APPENDIX 2: PROBABILITY

from Ellery Eells, *Probabilistic Causality*. Cambridge University Press, 1991, pp. 399-402.

In this appendix, I will present some of the basic ideas of the mathematical theory of probability. As in the case of Appendix 1, this will not be a comprehensive or detailed survey -- it is only intended to introduce the basic formal probability concepts and rules used in this book, and to clarify the terminology and notation used in this book. Here I will discuss only the abstract and formal calculus of probability; in Chapter 1, the question of interpretation is addressed.

A probability function, Pr, is any function (or rule of association) that assigns to (or associates with) each element X of some Boolean algebra B (see Appendix 1) a real number, Pr(X), in accordance with the following three conditions:

For all X and Y in B,

(1) Pr(X) ≥ 0;

(2) Pr(X) = 1, if X is a tautology (that is, if X is logically true, or X = 1 in B);

(3) Pr(XvY) = Pr(X) + Pr(Y), if X&Y is a contradiction

(that is, if X&Y is logically false, or X&Y = 0 in B).

These three conditions are the probability axioms, also called "the Kolmogorov axioms" (for Kolmogorov 1933). A function Pr that satisfies the axioms, relative to an algebra B, is said to be a probability function on B -- that is, with "domain" B (that is, the set of propositions of B) and range the closed interval [0,1]. In what follows, reference to an assumed algebra B will be implicit.

In Appendix 1, I explained how the propositional calculus is applicable to "propositions" understood as sentences or statements as well as to "propositions" understood as factors or properties -- and the same goes for the probability calculus. Roughly speaking, "Pr(X) = r" can be understood either as asserting that a sentence or statement X has a probability of r of being true (in a given situation), or as asserting that a factor or property has a probability of r of being exemplified (in a given instance or population). Specifying an interpretation of the propositions is part what must be done to "interpret" a probability function on an algebra; the other part is interpreting "Pr". Various interpretations of probability (such as frequency, degree of belief, and partial logical entailment interpretations) are discussed in Chapter 1; here, the focus is on the formal calculus.

Here are some easy consequences of the probability axioms.

(4) Pr(~X) = 1 - Pr(X), for all X.

Proof: By (1), Pr(Xv~X) = 1; and by (3), Pr(Xv~X) = Pr(X) + Pr(~X). So, 1 = Pr(X) + Pr(~X), and thus Pr(~X) = 1 - Pr(X).

(5) Pr(X) = 0, if X is a contradiction.

Proof: ~X is a tautology, so by (2), Pr(~X) = 1. By (4), Pr(~X) = 1 -Pr(X). So, 1 = 1 - Pr(X), and thus Pr(X)

= 0.

(6) Pr(X) = Pr(Y), if X and Y are logically equivalent.

Proof: X and ~Y are mutually exclusive and Xv~Y is a tautology. So by (2), (3), and (4), 1 = Pr(Xv~Y) = Pr(X) + Pr(~Y) = Pr(X) + 1 - Pr(Y). So, 1 = Pr(X) + 1 - Pr(Y), and 0 = Pr(X) - Pr(Y), and thus Pr(X) = Pr(Y).

(7) Pr(X) ≤ Pr(Y), if X logically implies Y.

(8) 0 ≤ Pr(X) ≤ 1, for all X.

(9) Pr(XvY) = Pr(X) + Pr(Y) - Pr(X&Y), for all X and Y.

The probability of Y conditional on (or given) X, written Pr(Y/X), is defined to be equal to Pr(X&Y)/Pr(X). Note that Pr(Y/X) is defined only when Pr(X) > 0. Since for any X and Y, Pr(X&Y) = Pr(Y&X) (by (6) above), an immediate consequence of the definition of conditional probability is what is often called the multiplication rule:

(9) Pr(X&Y) = Pr(X)Pr(Y/X) = Pr(Y)Pr(X/Y), for all X and Y.

From (9) follows this simple version of Bayes' theorem:

Pr(Y/X) = Pr(X/Y)Pr(Y)/Pr(X), for all X and Y.

A proposition Y is said to be probabilistically (or statistically) independent of a proposition X if Pr(Y/X) = Pr(Y). Alternatively, and equivalently, Y's being probabilistically independent of X can be defined as Pr(X&Y) = Pr(X)Pr(Y). Thus, probabilistic independence is symmetric: if Y is probabilistically independent of X, then X is probabilistically independent of Y, for all X and Y.

If propositions X and Y are not probabilistically independent, then there is said to be a probabilistic (or statistical) correlation (or dependence) between X and Y. The correlation is called positive or negative according to whether Pr(Y/X) is greater or less than Pr(Y). This is sometimes described by saying that X is positively or negatively probabilistically relevant to Y, or that X has positive or negative probabilistic significance for Y. It is easy to see that the following six probabilistic relations are equivalent:

Pr(Y/X) > Pr(Y);

Pr(X/Y) > Pr(X);

Pr(Y) > Pr(Y/~X);

Pr(X) > Pr(X/~Y);

Pr(Y/X) > Pr(Y/~X);

Pr(X/Y) > Pr(X/~Y).

Also, these six relations would remain equivalent if the ">"'s were all replaced with "<"'s, or with "="'s. Thus, the two kinds of probabilistic correlation (positive and negative), as well as probabilistic independence, are symmetric. If Pr(Y/Z&X) = Pr(Y/Z&~X), then Z is said to screen off any probabilistic correlation of Y with X.

Two propositions X and Y are called probabilistically equivalent if Pr((X&Y)v(~X&~Y)) = 1. Another way of putting this is as follows. A common propositional connective, not mentioned in Appendix 1, is the biconditional connective, "<->". The biconditional of two propositions X and Y is the proposition that is true just in case X and Y have the same truth value -- that is, either they are both true or they are both false. The biconditional of X and Y is often expressed as "X if and only if Y", or, for short, "X iff Y" (X if Y, and X only if Y). Then X and Y are probabilistically equivalent just when Pr(X<->Y) = 1. When two propositions X and Y are probabilistically equivalent, then they are "interchangeable in all probabilistic contexts". That is, given that X and Y are probabilistically equivalent, if (possibly truth-functionally complex) propositions Z(X,Y) and W(X,Y) result from any (possibly truth-functionally complex) propositions Z and W, respectively, by changing X's to Y's or Y's to X's, in any way, then Pr(Z/W) = Pr(Z(X,Y)/W(X,Y)).

A generalization of the common idea of an average is the statistical idea of expectation, or expected value. Given a variable N which can take on the possible values n_1, ..., n_s, and a probability Pr on propositions of the form "$N = n_i$", the expectation, or expected value, of N (calculated in terms of the probability Pr) is:

$$\mathrm{SUM}_{i=1}^{r}\ Pr(N = n_i)n_i.$$

If the probabilities in terms of which an expectation is calculated are conditional probabilities, then the expectation is a conditional expectation, or conditional expected value. For example, if R is a proposition that may be relevant to the value of N, then

$$\mathrm{SUM}_{i=1}^{r}\ Pr(N = n_i/R)n_i$$

is a conditional expectation.