- WinSwitch + 1
    }
  }
}
cat("Prob(Car | no switch) =", WinNoSwitch/sims, "\n")
cat("Prob(Car | switch) =", WinSwitch/sims, "\n")
```
22 2 Conceptualizing uncertainty and inference

distinguish between these two cases, I refer to the hypothetical parameter value as \( \hat{\theta} \) and the single unobserved true value as \( \theta \). In the next chapter, I will introduce \( \hat{\theta} \) as a point estimator for \( \theta \), based on the maximum of the likelihood with respect to \( \hat{\theta} \), \( \hat{\theta} \) is a number in a single experiment, but a random variable across hypothetical experiments.

The likelihood that a hypothetical model (summarized by the hypothetical parameter value \( \hat{\theta} \)) produced the data we observe, given \( M^* \), is denoted \( L(\hat{\theta}|y,M^*) \), where \( M^* \) may again be suppressed since it appears in all subsequent expressions. The likelihood axiom then defines this concept as follows:

\[
L(\hat{\theta}|y,M^*) \equiv L(\hat{\theta}|y) = k(y)Pr(y|\hat{\theta}) \propto Pr(y|\hat{\theta}). \tag{2.5}
\]

In the second line of this equation, \( k(y) \) is an unknown function of the data; since it is not a function of \( \theta \), it is treated as an unknown positive constant. In the third line, “\( \propto \)” means “is proportional to.” The third line is only a more convenient way of writing the second without the constant. For a given set of observed data, \( k(y) \) remains the same over all possible hypothetical values of \( \hat{\theta} \). However, \( k(y) \) is a function of \( y \) and therefore may change as \( y \) changes. The likelihood \( L(\hat{\theta}|y) \) is similar to the concept of inverse probability in that it permits one to measure and compare the uncertainty one has about alternative hypothetical values of \( \hat{\theta} \). However, the unknown value \( k(y) \) ensures that likelihood is a relative rather than an absolute measure of uncertainty. This likelihood axiom is but one way to make the measure explicitly relative. Indeed, one could use any monotonic function of \( Pr(y|\hat{\theta}) \). The choice represented in Equation (2.5) is arbitrary, just as is the choice of making the scale of probability range between 0 and 1. The advantage of likelihood is that it can be calculated from a traditional probability, whereas inverse probability cannot be calculated in any way.

If the data are continuous rather than discrete, the likelihood is calculated in the same way, except that the underlying probability distribution is now a density. Hence, a more general way to write the formula is as follows:
[gov2001-l] likelihood vs loglikelihood

Phillip Y. Lipsycy  lipsycy at fas.harvard.edu
Wed Mar 5 00:37:25 EST 2003

For 2b, if we use the product term of the negative binomial distribution to estimate the likelihood rather than the loglikelihood, am I right to assume that we should get the same result for the maximum point? i.e. we use loglikelihood instead of likelihood to get rid of the product term, but our results should not change?

Thanks,
Phillip.
[gov2001-l] HW 2 Code  Andrew Reeves
  o  [gov2001-l] HW 2 Code  Kosuke Imai

[gov2001-l] hw3 Question 2  Kosuke Imai

[gov2001-l] question 2b  Mee-Jung Jang
  o  [gov2001-l] question 2b  Kosuke Imai
    - [gov2001-l] question 2b  Stanislav Markus
    - (gov2001-l] question 2b  Olivia Lau
    - [gov2001-l] question 2b  Kosuke Imai

[gov2001-l] question 2b  Daniel Hopkins
  - [gov2001-l] question 2b  Kosuke Imai

[gov2001-l] back to basics: Question 1(b)  Ryan Thomas Moore
  o  [gov2001-l] back to basics: Question 1(b)  Kosuke Imai
    - [gov2001-l] back to basics: Question 1(b)  Phillip Y. Lipsy
    - [gov2001-l] back to basics: Question 1(b)  Kosuke Imai

[gov2001-l] 2(a)  Traci Burch
  o  [gov2001-l] 2(a)  Jennifer Fitzgerald

[gov2001-l] likelihood vs loglikelihood

Phillip Y. Lipsy  lipscy at fas.harvard.edu

Wed Mar 5 00:37:25 EST 2003

For 2b, if we use the product term of the negative binomial distribution to estimate the likelihood rather than the loglikelihood, am I right to assume that we should get the same result for the maximum point? i.e. we use loglikelihood instead of likelihood to get rid of the product term, but our results should not change?

Thanks,
Phillip.
What’s left to do in class?

Focus on what they don’t know!

(Intensely participatory) discussions on topics they think they’re confused about

Lecture on topics they are confused about

A version of “peer instruction”
What’s left to do in class? **Focus on what they don’t know!**

- (Intensely participatory) discussions on topics they think they’re confused about
What’s left to do in class?  **Focus on what they don’t know!**

- (Intensely participatory) discussions on topics they think they’re confused about
- Lecture on topics they are confused about
What’s left to do in class? **Focus on what they don’t know!**

- (Intensely participatory) discussions on topics they think they’re confused about
- Lecture on topics they are confused about
- A version of “peer instruction”
What’s left to do in class? Focus on what they don’t know!

- (Intensely participatory) discussions on topics they think they’re confused about
- Lecture on topics they are confused about
- A version of “peer instruction”
Assume the model below:

\[ Y_i \sim N(\mu, \sigma_i^2) \]

\[ \sigma_i^2 = x_i \beta \]

This model is useful for:

A. Nothing, the model is internally inconsistent.
B. To study whether as unemployment increases, agreement about whether the president is doing a good job goes down.
C. To study how predictable \( Y_i \) is.
D. To test if the variance is non-negative.
E. To study whether female unemployment is higher than male unemployment.

**Round 1**
- A. 12%
- B. 36%
- C. 51%
- D. 10%
- E. 10%

**Round 2**
- A. 12%
- B. 48%
- C. 83%
- D. 2%
- E. 8%

14 get it now
0 still don’t get it
A positively charged rod is held near a neutral conducting sphere as illustrated below. A positively charged particle is moved from point A to point B at constant speed. The mechanical work required to cause this motion is
A positively charged rod is held near a neutral conducting sphere as illustrated below. A positively charged particle is moved from point A to point B at constant speed. The mechanical work required to cause this motion is

Please discuss your response with:

- Brian Lukoff (to your left)
Social Science Principles
1 Social connections motivate
1. Social connections motivate
2. Teaching teaches the teacher
Social connections motivate

Teaching teaches the teacher

Instant feedback improves learning
Social Science Principles

1. Social connections motivate
2. Teaching teaches the teacher
3. Instant feedback improves learning
4. The advantages of large scale measurement and analysis