Philosophical Theories of Probability

- Administrative:
  - All papers due December 16th (at the latest).
  - New paper topics are up. These will be last (probably).
- Philosophical Theories of Probability
  1. Subjective Theories
     (a) Theories based on Dutch Book Arguments
     (b) Theories based on Representation Theorems
     (c) Theories based directly on Expected Utility Principles
  2. Objective Theories
     (a) Classical Theories
     (b) Logical Theories
     (c) Frequency Theories
     (d) Propensity Theories

Subjective Theories I – Dutch Book Arguments I

- For each proposition E (about which he has beliefs), Mr. B must announce a number $q(E)$ (called his *betting quotient* on E), and then Ms. A will choose the *stake* S. Mr. B must pay Ms. A \$ $[q(E) \times S]$ in exchange for S if E occurs (else, he gets $q(E) \times S$). S can be positive or negative, but $|S|$ should be small in relation to Mr. B's total wealth. Then, $q(E)$ is taken to be a measure of Mr. B's degree of belief in E.
- We say that $q(\cdot)$ is *coherent* iff it is impossible for Ms. A to choose stakes S such that she wins *no matter what happens*. That is, $q(\cdot)$ is *coherent* iff Ms. A cannot construct a "Dutch Book" against Mr. B.
- **Theorem**. $q(\cdot)$ is *coherent* iff $q(\cdot)$ is a *probability function*.
- **Claim** (*not* a theorem!). $q(\cdot)$ is *rational* iff $q(\cdot)$ is a *probability function*.
- **Theorem** $\overset{?}{\Longrightarrow}$ **Claim**? *Betting* behavior $\overset{?}{\approx}$ *rational* behavior?
- **Theorem***. Mr. B is open to a *sure win* iff $q(\cdot)$ is *non*-probabilistic!

Subjective Theories I – Dutch Book Arguments II

1. $\Pr(p \vee {\sim}p) = 1$. If Mr. B assigns $q(p \vee {\sim}p) = q < 1$, then Ms. A can just set $S < 0$, *guaranteeing* that Mr. B has a net gain of $S - qS < 0$. If Mr. B assigns $q(p \vee {\sim}p) = q > 1$, then Ms. A can just set $S > 0$, and Mr. B has a net gain of $S - qS < 0$. So, the only coherent betting quotient on $p \vee {\sim}p$ is $q(p \vee {\sim}p) = 1$, which is exactly what is mandated by the probability calculus.
2. $\Pr(E) \geq 0$. If Mr. B assigns $q(E) = q < 0$, then Ms. A can just set $S < 0$, so that Mr. B has a net gain of $S - qS < 0$ if E is true. Otherwise (if E is false), then Mr. B has a net gain of $qS < 0$. So, the only coherent betting quotient on E is such that $q(E) \geq 0$, which is exactly what is mandated by the probability calculus.
3. The only other axiom is: $\Pr(p \vee q) = \Pr(p) + \Pr(q)$, provided that p and q cannot both be true. The Dutch Book argument for this *additivity* axiom is somewhat more complex and controversial. See Patrick Maher's *Betting On Theories* for details and critique.

Subjective Theories II – Representation Theorems I

- Let f, g, h be acts which Mr. B could choose to perform. And, let $f(x)$ be the consequence (for Mr. B) which obtains if he chooses to perform act f, and the resulting state of the world is x. The *expected utility* of choosing to perform act f is defined as: $EU(f) = \sum_x q(x) \cdot u(f(x))$, where $q(\cdot)$ are Mr. B's degrees of belief, and $u(\cdot)$ are Mr. B's utilities.
- $f \precsim g$ iff Mr. B (weakly) *prefers* g to f. $f \precsim g$ iff either Mr. B *strictly* prefers g to f ($f \prec g$), or is *indifferent between* g and f ($f \sim g$). $f \precsim_A g$ iff Mr. B would prefer g to f, were he to learn that A is true.
  1. **Normality**. Either $f \precsim g$ or $g \precsim f$ (or both).
  2. **Transitivity**. If $f \precsim g$ and $g \precsim h$, then $f \precsim h$.
  3. **Sure-Thing Principle**. If $f \precsim_A g$ & $f \precsim_{\sim A} g$, then $f \precsim g$. And, if Mr. B thinks A is possible, then if $f \prec_A g$ & $f \precsim_{\sim A} g$, then $f \prec g$. [What you prefer at a node in a decision tree does not depend on what would have happened had you *not* reached that node.]

Subjective Theories II – Representation Theorems II

- **Theorem**. If Mr. B's preferences satisfy (1)–(3), then there is a *unique* representation of Mr. B's preferences in terms of his degrees of belief $q(\cdot)$ and his utilities $u(\cdot)$, such that $f \precsim g$ iff $EU(f) \leq EU(g)$. Moreover, the degrees of belief $q(\cdot)$ in this representation will obey the probability axioms. So, under these assumptions, Mr. B will act (if he acts rationally!) *as if* his degrees of belief are probabilistic.
- **Theorem***. If Mr. B's preferences satisfy (1)–(3), then there is a *unique* representation of Mr. B's preferences in terms of his degrees of belief $q(\cdot)$ and his utilities $u(\cdot)$, such that $f \precsim g$ iff $EU^*(f) \leq EU^*(g)$. Moreover, the degrees of belief $q(\cdot)$ in *this* representation will *not* obey the probability axioms (but, $EU^*(f)$ is defined *non*-standardly).
- Rational agents *can be represented as having* probabilistic degrees of belief $\overset{?}{\Longrightarrow}$ rational agents *do* have probabilistic degrees of belief?
- See Maher's *Betting On Theories* for discussion of DB's and RT's.

Subjective Theories III – A NEW Direct EU Approach

- Maher (unpublished) has recently shown that if one works directly with expected utility instead of preference structure + expected utility, then one can give a very simple argument that degrees of belief q must be probabilities. So, using the same notation as before, we have ...
- Assume that an agent has degrees of belief q and utilities u. And, assume that the agent calculates expected utility $EU(f)$ of acts f in the standard way. It can be shown that the following two simple and intuitive dominance principles are sufficient to guarantee that the degree of belief function q in EU is a *probability* function.
  - **Unconditional Dominance**: If $u(f(x)) > u(g(x))$ for all possible outcomes x, then $EU(f) > EU(G)$.
  - **Conditional Dominance**: If $EU(f \mid E) > EU(g \mid E)$, and $EU(f \mid {\sim}E) > EU(g \mid {\sim}E)$ (some possible E), then $EU(f) > EU(g)$.
- MUCH simpler and easier + makes clear how much work EU is doing!

Objective Theories 0 – The "Classical" Theory

- The assumption behind the finite version of the classical theory is that we have a set P of n *equiprobable* possible cases, and that the probability of any event E relative to P is just the number of possible cases in which E is true divided by the total number of possible cases n.
- Infinite cases involving continuous magnitudes can be paradoxical:
  - A square has been generated "at random." We are told that the length of its sides is somewhere between 10 and 20 feet. What is the probability that the square is between 10 and 15 feet on a side? It might seem natural to say that this probability is $\frac{1}{2}$. However, it is also true (given what we've been told) that the square will have a "random" area between 100 and 400 square feet. This way of describing the square makes it sound natural to say that the probability is $\frac{1}{2}$ that its area is between 100 and 250 square feet, which would seem to imply that the probability is $\frac{1}{2}$ that the square is between $10 = \sqrt{100}$ and $15.8 \approx \sqrt{250}$ feet on a side. Absurd!

Objective Theories I – Logical Theories

- Traditional logical approaches to probability start with a *language* $\mathcal{L}^n_m$ consisting of n predicates $P_1, \ldots, P_n$ and m individuals $a, b, c \ldots$.
- Two sentences of a language $\mathcal{L}$ are said to be *isomorphic* if they differ only with respect to where the names of individuals occur (*i.e.*, up to permutations of names). "$P_1 a \mathbin{\&} P_2 b$" is isomorphic to "$P_1 b \mathbin{\&} P_2 a$".
- A *state description* is a complete specification of which individuals have which predicates. For instance, consider the language $\mathcal{L}^2_2$, consisting of the predicates P_1, P_2 and the individuals a and b. In $\mathcal{L}^2_2$, the conjunction $P_1 a \mathbin{\&} {\sim}P_1 b \mathbin{\&} {\sim}P_2 a \mathbin{\&} P_2 b$ is a state description.
- A *structure description* is a disjunction of a state description and all of its isomorphs. "$P_1 a \mathbin{\&} {\sim}P_1 b \mathbin{\&} {\sim}P_2 a \mathbin{\&} P_2 b \vee P_1 b \mathbin{\&} {\sim}P_1 a \mathbin{\&} {\sim}P_2 b \mathbin{\&} P_2 a$" is a structure description in $\mathcal{L}^2_2$.
- There are $2^{m \cdot n}$ state descriptions in $\mathcal{L}^n_m$. And, there are at least $2^{m \cdot n} - m! + 1$ (but fewer than $2^{m \cdot n}$) structure descriptions in $\mathcal{L}^n_m$.

- Assigning equal probabilities to state descriptions yields different results from assigning equal probabilities to structure descriptions. [Analogy: if particles are *distinguishable* (state descriptions) in statistical mechanics, then we get classical statistics (Maxwell-Boltzmann). But, if particles are *indistinguishable* (structure descriptions), then we get quantum statistics (Fermi-Dirac or Bose-Einstein).] Carnap discusses both approaches in his *Logical Foundations of Probability*.
- Problem: Probabilities are language-dependent. If we choose a different (but equally expressive) set of predicates, then we can end-up with radically different probabilities assigned to (intuitively) the same propositions. Typical problem with syntactical approach.
- Problem: if we assume equal probabilities over either state or structure descriptions, we cannot have any analogical inferences. That is, we will never be able to say things like $\Pr(P_1 b \mid P_1 a) > \Pr(P_1 b)$ on such models.
- Challenge: How can "logical probability" be founded on *semantical* ideas? Some recent work has been done on this (Joe Halpern).

Objective Theories II – Actual Frequency Theory

- Say we are interested in determining probabilities involving certain factors X, Y, Z, etc. in a (finite) population P of size N.
- One suggestion is to define $\Pr(X)$ as the *actual frequency* of X's in P. Let $\#(\alpha)$ be the size of α. Then, the actual frequency account suggests:

$$\Pr(X) = \frac{\#(X)}{N}$$

- On this account, *all probabilities must be rational numbers*. What if the bias of a coin is $\frac{1}{\sqrt{2}}$ or $\frac{1}{\pi}$? Moreover, our physical theories (*e.g.*, QM) imply *irrational* probabilities for many events (*e.g.*, the probability of finding en electron in a certain region around a hydrogen nucleus).
- Actual frequencies can often be *misleading* about the underlying *causal* (or just *probabilistic*!) facts. Small samples may be *unrepresentative* of underlying structure. We think that a streak of 1000 heads in a row (with a fair coin) will eventually get "washed out" in the long run.

Objective Theories III – Hypothetical Frequency Theory I

- Instead of taking the *actual* frequency of X in a finite population (or sequence) P as definitive of the probability of X (relative to the chance set-up K which generated P), we could instead think about *hypothetical infinite extensions* of the sequence P.
- For instance, let $\Pr(X \mid K)$ be the probability of a coin landing heads in a particular chance set-up K. We could throw the coin N times (in K), and generate a sequence P of tosses. Then, we could look at the actual frequency of heads in P. But, would this give us $\Pr(X \mid K)$?
- Why not take the *hypothetical limiting frequency* of X in *hypothetical infinite extension*(s) of P, generated in K as $\Pr(X \mid K)$?
- Do we just use frequency in "the" sequence which would have obtained if P had been extended indefinitely (in K)? Or, do we average over frequencies of "the" sequence of *populations* of size N which would have obtained, had we re-generated P's over and over again (in K)?

Objective Theories III – Hypothetical Frequency Theory II

- Let us assume that the experiment is repeated on populations of the same size N as the actual population P. Specifically, let P_0 $(= P)$, P_1, P_2, … be the infinite sequence of hypothetical populations that would result from conducting this experiment infinitely may times.
- And, let Fr_0, Fr_1, Fr_2, … be functions giving the values of frequencies involving the factors X, Y, Z, etc. in the populations P_0, P_1, P_2 … respectively. Then, according to this sketch of a hypothetical frequency (plus propensity?) account, we have:

$$\Pr(X) = \lim_{n\to\infty} \frac{\sum_{i=0}^{n} Fr_i(X)}{n}$$

- Q: Why consider the different populations P_i one by one and then average their Fr_i's rather than combining the P's into one overall population and then looking for limits of frequencies in the combined population? A: we need to have an *ordering* over the P_i …

Objective Theories III – Hypothetical Frequency Theory III

- The limit (if it exists — see below!) depends on the *order* in which the (partial sums of the) frequencies are taken. So, we need to have an ordering of the P_i's, and the "temporal" ordering (of the hypothetical populations!) seems most natural (there would be no principled way of ordering the frequencies in the "combined" infinite set of populations).
- Also, the *size* of a population might *itself* be a relevant factor. So, we want to first identify the Fr_i — all defined on populations *of size* N — and then "average" them, rather than combining into an *infinite* population, and then looking for limiting frequencies within it.
- This approach avoids the two problems of the actual frequency account. On this account, probabilities may take on any real number on $[0, 1]$. And, intuitively, we think contingencies of the actual population will "wash out" as we go to the infinite (hypothetical) limit — especially, since we're assuming the initial conditions are *causally complete* with respect to the factors in which we are interested.

Objective Theories III – Hypothetical Frequency Theory IV

- We can see now why we need to specify what *kind* of population P is. Presumably, we are interested not just about the probabilities of X, Y, etc. in the actual population P. We want to know what the probabilities are in populations of a certain *kind*, which P exemplifies.
- Different kinds of populations would, presumably, have different sets of initial-condition-features, and so would (in an indeterministic world!) determine different probabilities for the properties in question. This can be generalized to cases in which the populations are of different sizes.
- Problems: Is there such a thing as *THE* sequence of (hypothetical) repetitions P_i that would have obtained, if we had repeated the experiment over and over again *ad infinitum*? This matters because even a different order of the same P_i's would undermine the limit definition. There are uncountably many sequences (just re-orderings of the P_i's) — and *any* limit value can be generated in infinitely many ways . . . See Eells' *Probabilistic Causality* chapter 2 for much more . . .

Objective Theories IV – Propensity Theories

- Propensity interpretations are similar to hypothetical frequency interpretations, in that they assume there is an experimental set-up which fixes (presumably, *via* fixing relevant causal factors) the limiting-frequencies that will be observed in repetitions of experiments.
- The difference is that propensity accounts do not define probability as limiting relative frequency. Rather, they take the probability to be the propensity or the disposition itself, and the frequencies are just emergent properties of the underlying structure in the set-up.
- Problems: Are propensities probabilities at all? Presumably, the experimental set-up S confers probabilities on factors, but is the converse also true? In probability theory, there is no asymmetry. If $\Pr(X \mid S)$ is well-defined in a probability space, then so is $\Pr(S \mid X)$. But this seems nonsensical on the propensity account (propensities have a directionality imposed by causal structure). How dow we learn about propensities? Is our only access through *frequencies*?

Objective Theories V – Reference Classes and Single Cases

- The reason *propensities* are needed is that there seems to be no other way to account for probabilities of *token events* or *single cases*. Probabilities are typically thought of as being relative to a *reference class*, and as being some kind of *frequencies* in a *population* of events.
- What is the probability that John Doe will get cancer in the next year? Well, that depends on what *reference class* you have in mind. If you choose a reference class too narrowly (things identical to John Doe), then it seems that the probability (relative to this reference class) is either 1 or 0 (depending on whether he does in fact contract cancer).
- If we choose too broadly, then we're leaving out relevant factors. However, if we take probability to be a propensity of John Doe himself (given his life-history up to this point, and everything about him), then we have no trouble thinking about the probability of this token event. This is just a disposition that John possesses (in his current context). [Relative frequencies always have to be taken wrt a reference class . . .]