In the Realm of Probability: Limitations of Standard Probability

Plenary Talk for ISIPTA07

(18 July 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)

To rephrase Mark Antony,
``I come to confine standard probability, not to condemn it."

My Debts to My Students (Comparative Probability)

- Michael Kaplan
- Peter Walley
- Marco Wolfenson
- Anurag Kumar

(Interval-Valued Probability)

- Peter Walley
- Anurag Kumar
- Yves-Laurent Grize
- Adrian Papamarcou
- Amir Sadrolhefazi

(Set-valued Probability)

- Pablo Fierens
- Leandro Rego

The `I' in SIPTA

- `Imprecise' is a negative word, suggesting a defect as compared to `precise'.
- In fact, SIPTA is the most concerned of the statistical societies with matching the precision of a mathematical model of probability to the accuracy inherent in a probability concept.

- If it were just about precision, then even the real numbers are imprecise compared to the nonstandard reals.

- The true issue is to match precision to the accuracy inherent in the concept at hand.

- This is a central concern to SIPTA and our focus this afternoon.

Overview

- We are motivated by some general views about applied mathematics and
- by their specific implications for mathematical models of a variety of concepts of probability.

Thoughts on Applied Mathematics

- Applied mathematics starts from applications and moves to mathematical models that are then worked out to provide understanding of, and working techniques for, these applications.
- A mathematical model lies in a formal domain, unlike most applications.
- There is a potentially treacherous shift of categories.

- Theory of measurement as developed by Suppes, Luce, and others views the application as being in an *empirical domain* and the model in a *formal or mathematical domain.*

- Important *relations* in the empirical domain are homomorphically mapped into convenient relations in the mathematical domain.

- The *expressiveness* of a model is its ability to host relations that preserve the significant distinctions in the application domain. A model may be *insufficiently* expressive and unable to host all of the empirical relations.

- We also desire that a model not be so expressive that it offers many distinctions that point only to phantom distinctions in the empirical domain. In such a case, the model is *excessively* expressive.

- We do not wish to be pedantic about this rendering.
- Distinctions should be ones of significance.
- Some excess expressiveness will have to be tolerated.
- It is not about a glove fitting a hand.

Make everything as simple as possible, but not simpler. (Albert Einstein)

- Nor should the goal be an overarching mathematical theory of probability that accommodates to all probability concepts.
- Excessive generality comes at the expense of saying very little of specific interest.
- We need mathematical theories closely attuned to individual probability concepts.

Accuracy and Precision

- In engineering and scientific terms, we talk about *accuracy* and *precision.*
- We might have an ``application'' or concept of, say, `expertise in probability theory' and be able to compare some pairs of individuals as to who is more expert than whom.
- We surely have no understanding that enables us to meaningfully compare all pairs.

- If we attach, say, integers to each individual to represent expertise, with a larger integer representing greater expertise, then we have an unrealistic total ordering.
- Furthermore, what are we to make of one level of expertise having twice the rating of another?
- The precision or expressiveness of the representation exceeds the inherent accuracy of the expertise concept.

Mathematical Caution

- Mapping a non-formal system into a formal system comes at a necessarily significant cost in translation
- We are establishing a correspondence between different categories of objects
- Mathematics has its roots in empirical phenomena but follows its own creative path

John von Neumann

- ...mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science.

But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration.

At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up.

In any event, whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the re-injection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future.

Focus on Probability

- We will apply the foregoing general considerations to a critical examination of the suitability of standard probability, that due either to Kolmogorov or to de Finetti, to a variety of examples of different concepts of probability.
- We will find the standard model can be either insufficiently or excessively expressive.

Illustrative Probability Concepts

- **Physically-determined probability in quantum mechanics.**
- The accuracy of finite relative frequency.
- Relative frequency and CP in the long run.
- Asymptotics and LLNs for relative frequency.
- Stationary random sequences.
- Epistemic probability for inductive support.
- Subjective probability.

Physically-Determined Probability

- Probability determined from physical properties, not from other probabilities
- Classical probability---approximate symmetry, classical mechanics, and sensitive dependence on initial conditions
- Statistical mechanics, e.g., Boltzmann's law
- Quantum mechanics

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

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 events corresponds to closed linear subspaces 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 Modularity

- If $\hat{A}$ and $\hat{B}$ do not commute and A_a and B_b are sets of possible measured values, then

$$\mathbf{P}(\mathbf{A_a} \cup \mathbf{B_b}) \neq \mathbf{P}(\mathbf{A_a}) + \mathbf{P}(\mathbf{B_b}) - \mathbf{P}(\mathbf{A_a} \cap \mathbf{B_b}).$$

- Hence, the Kolmogorov formulation of standard probability is inapplicable to quantum mechanics.
- This is a failure of the expressivity of standard probability.

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.
- Distributivity of union and intersection fails.
- A consequent failure is that QM probability does not obey the formula for the probability of a non-disjoint union of two events.

Physicists & Finite Frequencies

- Physicists interpret the theory-determined probabilities and expected values produced by QM through sample averages of finitely many repeated, unlinked experiments.
- They are satisfied with the variance of such estimates converging to zero at rate 1/n.

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.

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- **The accuracy of finite relative frequency.**
- Relative frequency and CP in the long run.
- Asymptotics and LLNs for relative frequency.
- Stationary random sequences.
- Epistemic probability for inductive support.
- Subjective probability.

Finite Relative Frequency

- For relative frequency r_n(A) based on observations, a change of 1/n suffices to render distinguishable relative frequencies. Precision of relative frequency is to within 1/n changes.

- However, the relative frequency is intended as an estimator of P(A) and the accuracy is only order of square-root of n.

Accuracy

- **Hoeffding's and Chebychev's inequalities demonstrate that accuracy is only to within $\sqrt{\mathbf{c}/\mathbf{n}}$.**
- **To match precision to accuracy, introduce the family of intervals**

$$\mathbf{I_k} = [\mathbf{k}\epsilon, (\mathbf{k}{+}\mathbf{3})\epsilon] \text{ for } \mathbf{k} = \mathbf{0}, \ldots, \frac{\mathbf{1}}{\epsilon}-\mathbf{2}.$$

- **For simplicity, we ignore "end effects" at the boundaries of the unit interval $[0, 1]$.**
- **Map a relative frequency**

$$\mathbf{r_n}(\mathbf{A}) \to \mathbf{I}_\rho,$$

where $(\rho + \mathbf{1})\epsilon \leq \mathbf{r_n}(\mathbf{A}) \leq (\rho + \mathbf{2})\epsilon.$

- It follows that

$$|\mathbf{r_n(A) - P(A)}| \leq \epsilon$$
$$\iff \mathbf{r_n(A)}-\epsilon \leq \mathbf{P(A)} \leq \mathbf{r_n(A)}+\epsilon$$
$$\Rightarrow \mathbf{P(A)} \in \mathbf{I}_\rho.$$

- Letting $\epsilon = \sqrt{\mathbf{c/n}}$,

$$\mathbf{1 - 2e^{-c}} \leq \mathbf{P}(|\mathbf{r_n(A) - P(A)}| \leq \sqrt{\mathbf{c/n}})$$
$$\leq \mathbf{P(P(A)} \in \mathbf{I}_\rho).$$

- **Values of relative frequencies that share the same subintervals of width $\sqrt{\mathbf{c/n}}$ can nearly equally accurately estimate the unknown $\mathbf{P(A)}$ in the sense provided by confidence intervals.**

- **This establishes that we have matched the precision of our measurement of relative frequencies to the accuracy of our resulting inference about the probability of an event.**
- **While there are $n+1$ possible different values of $r_n(A)$, there are no more than $\sqrt{n/c}$ distinct subintervals quantizing these relative frequency values.**

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- The accuracy of finite relative frequency.
- **Relative frequency and CP in the long run.**
- Asymptotics and LLNs for relative frequency.
- Stationary random sequences.
- Epistemic probability for inductive support.
- Subjective probability.

John Venn on the Long Run

- ...view of Probability adopted in this Essay. ...which regards it as taking cognisance of laws of things and not of the laws of our own minds in thinking about things,

- ...That the science of Probability, on this view of it, contains something more important than the results of a system of mathematical assumptions, is obvious.

- ...the fundamental conception which the reader has to fix in his mind as clearly as possible, is,...a series of a peculiar kind, one of which no better compendious description can be given than that which is contained in the statement that it combines individual irregularity with aggregate regularity. This is a statement which will probably need some explanation.

- 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.

- ---a few instances are not sufficient to display a law at all; a considerable number will suffice to display it; but it takes a very great number to establish that a change is taking place in the law.

Frequentist Comparative Probability

- Can the regularities in somewhat unstable long-run relative frequencies be modeled by comparative probability (CP) orders?
- Might the reduced precision of CP orders better match the reduced accuracy of unstable frequentist sources?
- Are there CP orders with no relation to standard probability?

CP Axioms

- $\mathbf{A} \succsim \mathbf{B}$ is read "A is at least as probable as B"
- de Finetti axiomatized CP on an algebra $\mathcal{A}$ of subsets of $\boldsymbol{\Omega}$
- CP1.(order) $\succsim$ is a total on order on $\mathcal{A}$.
- CP2.(nontrivial) $\boldsymbol{\Omega} \succ \emptyset$.
- CP3.(positive) $(\forall \mathbf{A} \in \mathcal{A})\ \mathbf{A} \succsim \emptyset$.
- CP4.(cancellation)

$$\mathbf{A} \succsim \mathbf{B} \iff \mathbf{A} - \mathbf{B} \succsim \mathbf{B} - \mathbf{A}.$$

Implications include:

(a) $\mathbf{A} \supset \mathbf{B} \Rightarrow \mathbf{A} \succsim \mathbf{B}$;

(b) $\mathbf{A} \succsim \mathbf{B} \iff \mathbf{B^c} \succsim \mathbf{A^c}$;

(c) $\mathbf{A} \succsim \mathbf{B}, \mathbf{C} \succsim \mathbf{D}, \mathbf{A} \perp \mathbf{C} \Rightarrow \mathbf{A} \cup \mathbf{C} \succsim \mathbf{B} \cup \mathbf{D}$

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.**

Partial CP Orders

- For convenience, we assume antisymmetric CP orders (no equivalent distinct events).
- A partial CP order is transitive, irreflexive, and satisfies CP2, CP3, CP4, and consequence (a) of monotonicity with respect to set inclusion.

Finite Frequentist CP

- Choose a minimal sample size n_0.
- As a first of four alternatives, define

$$\mathbf{A} \succ_1 \mathbf{B} \iff (\forall \mathbf{n_0} \leq \mathbf{j} \leq \mathbf{n})\, \mathbf{r_j}(\mathbf{A}) > \mathbf{r_j}(\mathbf{B}).$$

- We do not require "convergence" of relative frequencies.
- If the relative frequencies are converging to a measure P, then the CP order will have an agreeing representation

$$\mathbf{A} \succ_1 \mathbf{B} \iff \mathbf{P}(\mathbf{A}) > \mathbf{P}(\mathbf{B}).$$

- **There exists frequentist long run data for which $\succ_1$ is only a partial order.**

- **For the remaining alternatives, define**

$$\underline{\mathbf{P}}(\mathbf{A}) = \min_{\mathbf{n_0 \leq j \leq n}} \mathbf{r_j}(\mathbf{A}) \textbf{ and}$$

$$\mathbf{A} \succ_{\mathbf{2}} \mathbf{B} \iff \underline{\mathbf{P}}(\mathbf{A}) > \underline{\mathbf{P}}(\mathbf{B}).$$

$$\bar{\mathbf{P}}(\mathbf{A}) = \max_{\mathbf{n_0 \leq j \leq n}} \mathbf{r_j}(\mathbf{A}) \textbf{ and}$$

$$\mathbf{A} \succ_{\mathbf{3}} \mathbf{B} \iff \bar{\mathbf{P}}(\mathbf{A}) > \bar{\mathbf{P}}(\mathbf{B}).$$

$$\mathbf{A} \succ_{\mathbf{4}} \mathbf{B} \iff \underline{\mathbf{P}}(\mathbf{A}) > \bar{\mathbf{P}}(\mathbf{B}).$$

- **In the absence of equivalences, $\succ_{\mathbf{2}}, \succ_{\mathbf{3}}$ are total orders. However, these orders need not satisfy CP4 nor the consequence (b) of complementation reversing ordering.**

- $\succ_4$ is typically only a partial order.
- $\succ_4$ satisfies the complementation consequence (b)
- When $\succ_4$ is a total order, then it is additive.
- $\succ_2, \succ_3$ can be nonadditive CP orders.
- In this case, additive numerical probability is insufficiently expressive.
- When any of these orderings are additive, then numerical probability is excessively expressive as there need not be a unique representation.

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- The accuracy of finite relative frequency.
- Relative frequency and CP in the long run.
- **Asymptotics and LLNs for relative frequency.**
- Stationary random sequences.
- Epistemic probability for inductive support.
- Subjective probability.

Asymptotics: Laws of Large Numbers

- Mises postulated convergence of frequencies in his collective as necessary to a physically meaningful account of probability.
- He also imposed a randomness condition of invariance under ``place selection”.
- The assumption of convergence in the long run is a claim about the world that can never be verified or contradicted.

- ``The essentially new idea...was to consider probability as a science of the same order as geometry or theoretical mechanics..."
- ``...probability theory deals with mass phenomena and repetitive events."

- Jeffrey rebuts Mises by likening Mises limit of the hypothetical convergent infinite sequence to a measurement of the mass of an hypothetical tenth planet.

- Not only are the sequences hypothetical, but so is their postulated limit.

The Anthropic Principle

- A possible argument for approximate long-run stability.
- The eminent physicist Steven Weinberg defines this controversial notion as follows.
- ``Briefly stated, the anthropic principle has it that the world is the way it is, at least in part, because otherwise there would be no one to ask why it is the way it is. There are a number of different versions of this principle...”

- Can we use the anthropic principle in the form of a *universal conditioning event U?*
- This event is not a true event to which we can assign probability.
- It suggests that the long run phenomena of interest to us are likely ones we have co-existed with. They will have some form of long-range approximately stable behavior.
- A ``fact" that philosophy cannot justify.

LLNs without Convergence

- As observed by Venn and Jeffrey, the assumption of convergence in the long run is metaphysics and not even a plausible hypothesis.
- Accepting the hypothetical long-run frequentist data, we eliminate the assumption of a limiting relative frequency for ***P(A)*.**

Lower and Upper Probability

- Let $\mathbf{r_n}(\mathbf{A})$ denote the relative frequency with which event A is observed in the outcomes x^n of the first n repeated experiments.
- As in the discussion of the long run, define lower probability $\underline{\mathbf{P}}$ and upper probability $\bar{\mathbf{P}}$ through

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

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

- $\underline{\mathbf{P}}$ and $\bar{\mathbf{P}}$ will exist for all possible (hypothetical) sequences of outcomes x^∞.

- **The following are properties of lower and upper probability:**

(*a*) $1 \geq \bar{\mathbf{P}}(\mathbf{A}) \geq \underline{\mathbf{P}}(\mathbf{A}) \geq \mathbf{0}$;

(*b*) $\bar{\mathbf{P}}(\mathbf{A^c}) = \mathbf{1} - \underline{\mathbf{P}}(\mathbf{A})$;

(*c*) $\mathbf{A} \supset \mathbf{B} \Rightarrow \underline{\mathbf{P}}(\mathbf{A}) \geq \underline{\mathbf{P}}(\mathbf{B})$
and $\bar{\mathbf{P}}(\mathbf{A}) \geq \bar{\mathbf{P}}(\mathbf{B})$

(*d*) $(\forall \mathbf{A} \perp \mathbf{B})\ \underline{\mathbf{P}}(\mathbf{A} \cup \mathbf{B}) \geq \underline{\mathbf{P}}(\mathbf{A}) + \underline{\mathbf{P}}(\mathbf{B})$

(*e*) $\bar{\mathbf{P}}(\mathbf{A} \cup \mathbf{B}) \leq \bar{\mathbf{P}}(\mathbf{A}) + \bar{\mathbf{P}}(\mathbf{B})$.

- **Walley and Fine showed that $\underline{\mathbf{P}}$ is the *lower envelope* of the class $\mathcal{M}_{x^\infty}$ of all limits of pointwise convergent subsequences of relative frequencies of events calculated along the given sequence x^∞:**

$$\underline{\mathbf{P}}(\mathbf{A}) = \inf\{\mu(A) : \mu \in \mathcal{M}_{x^\infty}\}$$
$$\bar{\mathbf{P}}(\mathbf{A}) = \sup\{\mu(A) : \mu \in \mathcal{M}_{x^\infty}\}.$$

- **In addition, the class of $\underline{P}$ generated as above as we range over all possible x^∞ is precisely the class of all lower envelopes on the given event algebra $\mathcal{A}$.**

- **When the relative frequencies converge then**

$$\underline{\mathbf{P}}(\mathbf{A}) = \bar{\mathbf{P}}(\mathbf{A}) = \mathbf{P}(\mathbf{A}) = \lim_{\mathbf{n}\to\infty} \mathbf{r_n}(\mathbf{A}).$$

- **The room for difference is found in how we deal with the infinite data sequence x^∞ when relative frequencies do not converge.**
- **We can use this data in all circumstances by keeping track of the structure of persistent fluctuations and not just the terminal value of relative frequency.**

Sets of Measures

- If there is a known or learnable time variation in choice of measures on each trial, then you would not follow our suggestions.
- Fine, Fierens, Rego have introduced a ``sets of measures" model that allows for a fairly arbitrary choice of measure on each trial.
- Results obtained include conditions under which the set of measures being used can be estimated reliably from long finite data sequences.

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- The accuracy of finite relative frequency.
- Relative frequency and CP in the long run.
- Asymptotics and LLNs for relative frequency.
- **Stationary random sequences.**
- Epistemic probability for inductive support.
- Subjective probability.

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 theorem* of standard probability asserts that every stationary random process of bounded 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.

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.**

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- The accuracy of finite relative frequency.
- Relative frequency and CP in the long run.
- Asymptotics and LLNs for relative frequency.
- Stationary random sequences
- **Epistemic probability for inductive support.**
- Subjective probability.

Inductive Support: Epistemic Probability

- Kyburg and Levi are well-known for their efforts to develop theories of inductive support.
- They both rely on upper and lower probabilities to reflect their understanding that epistemic probability need not be as precisely defined as standard probability.
- We believe that uppers and lowers are too precise.

- Start with information expressed in a possibly natural language *L*.
- Translate or encode evidence and hypothesis statements in *L* into a formal language.
- Evidence is postulated to be known.
- An hypothesis is a statement of interest that we wish to partially justify on the basis of the evidence.

- Induction requires us to determine some form of expression of this justification or of the support lent by evidence to an hypothesis.

- Expressions of inductive support need not be part of our formal language.

- We hold that standard probability and upper and lower probability are too expressive to model inductive support.
- We seek instead a *partial ordering* of the support $h|e$ lent to h by e.
- We follow the lines of algorithmic or Kolmogorov complexity.
- Our account is a highly tentative one!

Syntax

- Our syntax is that of finite-length strings from a finite alphabet (that can be taken to be binary).

- This syntax permits concatenation of strings to form new strings.

- There is no Boolean logic associated with our syntax.

Semantics

- Strings are encodings or translations of expressions in another language *L*.
- *L* has a semantics perhaps including objective data, such as relative frequencies in Kyburg's epistemological probability.
- The semantic relations of *L* are no longer accessible to us.

- In the absence of an appropriate semantics, we cannot identify truth-preserving operations such as the usual rules of deductive logic.

- Our semantics is based instead upon the *evaluation* of a string p as producing another string q through a selected universal Turing machine (UTM) computation
$T(p)=q$.

- The choice of encoding and UTM seem arbitrary.
- We desire our eventual measures of support to be robust with respect to computable choices of translations from *L* and choices of the evaluation mechanism T.

Explanations

- A string p is an *explanation* of a string q if T(p)=q
- Given any evidence string e and hypothesis string h, there exist infinitely many strings p such that
h=T(pe)
- This is a consequence of T being a UTM.

The Explanatory Support Set

- The explanatory support set ***P***(h|e) is the set of all supplements p to the evidence e such that their concatenation is an explanation for h: T(pe)=h.
- Use of ***P***(h|e) follows the Epicurean prescription of keeping all arguments in favor of a conclusion, not just the ``best'' one.

- While ***P*** is an infinite set, we plausibly make it finite by limiting it to supplementary explanations p that are not much longer than those needed to generate h while ignoring the evidence e.

- More precisely, for a UTM T, there is a constant ***c*** such that for any evidence e and any h there exists a string h^*, a function only of h, whose length $|h^*| < \boldsymbol{c} + |h| + 2\log(|h|)$, and $T(h^*e) = h$.

- We omit supplementary explanations that are longer than a direct description.

Comparing $\boldsymbol{P}(h|e)$ to $\boldsymbol{P}(h'|e')$

- Motivated by Keynes, we compare the inductive support h|e, lent to h by e, with h'|e', lent to h' by e'.
- Given our minimal semantics and syntax, we ignore the ``content" of the strings and focus only on the *lengths* of strings in $\boldsymbol{P}(h|e)$ and $\boldsymbol{P}(h'|e')$.

Summarizing an Explanatory Support Set

- **Introduce the unnormalized cumulative distribution function**

$$\mathbf{F}_{\mathbf{h}|\mathbf{e}}(\mathbf{z}) = \|\{|\mathbf{p}| : \mathbf{p} \in \mathcal{P}(\mathbf{h}|\mathbf{e}) \text{ and } |\mathbf{p}| \leq \mathbf{z}\}\|.$$

- **The function $\mathbf{F}_{\mathbf{h}|\mathbf{e}}$ contains all the information relevant to induction.**

- $F_{h|e}(z)$ is nondecreasing in z, is zero for negative z and reaches a maximum for $z \geq |h^*|$.
- $F_{h|e}(z)$ is the number of explanations for h given e that have lengths no greater than z.
- A comparison of the inductive supports $h|e$ and $h'|e'$ is then a comparison between the two corresponding cumulative distribution functions $F_{h_1|e_1}$, $F_{h_2|e_2}$.

Partially Ordering Inductive Support

- **Occam's Razor suggests favoring shorter strings over longer ones as explanations.**
- **We recall the idea of *stochastic dominance* $\mathbf{X} \succ \mathbf{Y}$ of the random variable $\mathbf{Y}$ by the random variable $\mathbf{X}$ that is determined by their corresponding (normalized) cdfs through**

$$\mathbf{X} \succ \mathbf{Y} \iff (\forall \mathbf{z} \in \mathbb{R})\, \mathbf{F}_{\mathbf{X}}(\mathbf{z}) \leq \mathbf{F}_{\mathbf{Y}}(\mathbf{z}).$$

- **The binary relation $\succ$ is transitive but provides only a partial ordering between pairs of random variables.**

- **Partially order pairs $\mathbf{h_i|e_i}$ through**

$$\mathbf{h_1|e_1} \succ_\tau \mathbf{h_2|e_2} \iff (\forall \tau \leq \mathbf{z} \leq \min(|\mathbf{h_1^*}|, |\mathbf{h_2^*}|) \; \mathbf{F_{h_1|e_1}}(\mathbf{z}) \leq \mathbf{F_{h_2|e_2}}(\mathbf{z}).$$

- **It is evident that the partial ordering between pairs $\{\mathbf{h_i|e_i}\}$ is not one induced by some conditional probability $\mathbf{P}(\mathbf{h_i|e_i})$.**
- **Such a conditional probability would induce a total ordering.**

- **The expressiveness of $\succ_{\mathcal{T}}$ as a representation of strength of inductive support, is far short of that of a numerical representation.**
- **However, remarks by Keynes and difficulties encountered by others suggest that this might be a step in the right direction.**
- **There is little *a priori* or intuitive reason to think that inductive support can be identified between any pair of hypotheses given their individual supporting evidence.**

Robustness

- **Given the arbitrariness of translation of a statement $s \in \mathcal{L}$ in our base language into our syntax via an encoding E, and the arbitrariness in our choice of semantics through a choice of UTM T, we need to examine how robust the comparisons are to specific encodings of whatever our original information might be into our string syntax and to changes of UTM.**
- **This is a work in progress.**

Illustrative Probability Concepts

- Physically-determined probability in quantum mechanics.
- The accuracy of finite relative frequency.
- Relative frequency and CP in the long run.
- Asymptotics and LLNs for relative frequency.
- Stationary random sequences.
- Epistemic probability for inductive support.
- **Subjective probability.**

Subjective Probability

- There is no doubt that reasoning about uncertainty is an activity critical to our survival and well-being.
- And this includes its importance to much of the animal kingdom as well.
- Such skills are critical to the survival of individuals and of species.

- There is also no doubt that none of this reasoning relies upon mathematics---unless you want to give mice a lot of credit.

- Subjective probability is in peril when it addresses individual decision-making under uncertainty without serious reference to the actual mental capacities of individuals.

- At an extreme, we can ask for clairvoyance.

- Perhaps mathematics can assist us and free us from forms of error---from the `paradoxes' of decision theory.

- However, I doubt that this mathematics is that of standard or finitely additive probability.

- Coherence or Dutch book arguments that force numerical probability are compelling only in contrived circumstances.
- Such arguments invariably require you to extend your preferences from the gambles you care about to a full set of combinations you had no need to consider.

- There is a need for far more insight obtained by psychologists into how we reason, largely unconsciously, about situations of risk and opportunity when chance plays a role.

- We have prematurely adopted mathematical models before we understood sufficiently what it was we were modeling.

- The lack of sufficient understanding of reasoning under uncertainty has licensed a reign governed by mathematical taste and convenience.

- The company of mathematics is seductive to the adept, as von Neumann warned us.

- My doubts extend as well to upper and lower previsions---to buying and selling prices offered to all comers in all combinations.

- Having said this, I also think that Walley's *Statistical Reasoning with Imprecise Probability* is a masterpiece and one of the most important and innovative treatments of statistical thinking of the past thirty years.

In Summation

- Applications, and their thoughtful consideration, come first!
- They are the source of a number of well-entrenched distinct concepts of ``probability” for random, chance, and uncertain phenomena that are still evolving.
- It is essential to match model precision to the inherent accuracy of the concept.

- Moving from either the natural or subjective worlds into the formal mathematical world perforce requires some degree of distortion.
- Mathematics can produce invaluable development of our conceptual understanding.
- However, mathematics has its own dynamic which can (mis)lead us far from our entry point.

- Thus arise issues of precision/imprecision, accuracy, or of expressiveness:
- issues of how faithfully will selected models represent key aspects of probability in its natural and subjective realms;
- and of the extent to which these models introduce misleading distinctions pointing to non-existent real-world phenomena.

- We have focussed on the expressiveness failures of standard mathematical probability.
- These failures were identified in quantum mechanics, in the use of finite relative frequencies, moderately unstable frequentist data, laws of large numbers, stationary processes, and in a basic look at inductive support.

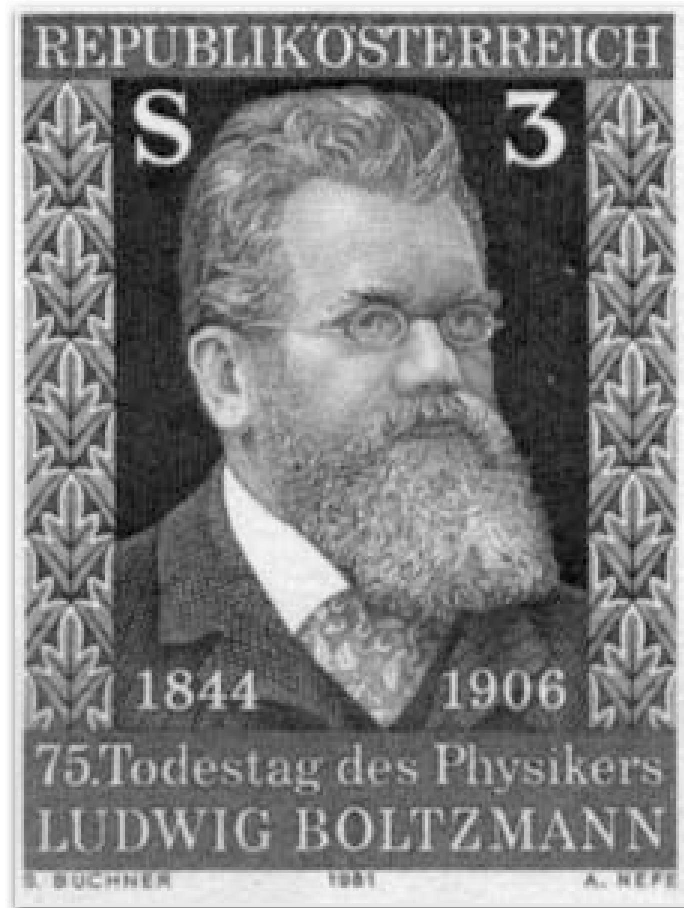
Availability of the Slides

- If you send a request to me at
  tlfine@ece.cornell.edu
  I will send you a copy of the slides for this talk together with a number of references.

Thanks!

- Again, I regret not being able to join you in person for the full round of professional and personal interactions that make ISIPTA meetings so valuable.
- My thanks to the Steering Committee for proposing and making this `high tech' talk possible,
- and I particularly thank our host Jirina who patiently arranged all this!

- Boltzmann was the first to express a physical law in terms of probabilities
- Maxwell-Boltzmann statistics for classical particles



- **Let F denote an external force acting on the particles (e.g., gravity).**
- **Let m be the common mass of an individual particle.**
- **The density f satisfies Boltzmann's collisionless equation**

$$\frac{\partial \mathbf{f}}{\partial \mathbf{t}} + \frac{\mathbf{p}}{\mathbf{m}}\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \mathbf{F}\frac{\partial \mathbf{f}}{\partial \mathbf{p}} = \mathbf{0}.$$

My General Intellectual Debts

- Andrei Kolmogorov for standard probability and the complexity approach
- Jimmie Savage for his deep insight into personalist probability and wide-ranging scholarship
- Glen Shafer for ``A Mathematical Theory of Evidence" and ongoing critical and creative historical studies
- Patrick Suppes for measurement theory and for bringing to bear a deep understanding of logic and philosophy of science on a very wide range of inquiry

- Henry Kyburg and Isaac Levi for conversations and their sustained examination of epistemic/epistemological probability using upper and lower probability

- Peter Walley for ``Statistical Reasoning...” and some years of collaborations.

- Ming Li and Paul Vitanyi for ``Kolmogorov Complexity Theory”

Hoeffding Bound

- **Let X_i be the $\{0,1\}$-valued random variable that is 1 if A occurs and 0 otherwise.**
- **The finite relative frequency estimator is defined by**

$$\mathbf{r_n(A) = \frac{1}{n}\sum_{i=1}^{n} X_i.}$$

- **The Hoeffding inequality provides us with a *confidence interval* $[\mathbf{r_n(A)} - \epsilon, \mathbf{r_n(A)} + \epsilon]$ that, in advance of being observed, will contain the unknown $\mathbf{P(A)}$ with probability at least**

$$\mathbf{P(r_n(A) + \epsilon \geq P(A) \geq r_n(A) - \epsilon)}$$
$$\mathbf{\geq 1 - 2e^{-n\epsilon^2}.}$$

- As with all confidence intervals, this is a claim made about the unknown $P(A)$ in advance of measuring or observing $r_n(A)$.
- Once we have observed $r_n(A)$, the case of real interest, we can no longer make this claim.
- We can rewrite the bound as

$$\mathbf{P}\left(\mathbf{r_n}(\mathbf{A}) + \frac{\mathbf{c}}{\sqrt{\mathbf{n}}} \geq \mathbf{P}(\mathbf{A}) \geq \mathbf{r_n}(\mathbf{A}) - \frac{\mathbf{c}}{\sqrt{\mathbf{n}}}\right)$$

$$\geq \mathbf{1} - \mathbf{2e^{-c}}.$$

Expressiveness of U/L Probability

- We can use $\underline{\mathrm{P}}$ and $\bar{\mathrm{P}}$ to define a variety of CP relations, exactly as was done in the long run case.
- Standard probability P is insufficiently expressive in that it cannot accommodate to the case of asymptotically divergent relative frequencies, except by making a model for how these relative frequencies change.

Relation to Comparative Probability

- **With this class of lower envelopes we can describe all possible comparative probability (CP) orders satisfying de Finetti's axioms.**
- **For all finite $\mathcal{A}$ and $\succsim$ a CP order relation there exists x^∞ such that $\underline{\mathbf{P}}$ as defined through the limit inferior along this sequence represents this order through**

$$(\forall A, B \in \mathcal{A})\; A \succsim B \iff \underline{\mathbf{P}}(\mathbf{A}) \geq \underline{\mathbf{P}}(\mathbf{B}).$$

- **All CP orders can now be inferred from data that is an infinite sequences of outcomes of repeated random experiments.**

- **Of course, in practice we will only observe a finite length data sequence.**
- **Walley and Fine used such sequences to estimate the unobtainable limiting versions.**

- **Of course, if the variations in relative frequency have a knowable structure, then one would use this structure.**
- **Use of possible structure is emphasized in a related approach of Cozman and Chrisman focusing on subsequences of x^{∞}.**
- **They consider the set of limiting relative frequencies of infinite subsequences of the data sequence x^{∞}.**
- **Walley and Fine emphasize the subsequences of relative frequencies calculated along all of x^{∞}.**