Probability

I. The Probability Calculus (Skyrms)

i) Axioms

ii) Key Theorems

iii) Bayes' Theorem

II. Interpretations of probability: overview

i) Criteria of adequacy

ii) Brief survey:
logical, subjective and objective interpretations

III. Two objective interpretations

A) Propensities: single-case probabilities (Giere)

B) Hypothetical limiting frequencies (Eells)

II. Interpretations of probability

- An interpretation of the concept of probability is a choice of some class of events (or statements) and an assignment of some meaning to probability claims about those events (or statements).

  Example: Drawings from a deck of cards (with replacement). *Pr*(A/B) is the number of possible drawings in which A occurs over the total number in which B occurs.

- Three criteria for any proposed interpretation (Salmon)

  1) *Admissible*: must satisfy the basic mathematical properties of the probability calculus. This is also called *coherence*.

  2) *Ascertainable*: must be values that we can determine (or else useless).

  3) *Applicable*: must be values that can be relied on as a “guide to life” (hook up with rational choice theory).

- Additional criteria may be required if an interpretation is to be suitable to model the relationship of probabilistic causation:

  Rough definition:

  A is a probabilistic cause of *B* iff $Pr(B / A) > Pr(B / {\sim}A)$.

1. Logical interpretation

Pr(*A* / *B*) is the degree to which *B* partially entails *A*.

Example: Predicates F, G; objects *a*, *b*. 16 possible states of the world.

States	State Descriptions	*m*	*m**	
F*a* · F*b* · G*a* · G*b*	All positive	1/16	1/5	
F*a* · F*b* · G*a* · ~G*b*		1/16	1/20	}
F*a* · F*b* · ~G*a* · G*b*	One negative	1/16	1/20	}
F*a* · ~F*b* · G*a* · G*b*		…	1/20	} total 1/5
~F*a* · F*b* · G*a* · G*b*			1/20	}
F*a* · F*b* · ~G*a* · ~G*b*			1/30	}
F*a* · ~F*b* · G*a* · ~G*b*			1/30	} total 1/5
~F*a* · F*b* · G*a* · ~G*b*			1/30	}
F*a* · ~F*b* · ~G*a* · G*b*	Two negative		1/30	}
~F*a* · F*b* · ~G*a* · G*b*			1/30	}
~F*a* · ~F*b* · G*a* · G*b*			1/30	}
F*a* · ~F*b* · ~G*a* · ~G*b*	Three negative		1/20	}
~F*a* · F*b* · ~G*a* · ~G*b*			1/20	} total 1/5
~F*a* · ~F*b* · G*a* · ~G*b*		…	1/20	}
~F*a* · ~F*b* · ~G*a* · G*b*		1/16	1/20	}
~F*a* · ~F*b* · ~G*a* · ~G*b*	All negative	1/16	1/5	

A: Everything is F   *B*: F*a*

- All states equi-probable (*m*):

  Pr(*A* / *B*) = *m*(*A* & *B*) / *m*(*B*) = (1/4) / (1/2)   = ½

- All state descriptions are equi-probable, and individual states are equi-probable within each state description (*m**):

  Pr(*A* / *B*) = *m**(*A* & *B*) / *m**(*B*) = (1/3) / (1/2)   = 2/3

Unsuitable for probabilistic causation:

1) Internal weaknesses (dependence on language; arbitrary use of P.I. R. in assignments of equi-probability: *m* is not unique)

2) Truth of claims about probabilistic causation is *empirical*, not *logical*.

3) Probabilistic causation relates *event types*, not *propositions.*

2. Subjective interpretations

a) Actual degree of belief

Pr(B / A) is your degree of belief in *B*, given the information *A*.

Unsuitable for probabilistic causation:

1) Not admissible (actual human degrees of belief violate the axioms of the probability calculus)

2) Vary from person to person (based on differing prior probabilities and variation in available evidence)

3) Probabilistic causation relates *events* (*event types*), not propositions.

b) Personalism/Bayesianism: idealized degree of belief

As in actual degree of belief, but assume *coherent* or admissible degrees of belief.

Problems:

1) Not obviously applicable
(no constraints beyond coherence)

R: best we have; convergence

2) Variability from person to person

R: convergence

3) *Event types*, not propositions

R: there are corresponding propositions

Best we can expect from subjective interpretations: an account of our beliefs about causation

3. Objective interpretations

1) Frequencies

Let *A* be a *reference class* or *population* and *B* an *attribute class*

a) Actual finite frequency

Pr(*B* / *A*) is the proportion of sampled *A*'s observed to be *B*'s

Problems: All connected with *applicability*.

1) *Freaky coincidences*. We want causal facts, not mere accidents of association.

Example 1: the rounded die (*should* favour 1, a run of trials actually favours 2-6).

Actual frequencies can take on *any* value from 0 to 1, consistent with the *true* probability strictly between 0 and 1.

Example 2: 'freaky' correlations (barometer/storm)

Here, actual frequencies do not sort out the misleading correlation.

2) *Graininess: a priori constraints*.

A coin only tossed an odd number of times and then destroyed *must* be biased for heads or tails. If tossed N times, prob. heads must be 0, 1/N, …

b) Hypothetical frequency

Intended to capture appropriate probability value that might not be reflected in the actual results.

Pr(*B* / *A*) is the *limiting frequency* of *A*'s observed to be *B*'s

Problems:

1) *Ascertainability*. (Any initial pattern is compatible with any limit, or with non-existence of a limit.)

2) *Problem of single-case*. Probability applied to the single case makes no sense. Choosing and justifying the appropriate population or reference class is difficult.

3) *Sensitivity to re-ordering*.

More below: Eells develops a sophisticated version.

2) Propensity

Pr(*B* / *A*) is the *single-case physical probability* for something that is *A* to be *B*.

Problem: What *is* a propensity?

More below: Giere develops a propensity interpretation.

III. Two Objective Interpretations

A) Giere: Objective Single-case Probabilities

1. Introduction

Fundamental question: what interpretation of probability is appropriate for scientific practice?

Carnap's answer:

Both the objective (frequency) and subjective (for Carnap, logical) interpretations are needed.

In reality, there is no easy compromise:

Most scientists (and some philosophers) are frequentists and don't trust the Bayesians.

Many philosophers (and some scientists) are Bayesians or subjectivists "all the way down", and many don't believe in objective probabilities.

Giere's strategy:

1. Take note of subjectivist criticisms of frequency interpretations
2. Define single-case probabilities or *propensities*
3. Show that this interpretation answers many of the objections to frequency interpretations, and illuminates scientific practice

Note: Primary concern is with scientific inference, but we're interested in application to scientific explanation and causation.

2. Frequency interpretations

a) Limiting frequency view

Pr(A / B) = limit of relative frequencies of A in a finite sequence of B's

Main problems:

i) *Problem of the single case.* Probability does not apply except relative to infinite sequences.

Best solution: Use the limiting frequency of the attribute in some infinite reference class as a 'weight'.

ii) *Problem of the reference class*. Which infinite reference class?

Ex: First toss of a newly minted coin. What is probability of heads?

- all tosses of this coin
- all first tosses of newly minted coins

iii) What is the *justification* for taking any limiting frequency as the probability in the single case?

b) Kolmogorov

Pr(E) = r, relative to a chance setup CSU, iff in a long series of repetitions of CSU, it is practically certain that the frequency of E will be approximately equal to r.

Main problems:

i) *Circularity*: "practically certain" is a probabilistic notion

[**Yet:** Eells' eventual definition is somewhat close to this.]

ii) *Confusion* of epistemology and metaphysics: how one asserts the existence of a probability is distinct from what it is.

3. Physical probabilities as single-case propensities

> *P(E)* = *r* iff the strength of the *propensity* of CSU (= a chance set-up) to produce outcome *E* on a particular trial *L* is *r*.

Note: Despite the notation, the probability is plainly relative to CSU and *L*.

a) Clarification and contrast with Popper

Giere:

> propensity *of* {CSU} to *produce* {E in a particular case}

Popper:

> propensity *of* {a repeatable experimental set-up under some description or specification} to *produce* {sequences with relative frequency *r*}.

4. Giere rejects the 'sequences' formulation because we still have no implication for the single case.

5. He rejects *relativization to specification* because we then have the old problem of the reference class to solve (*which* specification?)

Force analogy:

Giere's picture is that propensities are a *physical model* or *realization* of the formal abstract probabilities that appear in the laws of physics, just as particular forces are physical models of the abstract forces in Newton's Laws. Propensities are *basic* and irreducible.

Problems:

- VERY mysterious.
- apparent problem with ascertainability? (**§** 5)

[**Qu:** Does the equation on p. 477 express any sort of connection with finite frequencies? It is a propensity-propensity connection.]

- incomplete interpretation: inverse probabilities **can't** be propensities (Salmon's chocolate bar example)

4. Determinism and single-case probabilities

Objection to Bayesians: subjective probability lacks the resources to distinguish uncertainty due to lack of information from uncertainty that no possible increase in knowledge could eliminate (i.e., due to indeterminism).

In effect, Bayesians must assume that all uncertainty is epistemic, i.e., assume determinism.

Comments:
1. This objection, if valid, applies only to hard-core subjectivists.
2. Using evolving probabilities, you can make sense of the distinction.
3. Not clear why determinism is entailed by subjectivism.

Determinism vs. indeterminism

Under determinism, all single-case propensities are 0 or 1. (Not so for limiting frequencies.)

Paradigm example of propensity: particle decay.

Propensity is just the physical realization of the probability term in (e.g.) quantum mechanical laws.

5. Propensities and relative frequencies

What is the connection between propensities and relative frequencies?

1) Interpretation of probability

Propensity interpretation gives a natural interpretation of the theory of probability.

> [But see earlier comment on incompleteness wrt/ inverse probabilities.
>
> More generally: what sense can we make of *conditional probability*? If propensities are intrinsic properties of the CSU, how do we make sense of these relational types of probability?]

2) Connection with frequencies

i) Frequencies and propensities are not inter-definable (given indeterminism, any propensity is consistent with very different finite frequency).

> There is no propensity-frequency connection, but there is a propensity-propensity connection (LLN). But what use is it?

ii) Existence of propensities is inconsistent with a Humean metaphysics:

> - propensity statements cannot express Humean laws (no logical connection to actual regularities)
>
> - if one knows the entire history of the world, one knows the Humean laws, but not necessarily the single-case propensities [This is the problem of **ascertainability**.]

iii) Connection to finite relative frequencies.

> - propensity hypotheses are typically tested using large numbers
> - but this assumes certain propensity statements (e.g., all objects in the group have the same propensity)
> - this is not obviously unacceptable: same point goes for forces

6. Propensities and statistical practice

"Smoking causes lung cancer" = smoking increases, on average, one's propensity to get lung cancer

"Beautiful cases": some (possibly) deterministic systems (e.g., roulette wheels) simulate indeterministic systems, but the physical probabilities don't exist.

B) Eells, *Probabilistic Causality*, chapter 1

Objective interpretation: hybrid of hypothetical limiting frequency and propensity interpretations

Main claim: Type-level probabilistic causation is a **four-place** relation:

- a causal factor
- an effect factor
- a token population
- a *kind* of population associated with the token population

1.1 Populations

"C causes *E*", or "*C* raises the probability of *E*"

i) Such claims are true/false only **relative** to a **population**.

- there is no such thing as prob.(*E*) except as applying to an individual (possibly) or a population

But which individual or pop? Makes a big difference and can convert a positive factor into a neutral or negative one, or vice versa.

Example: Smoking causes lung cancer.

Claim: A property-level probabilistic causal claim must be made relative to a particular, token population.

ii) And only **relative** to a **kind**.

E.g., newly minted coins? Kind *Q*: all tosses of this coin. Kind *Q″*: all first tosses of new coins.

E.g., smoking/lung cancer (from introduction). Kind *Q*: simply human. Kind *Q″*: human + all hit by buses first.

E.g., complex coin toss machine.

In each case, the answers to questions about probability depend on the *kind*.

Upshot: "*C* causes *E* in population *P* taken as of kind *Q*."
[**Question:** Is choice of *Q* independent, or fixed by *C*, *E*, *P*?]

iii) **Restrictions on the kind *Q***

(1) Q must not *logically imply* the value of the frequency of E, or of E given C, or of E given $\sim C$.

(2) Q must not even *logically imply* even the equality or direction of inequality among the above three frequencies.

(3) Q must allow both for the possibility of C and the possibility of $\sim C$ being exemplified in token populations of type Q.

Arguments for (1) and (2):

1. Some probabilistic causal claims would then be logical truths.
2. Some implied probabilistic claims would clearly be wrong.

Example:

P = pairs of simultaneous (independent) tosses of two fair coins
Q = this kind of population
C = left coin comes up heads
E = right coin comes up heads

We should see: C is causally neutral for E.

Choose a sub-population P' in which:

- results on left and right always match
- half the coins come up heads

P' is a token population both of kind Q and of kind Q' = {type Q with half heads and matching on left and right}.

Relative to Q: C is causally *neutral* for E.
Relative to Q', C is *positive* for E, as $f(E / C) = 1$ and $f(E) = 0.5$.

Same problem even if exemplifying Q' only implies *comparative* relations among $f(E)$, $f(E / C)$, $f(E / \sim C)$.

Argument for (3):

Suppose P is a population in which everything is C. If we allow the kind Q to be such that all tokens of kind Q consists only of C's, then it becomes *a priori* impossible for C to be causally relevant to E.

<For $f(E / C) = f(E)$ within such populations, by definition.>

Concrete example: In a population in which everyone smokes, we still want to be able to say that smoking is a cause of lung cancer.

1.2 Probability

Key question: What interpretation of probability is appropriate for probabilistic causation?

An **objective**, **physical** relationship between event-types

Objective = independent of human conceptions (**and descriptions)
Physical = probabilistic facts are fixed by physical/causal factors

[Which is the fundamental concept: causation or probability?]

1. Review (and rejection) of alternative interpretations of probability

Special concern: rule out *actual frequency* as the correct interpretation.

Current "modal" frequency interpretations:

Popper/Kyburg: *limiting frequency* (or disposition to produce a limiting frequency) in a hypothetical infinite population

Giere/Fetzer: single-case *physical propensity*

Blanket criticism: A satisfactory account of objective probability should be adequate for the role probability plays in:

i) explanations of single events (e.g., particle decay)
ii) explanations of causal laws
*iii) rational choice theory in handling Newcomb cases

<Point iii) in particular poses problems for the single-case physical propensity view.>

Mixture: Eells' account is a "mixture" of propensity-based and hypothetical frequency approaches.

<Circularity/interdependence of causation and probability a possible problem?>

Eells' alternative interpretation

- Start with an actual finite population *P* with *N* members.

- Kind *Q*: N-membered populations in which some set of *initial conditions* I_1, …, I_n are *distributed* among the members. Call each such distribution a particular "experimental set-up" (ESU).

> These initial conditions are (*physical*) causal factors relevant to the possession of certain other factors *X*, *Y*, *Z*, … that are really of interest to us. The ESU determines each individual's *propensity* to have or lack *X*, *Y*, *Z*, …
>
> [**Qu:** What are these propensities? As for Giere, they are just irreducible single-case physical probabilities. But they are no longer defined intrinsically, so we can talk about conditional propensities, inverse propensities, etc.]

- We can repeat the experiment over and over, resulting in different populations where individuals have different propensities to have or lack *X*, *Y*, *Z*, etc.

  Consider the hypothetical sequence of populations that results:

  P_0 (= P), P_1, …

  And corresponding *frequencies* of the factors *X*, *Y*, *Z*, … given by

  Fr_0, Fr_1, …   where $Fr_i(X)$ is the frequency of *X* in P_i .

Define

$$Pr(X) = \lim_n \Sigma_{\{i=0 \text{ to } n\}} Fr_i(X) / n = \text{limiting average frequency of } X$$

Assessment

1. Virtues

1. solves ‘graininess’ problem: we can get any limit between 0 and 1 (note that there might be no limit)

2. the effect of ‘freakish coincidences’ will be dampened: with probability 1 (though not with certainty), you remove this effect

3. relativity to kind is crystal-clear. The kinds are specified in terms of the set of initial conditions $I_1, \ldots, I_n$.

[N.B. This **is** a problem with Giere’s propensities. Can we make sense of non-relativized objective single-case probabilities, even in quantum mechanics? If propensities are as he says, why are they *probabilities*?]

2. Extensions

1. allow variable population sizes.

[Esoteric discussion of two ways to define probability:

(1) as limit of average frequency
(2) as limit of cardinality ratio

with a defense of (1) for a select example.]

3. Difficulties

1. What meaning can be attached to the sequence of populations $P_0, P_1, \ldots$ that *would result* from repeated experiment?

   The limit may vary from one sequence to another.

   It won’t help to say that the limiting average should be the same in “almost all” cases. No clear solution here.

2. Propensities are still mysterious; and if defined in terms of hypothetical frequencies, you get the same difficulties with limits.