Limitations of Standard Probability and Several Alternatives

UC Berkeley
(28 November 2007)

Terrence L. Fine
School of Electrical & Computer Engineering
Cornell University
Ithaca NY 14853

Not everything that counts can be counted, and not everything that can be counted counts.

(attributed to Einstein)

Goethe

- Mathematicians are like Frenchmen: whenever you say something to them, they translate it into their own language, and at once it is something entirely different.

Kolmogorov

- My entirely general, half-philosophical reflections took more of my time and energy than it might seem at first sight. In elaborating quite general views, the outcome of one's efforts is not a formulation of precisely fixed ``results" but an overall restructuring of one's own thinking and the arrangement of all facts in due perspective. Therefore, outwardly such work may appear to be sheer waste of a lot of effort and time with nothing ``new" really discovered.

- My debt to Pat Suppes will become evident in the course of this talk.

- I speak against the prevailing ``unassailable" position that there is only one mathematical model for probability that can be given any of the usual ``interpretations".
- On the positive side, I will present several alternative mathematical models, based on orders, sets of measures, and upper and lower probabilities, that can be used for different applications and types of nondeterministic phenomena.

- My goal is to enlarge the mathematical space of probability representations to include ones that can more faithfully represent credibly possible types of non-determinate phenomena.

Spectrum of Meanings

- Different nondeterministic applications can require different *meanings* for the probability concept(s) needed to describe their nondeterministic sources of chance, indeterminacy, or uncertainty.
- I prefer to use ``meaning" as opposed to ``interpretation" that presumes a pre-existing set of axioms.
- Meaning *precedes* axiomatization.

- These different meanings have properties that require mappings from the ``empirical" realm of phenomena and meanings into possibly different mathematical relational systems in the formal realm.
- Mathematical probability is expressed by the chosen mathematical relational system.
- A key, usually neglected, question is the match between the *expressiveness* of the mathematical model and the nature of the application.

- A mathematical model is too expressive if it allows distinctions that are empirically meaningless---e.g., a real-valued IQ as a measure of some aspect of intelligence.
- The model is insufficiently expressive if it cannot represent important distinctions in the empirical realm---e.g., standard probability for quantum mechanics.

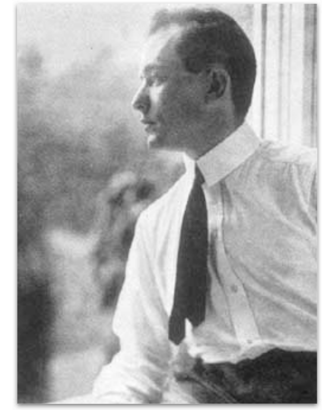
- Major meanings can be grouped under:
- *Physically-determined*: probability from physics rather than from pre-existing probabilities, e.g., through statistical or quantum mechanics, propensity
- *Chance:* objective claims about unobserved experimental outcomes based upon observed outcomes: e.g., frequentist---repeated, unlinked experiments---core of physical science relation of theory to experiment

- *Uncertainty, subjective, decision-oriented*---rational individual decision-making, degree of belief, propositional attitude
- *Epistemic, logical, indeterminate*---relations of inductive support between statements in a formal language, formal inductive reasoning, explicating indeterminacy.

Physically-Determined: Propensity Interpretation

- QM probability, applying as it does to the single case, is more understandable as a Popperian propensity.
- On the propensity account we can make sense of probability for a single experimental outcome as computed by QM.
- The display of propensity returns us to finite relative frequency.
- More needs to be done on the mathematical representation of propensity probability.

Quantum Mechanics



Max Born's 1926 interpretation of Schrodinger's wave function as yielding a probability density

Randomness in the physical realm is inextricably entrenched in QM

``for his fundamental research, especially for his statistical interpretation of the wave function."

Observables and State in QM

- The state space is a closed, infinite-dimensional Hilbert space.
- Observables correspond to Hermitian operators on the state space.
- The state and the operator yield the probabilities for possible measurements.
- Observation of an event corresponds to a projection of the state onto a closed linear subspace of the Hilbert space.

Heisenberg's Uncertainty Relation

- If $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, then the operators do not commute.
- Observables corresponding to non-commuting operators are not simultaneously measurable. The order of measurement matters.
- While we can measure each of A and B arbitrarily accurately, we cannot do so for both A and B.
- There is an irreducible minimum to the product of the variances of the two measurements.
- Boolean event logic fails and the event collection is a non-Boolean lattice.

Failure of Expressiveness of Standard Probability in QM

- The order of observation of canonically conjugate observables affects the state.
- The collection of observable events is non-Boolean (lattice of subspaces).
- Distributivity of union and intersection fails.
- A consequent failure is that QM probability does not always obey the formula for the probability of a non-disjoint union of two events.

- **Modularity**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This holds for events measured by *compatible experiments* in which projection is onto nested subspaces.

- **It does not hold in general, unlike the case for standard probability.**
- **See Jauch 6-3.**

Chance

- Specifying the *chance* behavior of a physical system through probability is done through statistical mechanics or quantum mechanics.
- Chance problematically links to experimental outcomes through frequencies of occurrence.

Chance and Frequency

- What properties of the past observed data are *reliably projectable* into claims about future data generated by the same source?
- Model of repeated, unlinked experiments.
- What properties of past trials are *stable* enough to be projected onto future trials?
- Future *relative frequency of occurrence* is the usual meaning assigned to chance phenomena in the physical realm.
- It is meant to be inferred from past or observed trials and to be *stable* over trials.

- Is there room for limited precision in objective physical probability that parallels that found for subjective probability through the use of upper and lower previsions?
- Are unstable relative frequencies an indicator of this imprecision?

The Old Masters on Irregularity

- Leibniz to Bernoulli: The difficulty in it seems to me to be that contingent things or things that depend on infinitely many circumstances cannot be determined by finitely many results, for nature has its habits, following from the return of causes, but only for the most part.

Who is to say that the following result will not diverge somewhat from the law of all the preceding ones—because of the mutabilities of things? New diseases attack humankind. Therefore even if you have observed the results for any number of deaths, you have not therefore set limits on the nature of things so that they could not vary in the future.

- De Moivre: ...if we should suppose the Event not to happen according to any Law, but in a manner altogether desultory and uncertain; for then the Events would converge to no fixt [sic] Ratio at all.

- Poisson: But one should not lose sight of the fact that he [Bernoulli] supposes that the chances remain constant, while, to the contrary, the chances of physical phenomena and of moral matters almost always vary continually, without any regularity and often to a great extent.

- Venn: So in Probability, that uniformity which is found in the long run, and which presents so great a contrast to the individual disorder, though durable is not everlasting.

Keep on watching it long enough, and it will be found almost invariably to fluctuate, and in time may prove as utterly irreducible to rule, and therefore as incapable of prediction, as the individual cases themselves.

- These citations suggest the existence of fluctuations in the very long run.
- We can go beyond the usual case of terminal relative frequency *r_n* after n trials as being the only projectable quantity.
- Measure of the range of recent (from trial m to trial n) fluctuations in relative frequencies?
- Leads us to *comparative probability orderings* and to *upper and lower probabilities* establishing probability intervals.

Frequentist Comparative Probability

- Consider speech, natural language text, usenet postings, economic data such as stock prices, social data on attitudes or mobility, and the weather.
- Huge quantities of available objective data.
- Frequentist concept should apply
- Do all such applications admit of accurate numerical probability models?

Comparative Probability

- $A \succsim B$ is read as "event A is as least as probable as event B".
- Axiomatized in the early 1930s by Bruno de Finetti for subjective probability.
- Studied in the 1970s by my students (especially Michael Kaplan) and me.

Order and Frequency

- Let $r_n(A)$ denote the relative frequency of occurrence of event A in n repeated experiments.
- Let $A \succsim B$ denote "event A is at least as probable as event B."
- Relate the two through,

$$A \succsim B \iff r_n(A) \geq r_n(B).$$

- Inductively projecting only the information on *comparisons* of relative frequencies makes for greater stability than projecting the numerical values themselves when they are sufficiently unequal.
- This correspondence implies the following properties of the comparative probability relation that are then taken to be defining.

- $\succsim$ is a complete order on an algebra $\mathcal{A}$ of events.
- False $\emptyset \succsim \Omega$.
- For all $A \in \mathcal{A}$, $A \succsim \emptyset$.
-

$$A \succsim B \iff$$

$$A - AB \succsim B - AB.$$

CP Orders and Probability

- The CP order $\succsim$ is ***additive*** if there exists a probability measure **P** (generally, not unique) and

$$\mathbf{A} \succsim \mathbf{B} \iff \mathbf{P}(\mathbf{A}) \geq \mathbf{P}(\mathbf{B}).$$

- $\succsim$ is ***almost additive*** if there exists P and

$$\mathbf{A} \succsim \mathbf{B} \Rightarrow \mathbf{P}(\mathbf{A}) \geq \mathbf{P}(\mathbf{B}).$$

- $\succsim$ is ***weakly additive*** if there exists P and

$$\mathbf{P}(\mathbf{A}) \geq \mathbf{P}(\mathbf{B}) \Rightarrow \mathbf{A} \succsim \mathbf{B}.$$

- $\succsim$ is (strictly) ***nonadditive*** if none of the above hold.

- CP-based conditions determining additivity, almost additivity, and nonadditivity were developed by Michael Kaplan.
- The additive CP orders are the only ones that admit of a joint order of ***independent type*** for any number of repetitions of a given CP order.
- These conditions underline the nontriviality of assuming the existence of joint experiments—in this case, of joint CP orders.

- Probability order relations can provide less precise summations of frequentist data that more accurately reflect the extent of instability in this data.
- We now show how frequentist data can yield probability order relations that are inconsistent with any probability measure.

Upper and Lower Probability

- **Relation to relative frequencies that fluctuate significantly.**

$$\bar{P}(A) = \max_{n_0 \le j \le n} r_j(A);$$

$$\underline{P}(A) = \min_{n_0 \le j \le n} r_j(A).$$

- **Asymptotic relative frequencies:**

$$\bar{P}(A) = \limsup_{n\to\infty} r_n(A);$$

$$\underline{P}(A) = \liminf_{n\to\infty} r_n(A).$$

- **With this definition, $\bar{P}$ and $\underline{P}$ always exist.**

- **Key new axioms of duality and sub- and superadditivity.**

(Duality) $\bar{P}(A) = 1 - \underline{P}(A^c)$.

(Subadditivity) $\bar{P}(A \cup B) \le \bar{P}(A) + \bar{P}(B)$.

(Superadditivity) $A \perp B \Rightarrow$
$\underline{P}(A \cup B) \ge \underline{P}(A) + \underline{P}(B)$.

- **Variety of ways that you can induce a comparative probability relation $\succsim$ from upper and/or lower probabilities.**

$$A \succsim B \iff \bar{P}(A) \ge \bar{P}(B);$$

$$A \succsim B \iff \underline{P}(A) \ge \underline{P}(B);$$

$$A \succsim B \iff \underline{P}(A) \ge \bar{P}(B).$$

- **The CP relations so induced will not generally obey the axioms given earlier for CP.**

Stationary Random Sequences

- Kumar, Grize, Papamarcou, and Sadrolhefazi developed lower and upper probability models for stationary random sequences of bounded random variables.
- The lower probability function is time shift invariant and therefore stationary.
- It is also monotonely continuous along convergent sequences of cylinder sets, the observable events.

- Particular attention was paid to the event of convergence of relative frequencies in this model.
- The *stationarity convergence* (*ergodic) theorem* of standard probability asserts that every stationary random process of finite mean random variables has time averages that converge almost surely (possibly to a non-degenerate random variable).

- The goal was to show that this was not necessarily true of lower probability models.
- That this was indeed the case demonstrated that the imposition of convergence in all cases by standard probability was too restrictive.
- In this sense, standard probability is insufficiently expressive and forces an unwarranted metaphysical commitment.

Background to Random Sequences

- Let $\mathcal{X} = \{X_i\}$ denote the set of doubly infinite sequence of random variables that are uniformly bounded.
- Let $\mathcal{C} = \{C\}$ be the set of cylinder sets in $\mathcal{X}$—those events whose outcomes are determined by finitely many of the random variables in the sequence.
- $\mathcal{D}_n$ is the subset of $\mathcal{C}$ of sets of span or diameter less than or equal to n.

- **$\mathcal{A}$ denotes the smallest σ-algebra containing all the cylinder sets.**
- **Let T denote the right-shift operator $(TX)_i = (X)_{i+1} = X_{i+1}$.**
- **The invariant sets are**

$$\mathcal{I} = \{A : TA = T^{-1}A = A\}.$$

- **The *tail events* $\mathcal{T}$ is the limit on n of the σ-algebras defined on $X_n, X_{n+1}, \ldots$ and includes convergence and divergence events.**

Stationarity

- **A set function ϕ is *(strictly) stationary* if**

$$(\forall A \in \mathcal{A})\ \phi(TA) = \phi(T^{-1}A) = \phi(A).$$

- **If P is a stationary probability measure on $\mathcal{A}$ then**

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(X_i) = Ef(X)|\mathcal{I}),\ \text{a.s.}$$

provided $Ef(X)$ is finite.

- **We will show that there exist stationary lower probabilities for which the preceding stationarity convergence theorem fails.**
- **If this is the case, then lower probabilities can generate models that cannot be accessed by standard probability, thereby proving it to be insufficiently expressive.**

LP Models Vacuous on Tail Events

- **Let $\mathcal{S}$ denote the set of lower probabilities that are stationary and monotonely continuous along $\mathcal{C}$.**
- **A lower probability $\underline{P}$ is vacuous on A if $\underline{P}(A) = 0, \bar{P}(A) = 1$.**
- **A theorem (5.8) by Sadrolhefazi asserts that given any lower probability $\underline{P}_0 \in \mathcal{S}$ and an integer $n \geq 1$ and $0 < \epsilon < 1$, there exists $\underline{P}_1 \in \mathcal{S}$ that is vacuous for all events in the tail algebra $\mathcal{T}$ and satisfies**

$$(\forall C \in \mathcal{D}_n)\ |\underline{P}_1(C) - \underline{P}_0(C)| \leq \epsilon,$$

and $|\bar{P}_1(C) - \bar{P}_0(C)| \leq \epsilon.$

- Given any lower probability $\underline{P}_0$ that is stationary and monotonely continuous on the cylinder sets, there exists a lower probability $\underline{P}_1$ that agrees with it, within any positive specified ϵ, on cylinder sets of span no more than n, yet $\underline{P}_1$ is vacuous or maximally noncommital on all tail events including those concerning convergence of time averages.
- $\underline{P}_0$ can be a standard stationary probability measure.

- Lower probability allows us to avoid assertions about what is, in principle, unobservable, while at the same time being able to mimic any other stationary and monotonely continuous lower probability on the fundamentally observable class of cylinder sets.
- Standard probability does not have this desirable option and must make specific commitments to unobservable events.

Uncertainty

- *Uncertainty* is faced in individual decision-making contexts ranging over career choices, choosing personal relationships, deciding when the ``experts" (e.g., doctors) disagree, car driving behavior, investments, etc.
- True of mice as well as men.
- The individual has developed over a lifetime much ``information" about each of these areas and little of this information can stated explicitly or stated truthfully when forced to make it explicit to satisfy an interrogator.

- Here we deal with *uncertainty*, *degrees of belief,* or *propositional attitudes.*
- Whatever the rational desirability (being *coherent* or *avoiding a Dutch book*) of assessing all of these kinds of uncertainty through real-valued assessments of standard probability, it strains credulity to assume that this is commonly possible.
- Look into your own heads on this one!

Dutch Book

- **A "horse race" with three horses H_1, H_2, H_3. Gambles offered are:**

$$G_1 = \begin{cases} 1 & \text{if } H_1 \text{ wins,} \\ -1 & \text{if } H_1 \text{ loses.} \end{cases}$$

$$G_2 = \begin{cases} 3 & \text{if } H_2 \text{ wins,} \\ -1 & \text{if } H_2 \text{ loses.} \end{cases}$$

$$G_3 = \begin{cases} 4 & \text{if } H_3 \text{ wins,} \\ -1 & \text{if } H_3 \text{ loses.} \end{cases}$$

- **Buy 100 of G_1, 50 of G_2, and 40 of G_3.**

- Peter Walley carefully and extensively developed a more realistic approach that uses a *set of probability measures* to represent Your state of knowledge.
- No one measure in this set is ``true".
- It is only the whole set that is a good description of your state of knowledge.
- This yields upper (selling) and lower (buying) prices for gambles or random variables and to upper and lower probabilities as a special case.

- **Set of measures $\mathcal{M} = \{\nu\}$**

$\bar{E}X = \sup\{E_\nu X : \nu \in \mathcal{M}\}$,

$\underline{E}X = \inf\{E_\nu X : \nu \in \mathcal{M}\}$.

- **If $\|\mathcal{M}\| = 1$ then**

$$\bar{E}X = \underline{E}X = EX.$$

- **Gamble X is preferred to Y if**

$$\underline{E}X > \bar{E}Y.$$

- **Defining properties of upper and lower expectation.**

Duality $\underline{E}X = -\bar{E}(-X)$;

Non-negativity $X \geq 0 \Rightarrow \bar{E}X \geq 0$;

Homogeneity $(\forall \lambda > 0)\ \bar{E}(\lambda X) = \lambda \bar{E}X$;

Sublinearity $\bar{E}(X + Y) \leq \bar{E}X + \bar{E}Y$.

- Is there an objective, frequentist counterpart to the sound set of measures characterization of uncertainty?
- An attempt to identify a counterpart is being made by Pablo Fierens, Leandro Rego, and myself under the label of ``chaotic probability".

Imprecision in Physical Probability

- Our motivation is the rational incorporation of limited precision in objective physical probability, as has been found to be appropriate for subjective probability.
- We seek a model for time series from irregular physical or socioeconomic phenomena.
- Such a view has also been identified in quantum mechanics.

A Game Between Two Agents

- We can think of the model $\mathcal{M} = \{\nu\}$ as a partial description of an Agent 1.
- At a given time i, Agent 1 chooses a measure ν_i on $\mathcal{X}$ and an observation $X_i = x_i$ is determined as the outcome of a random experiment described by ν_i.

- Agent 1's choice at time i can depend upon the past sequences of observations x^{i-1}.
- (We need to extend this to include the past sequence of chosen measures $\nu^{i-1} = (\nu_1, \ldots, \nu_{i-1})$.)
- Agent 1 might be "nature".

- Agent 2 at time $i-1$ knows x^{i-1} and $\mathcal{X}$, but not the sequence of measures ν^{i-1}.
- Consequences for Agent 2 depend upon the range $\mathcal{M}$ of actions available to Agent 1.
- We assume that the consequences depend upon the measures in $\mathcal{M}$ in a continuous fashion with respect to a metric on measures to be introduced later.

- The consequences to Agent 2 do not depend upon the observed outcomes $\{x_i\}$. These may only be informative about the sequence of measures.
- Agent 2, thus, has an interest in inferring $\mathcal{M}$, the set of measures from which Agent 1 is making his choices.

The Stochastic Process Model

- The standard probability of a given sequence $X^n = x^n \in \mathcal{X}^*$, for $\mathcal{X}$ finite, is given by

$$P(X^n = x^n) = \prod_{i=1}^{n} P(X_i = x_i | X^{i-1} = x^{i-1}),$$

where we define x^0 to be the empty string.
- We will refer to P as the ***process measure.***

- Given the observations x^n, we require that for each $n \geq i \geq 1$, $\mathcal{M}$ contains the measure

$$(\forall x \in \mathcal{X})\ \nu_i(x) = P(X_i = x | X^{i-1} = x^{i-1}).$$

The process measure can be written as

$$P(X^n = x^n) = \prod_{i=1}^{n} \nu_i(x_i),$$

- $\mathcal{M}^*$ denotes the set of possible process measures.

- **If we define $\mathcal{P}$ as the set of all probability measures over the algebra of all subsets of $\mathcal{X}$, then**

$$\mathcal{M} \subseteq \mathcal{P}.$$

- **Define the *measure selection function***

$F : \mathcal{X}^* \to \mathcal{M} \subseteq \mathcal{P}$ **and define**
$\nu_i = F(x^{i-1})$,
$\nu_i(x_i) = P(X_i = x_i | X^{i-1} = x^{i-1})$.

- **F provides a behavioral description of Agent 1.**

- **If F is too *simple*, we can infer it from long enough x^n with high process measure probability P.**

- **In this eventuality, we have an estimable standard stochastic process.**

- **If the selection function F is *very complex*, say, random, with selections made in an $i.i.d.$ manner according to some (prior) distribution on $\mathcal{M}$, we would not be able to distinguish whether x^n was produced by an $i.i.d.$ process according to some measure in $ch(\mathcal{M})$, the convex hull of $\mathcal{M}$, or by the chaotic probability model in question.**

- **The interest in *chaotic probability* resides in the little-explored *intermediate realm* where we cannot estimate F but can nonetheless reliably estimate $\mathcal{M}$.**

- **We deviate from the usual random process thinking by allowing the sequence of choices $\nu_1, \ldots, \nu_n$ to have a *computational complexity that grows with* n, when properly defined.**

- We do not expect that Agent 1 makes his choices through a computable function of sequences in $\mathcal{X}^*$.
- While noncomputability is not ruled out by the theory of random processes, what is unusual in our approach is that we care most about the case where the process measure P is ***not effectively computable.***

Indeterminacy

- A long-studied but less well understood notion relates to *indeterminacy*---the given facts or *evidence* (expressed in a formal language) only partially *deductively* support the *conclusion* or *hypothesis*.
- Numerical assessments of partial degrees of support were first proposed by Leibniz.

- This is the area of legal trials (Leibniz' concern) but more generally that of *inductive reasoning or inference,* including statistical inference.
- evidence: Of 500 observed birds of various colors, 25 are swans, and all swans were white.
- hypothesis: all swans are white.

- There is no logical deduction from the evidence *e* to the hypothesis *h*.
- Motivated by early work of Keynes, we let *h|e* represent the *degree of inductive support* lent to *h* by *e*.

- The "standard approach" is to assume that there exists a conditional probability P and

$$h|e \succsim h'|e' \iff P(h|e) \geq P(h'|e').$$

- Perhaps this conditional probability comes from an argument such as the one made by Solomonoff in the early 1960s when he derived a universal semicomputable prior.
- Assigning any numerical value, of whatever origin, to the support lent by evidence to hypothesis will automatically make accessible a complete ordering according to this support.

- Inductive relations have an inherent indefiniteness or imprecision that may not be expressible by standard numerical probability.
- We hold that standard probability and upper and lower probability are too expressive to model inductive support.
- Nor need all pairs *h*|e and *h'*|e' be comparable.

- We assume only the existence of a ***partial order*** $h|e \succsim h'|e'$ that is read "the inductive support lent by e to h is at least as great as that lent by e' to h'".

Summary

- We surveyed meanings of probability associated with applications in the domains of physically-determined, chance, uncertain, and indeterminate phenomena.
- For each such meaning or interpretation we offered alternative mathematical models that had possible advantages over the standard model of a measure.