

Probable Probabilities

John L. Pollock
Philosophy and Cognitive Science
University of Arizona

pollock@arizona.edu
<http://www.u.arizona.edu/~pollock>

The Epistemology of Probability

- It is generally supposed that the probability calculus completely characterizes the logical and mathematical structure of probabilities.
- It is supposed that familiar sorts of statistical inference provide us with our basic knowledge of probabilities.
- Then appeal to the probability calculus enables us to compute other previously unknown probabilities.

Sparse Probability Knowledge

- The probability calculus and our knowledge of probabilities.
 - Suppose we know that $\text{PROB}(P) = .7$ and $\text{PROB}(Q) = .6$.
 - All the probability calculus enables us to infer is that $.3 \leq \text{PROB}(P \ \& \ Q) \leq .6$.
 - Similarly, all we can conclude about $\text{PROB}(P \vee Q)$ is that $.7 \leq \text{PROB}(P \vee Q) \leq 1.0$.

Sparse Probability Knowledge

- **Complete probability distributions**
 - Suppose a problem is described by logical compounds of a set of simple propositions $P_1, \dots, P_n$.
 - In theoretical accounts of the use of probabilities in any discipline, it is generally assumed that we come to a problem equipped with a complete probability distribution — an assignment of unique probabilities to every logical compound of the simple propositions.
 - However, in real life this assumption is totally unrealistic.
 - Given n simple propositions, we would have to know and store 2^n logically independent probabilities.

Sparse Probability Knowledge

- As Gilbert Harman (1973) observed years ago, for a rather small number of simple propositions, there is a completely intractable number of logically independent probabilities.
 - » Given just 300 simple propositions, there will be 2^{300} logically independent probabilities.
 - » 2^{300} is approximately equal to 10^{90} .
 - » Recent estimates of the number of elementary particles in the universe put it at $10^{80} - 10^{85}$.
 - » Thus to have a complete probability distribution, we would have to know 5 to 10 orders of magnitude more logically independent probabilities than the number of elementary particles in the universe.

The Problem of Sparse Probability Knowledge

- **At any given time, our knowledge of probabilities is worse than just incomplete.**
- **The set of probabilities we know is many orders of magnitude smaller than the set of all true probabilities.**
- **How then can we be as successful as we are in applying probability to real-world problems?**

Statistical Independence

- It is noteworthy that in applying probabilities to concrete problems, probability practitioners commonly adopt undefended assumptions of statistical independence.
- $\text{PROB}(P)$ and $\text{PROB}(Q)$ are statistically independent iff $\text{PROB}(P \& Q) = \text{PROB}(P) \cdot \text{PROB}(Q)$.
- An equivalent definition is that $\text{PROB}(P/Q) = \text{PROB}(P)$.
- In the practical use of probabilities it is almost universally assumed, often apologetically, that probabilities are independent unless we have some reason for thinking otherwise.

Statistical Independence

- The independence assumption is a defeasible assumption, because obviously we can discover that conditions we thought were independent are unexpectedly correlated.
- The probability calculus can give us only necessary truths about probabilities, so the justification of such a defeasible assumption must have some other source.

Probable Probabilities

- I will argue that a defeasible assumption of statistical independence is just the tip of the iceberg.
- There are multitudes of defeasible inferences that we can make about probabilities, and a very rich mathematical theory grounding them.
- It is these defeasible inferences that enable us to make practical use of probabilities without being able to deduce everything we need via the probability calculus.
- I will argue that, on a certain conception of probability, there are mathematically derivable second-order probabilities to the effect that various inferences about first-order probabilities, although not deductively valid, will nonetheless produce correct conclusions with probability 1, and this makes it reasonable to accept these inferences defeasibly.

Two Kinds of Probability

- **Definite (single case) probabilities - $\text{PROB}(P)$**
versus
- **Indefinite (general, statistical) probabilities**
 - about properties
 - » formulated using free variables - $\text{prob}(Fx/Gx)$
 - need a theory of “direct inference”
 - » The basic idea is that if we want to know the definite probability $\text{PROB}(Fa)$, we look for the narrowest reference class (or strongest reference property) G such that we know the indefinite probability $\text{prob}(Fx/Gx)$ and we know Ga , and then we identify $\text{PROB}(Fa)$ with $\text{prob}(Fx/Gx)$.

Nomic Probability

- Finite frequency theories.
- Hypothetical frequency theories.
- Nomic probability
 - $\text{prob}(Fx/Gx)$ is *the proportion* of physically possible G's that are F's.
 - A physically possible G is an ordered pair $\langle w, x \rangle$ such that w is a physically possible world (one compatible with all of the physical laws) and x has the property G at w . Let G be the set of all physically possible G's
 - $\text{prob}(Fx/Gx) = \rho(F, G)$.
 - subproperty: $F \nabla G$ iff it is physically necessary that $(\forall x)(Fx \rightarrow Gx)$.
- I doubt that we can pick out the right proportion function without appealing to prob itself, so the postulate is simply that there is some proportion function related to prob as above.
- Rather than axiomatizing prob directly, we adopt axioms for the proportion function

Nomic Probability and Proportions

- I assume that ρ satisfies the standard “Boolean” axioms for the probability calculus.
- Indefinite probabilities satisfy further axioms deriving from the fact that they relate relations rather than closed formulas. E.g.,
$$\text{prob}(Rxy/Sxy \ \& \ y = a) = \text{prob}(Rxa/Sxa)$$
- A third set of assumptions is the main focus of this paper. These are “limit assumptions”.
 - There are various principles that hold approximately for proportions among finite sets, with the degree of approximation approaching perfection as the size of the sets goes to infinity.
 - The limit assumptions tell us that these principles hold exactly for proportions among infinite sets.
 - I will discuss these limit assumptions in detail below.
 - They provide the basis for the theory of probable probabilities.

The Statistical Syllogism

- Statistical Syllogism

If $r > 0.5$, then “ $Gc \ \& \ \text{prob}(F/G) \geq r$ ” is a defeasible reason for “ Fc ”, the strength of the reason being a monotonic increasing function of r .

- Subproperty Defeat

“ $Hc \ \& \ \text{prob}(F/G\&H) < \text{prob}(F/G)$ ” is an undercutting defeater for the inference by the statistical syllogism from “ $Gc \ \& \ \text{prob}(F/G) \geq r$ ” to “ Fc ”.

Indifference

- Set X contains 10,000,000 members.
- I am going to select a subset of X randomly.
- How many members will it have?

Indifference

- Set X contains 10,000,000 members.
- I am going to select a subset of X randomly.
- How many members will it have?
- 99% of the subsets have cardinalities differing from 5,000,000 by less than .08%.

Finite Indifference Principle

- If $\#U = n$, then if $r \leq n$, the number of r -membered subsets of U is
 - $\text{Bin}(n,r)$ for $n = 100$, $n = 1000$, and $n = 10000$.

Infinitary Indifference Principle

- **Finite Indifference Principle:**

For every $\varepsilon, \delta > 0$, there is an N such that if U is finite and $\#U > N$ then

- **Infinitary Indifference Principle**

If U is infinite then for every $\delta > 0$,

- **Probabilistic Indifference Principle**

For any property G and for every $\delta > 0$,

- **Indifference Principle**

For any properties F and G , it is defeasibly reasonable to assume that $\text{prob}(F/G) = 0.5$.

- **Expectable Indifference Principle**

For any properties F and G , the expectable value of $\text{prob}(F/G) = 0.5$.

Probability = 1

- The fact that a nomic probability is 1 does not mean that there are no counter-instances.
- In fact, there may be infinitely many counter-instances.
- Consider the probability of a real number being irrational.
- It follows from the calculus of nomic probabilities that this probability is 1, because the cardinality of the set of irrationals is infinitely greater than the cardinality of the set of rationals.
- But there are still infinitely many rationals.
- The set of rationals is infinite, but it has measure 0 relative to the set of real numbers.

Probability = 1

- In classical probability theory (which is about definite probabilities), conditional probabilities are defined as ratios of unconditional probabilities:
$$\text{PROB}(P/Q) =$$

If conditional probabilities are defined in this way, $\text{PROB}(P/Q)$ is undefined when $\text{PROB}(Q) = 0$.
- However, for indefinite probabilities, there are no unconditional probabilities, so conditional probabilities must be taken as primitive. (“Popper functions”)

$\text{prob}(F/G\&H)$ can be perfectly well defined even when $\text{prob}(G/H) = 0$.

If $\text{prob}(F/G) = 1$, it does not follow that $\text{prob}(F/G\&H) = 1$. Specifically, this can fail when $\text{prob}(H/G) = 0$. Thus, for example,

$$\text{prob}(2x \text{ is irrational} / x \text{ is a real number}) = 1$$

but

$$\text{prob}(2x \text{ is irrational} / x \text{ is a real number} \ \& \ x \text{ is rational}) = 0.$$

Independence

- **Finite Product Principle**

For $0 \leq r, s \leq 1$ and for every $\varepsilon, \delta > 0$ there is an N such that if U is finite and $\#U > N$, then

- **Infinitary Product Principle**

For $0 \leq r, s \leq 1$, if U is infinite then for every $\delta > 0$:

- **Probabilistic Product Principle**

For $0 \leq r, s \leq 1$ and for any property U , for every $\delta > 0$:

Independence

- **Principle of Statistical Independence**

“ $\text{prob}(A/C) = r$ & $\text{prob}(B/C) = s$ ” is a defeasible reason for “ $\text{prob}(A \& B/C) = r \cdot s$ ”.

- **Principle of Expectable Statistical Independence**

If $\text{prob}(A/C) = r$ and $\text{prob}(B/C) = s$, the expectable value of $\text{prob}(A \& B/C) = r \cdot s$.

Probable Probabilities Theorem

- *Linear constraints* either state the values of certain proportions, e.g., stipulating that $\rho(X,Y) = r$, or they relate proportions using linear equations. For example, if we know that $X = Y \cup Z$, that generates the linear constraint

$$\rho(X,U) = \rho(Y,U) + \rho(Z,U) - \rho(X \cap Z, U).$$

- Probable Probabilities Theorem:

Let $U, X_1, \dots, X_n$ be a set of variables, and consider a set LC of linear constraints on proportions between truth-functional compounds of those variables. If LC is consistent with the probability calculus, then for any pair of truth-functional compounds P, Q of $U, X_1, \dots, X_n$ there is a real number r between 0 and 1 such that for every $\varepsilon, \delta > 0$, there is an N such that if U is finite and $\#U > N$, then

Probable Probabilities Theorem

Expectable Probabilities Principle:

Let $U, X_1, \dots, X_n$ be a set of variables , and consider a set LC of linear constraints on probabilities between truth-functional compounds of those variables. If LC is consistent with the probability calculus, then for any pair of truth-functional compounds P, Q of $U, X_1, \dots, X_n$ there is a real number r between 0 and 1 such that for every $\delta > 0$,

In other words, given the constraints LC , the expectable value of $\text{prob}(P/Q) = r$.

Nonclassical Direct Inference

- Principle of Agreement

For $0 \leq r \leq 1$ and for every $\varepsilon, \delta > 0$, there is an N such that if U is finite and $\#U > N$, then:

- Nonclassical Direct Inference

If $\text{prob}(A/B) = r$, the expectable value of $\text{prob}(A/B \& C) = r$.

- Independence and Agreement Theorem

$\text{prob}(A/B \& C) = \text{prob}(A/B)$ iff A and C are independent relative to B .

- Subproperty Defeat for Nonclassical Direct Inference

If $C \neq D$, $\text{prob}(A/B \& D) = s$, and $\text{prob}(A/B) = r$, then the expectable value of $\text{prob}(A/B \& C) = s$ (rather than r).

Classical Direct Inference

- Direct inference is normally understood as being a form of inference from indefinite probabilities to definite probabilities rather than from indefinite probabilities to other indefinite probabilities.
- Let $\mathbf{K}$ be the conjunction of all the propositions the agent knows to be true, and let K be the set of all physically possible worlds at which $\mathbf{K}$ is true (“ $\mathbf{K}$ -worlds”). I propose that we define the definite probability $\text{PROB}(P)$ to be the proportion of $\mathbf{K}$ -worlds at which P is true. Where P is the set of all P -worlds:
- $\text{PROB}(P) = \rho(P, K)$.
- More generally, where Q is the set of all Q -worlds, we can define:
- $\text{PROB}(P/Q) = \rho(P, Q \cap K)$.

Classical Direct Inference

- Representation Theorem for Definite Probabilities

(1) $\text{PROB}(Fa) = \text{prob}(Fx/x = a \ \& \ \mathbf{K})$;

(2) If it is physically necessary that $[\mathbf{K} \rightarrow (Q \leftrightarrow Sa_1...a_n)]$ and that $[(Q \& \mathbf{K}) \rightarrow (P \leftrightarrow Ra_1...a_n)]$, and Q is consistent with $\mathbf{K}$, then $\text{PROB}(P/Q) = \text{prob}(Rx_1...x_n/Sx_1...x_n \ \& \ x_1 = a_1 \ \& \ ... \ \& \ x_n = a_n \ \& \ \mathbf{K})$.

(3) $\text{PROB}(P) = \text{prob}(P \ \& \ x=x/x = x \ \& \ \mathbf{K})$.

- Classical Direct Inference

“ Ga is known and $\text{prob}(Fx/ Gx) = r$ ” is a defeasible reason for “ $\text{PROB}(Fa) = r$ ”.

- Subproperty Defeat for Classical Direct Inference

“ $G \neq H$, Ha is known, and $\text{prob}(Fx/ Gx \ \& \ Hx) \neq r$ ” is an undercutting defeater for the inference by classical direct inference from “ Ga is known and $\text{prob}(Fx/ Gx) = r$ ” to “ $\text{PROB}(Fa) = r$ ”.

Classical Direct Inference

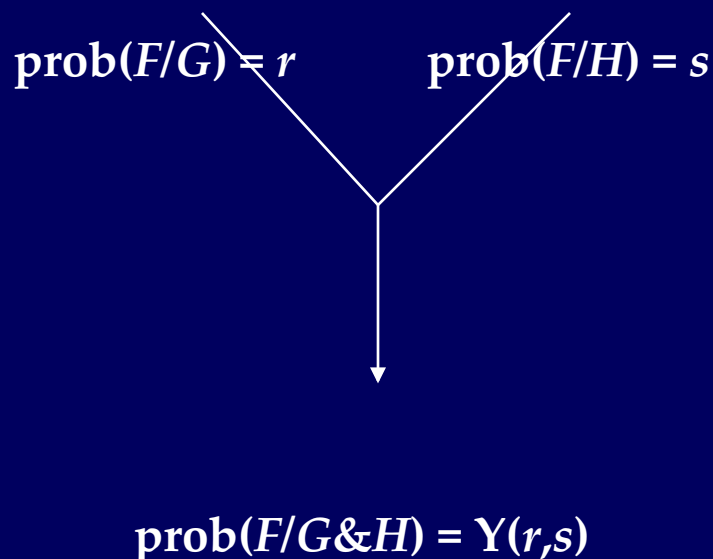
- Because definite probabilities are indefinite probabilities in disguise, we can also use direct inference to infer definite probabilities from definite probabilities.
- Thus $\text{£PROB}(P/Q) = r$ gives us a defeasible reason for expecting that $\text{PROB}(P/Q \& R) = r$.
- We can employ principles of statistical independence similarly. $\text{£PROB}(P/R) = r$ & $\text{PROB}(Q/R) = s$ gives us a defeasible reason for expecting that $\text{PROB}(P \& Q/R) = r \cdot s$.

Computational Inheritance

- Suppose we have two diagnostic tests for a disease.
 $\text{prob}(D/T_1) = r$ and $\text{prob}(D/T_2) = s$.
- Bernard tests positive on both tests.
 - What is the probability that Bernard has the disease?
 - What is $\text{prob}(D/T_1 \& T_2)$?
- Neither the probability calculus nor direct inference is any help.
- Intuitively, it seems that knowing that Bernard tests positive on both tests should raise the probability of his having the disease, but how much, and how do we justify this?

Computational Inheritance

- Knowledge of indefinite probabilities would be vastly more useful in real application if there were a function $Y(r,s)$ such that, in a case like the above, when $\text{prob}(F/G) = r$ and $\text{prob}(F/H) = s$, we could defeasibly expect that $\text{prob}(F/G\&H) = Y(r,s)$, and hence (by classical direct inference) that $\text{PROB}(Fa) = Y(r,s)$.



Computational Inheritance

- It has generally been assumed that there is no such function as the Y-function.
- One can certainly not make such an inference deductively, but could there be a function such that we can defeasibly expect $\text{prob}(F/G\&H)$ to be $Y(r,s)$? It follows from the probable probabilities theorem that there is. However, it is more useful to look at a simpler case, in which we also take account of the base-rate of F :
- Define:

$$Y(r,s:a) =$$

- Y-Principle:

If $B, C \not\supset U$, $\text{prob}(A/B) = r$, $\text{prob}(A/C) = s$, and $\text{prob}(A/U) = a$, then the expectable value of $\text{prob}(A/B \& C) = Y(r,s:a)$.

The Y-function

$Y(z, x : .5)$

$Y(z, x : .7)$

$Y(z, x : .3)$

$Y(z, x : .1) \quad Y(z, x : .01)$

Computational Inheritance

- **Applying this to Bernard,**
 - suppose the base-rate with which people having Bernard's symptoms have the disease is .6.
 - Suppose the probability of having the disease given that one tests positive on the first test is .7,
 - and the probability of having the disease given that one tests positive on the second test is .75.
- Then we can defeasibly estimate that the probability of Bernard having the disease given that he tests positive on both tests is $Y(.7,.75|.6) = .875$.

Inverse Probabilities

- Where $A, B \in U$, suppose we know the value of $\text{prob}(A/B)$. If we know the base rates $\text{prob}(A/U)$ and $\text{prob}(B/U)$, the probability calculus enables us to compute the value of the *inverse probability* $\text{prob}(\sim B/\sim A \& U)$:

Theorem: If $A, B \in U$ then
 $\text{prob}(\sim B/\sim A \& U) =$

Inverse Probabilities

- However, if we do not know the base rates, the probability calculus imposes no constraints on the value of the inverse probability.
 - In my (1990), I took that to indicate the need for a second form the statistical syllogism that is related to the above form as modus tollens is related to modus ponens:
- Inverse Statistical Syllogism:
 - If $r > .5$, then “ $\sim A_c$ & $\text{prob}(A/B) = r$ ” is a defeasible reason for “ $\sim B_c$ ”, the strength of the reason being a monotonic increasing function of r .
- However, this is unnecessary:
 - It can be shown that there are expectable values for the inverse probability, and generally, if $\text{prob}(A/B)$ is high, so is $\text{prob}(\sim B/\sim A \& U)$.
 - Thus rather than applying some form of inverse statistical syllogism to $\text{prob}(A/B)$, we can apply the standard form of the statistical syllogism to the inverse probability.

Inverse Probabilities

- Inverse Probabilities I

If $A, B \in U$ and we know that $\text{prob}(A/B) = r$, but we do not know the base rates $\text{prob}(A/U)$ and $\text{prob}(B/U)$, the following values are expectable:

$$\text{prob}(B/U) =$$

$$\text{prob}(A/U) =$$

$$\text{prob}(\sim A/\sim B \& U) = .5;$$

$$\text{prob}(\sim B/\sim A \& U) =$$

Inverse Probabilities

- Inverse Probabilities II

If $A, B \in U$ and we know that $\text{prob}(A/B) = r$ and $\text{prob}(B/U) = s$, but we do not know the base rate $\text{prob}(A/U)$, the following values are expectable:

$$\text{prob}(A/U) = .5(1 - (1 - 2r)s);$$

$$\text{prob}(\sim A/\sim B \& U) =$$

$$\text{prob}(\sim B/\sim A \& U) =$$

For low values of $\text{prob}(B/U)$,
 $\text{prob}(\sim B/\sim A \& U)$ can be expected to be
higher than $\text{prob}(A/B)$, and for all values
of $\text{prob}(B/U)$, $\text{prob}(\sim B/\sim A \& U)$ will be
fairly high if $\text{prob}(A/B)$ is high.

Furthermore, $\text{prob}(\sim B/\sim A \& U) > .5$ iff

$$\text{prob}(B/U) <$$

Inverse Probabilities

- Inverse Probabilities III

If $A, B \in U$ and we know that $\text{prob}(A/B) = r$ and $\text{prob}(A/U) = s$, but we do not know the base rate $\text{prob}(B/U)$, then:

(a) where b is the expectable value of $\text{prob}(B/U)$,

(b) The expectable value of $\text{prob}(\sim B/\sim A \& U) =$

When $\text{prob}(A/B) = \text{prob}(A/U)$, the expected value for $\text{prob}(\sim B/\sim A)$ is .5, and when $\text{prob}(A/B) > \text{prob}(A/U)$, $\text{prob}(\sim B/\sim A \& U) > .5$. If $\text{prob}(A/U) < .5$, the expected value of $\text{prob}(\sim B/\sim A \& U)$ is greater than $\text{prob}(A/B)$.

Conclusions

- **The problem of sparse probability knowledge: in the real world we lack direct knowledge of most probabilities.**
- **We must have ways of making defeasible estimates of their values even when those values are not computable from known probabilities using the probability calculus.**
- **Limit theorems from combinatorial mathematics provide the necessary bridge for these inferences.**

Conclusions

- Under very general circumstances, there will be expectable values for otherwise unknown probabilities.
 - These are described by principles telling us that although certain inferences from probabilities to probabilities are not deductively valid, nevertheless the second-order probability of their yielding correct results is 1.
 - This makes it defeasibly reasonable to make the inferences.
- indifference
- statistical independence
- classical and nonclassical direct inference
- computational inheritance
- inverse probabilities
- There are a huge number of useful principles of probable probabilities, some of which I have

Conclusions

- **These principles of probable probability are reminiscent of Carnap's logical probabilities.**
- **Historical theories of objective probability required probabilities to be assessed by empirical methods, and because of the weakness of the probability calculus, they tended to leave us in a badly impoverished epistemic state regarding probabilities.**
- **Carnap tried to define a kind of probability for which the values of probabilities were determined by logic alone.**

Conclusions

- **Nomic probability and the theory of probable probabilities lies between these two extremes.**
 - **It still makes the values of probabilities contingent rather than logically necessary, but it makes our limited empirical investigations much more fruitful by giving them the power to license defeasible, non-deductive, inferences to a wide range of further probabilities that we have not investigated empirically.**
 - **Unlike logical probability, these defeasible inferences do not depend upon ad hoc postulates.**
 - **Instead, they derive directly from provable theorems of combinatorial mathematics.**

Conclusions

- So even when we do not have sufficient empirical information to deductively determine the value of a probability, purely mathematical facts may be sufficient to make it reasonable, given what empirical information we do have, to expect the unknown probabilities to have specific and computable values.
- Where this differs from logical probability is that the empirical values are an essential ingredient in the computation.

The End

**(This paper is available at
<http://www.u.arizona.edu/~pollock>)**