Describing dynamics in classroom education using simple teaching games

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

We describe a model of classroom teaching that construes teaching as communication to a heterogeneous audience. A number of basic educational results fall out of this construal, including (1) decreasing mean performance with the increasing size and variability among students in a class, (2) increases in performance based on grouping students by abilities, and (3) the value of formative evaluation to enhance teachers' knowledge of student ability.

Keywords: Bayesian modeling; agent-based modeling; education; pragmatics; teaching

Introduction

To fill this need, we describe a model of classroom teaching. This model captures phenomena that have to do with the informational dynamics of the classroom—how much information can be transferred between a teacher and a group of students with certain abilities and prior beliefs. It has nothing to say about another—perhaps ultimately more important—part of the classroom experience, its motivational dynamics.

What are the functions of such a model?

Teaching games

The basic unit of our analysis is a teaching game. In such a game, teacher T attempts to provide information to students $S = s_1 ... s_m$. Teacher conveys information by choosing examples $E = e_1 ... e_n$ to illustrate an underlying concept C, based on some estimate of the students' prior knowledge and abilities $\hat{S} = \hat{s_1} ... \hat{s_m}$. Learners in turn attempt to recover C with maximal fidelity. The teacher's payoff is determined by a test, administered to each of the learners, with some payoff function.

We will begin by considering a very simple form of this sort of game, with only one student, one example, and perfect teacher knowledge of that student's abilities. We think of this as the "optimal tutoring" regime. In this simple game, the concept to be learned is the parameter on a Bernoulli variable (the weight on a coin, e.g.). The teacher must choose a single observation to give to the student.

Students

We model the student here as a Bayesian (optimal) estimator of this Bernoulli distribution, using a conjugate Beta-Bernoulli distribution. This model is very convenient: the form of the prior distribution is $Beta(\alpha, \beta)$, and the form of the posterior can be written $Beta(\alpha + t, \beta + h)$ where t and h represent the number of heads (0) and tails (t) observed in the data respectively. In this sense, if t and h are the *counts* of observed data, then α and β can be referred to as *pseudo-counts*.

This formulation also gives us a way to model both the student's abilities and their prior knowledge about the situation. Consider the example beta-Bernoulli distributions shown in Figure 1. Symmetric priors of $\alpha = \beta = .5$ lead to a bias that the target coin weight is either 0 or 1, while $\alpha = \beta = .5$ leads to a bias towards fairer coins. As the strength of the prior grows, the effect of observing a single coin flip becomes weaker.

Under this formulation, the prior controls both the speed at which a student will learn and their overall bias. For example, as α and β both go towards 0, the student's estimate converges to a maximum-likelihood estimate based on the observed data alone. In contrast, as α and β both get larger, the student makes less and less use of the data and is more and more reliant on the shape of the prior distribution. The relative weights of α and β control the student's bias—greater pseudo-counts on one or the other will lead to greater bias to believe that the correct parameter is lower or higher. We explore each of these scenarios—learning speed and bias—below.

$$P(S = c) = \frac{C^{\alpha-1}(1-C)^{\beta-1}}{B(\alpha, \beta)} \quad (1)$$

where B is the beta function, and hence

$$P(S' = C) = \frac{C^{\alpha+h-1}(1-C)^{\beta+t-1}}{B(\alpha+h, \beta-t)} \quad (2)$$

Evaluation

We define a test as a set of examples drawn randomly from c that must be guessed.

$$\begin{aligned} L(S', C) &= P(S' = C) && (3) \\ &= && (4) \end{aligned}$$

Teachers

Simulations

Classroom size

Tracking

Testing for better knowledge

Decision-theoretic analyses

Acknowledgments

Place acknowledgments (including funding information) in a section at the end of the paper.

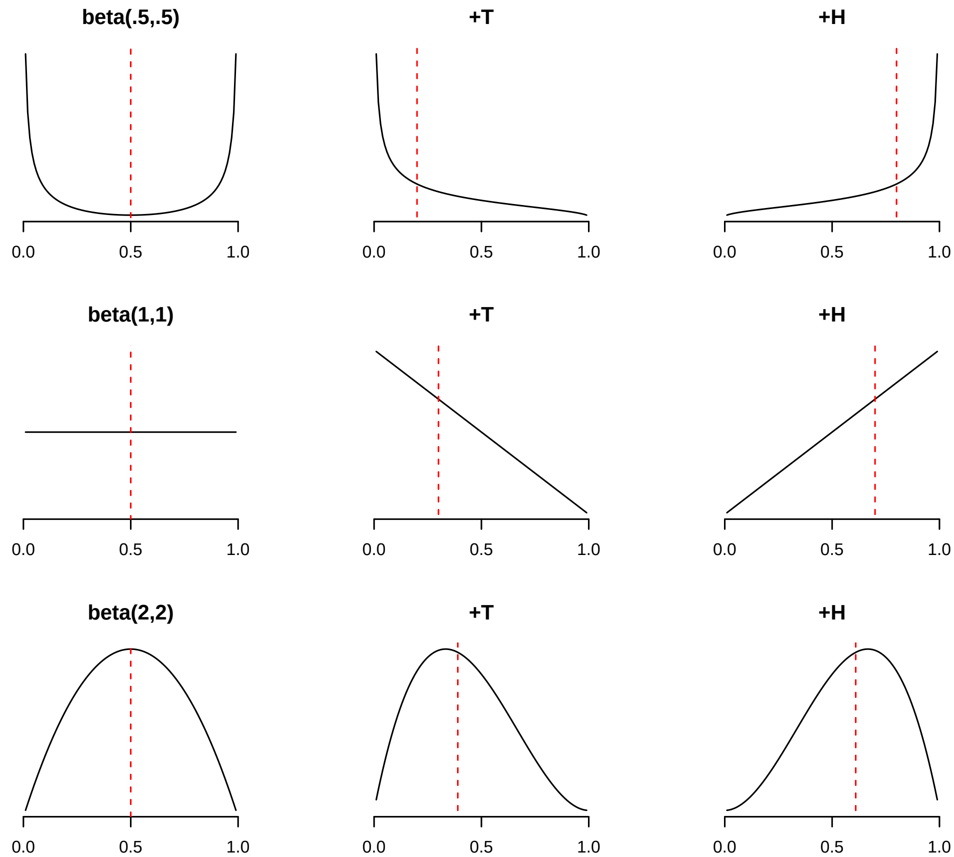

Figure 1: Examples of the beta-bernoulli distribution with different priors and patterns of evidence. Black curves show probability distribution with a given prior (left column) and after observing a single tail or head (middle and right columns). Red lines show posterior mean.

References Instructions

Follow the APA Publication Manual for citation format, both within the text and in the reference list, with the following exceptions: (a) do not cite the page numbers of any book, including chapters in edited volumes; (b) use the same format for unpublished references as for published ones. Alphabetize references by the surnames of the authors, with single author entries preceding multiple author entries. Order references by the same authors by the year of publication, with the earliest first.

Use a first level section heading for the reference list. Use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 1/8 inch. Below are example references for a conference paper, book chapter, journal article, technical report, dissertation, book, and edited volume, respectively.

References

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287–317.

Lewis, C. (1978). *Production system models of practice effects*. Doctoral dissertation, Department of Psychology, University of Michigan, Ann Arbor.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

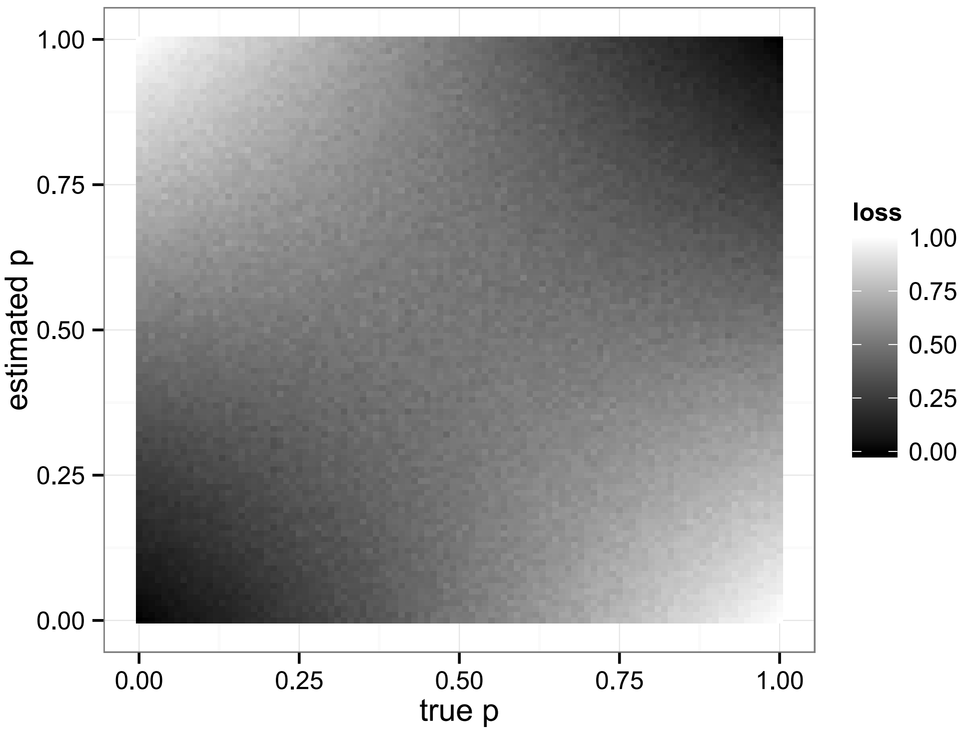

Figure 2: Map of the loss function for testing-based evaluation, assessed numerically.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. No. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.