Randomness

Philosophy of Probability
May 6th, 2013

Outline

1. From Finite to Infinite Frequentism
   - Review
   - Infinite Frequentism

Outline

1. From Finite to Infinite Frequentism
   - Review
   - Infinite Frequentism
2. Randomness
   - Von Mises

Outline

1. From Finite to Infinite Frequentism
   - Review
   - Infinite Frequentism
2. Randomness
   - Von Mises
3. Computability Theory

Outline

1. From Finite to Infinite Frequentism
   - Review
   - Infinite Frequentism
2. Randomness
   - Von Mises
3. Computability Theory
4. Definitions of Randomness
   - Kolmogorov Complexity
   - Betting and Martingales

Outline

1. From Finite to Infinite Frequentism
    - Review
    - Infinite Frequentism
2. Randomness
    - Von Mises
3. Computability Theory
4. Definitions of Randomness
    - Kolmogorov Complexity
    - Betting and Martingales
5. Extras
    - Technical Snags

Differences among Frequency Theories

Three Distinctions Among Finite Theories:

- Probability is a measured proportion vs. Probability is a proportion in a wider population, some units of which have been observed.

Differences among Frequency Theories

Three Distinctions Among Finite Theories:

- Probability is a measured proportion vs. Probability is a proportion in a wider population, some units of which have been observed.
- Probability is an actual proportion vs. Probability is a hypothetical proportion

Differences among Frequency Theories

Three Distinctions Among Finite Theories:

- Probability is a measured proportion vs. Probability is a proportion in a wider population, some units of which have been observed.
- Probability is an actual proportion vs. Probability is a hypothetical proportion
- Sample space (and algebra) generated by a sequence vs. Sample space is an arbitrary unordered set.

Infinite Frequencies

For reasons that may or may not have been well-motivated, many gave up on finite frequency theories, and turned to infinite ones.

Infinite Frequencies

Infinite frequentists cannot make the three distinctions we previously employed. Why?

Infinite Frequencies

1. In the cases in which human beings cannot perform infinitely many measurements (all cases?), an infinite frequency theory must define probability to be a proportion in some in a wider population, some units of which have not been observed.

Infinite Frequencies

2. It's possible the universe contains finitely many objects (in total and/or of the type in which one is interested), in which case probability is a hypothetical (rather than actual) proportion.

Formalizing Infinite Frequentism

If probability is a hypothetical proportion involving objects that we cannot measure, then questions about ascertainability arise: how can we ever discover or even approximate the probability of some event?

This is ironic because one of the chief virtues of finite frequntism was supposed to be that probabilities are easily learned via experiments.

Infinite Frequencies

3. Finally, it's not clear how to define proportions among unordered sets when infinitely many objects are under investigation). The symbol $\frac{\infty}{\infty}$ does not denote a real number.
    - I.e., The only way of making a proportion precise is by taking the sample space to be generated by an ordered sequence.

Formalizing Infinite Frequentism

How can we make infinite frequentism precise?

Formalizing Infinite Frequentism

Rough idea: Take an infinite sequence of coin flips $\langle T, H, T, T, T, H, T, H, H, T, \ldots \rangle$

- Let E be any subset of the range of the sequence, e.g., “Heads” or $\{H\}$.
- After every finite number of flips, there is some proportion that have been heads. E.g., Above:
    - After 1 flip: 0
    - After 2 flips: $\frac{1}{2}$,
    - After 3 flips $\frac{1}{3}$, Etc.

Formalizing Infinite Frequentism

- In some sequences, these proportions approach a limiting value. Then one can define the probability of the set E to be this limiting proportion.

FORMALIZING INFINITE FREQUENTISM

- In some sequences, these proportions approach a limiting value. Then one can define the probability of the set E to be this limiting proportion.
- There are one or two technical snags (see **Extras**), but this definition turns out to satisfy Kolmogorov's axioms.

Infinite Frequencies

- The infinite sequences are intended to represent the outcome of repeating an experiment over and over again.

Infinite Frequencies

- The infinite sequences are intended to represent the outcome of repeating an experiment over and over again.
- In order to attribute an event some **unique** probability, the relative frequency in the sequence must approach some fixed value.

Infinite Frequencies

- Moreover, if a sequence exhibits patterns (e.g. $\langle T, H, T, H, T, H \ldots \rangle$) that are repeated indefinitely, then it seems wrong to interpret the sequence as an outcome of the same process.
    - Instead, one might assert the probability of an outcome depends upon **when** it occurs.

Infinite Frequencies

- Moreover, if a sequence exhibits patterns (e.g. $\langle T, H, T, H, T, H \ldots \rangle$) that are repeated indefinitely, then it seems wrong to interpret the sequence as an outcome of the same process.
    - Instead, one might assert the probability of an outcome depends upon **when** it occurs.
- So philosophers, mathematicians, statisticians, and computer scientists became interested in characterizing random sequences, i.e., ones in which the probability of an event is fixed and constant across time.

Infinite Frequencies

What makes a sequence random?

Let's restrict our attention to **binary** sequences.

Von Mises

Von Mises' definition of randomness is vague, but the informal idea is clear:

If a sequence is random, then one should not be able to place bets on values of the sequence in a way that guarantees winning.

Von Mises

Here's an example to motivate the informal idea:

- Suppose the sequence is alternating heads and tails: $\langle H, T, H, T, H, T \ldots \rangle$

Von Mises

Here's an example to motivate the informal idea:

- Suppose the sequence is alternating heads and tails: $\langle H, T, H, T, H, T \ldots \rangle$
- Then one could detect the pattern and bet "Heads" on odd throws and "Tails" on even ones.

Von Mises

Here's an example to motivate the informal idea:

- Suppose the sequence is alternating heads and tails: $\langle H, T, H, T, H, T \ldots \rangle$
- Then one could detect the pattern and bet "Heads" on odd throws and "Tails" on even ones.
- Doing so would guarantee winning.

Von Mises

As soon as on starts asking for ways of making that precise, however, things get hairy . . .

Von Mises

What counts as guaranteed winning?

- Suppose the coin lands heads $\frac{1}{3}$ of the time in the limit. Recall, we are only considering randomness for sequences in which the event in question has a limiting frequency.

Von Mises

What counts as guaranteed winning?

- Suppose the coin lands heads $\frac{1}{3}$ of the time in the limit. Recall, we are only considering randomness for sequences in which the event in question has a limiting frequency.
- If you use the observed frequency to estimate the limiting frequency, then you'll eventually know the coin is biased towards tails.

Von Mises

What counts as guaranteed winning?

- Suppose the coin lands heads $\frac{1}{3}$ of the time in the limit. Recall, we are only considering randomness for sequences in which the event in question has a limiting frequency.
- If you use the observed frequency to estimate the limiting frequency, then you'll eventually know the coin is biased towards tails.
- On each throw, bet $1 on tails.

Von Mises

What counts as guaranteed winning?

- You win $\frac{2}{3}$ of the time, and so you win \$2 for every dollar that you lose.

Von Mises

What counts as guaranteed winning?

- You win $\frac{2}{3}$ of the time, and so you win \$2 for every dollar that you lose.
- So if you start with a big enough bank account to support your bets, then as the game goes on, you win an arbitrarily large amount of money for sure.

Von Mises

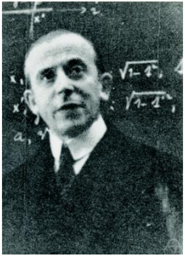
What counts as guaranteed winning?

- You win $\frac{2}{3}$ of the time, and so you win \$2 for every dollar that you lose.
- So if you start with a big enough bank account to support your bets, then as the game goes on, you win an arbitrarily large amount of money for sure.
- This argument only fails for events with probability (exactly) one half; there, your expected earnings are zero.

Von Mises

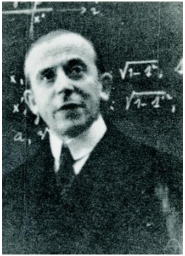
Revision: There is no betting rule that is guaranteed to win particular bets.

Von Mises



- Informal idea: One can guarantee winning on particular bets if one can identify subsequences that are exclusively ones or exclusively zeroes.

Von Mises



- So Von Mises defines what he calls a “Kollektiv”, which is a sequence in which one cannot pick subsequences that are exclusively ones or zeros without already knowing the values of the sequence.

Von Mises

One (very bad) attempt to formalize this idea is as follows:

- **Definition:** A sequence is random if every infinite subsequence has the same (limiting) proportion of zeroes as the whole sequence.

Von Mises

One (very bad) attempt to formalize this idea is as follows:

- **Definition:** A sequence is random if every infinite subsequence has the same (limiting) proportion of zeroes as the whole sequence.
- **Problem:**
  - Suppose the limiting proportion of zeroes is between 0 and 1.

Von Mises

One (very bad) attempt to formalize this idea is as follows:

- **Definition:** A sequence is random if every infinite subsequence has the same (limiting) proportion of zeroes as the whole sequence.
- **Problem:**
  - Suppose the limiting proportion of zeroes is between 0 and 1.
  - Then the sequence has infinitely many zeroes and infinitely many ones.

Von Mises

One (very bad) attempt to formalize this idea is as follows:

- **Definition:** A sequence is random if every infinite subsequence has the same (limiting) proportion of zeroes as the whole sequence.
- **Problem:**
  - Suppose the limiting proportion of zeroes is between 0 and 1.
  - Then the sequence has infinitely many zeroes and infinitely many ones.
  - So there are some infinite subsequences with only zeroes, and some with only ones.

Von Mises

This objection is really unfair to Von Mises, but it was a standard reason for dismissing discussions of randomness.

Randomness

Today:

- We'll discuss several mathematically precise definitions of randomness that have become accepted since Von Mises work.
- We'll also discuss the argument that they capture Von Mises intuitive ideas about randomness.

What we should discuss but won't:

- "Failed" definitions of randomness (e.g., Church's) and why they were dismissed (e.g., Ville's objection)
- Why do we need a theory of random sequences anyway?

Hilbert's Problems

Hilbert's Problems: In 1900, David Hilbert outlined 23 open problems in mathematics and physics at his address at the International Congress of Mathematicians in Paris.

Hilbert's Problems

HIlbert's 10th Problem: Find an "algorithm" that generates integer solutions to Diophantine Equations, if any exist. For example:

$$x^2 + y^2 = z^2$$

$$ax + by = c$$

$$x^2 - ny^2 = 1$$

$$x^3 + y^3 = z^3$$

are all Diophantine equations.

Entscheidungsproblem

Entscheidungsproblem: Find an “algorithm” that determines whether any given formula of first- order logic is valid. For example:

$$(\forall x)(P(x) \vee \neg P(x))$$

vs.

$$(\exists x)(P(x) \wedge \neg Q(x))$$

Algorithms and Mechanical Computability

- The concept of an “algorithm” was not made precise in either of Hilbert’s challenges.

Algorithms and Mechanical Computability

- The concept of an “algorithm” was not made precise in either of Hilbert’s challenges.
- Philosophical Problem: Find a precise notion of “algorithm” that characterizes what we mean by a “mechanical” step-by-step process.

Algorithms and Mechanical Computability

- The concept of an “algorithm” was not made precise in either of Hilbert’s challenges.
- Philosophical Problem: Find a precise notion of “algorithm” that characterizes what we mean by a “mechanical” step-by-step process.
- Mathematical Problem: Show there are (or are not) algorithms, in the precisely defined sense, meeting Hilbert’s challenges.

Algorithms and Mechanical Computability

- In the early 20th century, logicians developed several definitions of “mechanical” or “effectively” or “algorithmically” computable.
- They also gave several circular arguments that certain formal definitions capture the intuitive concept of an “algorithm” [Sieg, 2009]

Algorithms and Mechanical Computability

- In the early 20th century, logicians developed several definitions of “mechanical” or “effectively” or “algorithmically” computable.
- They also gave several circular arguments that certain formal definitions capture the intuitive concept of an “algorithm” [Sieg, 2009]
- Turing [1936] gives a rather ingenious argument . . .

Turing's Argument

Following [Sieg, 2008]'s reconstruction, Turing argues that humanly-executable algorithms on symbols satisfy two types of constraints.

Turing's Argument

Boundedness:

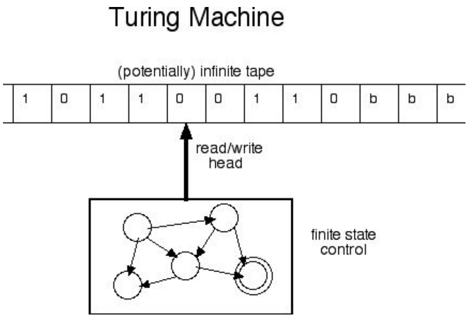
- Capable of immediately differentiating only finitely many symbols
- Capable of immediately using only finite amounts of memory

Turing's Argument

Locality:

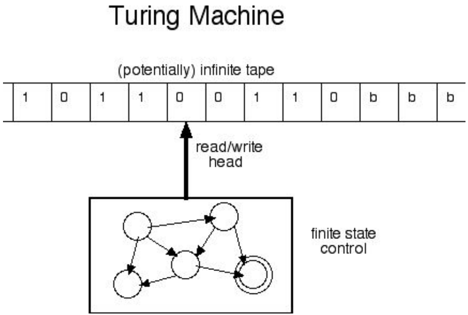
- Only elements of observed symbols can be changed
- Which symbols are observed can be changed, but each of the new observed symbols must be within a bounded distance L of a previously observed square.

Turing Machines



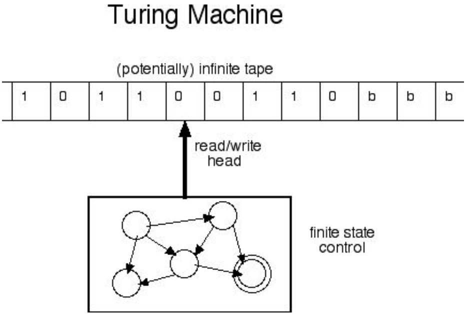
Turing then introduces a model of a computing device, now called a Turing Machine.

Turing Machines



- Turing gives an informal argument that his machines satisfy the boundedness and locality conditions, **and**
- He argues that all devices satisfying the conditions are Turing machines.

Turing Machines



1	0	1	1	0	0	1	1	0	b	b	b

Church Turing Thesis: The set of "algorithmically" computable functions are equal to those that can be executed on Turing machines.

Turing's Argument

Sieg [1994] formalizes the boundedness and locality axioms and proves the set of functions satisfying them are precisely the Turing machine computable functions.

Back to the topic

What does all of this have to do with probability and randomness?

Two Aspects of Randomness

Here are two intuitive features of “random” sequences:

- They exhibit no pattern.
- They cannot be predicted.

Patterns and Complexity

How can we make the notion of a “pattern” precise?

Patterns and Complexity

Consider the following two sequences:

Sequence 1: 001001001001001001001001001001

Sequence 2: 110010100011111100110010011100

Patterns and Complexity

Consider the following two sequences:

Sequence 1: 001001001001001001001001001001

Sequence 2: 110010100011111100110010011100

- The first one can be described succinctly: its 001 repeated ten times.

Patterns and Complexity

Consider the following two sequences:

Sequence 1: 001001001001001001001001001001

Sequence 2: 110010100011111100110010011100

- The first one can be described succinctly: its 001 repeated ten times.
- Describing the second is harder . . .

Patterns and Complexity

- How can we capture the notion of a simple “description” of a sequence of symbols?

Patterns and Complexity

- How can we capture the notion of a simple "description" of a sequence of symbols?
- **Key Idea**: Identify a description with the length of a Turing machine program, as Turing machines represent the symbolic manipulations that agents with limited computational abilities (like us) can perform.

Patterns and Complexity

Suppose you had to write a computer program to print the two strings:

001001001001001001001001001001

110010100011111100110010011100

Patterns and Complexity

Suppose you had to write a computer program to print the two strings:

001001001001001001001001001001

110010100011111100110010011100

- **Program 1:** Print 001 repeated ten times.

Patterns and Complexity

Suppose you had to write a computer program to print the two strings:

001001001001001001001001001001

110010100011111100110010011100

- **Program 1:** Print 001 repeated ten times.
- **Program 2:** Print 110010100011111100110010011100.

Patterns and Complexity

Suppose you had to write a computer program to print the two strings:

001001001001001001001001001001

110010100011111100110010011100

- **Program 1:** Print 001 repeated ten times.
- **Program 2:** Print 110010100011111100110010011100.

The second program is as long as the string itself!

Patterns and Complexity

- Strings with patterns can be compressed: they can be "encoded" in a short computer program that prints them.

Patterns and Complexity

- Strings with patterns can be compressed: they can be "encoded" in a short computer program that prints them.
- Conversely, strings that can be compressed must exhibit some sort of pattern.

Patterns and Complexity

- Strings with patterns can be compressed: they can be "encoded" in a short computer program that prints them.
- Conversely, strings that can be compressed must exhibit some sort of pattern.
- Ergo, random strings are incompressible ones, i.e., ones for which there is no short program.

Patterns and Complexity

A little more formally . . .

- The Kolmogorov Complexity $K(s)$ of a string s is the length of the shortest Turing machine program that prints s.
    - I'm omitting one technical part of this definition that won't concern us now.

Patterns and Complexity

A little more formally . . .

- The Kolmogorov Complexity $K(s)$ of a string s is the length of the shortest Turing machine program that prints s.
    - I'm omitting one technical part of this definition that won't concern us now.
- Let n be some whole number. A string s is called n-incompressible if $K(s) \geq l(s) - n$, where $l(s)$ is the length of s.

Patterns and Complexity

Kolmogorov complexity is defined for strings that are finitely long. How can it be extended to the infinite sequences?

Patterns and Complexity

- **Definition 1:** An infinite sequence *s* is random if there is some constant *n* such that all of the initial segments of *s* are *n*-incompressible.

Patterns and Complexity

What about formalizing the notion of prediction?

Patterns and Complexity

- The exact mathematical ideas are rather hard. Here's the brief overview. For details, see Durrett [2010]

Patterns and Complexity

- The exact mathematical ideas are rather hard. Here's the brief overview. For details, see Durrett [2010]
- In the late 1930s, the French mathematicians Lévy and Ville introduced what are called martingales.

Patterns and Complexity

- The exact mathematical ideas are rather hard. Here's the brief overview. For details, see Durrett [2010]
- In the late 1930s, the French mathematicians Lévy and Ville introduced what are called martingales.
- For our purposes, think of a martingale is a particular way to place bets - say on the flips of a coin.

Patterns and Complexity

- The exact mathematical ideas are rather hard. Here's the brief overview. For details, see Durrett [2010]
- In the late 1930s, the French mathematicians Lévy and Ville introduced what are called martingales.
- For our purposes, think of a martingale is a particular way to place bets - say on the flips of a coin.
- A martingale has the property of being fair:
  - If you place your bets according to a martingale, then your expected earnings at stage $n+1$ are equal to your earnings at stage n. So if you start with no earnings, then your expected earnings on any stage are exactly 0.

Patterns and Complexity

- Some martingales are not Turing computable.

Patterns and Complexity

- Some martingales are not Turing computable.
- Since we are interested in what can be predicted by computationally bounded agents like ourselves, we should confine ourselves to computable martingales.

Patterns and Complexity

- Some martingales are not Turing computable.
- Since we are interested in what can be predicted by computationally bounded agents like ourselves, we should confine ourselves to computable martingales.
- The definition of a computable martingale is a bit different but satisfies a similar "fairness" condition.

Patterns and Complexity

- Some martingales are not Turing computable.
- Since we are interested in what can be predicted by computationally bounded agents like ourselves, we should confine ourselves to computable martingales.
- The definition of a computable martingale is a bit different but satisfies a similar "fairness" condition.
- **Definition 2:** A sequence is random if there is no computable martingale that earns money while betting on it.

Patterns and Complexity

Theorem (Schnorr [1971])

The two definitions of randomness are equivalent.

In fact, both are equivalent to a third notion called Martin-Löf randomness, which is supposed to formalize the idea that sequences exhibiting patterns for infinitely long are rare.

Randomness

Morals:

- There are several precise mathematical notions of randomness.

Randomness

Morals:

- There are several precise mathematical notions of randomness.
- Several definitions are equivalent, and each captures some intuitive features about our judgments of randomness.

Randomness

Morals:

- There are several precise mathematical notions of randomness.
- Several definitions are equivalent, and each captures some intuitive features about our judgments of randomness.
- Finally, some of the most useful theorems from probability theory (laws of large numbers, law of iterated logarithm, Borel Cantelli Lemma, Hoefding's inequality etc.) hold for such random sequences.
  - Note: I'm not sure whether the central limit theorem has been proven for such random sequences.

Randomness

But what's the philosophical upshot for philosophy of probability?

To be honest, I'm not completely sure.

Random Sequences and Criteria for Adequacy

How do they fair with respect to Salmon's criteria for adequacy?

- **Admissibility:** As Suppes argues, the infinite frequency interpretation does satisfy Kolmogorov's axioms, plus countable additivity.

Random Sequences and Criteria for Adequacy

How do they fair with respect to Salmon's criteria for adequacy?

- **Admissibility:** As Suppes argues, the infinite frequency interpretation does satisfy Kolmogorov's axioms, plus countable additivity.
- **Applicability:** As I said, some of the most useful theorems from probability theory hold for such random sequences. But one might still have questions:
    - Do data observed in scientific practice have high Kolmogorov complexity? If not, then are such data readily dismissed as not random?

Random Sequences and Criteria for Adequacy

Ascertainability: One might think that because probabilities are defined as relative frequencies in the infinite limit, that there's no finite time by which we can be sure of an event's probability.

This is a standard philosophical worry [Hájek, 2009], but I'm not sure it's true.

Random Sequences and Criteria for Adequacy

Ascertainability:

- In certain cases in which we know a lot about the process, results like Hoeffding's inequality can make it "probable" that our estimates of probability are close.
  - Formally, the measure of random sequences deviating from the limiting proportion decreases exponentially as a function of sample size.

Random Sequences and Criteria for Adequacy

Ascertainability:

- In certain cases in which we know a lot about the process, results like Hoeffding's inequality can make it "probable" that our estimates of probability are close.
  - Formally, the measure of random sequences deviating from the limiting proportion decreases exponentially as a function of sample size.
  - Whether such measure ought to (can be?) interpreted as a probability itself is open for grabs.

Ascertainability:

- In certain cases in which we know a lot about the process, results like Hoeffding's inequality can make it "probable" that our estimates of probability are close.
    - Formally, the measure of random sequences deviating from the limiting proportion decreases exponentially as a function of sample size.
    - Whether such measure ought to (can be?) interpreted as a probability itself is open for grabs.
- If we don't know a lot about the process, then our estimates may be really far off. But I am not sure this is a consequence of the infinite frequency interpretation of probability theory; it may just be an artifact of the mathematics of probability theory.

Random Sequences and Criteria for Adequacy

Ascertainability:

- In certain cases in which we know a lot about the process, results like Hoeffding's inequality can make it "probable" that our estimates of probability are close.
  - Formally, the measure of random sequences deviating from the limiting proportion decreases exponentially as a function of sample size.
  - Whether such measure ought to (can be?) interpreted as a probability itself is open for grabs.
- If we don't know a lot about the process, then our estimates may be really far off. But I am not sure this is a consequence of the infinite frequency interpretation of probability theory; it may just be an artifact of the mathematics of probability theory.
- If you want a final paper topic that mixes technical questions with philosophical ones, there's plenty to study in this area.

References I

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge University Press.

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

Schnorr, C.-P. (1971). A unified approach to the definition of random sequences. *Mathematical Systems Theory*, 5(3):246258.

Sieg, W. (1994). Mechanical procedures and mathematical experience. *Mathematics and mind*, page 71117.

Sieg, W. (2008). Church without dogma: Axioms for computability. In *New computational paradigms*, page 139152. Springer.

Sieg, W. (2009). On computability. In Irvine, A., editor, *Handbook of the Philosophy of Mathematics*, pages 535—630.

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

Turing, A. M. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 42(2):230265.

Infinite Frequencies

Let s be an finite sequence and E be any subset of the range of s.

- **Technical snag**: The set of all E with limiting frequencies (i.e., probabilities) may not form an algebra because it is not closed under union Suppes [2002].
- But if the number of sequence elements in a set E has a limiting value, then there is some algebra (namely, $\{\emptyset, E, E^c, \Omega\}$) over which one can define a probability measure. It's not clear how useful such a definition is.