On Randomness and Probability

Ricardo Dahis

Northwestern University

May 12, 2019

Abstract

This essay provides coherent definitions of two bedrock concepts in philosophy and statistics: randomness and probability. When constructing the first, I define repeatability, the measurement set, and distinguish between frequency vs. value prediction. The definition of randomness proposed, namely of a random variable not being perfectly value-predictable for any given information set, is stronger than those commonly used in the literature. Second, after defining probability as a theory about a variable's potential frequency distribution, I argue that the dichotomy between frequentist and bayesian interpretations is illusory. I conclude with remarks about knowledge and determinism.

I Introduction

As Wittgenstein (1953) put it, "Philosophy is a battle against the bewitchment of our intelligence by means of our language." And few concepts illustrate this point better than randomness and probability. These are often either misunderstood, or built upon other complicated mathematical ideas, such as sigma-algebras or the Axiom of Choice.

In this essay I provide definitions for these two terms that are both coherent and based purely on simple empirical notions. While working the way up to the main ideas, I also define auxiliary but useful concepts, such as repeatability, the measurement set, scope and measurement uncertainty, frequency vs. value prediction, among others.

Clarifying the meaning of randomness not only improves our understanding of key ideas in statistics, but also contributes to making a substantive philosophical claim. The definition developed below, namely of a random variable not being perfectly value-predictable for any given information set, raises the bar for refuting determinism. In particular, it shows that insufficient information and knowledge may disguise determinism as perceived randomness.

Moreover, the framework laid below helps undo an old interpretation divide about probability, held between frequentists and bayesians. I define probability as a theory about a variable's potential frequency distribution, which can be measured or not, depending on the variable's degree of repeatability. This helps clarifying that probability can be both a property of the object and of the observer.

This essay proceeds as follows. In Section II I define randomness after laying the necessary logical groundwork. In Section III I discuss why the probability-as-theory definition breaks the dichotomy of standard interpretations. Section IV concludes.

II Randomness

Our first goal is to define the term "randomness". In search for clarity, I start from first principles and move forward step by step. For simplicity of exposition, I keep mathematical notation to a necessary minimum.

Definition 1. A *variable* is a correspondence $f : \Theta \to Y$.

We define a variable in abstract, to allow for a logical representation of real-world variables. Examples are the a coin toss, the inflation rate observed next month, our mental mood after breakfast, and so on. First, notice that a variable is a correspondence, and not necessarily a function. Second, it maps the set of parameters Θ into a set of outcomes Y.

The set Θ has finite L dimensions. Each element $\theta \in \Theta$ describes the relevant parameters, or characteristics, or conditions, or context, or "state of the world", of the variable. Fully describing these for any particular variable may be cumbersome, if not impossible. As an example, the set of parameters describing a coin toss contains the coin's weight, wind speed, the person's hand position, the surface's inclination, etc. Moreover, a variable takes values, or outcomes, $y \in Y$. The set Y may be finite or not, and may contain symbols, numbers, or letters. The relevant outcomes for a coin toss are $Y = \{\text{heads, tails}\}$. For economic growth next year, outcomes are all real numbers $\mathbb{R}$.

Definition 2. The *measurement set* M_Θ is a set containing subsets of Θ, such that every $\theta \in \Theta$ is in at least one $m \in M_\Theta$.

This set represents how finely one can observe parameters. Coarseness may result, for example, from imperfect measuring instruments, or from complexity in even understanding what the relevant dimensions are. For example, when tossing a coin, one's instruments for measuring wind speed may be precise only up to 0.5 km/h. Or, when measuring cultural traits, it might not be straightforward to know what to measure.

Definition 3. A *draw*, *roll*, *experiment*, *realization* or *observation*, is an instantiation of a variable. It is represented by f at a specific parameter measurement, i.e. $f(m)$.[1]

There are several types of uncertainty associated with variables, some of which are discussed in previous literature (Knight, 1921; Zeckhauser, 2010).[2] To avoid confusion created by bad terminology, I have rephrased several terms to fit the framework introduced above.

First, there is *outcome uncertainty*, when a variable's probabilities are known, but not the outcome $y \in Y$ of each specific realization (named *risk* in the literature). Second, there is *probability uncertainty*, when probabilities are unknown (named simply *uncertainty* in the literature). Third, there is *parameter uncertainty*, when the state of the world $\theta \in \Theta$ is unknown (also called *ignorance* in previous literature). I introduce two new important types of uncertainty. There is *scope uncertainty*, when one does not know what the relevant dimensions of Θ are. Additionally, there is *measurement uncertainty*, when one observes a measurement $m \in M_\Theta$, but not the specific $\theta \in m$ generating the variable's realization.

Definition 4. A variable's *degree of repeatability*, or *degree of controllability*, is a measure of how fine M_Θ is. It is abstractedly represented by $c_f : M_\Theta \to [0, 1]$.

[1] Here I abuse notation. It could be that $f(\theta) = f(\theta') \; \forall \theta, \theta' \in m$, or simply that one cannot empirically differentiate what $\theta \in m$ corresponds to $f(m)$.

[2] Some of the concepts mentioned in this paragraph, such as *prediction* and *probability*, will only be properly defined in later paragraphs.

This measures both how well can one can observe parameters, and how well one can hold conditions constant across draws. Some simple properties follow. A variable is not repeatable when $M_\Theta = \Theta$, with $c_f(M_\Theta) = 0$. A variable is perfectly repeatable when M_Θ is a partition of Θ containing only singletons, with $c_f(M_\Theta) = 1$. Any two variables f and f' may be compared with c_f and $c_{f'}$. Two examples are (1) dice rolling, which is in principle highly repeatable, and (2) a given historical war, which is virtually unrepeatable, given how specific and complex the context of such events are.

Definition 5. An *empirical relative frequency* (ERF) is the share of observed realizations equal to a certain outcome, for a given number of draws n and a measurable set of parameters $m \in M_\Theta$. Formally, write $p_n(y|m) = \frac{1}{n}\sum_{i=1}^{n} 1[f(m) = y]$.

Definition 6. An *empirical relative distribution* (ERD) is the vector of ERFs across values of $y \in Y$ for a given $m \in M_\Theta$. It is represented as $ERD_f : M_\Theta \to [0,1]$.

Some remarks follow. First, the ERF and ERD are not meaningful for variables that are not repeatable, i.e. when $n = 1$. Second, counting frequencies does not inform us on *why* the variable behaves the way it does.

Until this point, we discussed only properties of variables that are objective, or external to human affairs. Now, let us move to how agents interact with variables.

Definition 7. A *theory*, *information set*, or *hypothesis*, is a set of claims with propositional value.

Such statements may be measurements of the environment (be it quantitative or qualitative). It may contain not-necessarily-true, or even unverifiable, claims. It may also be incomplete. [3]

A theory makes predictions, and one may summarize the former into a function $g : \Theta \to Y$. A theory may make sense only for a subset of Θ[4], and predictions may be wrong (i.e. not corresponding with observation, or nature, or the variable's outcome). In particular, theories can generate two types of prediction, in ascending order of strength.

Definition 8. *Frequency prediction* is a statement about a variable's ERD, for a given n and conditions measurement m, and represented as $fp_g : M_\Theta \to [0,1]^{card(Y)}$.

Definition 9. *Value prediction* is a prediction for a variable's next realization, and is possible when g is a function (not a correspondence). In this case, it is written simply as $g(m)$.[5]

[3] See Dahis (2018) for a more complete exposition on theories and their properties.

[4] Newtonian physics makes sense only for the not very small, or very big, or very fast, or very slow.

[5] A more general definition would allow for g to be a correspondence. In this case, for each m, the prediction would also include a frequency prediction for each $y \in g(m)$.

The distinction among types of prediction motivates a distinction between two types of accuracy.

Definition 10. *Frequency accuracy* is the difference between a theory g's frequency prediction and the observed ERD, for a given n. It may be defined for a fixed m, as $fa_n(m) = |fp_g(m) - ERD_f(m)|$, or across M_Θ, as $\sum_{m \in M_\Theta} fa_n(m)' fa_n(m)$.

Definition 11. *Value accuracy* is the difference between a theory g's value prediction and the variable's realization, for a given n. It may be defined for a fixed m, as $va_n(m) = \sum_n |g_n(m) - f_n(m)|$, or across M_Θ, as $\sum_{m \in M_\Theta} va_n(m)$.

Two remarks here. First, naturally, a theory's accuracy is always relative to the variable's degree of controllability. Even holding a theory's claims constant, it may suddenly become more accurate after better measurement instruments are introduced. Moreover, perfectly accurate theories are consistent with measurement uncertainty. We may be able to predict outcomes for any given m even if not knowing what particular $\theta \in m$ generated it.

Before I introduce a definition of randomness, we need a final pair of definitions about predictability.

Definition 12. A variable is *δ-frequency-predictable* if there exists a theory g such that $fa_n(m) \geq \delta$ for a given m, or such that $\sum_{m \in M_\Theta} fa_n(m)' fa_n(m) \geq \delta$ across M_Θ.

Definition 13. A variable is *δ-value-predictable* if there exists a theory g such that $va_n(m) \geq \delta$ for a given m, or such that $\sum_{m \in M_\Theta} va_n(m) \geq \delta$ across M_Θ.

Notice the following. First, we may loosely refer to a variable as δ-predictable, not differentiating frequency vs. value. Second, predictability, through the definition of accuracy, depends on the number of realizations n observed[6] and the measurement set M_Θ. Third, as with accuracy, variables may be more or less predictable for different values of measurements m. Fourth, uninformed theories may predict frequencies successfully as n grows, but rarely do so when n is small. Fifth, one could propose more sophisticated measures of predictability, e.g. one that accounted for continuous variables and that computed numerical errors between prediction and observations. For simplicity of exposition I maintain a true vs. false measure. Sixth, we can label a variable as *predictable* if is is 100% (frequency- or value-) predictable.

Definition 14. A *demarcating information set* (DIS) is the smallest, or weakest, set of claims necessary for a given δ% of predictability.

[6]This provides a simple representation of Hume's problem of induction. Predictive theories may perform well only with the observations recorded so far, but there is no guarantee they would continue to do so in future realizations.

Notice, naturally, that, for a given level of predictability, there may still be larger, or stronger theories available. These may yield the same prediction accuracy, or even more. One obvious case is the uninformed theory, which delivers good frequency-prediction by definition.

We are finally ready to introduce a meaningful definition of randomness.

Definition 15. A variable is *random*, or *not value-predictable*, if there exists a $\theta \in \Theta$ such that there is no theory g that makes value-predictions with 100% accuracy.

In other words, true randomness means that one cannot, in principle, predict the next realization of a variable with perfect accuracy. Moreover, there is no amount of information about the current state of conditions, or about the present, that allows one to predict the next event. A couple interesting points can be brought to bear.

First, a variable may be 100% frequency-predictable and still be random. As discussed above, trivially, uninformed theories may result in great frequency-accuracy, but still be empty of true knowledge of how a variable is determined.

Second, we cannot prove that a variable is random. All we can safely say is that a variable is *currently* not-value-predictable. The appropriate analogy is with statistical hypothesis testing. One may assume a variable to be random as a null hypothesis, and then search for enough evidence to the contrary. Besides, we may not even know what potential information sets exist. Our knowledge of the Universe is limited and imperfect.

Third, perhaps unintuitively, it is not possible to compare levels of randomness. Randomness is a discrete property of variables. Nevertheless, one may compare levels of predictability. Different measures exist, but some common ones are the following. A variable is more predictable than another if (1) it takes a smaller information set to predict it at a same level of accuracy, or (2) given the same information set, one can predict it more accurately than the other.[7]

Fourth, as indicated above, there are several qualifications of predictability. A variable may be *currently* not predictable if we currently do not have theories that generate suitable accuracy. A variable may be predictable *in principle* if we know a theory exists that provides perfect predictability, but it simply is too costly to calculate it (for data availability or computational issues). Finally, a variable may be *already* predictable if we currently know its demarcating information set.

Fifth, there are weaker definitions of randomness which are commonly found in the literature. For example, one may define a variable as random if 100% value-predictability is impossible for all $\theta \in \Theta$. Put it differently, a variable would not be random if perfect value-predictability is possible for some $\theta \in \Theta$, which is much weaker

[7] A concrete concept in mathematics on this topic is stochastic dominance.

than the definition given above. Additionally, there are pseudo-random numbers. These are deterministically generated, but result from such complicated algorithms that value-prediction becomes virtually impossible.

III Probability

With this framework in place, we can proceed to tackle an old problem in statistics, namely how to interpret probability.[8] The two most common views imply an apparent irreconcilable dichotomy. Let us discuss each view in more detail and then argue that this dichotomy is a false one. I will argue that both interpretations are consistent with the ability to make value-predictions.

Frequentists interpret probability as a variable's ERF for large n – sometimes fixing m and sometimes not. According to this view, probability is a property of the object. It is what would happen if only we could repeat many draws of the same variable. The numbers used for the probability of each event are abducted from observation, and not known a priori.

The problem with frequentism is obvious: how to make sense of probability for variables that are not repeatable? What would it even mean to think of re-rolling the dice of history? How can one repeat an election that has just come to pass? And would we expect different results? Such events may be virtually random, since predicting it is difficult, and also not-repeatable, but one could potentially do away with notions of probability in such cases.

On the other hand, Bayesians interpret probability as a subjective belief about outcomes of events, which can be updated via Bayes rule as more evidence is accumulated. In this case, probability is a property of the observer. A famous concept in Bayesian statistics is the Principle of Indifference, or the Principle of Insufficient Reason, which assigns an uninformed prior (or theory) to a variable's outcomes Y when no information is available.

Here I would like to make two important points. First, nothing in the discussion above depends on the relevant variables being truly random. Statements about probability, be it seen as an objective or subjective measure, are simply claims about frequencies. In practice it may make little difference whether something is random or not. Dealing with levels of predictability alone is sufficient to build whole fields of knowledge around statistics.

Second, however, philosophical precision matters. So I propose a definition of probability that is consistent with both views, and undoes the apparent dichotomy.

[8] The history of interpretations of probability is long and interesting. For more detail, see Hájek (2012).

Definition 16. *Probability* is a theory, or claim, or statement of fact, about a variable's frequency distribution.

As any other theory, statements about probability may be true or false, in the correspondence sense. Probability is an abstraction, which describes a property of variables if repeated countless times. Repetition, however, is not always feasible. When it is, probability claims may be directly tested, if measurement conditions are good enough. In other words, a probability claim may be taken as a property of the object, if one supposes the statement to be true. On the other hand, it may also be taken as a property of the observer, in the sense of theories being held as beliefs by agents. Moreover, an observer may update its statement of probability as new evidence is collected, just as one does with any other (scientific) theory.

IV Concluding Remarks

In this essay I attempt to provide coherent definitions of randomness and probability. To do so, it was necessary to slowly build vocabulary and to dispel common myths and confusions around terms. An advantage of this construction is to deliver all necessary intuition in a precise manner, with no recourse to abstract concepts such as sigma-algebras, data-generating processes, etc. I close with miscellaneous remarks on related topics.

A good measure of how much knowledge we have is how well we can predict events. And having predictive power is a necessary condition for a theory, or information set, to be true. Obviously, other properties also matter, such as being verifiable, realistic, etc. (Dahis, 2018).

On a separate note, the arguments developed in this essay raise the bar for refuting the concept of determinism, which states that events are determined by previous causes. Since determinism is only opposite to true randomness, and since most of what we commonly deem random is simply unpredictability caused by lack of information, then the burden of proof for refuting determinism is with those arguing against it.

References

Dahis, R. (2018). Is Economics a Science? Well, Not Yet. *SSRN Electronic Journal*, pages 1–22.

Hájek, A. (2012). Interpretations of Probability.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*.

Wittgenstein, L. (1953). *Philosophical Investigations*. Macmillan Publishing Company, Oxford.

Zeckhauser, R. (2010). Investing in the Unknown and Unknowable. In *The Known, the Unknown, and the Unknowable in Financial Risk Management: Measurement and Theory Advancing Practice*, pages 304–346. Princeton University Press.