Randomness: Rethinking the Foundation of Probability

Yan Liu
Vanderbilt University

Patrick Thompson
Vanderbilt University

In this article, we highlight a series of tensions inherent to understanding randomness. In doing so, we locate discussions of randomness at the intersections of a broad range of literatures concerned with the ontology of stochastic events and epistemology of probabilistic ideas held by people. Locating the discussion thus has the advantage of emphasizing the growth of probabilistic reasoning and deep connections among its aspects.

Although we find no explicit definition of *randomness*, we might catch a glimpse of its meaning by looking at what texts say are the meanings of *random phenomenon*, and *random sampling.* Yates, Moore & McCabe (1998, p314) define a *random phenomenon* as one in which “individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.” Bluman (2001, p. 632) defines a *random sampling* as one in which all possible samples of a certain size “must have an equal chance of being selected from the population.” As we shall elaborate below, these definitions are circular. In order for a student to understand them, they must already understand the idea of random variable and the distribution of its values, which itself entails an understanding of randomness. The problem, then, is that instruction is often designed to introduce students to ideas in ways that presume, essentially, that students already understand them.

Yet, surprisingly, we found little attention paid in statistics textbooks and research studies to what it means to understand randomness. Our analysis of popular secondary and college statistics and probability textbooks (Thompson & Liu, 2002) revealed that 34-100% of questions concerning probability are stated, grammatically speaking, as if about a single, unrepeated (and hence non-probabilistic) event. Moreover, if students understand the phrase “at random” simply as “cannot predict ahead of time” or “unexpected”, then 92-100% of the questions can be interpreted as if about a single, unrepeated (and hence non-probabilistic) event. Yet, helping students understand “at random” productively are not explored. As Falk and Konold (1994) state, the complex nature of randomness poses an educational dilemma: “Shall we forgo discussing the meaning of the concept in our teaching (relying on students’ existing intuitions)? Or shall we endeavor to find a satisfactory way of presenting randomness, undertaking the challenge of bringing up the doubts and difficulties students will predictably have?” (Falk & Konold, 1994, p. 2).

Randomness and Determinism

Ideas of randomness range along an ontological and epistemological spectrum. On one side of the spectrum, probabilities are regarded as not being inherent in objective nature but as reflecting human ignorance of a true determinist course of events. This view of probability was a product of the thinking of the European Enlightenment, the admiration for Newtonian mechanics, and the consequent belief in universal determinism. As Laplace (1814) remarked, “We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow.”(p. 4). It is only a short step from this to the conclusion that absolute randomness does not exist, and therefore that all probabilities will be 0 or 1. From a deterministic perspective, “Chance then exists not in nature, and cannot coexist with knowledge; it is merely an expression, as Laplace remarked, for our ignorance of the causes in action…Probability belongs wholly to the mind” (Jevons 1958, as quoted in McShane 1970, p.37). The other end of the spectrum comes from the renunciation of determinism that was brought by quantum mechanics in the twentieth century. The understanding of Brownian motion as an expression of constant internal agitation of microscopic particles without external causes (von Plato 1994), was taken as proving the existence of “irreducible chance” and therefore that randomness is an inherent feature of nature. “Causality, long the bastion of metaphysics, was toppled, or at least tilted” (Hacking 1990, p. 1). von Mises concluded, “there exist genuinely statistical sequences in nature” (von Mises 1957). However, controversies still exists as to whether randomness is a characteristic of a sequence itself or the process from which the sequence was generated. In *The Logic of Chance*, Venn said that it is ‘the nature of a certain arrangement,” but not ‘the particular way in which it is brought about,” that should be considered when judging a random arrangement, and the arrangement must be judged by what would be observed in the long run (Bennett 1998, p.166). Venn hoped to illustrate randomness by building a graph using the decimal expansion of π. From a deterministic perspective, however, even though the decimal expansion of numbers such as π, e, and $\sqrt{2}$ might exhibit disorderliness, the process from which each is generated is completely predetermined. Hence, for those who believe that randomness is a property of the process there exists no randomness in this situation. On the other hand, a sequence obtained from a random process may exhibit certain visible pattern in appearance. Thus, there is a necessary distinction between the idea of a random sequence and the idea of a sequence whose elements are chosen by a random process.

Formalization of randomness

von Mises' formulation of the notion of randomness was based on the intuition of "the impossibility of a gambling system". The randomness in a sequence lies in the "impossibility of devising a method of selecting the elements so as to produce a fundamental change in the relative frequencies"(von Mises 1957). This definition of randomness had long been criticized as being "not mathematically expressible" and being "too inexact to serve satisfactorily as the basis of a mathematical theory" (McShane 1970; Gillies 2000). Perhaps von Mises defined randomness as a property of a sequence because the concept of collective played such an important role in frequentist definitions of probability.

Kolmogorov and others later proved that von Mises-type sequences would exist if only simple formulas, rules, or laws of prediction are allowed, in other words, rather than requiring the randomness of a sequence to be judged by absolute unpredictability, Kolmogorov would require only unpredictability by a small set of simple rules. Kolmogorov further constructed a definition of random sequence based on the notion of complexity in information theory: a random sequence is one with maximal complexity, in other words, a sequence is random if the shortest formula which computes it is extremely long (von Plato, 1994). The modern mathematical treatment of random process is based on Kolmogorov's measure theoretical probability. In his approach a random process (or stochastic process) is a family of random variables having time t as its independent variable. Kolmogorov distinguished random process from determined process by considering time and distribution of outcomes simultaneously: "If the state y of a system at time t is uniquely defined by its state x at an arbitrary moment t_0 through a unique function f such that $y=f(x, t_0, t)$, situations of this general type is called *schemes of a well-determined process*. On the contrary, if the state x at time t_0 only determines a probability distribution for the possible future state y, these are called *schemes of a stochastically definite process*" (von Plato, 1994). Kolmogorov's solution to distinguishing between randomness and determinacy rests, however, on a troubling assumption: That the function $f(x, t_0, t)$ exists in its own right, which means that processes *have* certain properties timelessly, even if, over time, they vary.

Implications for teaching and learning

Philosophical and mathematical debates on what "randomness" means highlight its deeply problematic nature and therefore highlight an equally problematic question of what "probability" means. Two approaches to addressing the issue are (1) ignore it and (2) try to bring the debate to students' levels. We propose a third approach, which addresses the problem by cutting the Gordian knot. We note that both ends of the spectrum presume that the question is ontological, that it is about what randomness *is*, and approach it by speaking of it as a property of something that exists, be it sequence, process, or distribution. We propose that instructional design finesse the question, focusing instead on what might we *mean* by randomness. That is, focus on the imagery and operations that support coherent thinking and reasoning about situations that we see as entailing randomness and that we hope students come to see as entailing randomness. One scheme of images and operations that we will describe in our presentation is that students build an image of a loosely-coupled process with imprecisely-determined inputs that generates collections of outcomes that have predictable distributions in the long run but unpredictable distributions in the short run. We hasten to point out that we are not talking about adopting a perspective on the ontology of processes or collections. Rather, we are talking about helping students *imagine* processes and collections in a way that supports coherent stochastic reasoning.

References

Bennett, D. J. (1998). *Randomness*, Harvard Univeristy Press.
Bluman, A. G. (2000) *Elementary Statistics - A Step by Step Approach.* 3rd edition. Boston: McGraw-Hill.
Falk, R. and C. Konold (1994). "Random means hard to digest." *Focus on Learning Problems in Mathematics*(16): 2-12.
Gillies, D. (2000). *Philosophical Theories of Probability.* London and New York, Routledge.
Good, I. J. (1983). *Good Thinking: The Foundations of Probability and it Application.* Minneapolis, University of Minnesota Press.

Hacking, I. (1965). *The logic of statistical inference*. Cambridge, UK, Cambridge University Press.
Hacking, I. (1990). *The Taming of Chance.* Cambridge, Cambridge University Press.
Jevons, W. S. (1958). *The Principles of Science*, New York.
Laplace, P. S. (1814) *A Philosophical Essay on Probabilities*, English Translation of the 6th French Edition, Dover, 1951.
McShane, P. (1970). *Randomness, Statistics, and Emergence*, University of Notre Dame Press.
Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company.
Parzen, E. (1962). *Stochastic Processes*, Holden-Day, Inc.
von Mises, R. (1957). *Probability, statistics, and truth.* London, Allen & Unwin.
von Plato, J. (1994). *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*, Cambridge University Press.
Yates, Moore, & McCabe. (2000) *The Practice of Statistics.* W H Freeman.