Frequentism and the Best Systems Account

Kevin Li

May 15, 2020

The notion of chance plays a central role in many fields of scientific endeavor, as well as everyday conversation. This paper will discuss two different interpretations of chance and probability, and the problems that they face. I will first outline the frequentist account of chance and discuss some of its shortcomings. I will then discuss how Lewis' Best Systems Account resolves some of these issues and pose possible further objections.

Frequentism approaches the notion of chance as strictly a reflection of actual events. Probability, in this view, is simply a record of how frequently certain outcomes have occurred in the past, without consideration for what series of outcomes are theoretically possible. For example, if I flip a (fair) coin 3 times and it turns heads twice, the relative frequency of heads is $\frac{2}{3}$, which, for a frequentist, means that the probability of the coin turning heads is $\frac{2}{3}$. With this approach, there is no notion that a coin being fair gives it any particular probability—its probability is only a matter of relative frequency, which is based on empirical outcomes rather than theoretical ones.

For a frequentist, probability does not mean anything aside from relative frequency. However, in order to make sense of the idea of relative frequency, there must be some notion of what frequency is relative *to*. Frequentism uses the idea of a reference class, or some defining boundaries of not only what events count as an outcome being realized, but what events count as an outcome not occurring. To use the earlier coin flipping example, we understand the relative frequency to be $\frac{2}{3}$—that is, the coin flipped heads twice, out of the 3 times that we flipped it. The reference class in this case is this specific coin being flipped. In contrast, if we instead were counting the frequency of this coin being flipped relative to the reference class of all coins of the same diameter being flipped, the relative frequency would be different.

This is the first problem for frequentists. On face, there is no "inherent" reference class that a set of outcomes is tied to. In order to obtain meaningful probabilities (or to say that one fact of probability is more relevant than another), frequentists have to commit themselves to an idea of certain reference classes being more appropriate than others, which there is no basis for. For certain problems like coin flipping, the reference class seems obvious to us, but there are cases where the reference class is a lot more unclear— for example, the likelihood of catching infectious diseases could be an important probability to know, but it's unclear whether it would be best to measure the frequency of infection relative to all people in a geographic area, all people in a certain age group, or all people that watch football. One might argue that it is not the job of a theory of chance to tell you which probabilities to use, but if we end up using probability to help us understand the world (and use it for important things like science) it would be useful if frequentism could be revised and adapted to at least say more on which reference classes are actually useful or meaningful.

Another shortcoming of frequentism is that it's measurements of probability are always discrete. In other words, since frequentists base probabilities off of the idea of counting frequency, the probabilities will always lack certain kinds of precision and are held back by their "finiteness". There are certain events that, in reality, behave along continuous probability distributions that frequentism would never be able to capture because of its inability to reflect continuity. To contextualize this argument, consider a radioactive atom whose time of decay may vary along some exponential probability distribution. Even though this is true, a single atom will only ever decay once—let's say it does so at time t. Since it decayed once at time t, and never decayed at any other time, the frequentist, who only looks at actual occurrences, would say that the atom has a probability of decaying at time t of 1, and at any other time, 0. This is misleading—it implies that if another atom of this kind were to decay, it would also decay at exactly time t, with 100% certainty. Let's say that it doesn't, and instead decays at a different time s. The frequentist would then revise their probability, and say that the probability of decay is $\frac{1}{2}$ at time t, and $\frac{1}{2}$ at time s. Given another 1,000 trials, the frequentist would continually revise their probability distribution, and it would start to look exponential, but the point is that however many trials the frequentist may observe, they will never observe an infinite number of trails. Therefore, the frequentist will never be able to capture this unique idea of continuous probability, where the atom's probability of decay at any time is infinitesimally small.

Lewis's Best Systems Account (BSA) offers a revision to frequentism. Lewis' BSA, more generally, argues that, out of all the truths and systems of truth possible, the laws of nature are the particular set of truths that offer the best balance between simplicity and informativeness. He says that this system extends to the idea of probability, because there are matters that make the most sense to summarize probabilistically. For example, if some system X sometimes will evolve into Y but other times will evolve into Z, a theory that best summarizes these non-clearcut scenarios will involve some kind of probability, whose distribution gives the most informative and simple answer to what happens from state X. Chance, in this case, is accounted for in the probabilistic laws of the best system. In addition to simplicity and informativeness, Lewis also considers *fit* to be a third condition for deciding which truths are part of the BSA. In other words, a theory not only needs to be informative and simple but should also mold to the actual likelihood of the history of the world. The theory should imply that the world we actually live in is highly probable. This allows it to keep true to frequentism— Lewis's BSA generates probabilities based solely on actual past outcomes because the laws of nature need to be directly fitting of what is in fact true about the world. What is different, however, is that the BSA has more to say about what is the most economical way to think about and summarize these probabilities.

This solves the frequentist's problem of reference classes— Lewis's BSA would say that we should use certain reference classes over others if they offer us a good balance of information, simplicity, and fit. Take two statements of chance based on the same event, relative to two different reference classes: (1) infection is more likely among elders than other age groups and (2) infection is more likely among people who watch football. These are equally meaningful to our initial frequentist, but the BSA would say that statement (1) offers a better summary because it is more informative and fitting than statement (2).

Lewis' BSA also gives the frequentist the tools to describe non-discrete probabilities because of its commitment to simplicity. The radioactive decay of an atom based on relative frequency, after many, many trials, starts to look kind of like an exponential function. At this point, the frequentist must still say that it's probability of decay is relative to whatever specific ratio of frequencies occurred in the past, without consideration for any exponential function it seems to follow. The BSA, on the other hand, could say that a continuous exponential function is a good way to summarize a bunch of intermediate recorded probabilities that are otherwise really complicated, and therefore that should be the probabilistic law. It sacrifices a small amount of fit to obtain a much greater amount of simplicity.

This sacrifice, however, may be a point of objection against using the BSA to determine probabilistic laws. It is a good idea to let the BSA sacrifice fit for the sake of simplicity? Probability, which deals with things that are chancy by nature, may not benefit from simplicity or informativeness if it sacrifices accuracy. The BSA seems to presume that probabilistic laws should be simplistic and informative the way that natural laws are, even if it means ignoring small inconsistencies. However, there are cases where the small inconsistencies that are driven out by simplistic laws are in fact important to our understanding and use of probability. For real world example: in the self-driving car industry, probabilistic instances where a vehicle fails to properly detect its surroundings and causes human injury are critically important, although rare. For the scientists that train these intelligent systems, the scenarios that are least probable are the most important, because the cars are more likely to act unpredictably. Probability, in this case, does not benefit from simplicity— instances of failure that are the most rare and seemingly contradictory to the general behavior of the vehicle are the most important, and actually indicate that the general laws that define the intelligence systems may be wrong. Lewis' BSA seems to imply that simplicity and informativeness are good virtues because they avoid the initial problems of frequentists, but they are not independently justified as virtues for probabilistic law.