Determining the Independence of Random Variables

[Reprinted from *Proc. IEEE Int. Symp. on Info. Th.*, 1995, p. 11.]

James L. Massey

Signal & Info. Proc. Lab., Swiss Federal Inst. Tech., CH-8092 Zurich, Switzerland

Abstract — **A graphical calculus is presented for determining the independence and conditional independence of random variables in a specified probabilistic setting. The calculus is developed first for the case of random variables that form a Markov chain. The calculus is then extended to the "general causal case" where the random variables are obtained from a sequence of random experiments in which each experiment can be carried out in full when the results of specified previous experiments are made available to it.**

I. Introduction

Because mutual information is essentially a measure of probabilistic dependence, information theory can be used to devise a convenient calculus for reasoning about probabilistic dependence. For example, because $I(X;Y) \geq 0$ with equality if and only if the random variables X and Y are independent, it follows that the determination of whether X and Y are independent is equivalent to determining whether $I(X;Y)$ vanishes. Moreover, the vanishing of $I(X;Y)$ can alternatively and conveniently be taken as the definition of (probabilistic) independence. Similarly, the vanishing of the conditional mutual information $I(X;Y \mid Z)$ can be taken as the definition of the independence of X and Y when conditioned on knowledge of Z.

Conditional independence will be seen to play an important role in the study of probabilistic dependence. Independence and conditional independence are in general unrelated properties of random variables in the sense that X and Y can be independent but dependent when conditioned on Z and, conversely, X and Y can be dependent but independent when conditioned on Z.

II. Markov Chains

A Markov chain can alternatively and conveniently be defined as a sequence $X_1, X_2, \ldots X_n$ of random variables such that, for all i strictly between 1 and n, $[X_1, X_2, \ldots X_{i-1}]$ and $[X_{i+1}, X_{1+2}, \ldots X_n]$ are independent when conditioned on X_i. An immediate consequence of the symmetry of mutual information, i.e., of the fact that $I(X;Y \mid Z) = I(Y;X \mid Z)$, is that the reversed sequence $X_n, X_{n-1}, \ldots X_1$ is also a Markov chain, which is a well-known fact but one that is awkward to prove from the usual definition of a Markov chain. Another immediate consequence of this alternative definition of a Markov chain is that any subsequence of a Markov chain $X_1, X_2, \ldots X_n$ is also a Markov chain, which again is a well known fact that is awkward to prove from the usual definition.

The following result is as useful in formulating a calculus of dependence as it is trivial to prove.

Proposition 1 *(Independence Inheritance)*
If $I(WX;Z \mid Y) = 0$, then also $I(X;Z \mid Y) = 0$ and $I(X;Z \mid WY) = 0$.

In other words, if some (possibly conditional) mutual information is zero, then any random variable *not in the conditioning* can be discarded or moved into the conditioning with the mutual information remaining zero.

The above proposition is the basis for the following calculus of independence for Markov chains: The random variables $X_1, X_2, \ldots X_n$ in the Markov chain are used to label in the natural order the nodes of a simple (undirected) linear graph with n nodes. Then any (possibly conditional) mutual information involving only the random variables $X_1, X_2, \ldots X_n$ is zero if, for every pair of random variables with one to the left and one to the right of the semicolon in the mutual information expression, there is a random variable in the conditioning whose node in the graph lies between the nodes for these two random variables. Moreover, this is the strongest statement that can be made in general about the (conditional) independence of the random variables in a Markov chain in the sense that there are chains for which the given mutual information is non-zero when this condition is not fulfilled. It is thus natural from the graphical viewpoint to think of conditioning as "blocking" dependence between the random variables in a Markov chain.

III. General Causal Systems

The graphical calculus of independence developed for Markov chains can be extended to apply to any random variables that can be described as the results of a sequence of random experiments in which the results of only previous experiments affect the results of following experiments, i. e., the random variables in the sequence have a well defined defined *causal dependence*. The distinction between causal dependence, which is directed, and probabilistic dependence, which is undirected, is crucial to the formulation of this extended graphical calculus. In contrast to the Markov chain case, conditioning can in general create probabilistic dependence between random variables that would be independent without this conditioning.

The real utility of the information–theoretical calculus for analyzing probabilistic dependence becomes evident when considering networks of information sources, channels, encoders and decoders. Precise definitions of all these devices together with the rules for their valid interconnection in neworks are required for the precise formulation of the calculus. Examples will be given in the presentation of this paper to illustrate the utility of the calculus in rather complicated networks.