LEARNING PROBABILITY THROUGH BUILDING COMPUTATIONAL MODELS

URI WILENSKY  
Center for Connected Learning  
Northwestern University  
Annenberg Hall 311  
2120 Campus Drive  
Evanston, IL 60208  
uriw@media.mit.edu  
847-647-3818

Epistemology & Learning Group  
Learning & Common Sense Section  
The Media Laboratory  
Massachusetts Institute of Technology  
20 Ames Street Room E15-315  
Cambridge, MA 02139  
uriw@media.mit.edu

Abstract

While important efforts have been undertaken to advancing understanding of probability using technology, the research herein reported is distinct in its focus on model building by learners. The work draws on theories of Constructionism and Connected Mathematics. The research builds from the conjecture that both the learner's own sense making and the cognitive researchers' investigations of this sense-making are best advanced by having the learner build computational models of probabilistic phenomena. Through building these models, learners come to make sense of core concepts in probability. Through studying this model building process, and what learners do with their models, researchers can better understand the development of probabilistic learning. This report briefly describes two case studies of learners engaged in building computational models of probabilistic phenomena.

Introduction

In the Connected Probability project (Wilensky, 1993; 1994), we explore ways for learners (both secondary and post-secondary) to develop intuitive conceptions of core probabilistic concepts. Computational technology can play an important role in enabling learners to build intuitive conceptions of probability. Through building computational models of everyday and scientific phenomena, learners can build mental models of the underlying probability and statistics. Even learners not usually considered good at mathematics and science can build models that demonstrate a qualitatively greater level of mathematical achievement than is usually found in mathematics classrooms. "Emergent phenomena", in which global patterns emerge from local interactions, are authentic contexts for learners to build with probabilistic parts. By giving probabilistic behavior to distributed computational agents, stable structures can emerge. Thus, instead of learning probability through solving decontextualized combinatoric formulae or being consumers of someone else's black box simulations, learners can participate in constructionist activities - they design...
and build with probability.

As part of the Connected Probability project, we have extended the StarLogo parallel modeling language (Resnick, 1992; Wilensky, 1993) and tailored it for building probabilistic models. The StarLogo language is an extension of the computer language Logo that allows learners to control thousands of screen "turtles". These turtles or computational agents have local state and can be manipulated as concrete objects. Through assigning thousands of such turtles probabilistic rules, learners pursue both forwards and backwards modeling. Forwards modeling involves exploring the effects of various sets of local rules to see what global pattern emerges, while in backwards modeling learners try to find an adequate set of local rules to produce a particular global effect. In this report, two case studies of probabilistic modeling projects are presented.

**Theoretical Framework: Constructionism and Connected Mathematics**

This research is organized and structured by a theory of Connected Mathematics (Wilensky, 1993). Connected Mathematics responds to a prevailing view that mathematics must be seen as "received" or "given" and graspable only in terms of formalism per se.

Connected Mathematics is situated in the constructionist learning paradigm (Papert, 1991). The Constructionist position advances the claim that a particularly felicitous way to build strong mental models is to produce physical or computational constructs which can be manipulated and debugged by the learner. As described by Wilensky (1993), Connected Mathematics also draws from many sources in the mathematics reform movement (e.g., Confrey, 1993; Dubinsky & Leron, 1993; Feurzeig, 1989; Hoyles & Noss, 1992; Lampert, 1990; Schwartz, 1989; Thurston, 1994).

A Connected Mathematics learning environment focuses on learner-owned investigative activities followed by reflection. Thus, there is a rejection of the mathematical "litany" of definition-theorem-proof and an eschewal of mathematical concepts given by formal definitions. Mathematical concepts are multiply represented (Kaput, 1989; Mason, 1987; von Glaserfeld, 1987) and the focus is on learners designing their own representations. Learners are supported in developing their mathematical intuitions (Wilensky, 1993) and building concrete relationships (Wilensky, 1991) with mathematical objects. Mathematics is seen to be a kind of sense-making (e.g., Schoenfeld, 1991) both individually and as a social negotiation (e.g., Ball, 1990; Lampert, 1990). In contrast to the isolation of mathematics in the traditional curriculum, it calls for many more connections between mathematics and the world at large as well as between different mathematical domains (e.g., Cuoco & Goldenberg, 1992; Wilensky, 1993).

**The Role of Technology**

The idea that mathematics is not simply received and formal implies a vision for how technology can be used. Not to simply animate received truth (e.g., by running black-box simulations) but instead as a medium for the design of models by learners. Under a traditional formalistic framework, mathematics is "given" and technology is seen as simply animating what is already known. In Connected Mathematics, knowing is situated and technology provides an environment in which understanding can develop. Learners literally construct an environment in which they then construct their understanding.

Because, when learners build computational models, they articulate their conceptual models through their
design, researchers can gain access to these conceptual models (see e.g., Collins & Brown, 1985; Pea, 1985). The researcher is given insight into the thinking of the learner at two levels: as model builder and as model consumer.

**Building Models vs. Running Simulations**

Computer based simulations of complex phenomena are becoming increasingly common (see e.g., Rucker, 1993; Stanley, 1989; Wright, 1992a, 1992b). In a simulation, the learner is presented with and explores a sophisticated model (built by an expert) of a subject domain. The user can adjust various parameters of the model and explore the consequences of these changes. The ability to run simulations (or pre-built models) interactively is a vast improvement over static textbook-based learning with its emphasis on formulae and the manipulation of mathematical tokens. Stanley (e.g., Shore et al, 1992) has demonstrated that curricular materials based on simulations of probabilistic phenomena can be very engaging to secondary students and teachers. But, in simulations, generally, learners do not have access to the workings of the model. Without access to the underlying structures, learners may perceive the model in a way quite at variance with the designer's intentions. Furthermore, learners cannot explore the implications of changing these structures. Consequently, their ability to develop robust mental models of these structures is inhibited. A central conjecture of this research is that for learners to make powerful use of models, they must first build their own models and design their own investigations. It is only by exploring the space of possible models of the domain that learners come to appreciate the power of a good model. To support users in building useful models, a number of powerful modeling environments have been designed. (e.g., STELLA - Richmond & Peterson, 1990, Roberts, 1978; Starlogo - Resnick, 1992; Wilensky, 1993; Agentsheets - Repenning, 1993; KidSim - Smith, Cypher & Spohrer, 1994).

**Extensible models**

In the spirit of Eisenberg's use of extensible application (Eisenberg, 1991), extensible models are pre-built models or simulations that are embedded in a general purpose modeling language. This combined approach has many of the advantages of both simulation and model building: there is a rich domain model to be investigated, access is given to the structure of the model, users can modify this structure, and even use it as a basis for building their own models and tools.

The challenge for such an approach is to design the right middle level of primitives so that they are neither (a) too low-level, so that the extensible model becomes identical to its underlying modeling language, nor (b) too high-level, so that the application turns into an exercise of running a small set of pre-conceived experiments.

**Probability**

The domain of probability (and statistics) has been an ongoing focus of research within the Connected Mathematics program. There are many reasons to recommend probability as a content domain. Among these are:

- There is a considerable literature attesting to the difficulty people have with understanding probability (e.g., Kahneman & Tversky, 1982, Nisbett et al, 1983, Konold, 1991). Standard instruction has been
shown to provide little remedy. Educators have responded to this research by advising students not to trust their intuitions when it comes to probability and to rely solely on the manipulation of formalisms. As a result, learners construct brittle formal models of the core probabilistic concepts and fail to link them to everyday knowledge. Connected Mathematics provides an alternative to this formalistic stance. It asserts that powerful probabilistic intuitions can be constructed by learners (Wilensky, 1993; 1994; forthcoming). By taking up such a challenging domain, a strong proof of the value of Connected Mathematics can be demonstrated.

- Computational environments can open doors to new ways of thinking about probability. Computational environments allow users to construct stable products (e.g., normal distributions, see below) using random components. This construction would be very difficult to do without computational environments. From a constructionist perspective, this ability to build meaningful products from random components is a prerequisite for making sense of the core notion of randomness.

- Particularly in the area of probability and statistics, the educational goal should emphasize interpreting (and designing) statistics from science and life rather than mastery of curricular materials. In order to make sense of scientific studies, it is not sufficient to be able to verify the stated model; one needs to see why those models are superior to alternative models. In order to understand a newspaper statistic, one must be able to reason about the underlying model used to create that statistic and evaluate its plausibility. For these purposes, building probabilistic and statistical models is essential.

- Many everyday phenomena exhibit emergent behavior: the growth of a snowflake crystal, the perimeter pattern of a maple leaf, the advent of a summer squall, the dynamics of the Dow Jones or of a fourth grade classroom. These are all systems which can be modeled as composed of many distributed but interacting parts. They all exhibit non-linear or emergent qualities which place them well beyond the scope of current K-12 mathematics curricula. Yet, through computational modeling, especially with parallel languages such as Starlogo, pre-college learners can gain mathematical purchase on these phenomena. Modeling these everyday complex systems can therefore be a motivating and engaging entry point into the world of probability and statistics.

**The Case of Normal Distributions**

As part of my efforts to create learning environments for probability, I have used a carefully selected set of materials (consisting of newspaper clippings, probability puzzles and paradoxes, core probability concepts and computational tools) to stimulate learners to pursue their own investigations and design their own computational tools for pursuing their inquiry.

One such example is Alan, a student with a strong mathematical background who nevertheless felt that he "just didn't get" normal distributions. Using a version of the parallel modeling language Starlogo which was enhanced for focusing on probability investigations (Wilensky, 1993; 1994), Alan developed a model for explaining his question,"Why is height (in men) normally distributed?" Alan's they was that perhaps "Adam" had children which were either taller or shorter than him with a certain probability. If this process was repeated with the children, then a distribution of heights would emerge. To explore what kinds of distributions were possible from this model, Alan built a "rabbit jumping" microworld. Taking advantage of the parallel modeling environment, Alan placed sixteen thousand rabbits in the middle of a computer screen. He then gave each rabbit a probabilistic jumping rule. (In Alan's model, the location of the rabbit corresponds to a person's height and a jump corresponds to a set deviation in height). The first such rule he explored was to tell each rabbit to jump left one step or right one step each with probability 1/2. After a number of steps, the classic symmetric binomial distribution became apparent. Alan was pleased with that
outcome but then asked himself the question: what rule should I give the rabbits in order to get a non-symmetric distribution? His first attempt was to have the rabbits jump two steps to the right or one step to the left with equal probability. He reasoned that the rabbits would then be jumping more to the right so the distribution should be skewed right. His surprise was evident when the distribution stayed symmetric while moving to the right. It didn't take too long though before he realized that it was the different sized probabilities not the different sized steps that made the distribution asymmetric. This example, while seemingly elementary, captures many facets of the model building approach to learning about complexity:

- The question was owned by the learner
- Theories were instantiable and testable
- Buggy theories could be successively refined
- The modeling environment did not limit the directions of inquiry

The environment provides a suitable set of syntactic primitives so that his model was easily built. It provided a suitable set of conceptual primitives that guided Alan’s investigation. In particular the parallelism of the modeling environment puts a focus on the relationship between micro- and macro- aspects of the problem. Typically, distributions are learned and classified by their macro- features (e.g., mean, standard deviation, variance, skew, moments) but the realization that distributions are emergent effects of numerous micro-level interactions is lost. This is a key point since 1) the concept of distribution is central to probability and statistics and 2) this failure to connect levels makes distributions seem like formal received mathematics, mathematics to be memorized and understood solely through formulae. In contrast, Alan constructs distributions and is able to link their macro- properties to the micro- rules he has given them.

**GPCEE - The Case of the Gas in a Box**

Harry is a science and mathematics teacher in the Boston public schools. He was very interested in the behavior of gas particles in a closed box. He remembered from school that the energies of the particles when graphed formed a stable distribution called a Maxwell-Boltzman distribution. Yet, he didn’t have any intuitive sense of why they might form this stable asymmetric distribution. He decided to build a model of gas particles in a box using the Starlogo modeling language.

The model Harry built is initialized to display a box with a specified number of gas "molecules" randomly distributed inside it. The user can then perform "experiments" with the molecules.

The molecules are initialized to be of equal mass and start at the same speed (that is distance traveled in one clock tick) but at random headings. Using simple collision relations, Harry was able to model elastic collisions between gas molecules, (i.e., no energy is "lost" from the system). The model can be run for as many ticks as wanted.

By using several output displays such as color coding particles by their speed or providing dynamic histograms of particle speeds/energies, Harry was able to gain an intuitive understanding of the stability and asymmetry of the Boltzman distribution.

Harry's story is told in greater detail elsewhere (Wilensky, forthcoming). {Originally, Harry had thought that because gas particles collided with each other randomly, they would be just as likely to speed up as to slow down. But now, Harry saw things from the perspective of the whole ensemble of particles. He saw that high velocity particles would "steal lots of energy" from the ensemble. The amount they stole would be
proportional to the square of their speed. It then followed that, since the energy had to stay constant, there
had to be many more slow particles to balance the fast ones.

This new insight gave Harry a new perspective on his original question. He understood why the Boltzman
distribution he had memorized in school had to be asymmetric. But it had answered his question only at the
level of the ensemble. What was going on at the level of individual collisions? Why were collisions more
likely to lead to slow particles than fast ones? This led Harry to conduct further productive investigations
into the connection between the micro- and macro- views of the particle ensemble.

GPCEE as an Extensible Model

Harry's story is not the end of the tale. Harry collaborated with me in making his model into an extensible
application. We call the model GPCEE (Gas Particle Collision Exploration Environment). Once GPCEE
became a publicly accessible model, we were struck by its capacity to attract, captivate and engage
"random" passers-by. People whose idea of fun did not include working out physics equations nonetheless
were mesmerized by the motion of the shifting gases, their pattern and colors. And they were motivated to
ask questions - why do more of them seem to slow down than speed up? What would happen if they were
all released in the center? in the corner?

As a result, many more learners used the GPCEE model and, since it was an extensible model, they
extended it. Among the extensions that users built were tools for measuring pressure in the box and viscosity
of the gas, pistons to compress the gas, different shapes for the container, different dimensional spaces for
the box, diffusion of two gases, different geometries for the molecules (e.g., diatomic molecules with
rotational and vibrational freedom), and sound wave propagation in the gas. In order to build the
computational tools for these extensions, users had to build conceptual models. They came to ask such
questions as: What kind of thing is pressure? How would you build a tool to measure it?

It is clear that GPCEE is both a physics simulation, one in which experiments difficult or impossible to do
with real gases can be easily tried out, and an environment for strengthening intuitions about the statistical
properties of ensembles of interacting elements. Through creating experiments in GPCEE, learners can get a
feel for both the macro- level, the behavior of the ensemble as an aggregate, and its connections to the
micro-level what is happening at the level of the individual gas molecule. In the GPCEE application,
learners can visualize ensemble behavior all at once, sometimes obviating summary statistics. Furthermore,
they can create their own statistical measures and see what results they give at both the micro- and the
macro- level. They may, for example, observe that the average speed of the particles is not constant and
search for a statistical measure which is invariant. In so doing, they may construct their own concept of
energy. Their energy concept, for which they may develop a formula different than the formula common in
physics texts, will not be an unmotivated formula the epistemological status of which is naggingly
questioned in the background. Rather, it will be personally meaningful, having been purposefully designed
by the learner.

The necessity of creating his own summary statistic led one learner to shift his view of the concept of
"average". In the GPCEE context, he now saw "average" as just another method for summarizing the
behavior of an ensemble. Different averages are convenient for different purposes. Each has certain
advantages and disadvantages, certain features which it summarizes well and others that if doesn't. Which
average we choose or construct depends on how we wish to make sense of the data.

Having shown the GPCEE environment to quite a few professional physicists, I can attest to the fact that although they knew that particle speeds fell into a Maxwell-Boltzman distribution, most were still surprised to see more blue particles than red -- they had formal knowledge of the distribution, but the knowledge was not well connected to their intuitive conceptions of the model. In a typical physics classroom, learners have access either only to the micro level - through exact calculation of the trajectories of two colliding particles, or only to the macro-level, but in terms of pre-defined summary statistics selected by the physics canon. Based on this example, it would seem that it is in the interplay of these two levels of description that powerful explanations and intuitions develop.

References


http://ccl.northwestern.edu/papers/pme19/


As we know, computer scientists can now build extremely sophisticated computing machines. All these machines implement something resembling Turing computation, although the details differ from Turing’s simplified model. 2. Artificial intelligence. Uncertainty is codified through probability. Precise rules dictate how to update probabilities in light of new evidence and how to select actions in light of probabilities and utilities. (See the entries Bayes’s theorem and normative theories of rational choice: expected utility for details.) CCTM holds that a suitable abstract computational model offers a literally true description of core mental processes. It is common to summarize CCTM through the slogan “the mind is a Turing machine.” Generally, most models that do give probabilities are usually using a logistic function, so it can be hard to compare. It just tends to work well in practice, Bayesian networks are an alternative. Naive Bayes just makes too simplistic an assumption for its probabilities to be any good - and that is easily observed on any reasonably sized data set. In the end, its usually easier to increase the quality of your probability estimates by picking the model that can represent the data better. In this sense, it doesn't matter too much how you get the probabilities. Probabilistic models and learning. In probabilistic modeling we treat all our knowledge in terms of probabilities. We want to answer questions like: what is chance of a certain temperature tomorrow at noon, given recorded weather data y from the last week? We are not primarily interested of a point estimate of y, but rather a (mathematical) description of the chance for all possible outcomes of y. More formally, the question would be: what is the probability density p(y | y) for a prediction y conditional on the data y? To answer such questions, we need to reason about uncertainties in the If you’re learning about machine learning, that also means you’re learning about computational statistics, which in turn means you’re learning about traditional statistics. If you still want to start out by reading up. Continue Reading. None. With modern-day computers, you can iterate over millions of different models on gigantic datasets. That’s the only fundamental difference from traditional statistics. If you’re learning about machine learning, that also means you’re learning about computational statistics, which in turn means you’re learning about traditional statistics. If you still want to start out by reading up on a few concepts in statistics, I would suggest these: Correlation (and how it doesn’t imply causation).