Propensity Theories

Conor Mayo-Wilson

Philosophy of Probability
May 13th, 2014

Review

Last Month: Subjective Probability

Definition: Probability is a measure of strength of degree of belief.

Review

How is probability measured?

- Betting Behavior
- Qualitative verbal comparisons of likelihood
- Preferences among compound lotteries
- Preferences among acts

Review

Why does it satisfy the axioms?

Measurement Procedure	**Rationality Criteria**
Betting Behavior	Avoid Sure Loss
Qualitative verbal comparisons	Qualitative probability axioms
Preferences among lotteries	V&M and Anscombe-Aumann axioms
Preferences among acts	Savage axioms

Theorem

If probability is elicited via the measurement procedure, then

Rationality criteria $\Leftrightarrow$ *Satisfy Probability Axioms*

Review

Why is it useful?

Answer: Probability allows us to calculate expected utility maximizing actions.

Choice of statistical estimates will be one such action.

Today

Today: We will discuss one objective interpretation of probability.

Motivation for Objective Interpretations

Motivation:

- Suppose today I pass around what appears to be a standard six-sided die.
- Each of you do some measurements and determine the die is not weighted in any funny manner.
- I ask you, "What is the probability that when I roll the die, it will land on three?"
- Three of you answer as follows:
  - Student 1 says, "One-sixth."
  - Student 2 says, "$\frac{7}{247}$."
  - Student 3 says, "$\frac{999}{1000}$"

Motivation for Objective Interpretations

- None of these guesses violate the probability axioms: all three could be coherent.
- However, many think Student 2 and Student 3 have strange, if not irrational, degrees of belief.

There are two common solutions . . .

Objective Bayesianism

Solution 1: Objective Bayesianism

- Probability = Degrees of belief, but
- Only certain degrees of belief are rational in light of
  - Ignorance, or
  - Evidential symmetry, or
  - Physical symmetry
- E.g., In the die case, some think only $\frac{1}{6}$ is warranted by evidential symmetries.

Motivation for Objective Interpretations

Solution 2: Objective Interpretations

- Probability = "Feature of the world" not your beliefs.
- Three common objective interpretations
  - Logical
  - Frequentist
  - Propensity

Objective Interpretations

In this class, we won't discuss the logical interpretation. Why?

- Few (if any) contemporary statisticians or scientists, to my knowledge, endorse the logical view.
- In contrast, many statisticians seem to endorse the frequency view.

Frequentist Statistics

> *The statistical methods that we have discussed so far are known as frequentist (or classical) methods. The frequentist point of view is based on the following postulates:*
>
> *F1 Probability refers to limiting relative frequencies. Probabilities are objective properties of the real world . . .*

Wasserman [2004], page 188.

Why propensities, then?

Question: Conor, why is this class about the propensity interpretation?

Answer:

1. I cannot make any sense of the frequency view.
2. I conjecture that most frequentist statisticians would endorse something like a propensity interpretation when pushed.

A brief digression about teaching and philosophy:

- Normally, I am opposed to teaching my own views; I think it puts students in an awkward position.
- Moreover, I hate arguments that are of the form "I cannot make sense of X. Therefore, X is wrong."
  - A lack of creativity and insight is not an argument.
- So let me briefly describe some significant technical and philosophical problems for the the frequency interpretation.
- I would be extremely happy if anyone (including you!) could rectify them.

A common formalization

Here is the most common formalization of the frequency view (Suppes [2002], pp. 167-170).

- Let ω be any countable sequence $\omega = \langle \omega_1, \omega_2, \ldots \rangle$.
- Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be the range of ω.
- For any set $E \subseteq \Omega$, define:

$$f_\omega(E, n) = \frac{|\omega_i : \omega_i \in E \text{ and } i \leq n|}{n}$$

  to be the relative frequency of the event E through stage n.
- Let $\mathcal{F}$ be an algebra of sets on Ω such that $\lim_{n\to\infty} f_\omega(E, n)$ exists for all $E \in \mathcal{F}$.

A common formalization

Easy Fact: $\langle \Omega, \mathcal{F}, P \rangle$ is a finitely-additive probability space where

$$P(E) = \lim_{n \to \infty} f_\omega(E, n)$$

A common formalization

Example:

- Let ω be a sequence of coin tosses $\omega = \langle H, H, T, H, T, T, T \ldots \rangle$.
- Let $\Omega = \{H, T\}$ be the range of ω.
- Suppose that the relative frequency of heads in the sequence ω approaches a fixed value.
- Then $\mathcal{F}$ is the power set of $\{H, T\}$, and the probability of heads is defined as the limiting relative frequency.

Countable Additivity

Minor Problem: Probability is not countable additive when formalized this way, contrary to what many frequentist statisticians assume.

- Consider the sequence $\omega = \langle 1, 2, 3, 4, \ldots \rangle$.
- Then for every natural number k, we have $P(k) = 0$.
- But $P(\mathbb{N}) = 1$.

Solution? Just ditch countable additivity.

Major Problem: Probability defined on such a space is inconsistent with the first assumption of basically every theorem that appears in a statistics textbook.

IID Random variables?

- Consider any theorem that begins, "Let $X_1, X_2 \dots$ be iid random variables." Why is this a strange assumption to make in the coin flipping example as formalized above?
- Random variables are (measurable) functions from Ω to $\mathbb{R}$.
- Suppose X_i represents that i^{th} flip of the coin.
- So each X_i takes values in the set $\Omega = \{H, T\}$. So each X_i maps Ω to itself.

IID Random variables?

- Ω is finite. So there are only finitely many functions from Ω to Ω.
- So $X_i = X_j$ (as random variables!) for distinct numbers i and j.
- So X_i and X_j are not independent (unless both take the value H (or T) with probability one).

Moral: The most obvious way of trying to formalize the frequency interpretation is problematic for statistical purposes.

Other obvious ways of trying to formalize the interpretation (that I have tried, at least) also lead to inconsistencies or they collapse distinctions that are common in probability theory.

Philosophical Worries

There are also several major philosophical worries.

Here, it is helpful to distinguish between two versions of the frequency view:

- Finite: Frequency is defined relative to some fixed, finite population.
- Infinite: Frequency is defined as a (hypothetical) limit in an infinite population.

Philosophical Worries

Dilemma:

- Actual ⇒ Probability is ascertainable but useless: once we've determined the probability of an event, we cannot use it to predict anything we don't already know.
- Hypothetical ⇒ Probability is both un-ascertainable and useless: even if we knew the infinite limit, it tells us nothing about finite samples.

So the frequency interpretation has trouble meeting Salmon's three criteria.

More on the Dilemma

Although Hajek's main aim is to investigate whether the frequency theory matches certain "pre- theoretic intuitions", [**?**] and [Hajek, 2009] provide further arguments that outline this dilemma.

Commitments of Hypothetical Frequentism

- Popper [1959] argues that hypothetical frequentists are already committed to a "propensity" view.
- Roughly, his argument is that in order to pick which sequence to use in the definition of probability, one appeals to certain physical facts about an experiment. Here's his example . . .

Commitments of Hypothetical Frequentism

Popper [1959] imagines alternating flipping two coins: one standard and one with the center of mass towards tails.

There are two obvious sequences that one might use to define the probability of heads on the tenth throw:

- The sequence of flips of **both** coins together.
- The sequence of flips of the second, biased coin.

Commitments of Hypothetical Frequentism

- Intuitively, many want to say the second sequence is the "correct" one to use.
- The limiting frequencies of the even and odd sequences are different because different physical properties of the two coins are different.
- One ought to define the probability of an event as a the limiting relative frequency in a sequence of repeated experiments, where experiments are repeated if the physical properties are relevantly similar.

Commitments of Hypothetical Frequentism

So it looks like the hypothetical frequentism is really the conjunction of the following thesis and definition:

Let E be an experiment.

- **Thesis:** If E were repeated a large number of times, the relative frequency of some events will approach a limiting in virtue of the properties (physical, chemical, etc.) of E.
- **Definition:** The probability of an event in E is this limiting value.

The Propensity Interpretation

But this is what the propensity theory of probability asserts Popper [1959].

- **Caveat:** Just as there are several different frequency theories, there are several different "varieties" of propensity theories. See Gillies [2000]

. The properties (or "generating conditions" or "causes") of the experiment are called propensities.

Why give this view a new name? Why does Popper not just say that he has clarified hypothetical frequentism?

There are two reasons, which nicely correspond to two of Salmon's criteria . . .

Reason 1: Propensities may be ascertainable (i.e., measurable) even if the limiting relative frequencies are not.

- E.g., We can see whether a die or coin is evenly weighted or whether they are tilted.

Reason 2:

- Many frequentists argue that, if an experiment is not repeated, it is meaningless to talk about the probability of an event.
- So, according to a frequency view, probability theory is inapplicable if an experiment is not repeated.
- In contrast, if the limiting relative frequencies arise due to properties of the experiment, one can define the probability of an event were it repeated even if it is not.
- This is what Popper [1959] does.

Problem: The central problem for the propensity view is Salmon's admissibility criterion [Suppes, 2002].

Why think that the limiting frequencies of results of repeated experiments exist (and, hence, form a probability space)?
Are the limiting frequencies arising from repeated experiments countably additive?

Today: The method of arbitrary functions as an argument for the admissibility of propensity theory.

Outline

Suppes' Representation Theorems

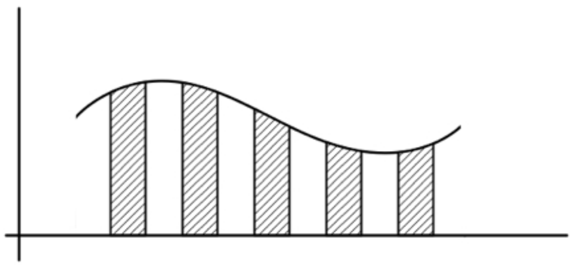
Suppes [1987] and Suppes [2002] try to characterize four types of experiments in which one should expect mathematical probabilities to emerge.

Suppes' Representation Theorems

One of Suppes [1987]'s theorems is structually similar to an argument given by Poincare, Reichenbach, and now Strevens.

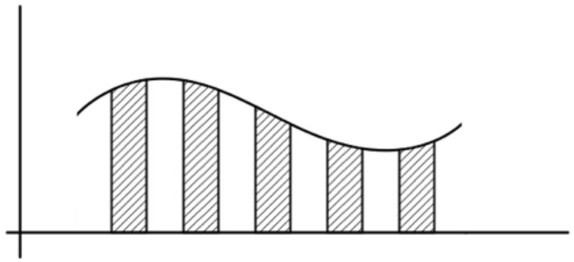
See [Von Plato, 1983] and Glymour and Eberhardt [2012] for references.

Strike Ratios



- Imagine a coin is tossed and that whether it lands heads or tails depends exclusively on its angular velocity.
- Imagine that small differences between angular velocities correspond to changes in the outcome of the toss, and
- The intervals corresponding to heads and tails tosses alternate and are of the same width.

Strike Ratios



- Finally, assume that there is some probability distribution $P(\omega)$ of obtaining a given angular velocity ω.
- Then whatever the frequency distribution over angular velocities is, the coin will land heads $\frac{1}{2}$ of the time.
- Suppes [2002]'s argument is similar because he lets the velocity of the coin become arbitrarily fast.

Method of Arbitrary Functions

- This argument is called the method of arbitrary functions because it does not matter what the original probability distribution P is.
  - The probability of heads is $\frac{1}{2}$ under almost any probability distribution.
- Similar arguments can be given for uniform probability distributions over roulette wheels, dice rolls, etc.
  - Strevens [1998] claims that similar arguments are available for propensities in other domains (e.g., propensities as Darwinian fitnesses or as properties of molecules in a gas). I don't understand this.

Strevens and Von Plato's Questions

Question 1: Isn't it too strong to argue that the uniform probability distribution is the unique propensity that a coin, die, or roulette wheel can have? Isn't our experience with loaded dice and trick coins enough to be reductio of this argument?

Question 2: How should the probability distribution P be interpreted (e.g., as a propensity, subjectively, frequency), etc.)? Is this argument question-begging?

Strevens and Von Plato's Questions

Question 1: Isn't it too strong to argue that the uniform probability distribution is the unique propensity that a coin, die, or roulette wheel can have? Isn't our experience with loaded dice and trick coins enough to be reductio of this argument?

Answer: Yes and no. We can use similar arguments to derive the probabilities that weighted dice, trick coins, etc. land in particular ways give physical knowledge [Von Plato, 1983].

Question 2: How should the probability distribution P be interpreted (e.g., as a propensity, subjectively, frequency), etc.)? Is this argument question-begging?

Subjective Probability Again?

- Savage has argued that one should interpret P as a subjective probability.
- The theorems then show that many individuals, with a wide variety of priors, will assign the same probability to outcomes of coin tosses, wheel spins, etc. when they are given certain physical knowledge.
- So these shared probabilities are only "objective" in the sense that they are agreed upon "agreed upon given sufficient evidence."
  - This is a typical Savage move.

Objective Interpretation

Strevens [1998] argue that P can be interpreted as a frequency.

> *In the case of the roulette wheel, some kind of enumerative induction is probably at work. We simply know from experience that human actions such as the twirl of a wheel produce results that are smoothly heaped around some average value. (The experience is often gained while trying to produce results that require a far more finely honed distribution.)*
>
> *We might also be able to arrive at a non-enumerative inference to the effect that $q(\omega)$ is smooth, if we knew enough about human physiology that we had some idea about the causes of variation in spins. This inference would depend on the distribution of some set of physiological initial conditions, which would have to be inferred in turn. At some point some enumerative induction must be done to get non-enumerative induction off the ground.*

An Inheritance Problem

Problem: Does the propensity theory inherit all the problems of the frequency theory? It seems that justifying the propensity interpretation requires using the frequency theory.

For instance, why should we expect human spins of the roulette wheel have some stationary frequency distribution over angular velocities?

Response: Inference to the Best Explanation

Problem: Does the propensity theory inherit all the problems of the frequency theory? It seems that justifying the propensity interpretation requires using the frequency theory.

Answer 1:

> *Our experience of the random and stationary distribution of results in games of chance of the mechanical type can be taken as evidence for the reality of a continuous initial distribution, because only that assumption has explanatory relevance.*

[Von Plato, 1983].

Response: Employ some ergodic theory

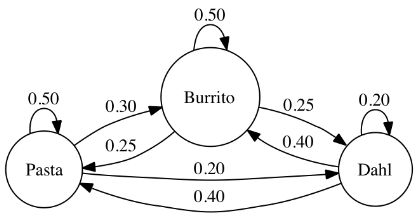
Answer 2: A quick (and cryptic) summary of what [Von Plato, 1983] is trying to do:

- Take measurements of some (deterministic) ergodic system over time.
- Chunk up your measurements into units of a given length n.
- Define the probability of an event to be its limiting frequency in the chunks.
- Use ergodicity to argue that that said frequency is (almost always) well-defined.
- Argue that the ergodicity assumption is empirically grounded, and justify the use of "almost-always" by an appear to inference to the best explanation.

Let me try (!) to give the argument in more detail, although I'm truthfully not sure exactly what Von Plato is saying because the argument is entirely in prose.

Markov Processes

How do students decide what to eat?

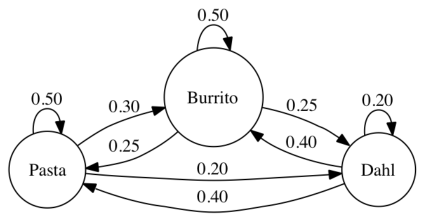


Markov Process = The current state of a system depends only upon its recent past.

Transition Matrices

Markov processes can be described by transition matrices:

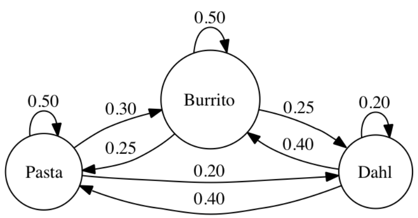
	Pasta	Burrito	Dahl
Pasta	.5	.3	.2
Burrito	.25	.5	.25
Dahl	.4	.4	. 2



Transition Matrices

Markov processes can be described by transition matrices:

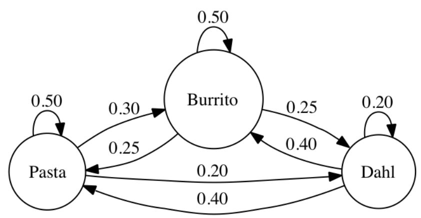
$$T = \begin{pmatrix} .5 & .3 & .2 \\ .5 & .25 & .25 \\ .5 & .25 & .25 \end{pmatrix}$$



Transition Matrices

The transition matrix for two stages is obtained by squaring the original matrix:

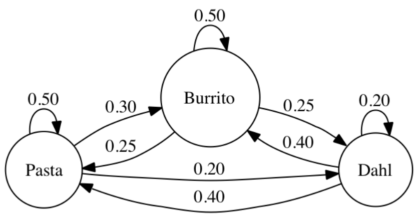
$$T^2 = \begin{pmatrix} .5 & .275 & .225 \\ .5 & .275 & .225 \\ .5 & .275 & .225 \end{pmatrix}$$



Stationary Limits

The transition matrix for n many stages is obtained by taking the n^{th} power of the original matrix:

$$T^n = \begin{pmatrix} .5 & .275 & .225 \\ .5 & .275 & .225 \\ .5 & .275 & .225 \end{pmatrix}$$



Stationary Limits

Curious: The transition matrix acquired a fixed value, and its rows are identical ...

Stationary Limits

Theorem

Under a wide variety of conditions, the T^n approaches a fixed, positive limiting matrix T^∞ with one row. The single row represents the probability of where the process will be in the limit, regardless of its starting point.

Stationary Limits

What conditions guarantee the above theorem?

- Aperiocity
- Ergodicity

References I

Gillies, D. (2000). Varieties of propensity. *The British journal for the philosophy of science*, 51(4):807–835.

Glymour, C. and Eberhardt, F. (2012). Hans reichenbach. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

Hajek, A. (2009). Fifteen arguments against hypothetical frequentism. *Erkenntnis*, 70(2):211235.

Popper, K. R. (1959). The propensity interpretation of probability. *The British journal for the philosophy of science*, 10(37):25–42.

Strevens, M. (1998). Inferring probabilities from symmetries. *Nous*, 32(2):231–246.

Suppes, P. (1987). Propensity representations of probability. *Erkenntnis*, 26(3):335–358.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures*. CSLI Publications Stanford.

Von Plato, J. (1983). The method of arbitrary functions. *British Journal for the Philosophy of Science*, 34(1):37–47.

Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. Springer.