CAUSAL LINEAR STOCHASTIC DEPENDENCIES: THE FORMAL THEORY

Rolf Steyer
University of Trier
Trier, Federal Republic of Germany

The formal background of the theory of causal linear stochastic dependence is provided, which was introduced by Steyer (1984). The theory presented is concerned with those kinds of dependencies which can be described by specifying the functional form of a conditional expectation $E(Y|X)$. This includes also those situations in which $X$ is a multidimensional random variable. The main concepts of the theory are causal and weak causal linear stochastic dependencies, the definition of which is based on the pre- and equiorderedness relations of sigma-fields and stochastic variables, on the notion of potential disturbing sigma-fields and variables, as well as on the invariance and on the average conditions. These concepts are formally defined and their properties are studied in some detail. Causal linear stochastic dependence is defined by the preorderedness condition that the influencing variable is antecedent to the influenced variable and by the invariance condition, whereas weak causal linear stochastic dependence is defined by the preorderedness and average conditions. Both, the invariance and the average conditions, and therefore both kinds of causal hypotheses, can empirically be tested in experimental as well as in nonexperimental observational studies.

1. INTRODUCTION

"Correlation does not prove causality", is a statement univocally found in
textbooks on applied statistics. However, whenever one looks for a definition of causal dependence, one either finds treatments on philosophical theories only loosely related to the concepts of correlation, regression, and the experimental control techniques such as randomization (see e.g. Bagozzi (1980) or Heise (1975)), or treatments of experimental and quasi-experimental control techniques (see e.g. Cook & Campbell (1979)). Although these discussions are very useful and instructive in many respects, there is no formal theoretical connection between the correlation, regression, and analysis of variance models applied and the control techniques discussed. What is, in formal terms, the difference between an analysis of variance model in a purely observational study at one hand and in a randomized experiment at the other?

Steyer (1983, 1984) has taken some first steps to bridge the gap between ideas of causal dependence, stochastic models, and experimental control techniques. The basic question raised is, which are the formal properties that make a stochastic model a causal one? Restricting the discussion to linear stochastic dependencies, i.e. to those dependencies which can be described by conditional expectations, Steyer (1984) proposed such properties defining two types of nonmurious or causal linear stochastic dependencies, a weak and a strong one, each of which are distinguished from non-causal linear stochastic dependencies by two conditions. Beside the pre-orderedness condition that the influencing variable $X$ is antecedent to the influenced variable $Y$, the crucial condition for causal linear stochastic dependence of a random variable $Y$ on a, possibly multidimensional, random variable $X$ is the invariance condition postulating that $E(Y|X, W) = E(Y|X) + W_h X$ holds for all potential disturbing variables $W$, where $W_h$ is a $W$-measurable composition of a function $H$ with $W$. The invariance condition states that, if any potential disturbing variable $W$ is included as another conditioning variable, it only adds to $E(Y|X)$ in the equation for $E(Y|X, W)$, leaving $E(Y|X)$ invariant.

The crucial condition for weak causal linear stochastic dependence, on the other hand, is the average condition postulating that, for all potential disturbing variables $W$, $E(Y|X=x) = \int E(Y|X=x, W)P(W)dw$ holds, for $P_W$ almost all $E(Y|X=x)$. This condition means that $E(Y|X=x)$ is the average of the conditional expectations $E(Y|X=x, W)$ across the values $w$ of $W$, for $P_W$-almost all $E(Y|X=x)$. If $W$ is a discrete stochastic variable, this equation can be written:

$$E(Y|X=x) = \sum_{W} E(Y|X=x, W)P(W).$$

It will be shown that this average condition holds, for example, if $X$ and all potential disturbing variables $W$ are stochastically independent, a condition which may be assumed to hold in randomized experiments, where $X$ indicates group membership and the potential disturbing variables $W$ represent proprieties of the experimental units (subjects) before or at the time of the treatment. This provides the desired link to the experimental control techniques such as randomization and matching.

In the formulation of the preorderedness, invariance and average conditions above, formally undefined terms have been used that have to be eliminated in order to construct a formal theory consisting only of mathematically well-defined terms. So far, we have no formal criteria to decide whether or not a specified stochastic variable $X$ is antecedent to a specified stochastic variable $Y$, and whether or not a specified stochastic variable $W$ is a potential disturbing variable. Can $H_1 = X^2$ be a potential disturbing variable with respect to the dependence of $Y$ on $X$? Can $H_2 = Y+X$ be a potential disturbing variable? Are variables mediating between $X$ and $Y$ potential disturbing variables, or variables which are influenced by $Y$?

In the following sections, we first summarize some basic concepts of probability theory and introduce notational conventions. Then, we give a formal definition of the pre- and equiorderedness relations and study their properties. Next, we define potential disturbing sigma-fields and variables, which will make the average and invariance conditions discussed above well defined formal concepts. Then, we give formal definitions of causal and weak causal linear stochastic dependencies, and investigate the properties of these concepts.

2. SOME BASIC CONCEPTS OF PROBABILITY THEORY AND NOTATION

In this section, a brief summary and notational conventions of some basic concepts of probability theory are given, which seem to be essential for a proper understanding of the theory proposed. For detailed introductions, the reader is referred to Bauer (1974), Breiman (1968), Gänßer and Stute (1977), Halmos (1969), or Loève (1977, 1978).
The fundamental assumption of every stochastic substantive (i.e., psychological, sociological, etc.) model is that the experiment, or more generally, the part of reality to be described, can be represented by a probability space, the definition of which is based on the following concepts.

Let \( \mathbb{N} = \{1, 2, \ldots\} \) be the set of natural numbers. A **sigma-field** \( \mathcal{A} \) on \( \Omega \) is defined to be a set of subsets of \( \Omega \) with the following three properties:

(a) \( \emptyset \in \mathcal{A} \).
(b) If \( A \in \mathcal{A} \), then \( A^c \in \mathcal{A} \), where \( A^c : = \Omega - A \) denotes the complement of \( A \).
(c) If \( (A_i, i \in \mathbb{N}) \) is a sequence of elements \( A_i \) of \( \mathcal{A} \), then their union \( \bigcup_{i \in \mathbb{N}} A_i \) is an element of \( \mathcal{A} \).

The intersection of a family \( (A_i, i \in I) \) of sigma-fields \( A_i \) on \( \Omega \) is also a sigma-field on \( \Omega \). If \( \mathcal{E} \) is a set of nonempty subsets of \( \mathcal{A} \), then the sigma-field \( \mathcal{A}(\mathcal{E}) \) generated by the set system \( \mathcal{E} \) is defined to be the intersection of all those sigma-fields on \( \mathcal{A} \), which contain \( \mathcal{E} \) as a subset.

A **measurable space** is defined to be a pair \((\Omega, \mathcal{A})\) of a set \( \Omega \) and a sigma-field \( \mathcal{A} \) on \( \Omega \).

Let \((\Omega, \mathcal{A})\) be a measurable space. A **probability measure** \( P : \mathcal{A} \to [0,1] \) is defined to be a function assigning each \( A \in \mathcal{A} \) a real nonnegative number with the following three properties:

(a) \( P(\emptyset) = 0 \).
(b) If \( (A_i, i \in \mathbb{N}) \) is a sequence of elements of \( \mathcal{A} \), \( A_i \cap A_j = \emptyset \) for \( i \neq j \), then \( P(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i) \).
(c) \( P(\Omega) = 1 \).

A probability space can now be defined to be a triple \((\Omega, \mathcal{A}, P)\) of a set \( \Omega \), called the set of elementary events, a sigma-field \( \mathcal{A} \) of subsets of \( \Omega \), where the elements \( A \in \mathcal{A} \) are called events, and a probability measure \( P : \mathcal{A} \to [0,1] \) which assigns the probability \( P(A) \) to each event \( A \in \mathcal{A} \).

Let \((\Omega, \mathcal{A})\) and \((\Omega', \mathcal{A}')\) be measurable spaces. A mapping \( X : \Omega \to \Omega' \) is called \((\mathcal{A}, \mathcal{A}')\)-measurable, if, for all \( A' \in \mathcal{A}' \),

\[
X^{-1}(A') := \{ \omega \in \Omega : X(\omega) \in A' \} \in \mathcal{A}.
\]

The set \( X^{-1}(A') \) is called the **sigma-field generated by** \( X \) and \( A' \). \( X^{-1}(A') \) may also be denoted by \( \mathcal{A}(X, A') \). If \( X \) is real-valued and \( \mathcal{N}(1) \)-dimensional, \( \mathcal{A}(X, A') = X^{-1}(A') \) is also denoted by \( \mathcal{B}(X) \) and \( A' \) is understood to be the borel sigma-field \( \mathcal{B}(\mathcal{N}(1)) \) on \( \mathcal{R}^{\mathcal{N}(1)} \), which is defined to be the sigma-field generated by the set system of all open intervals of \( \mathcal{R}^{\mathcal{N}(1)}, \mathcal{N}(1) \in \mathcal{N} \).

Let \((\Omega, \mathcal{A})\), \( i \in I \), be a finite or infinite sequence of measurable spaces. A **sigma-field generated by** a sequence \((X_i, i \in I)\) of \((\mathcal{A}_i, \mathcal{A}_i')\)-measurable mappings \( X_i : \Omega_i \to \Omega_i \) is defined to be the sigma-field generated by the union of the sigma-fields \( X_i^{-1}(A_i) \), i.e.

\[
\mathcal{A}(\{X_i, A_i\}, i \in I) := \bigcup_{i \in I} X_i^{-1}(A_i).
\]

If \( I = \{1, \ldots, N(1)\} \) is finite and \( \Omega_i = \mathcal{R} \), for all \( i \in I \), we also use the alternative notation \( \mathcal{A}(X_1, \ldots, X_N(1)) \) instead of \( \mathcal{A}(X_i, A_i), i \in I \).

Let \((\Omega, \mathcal{A})\) and \((\Omega', \mathcal{A}')\) be measurable spaces. An \((\mathcal{A}', \mathcal{A}')\)-stochastic variable \( X \) on the probability space \((\Omega, \mathcal{A}, P)\) is defined to be a mapping \( X : \Omega \to \Omega' \) that is \((\mathcal{A}', \mathcal{A}')\)-measurable. A \( \mathcal{N}(1) \)-dimensional real-valued stochastic variable, for example, is a \((\mathcal{B}(\mathcal{R}^{\mathcal{N}(1)}), \mathcal{N}(1))\)-stochastic variable.

The measurability condition in the definition of a \((\mathcal{A}', \mathcal{A}')\)-stochastic variable implies that

\[
X^{-1}(A') := \{ \omega \in \Omega : X(\omega) \in A' \} \in \mathcal{A}, \text{ for all } A' \in \mathcal{A}'.
\]

This allows to define the **distribution** \( P_X : \mathcal{A}' \to [0,1] \) of a \((\mathcal{A}', \mathcal{A}')\)-stochastic variable \( X \) by

\[
P_X(A') = P(X^{-1}(A')), \text{ for all } A' \in \mathcal{A}'.
\]

The distribution \( P_X \) of \( X \) is a probability measure on \( \mathcal{A}' \).

3. PRE- AND EUQUOREDNESS

A necessary condition for a stochastic dependence of \( Y \) on \( X \) to be causal is that \( X \) is antecedent or, synonymously, preordered to \( Y \). In experiments, for example, the treatment variables are manipulated before the effects on the dependent variables are assessed, and in nonexperimental studies, too, the influencing variables have to be antecedent to the influenced variables, if a causal statement should make sense at all. Even reciprocal causal dependence can much better be thought of as a process of mutual influencing of the variables involved (see e.g., Steyer (1982)), where, at each point of the process, the causing variable is preordered to the caused one. Thus, representing reciprocal causal relations by dynamic models not only allows the preservation of asymmetry of influence (see e.g., Simon (1952)) for reciprocal causality, but is also much more congruent with the dynamic nature.
of reality. Preorderedness of a stochastic variable $X$ to a second one, $Y$, is the first concept to be defined which draws on the fact that $(\Omega, A, P)$ is assumed to represent a process.

A $(\Omega', A')$-stochastic process on a probability space $(\Omega, A, P)$ is defined to be a family $(X_t, t \in T)$ of $(\Omega', A')$-stochastic variables on $(\Omega, A, P)$. Hence, one might think of defining $X_t$ to be reordered to $X_s$, if $s < t$, $s, t \in T$. However, such a concept would be too restricted. Oftentimes, variables have to be considered that are defined, for example, by $X := X_s \cdot X_t$, $s \neq t$. Obviously, the definition of preorderedness considered above could not be applied to $X$, because not only one but two points of time are involved in the definition of $X$.

These problems are not only of theoretical, but also of much practical interest. If, for example, the effect of a psychological intervention (therapy, training, etc.) is to be investigated, such a treatment usually extends over several sessions. Hence, if we let $T$ be the index set of the sessions, then the treatment variable, indicating whether or not a person receives a treatment, cannot be assigned to one single $t \in T$. The same problem occurs, if we let $T$ denote the continuous time covered by one session. An intervention of the psychotherapist cannot be assigned to a single $t \in T$ in the way outlined above. Similarly, repeated events, such as being cheated repeatedly by one's mother, may be a cause of someone being distrustful. Again, the causing variable cannot be assigned to a single point of time, a problem which Suppes (1970) seems to neglect (cf. Stegmüller (1983) p. 602).

In order not to run into these difficulties, we basically follow the approach proposed by Steyer (1983), which is, however, corrected in some points and studied in more detail. This approach is not based on a process $(X_t, t \in T)$, but on a monotonically increasing family $(A_t, t \in T)$ of sigma-fields, (see Figure 1). Let $T$ be a subset of the set $\mathbb{R}$ of real numbers. A monotonically increasing family $(A_t, t \in T)$ of sigma-fields is defined to be a family of sigma-fields $A_t$ on a set $\Omega$, $t \in T \subseteq \mathbb{R}$, with the property that, if $s \leq t$, $s, t \in T$, then $A_s \subseteq A_t$. Such a monotonically increasing family of sigma-fields is obtained, for example, if we define $A_t$ to be the sigma-field generated by all $X_s$, $s \in T$, $s \leq t$, where $(X_t, t \in T)$ is assumed to be a real-valued stochastic process. However, we do not need a stochastic process $(X_t, t \in T)$ to construct a monotonically increasing family $(A_t, t \in T)$ of sigma-fields, as is shown in the following example.

3.1. EXAMPLE

Consider an experiment in which two coins, each having one metal and one plastic side, are tossed onto a plate having the properties of an electro magnet which is on or off while the two coins are tossed. (For a more detailed description of this example, see Steyer (1984).) In this application, we may choose $\Omega$ to be the set product of the sets $\Omega_1 = \{a_1, a_2\}$, $\Omega_2 = \{b_1, b_2\}$, and $\Omega_3 = \{c_1, c_2\}$. Hence, each element $(a_1, b_2, c_1)$ of $\Omega$ denotes one of the eight elementary events that Coin 1 shows side $i$, Coin 2 shows side $j$, and the electro magnet is off ($k=0$) or on ($k=1$). Furthermore, we choose the sigma-field $A$ on $\Omega$ to be the set of all subsets of $\Omega$. If $X$ denotes the stochastic variable indicating the state of the electro magnet and the variables $Y_i$, $i=1, 2$, indicate the outcome of tossing coin $i$, a monotonically increasing family $(A_t, t \in T)$ of sigma-fields is easily obtained if we define $T := \{1, 2\}$, $A_1 := A(X)$, and $A_2 := A(X, Y_1, Y_2)$, where $A(X, Y_1, Y_2)$ denotes the sigma-field generated by the real-valued stochastic variables $X, Y_1, Y_2$.

The example above, involving only three variables $X, Y_1$ and $Y_2$ and their generated sigma-fields, is very simple. However, it will help to illustrate the concept of preorderedness that is defined as follows.

3.2. DEFINITION

Let $(\Omega, A, P)$ be a probability space, let $T$ be a subset of the set $\mathbb{R}$ of real
numbers, let \((A_t, t \in T)\) be a monotonically increasing family of sigma-fields with \(A(\cup_{t \in T} A_t) \subseteq A_t\). Let \(X, Y, Z, W\) be \((n, A')\)-, \((n', A')\)-stochastic variables on \((n, A, P)\), respectively, and finally, let \(C, D \subseteq A\) be two sigma-fields.

(i) We say that \(C\) is ordered to \(D\) with respect to \((A_t, t \in T)\), iff ('iff' is an abbreviation for 'if and only if')

(a) there is a \(s \in T\), with \(C \subseteq A_s\), \(D \not\subseteq A_s\) and
(b) there is an element \(t \in T\), with \(D \subseteq A_t\), \(D \not\subseteq A_s\), for all \(s \in T, s < t\).

(ii) We say that \(X\) is ordered to \(Y\) with respect to \((A_t, t \in T)\), iff Conditions (a) and (b) hold with

(c) \(C = X^{-1}(A')\), \(D = W^{-1}(A')\).

If no confusion is possible, the explicit reference to \((A_t, t \in T)\) may be omitted. According to Conditions (a) and (b) of Definition 3.2, there is a smallest \(t \in T\) for which \(D\) is a subset of \(A_t\) and that there is an element \(s \in T, s < t\), such that \(C\) is a subset of \(A_s\). In the example discussed above, where all variables involved are real-valued (i.e. \(n' = n'' = \mathbb{R}\) and \(A' = A'' = \mathbb{B}\), where \(\mathbb{B}\) denotes the Borel sigma-field on \(\mathbb{R}\)), \(X\) is ordered to \(Y_1\) and \(Y_2\), because \(A(X) = X^{-1}(\mathbb{B}) \subseteq A_1\), whereas \(A(Y_1) = Y_1^{-1}(\mathbb{B}) \not\subseteq A_1\) and \(A(Y_2) = Y_2^{-1}(\mathbb{B}) \not\subseteq A_2\), but \(A(Y_1), A(Y_2) \subseteq A_2\). In this example, \(Y_1\) and \(Y_2\) are equiordered in the following sense.

3.3. DEFINITION

Let the assumptions and notations of 3.2 be valid.

(i) We say that \(C\) and \(D\) are equiordered with respect to \((A_t, t \in T)\), iff

(a) there is an element \(t \in T\) with \(C, D \subseteq A_t\) and
(b) there is no element \(s \in T, s < t\), with \(C \subseteq A_s\) or \(D \subseteq A_s\).

(ii) We say that \(X\) and \(Y\) are equiordered with respect to \((A_t, t \in T)\), iff Conditions (a) and (b) hold with

(c) \(C = X^{-1}(A')\) and \(D = W^{-1}(A')\).

According to Conditions (a) and (b) of Definition 3.3, there is a smallest element \(t \in T\) with \(C\) and \(D\) being both subsets of \(A_t\), and there is no element \(s \in T, s < t\), such that \(C\) or \(D\) are subsets of \(A_s\).

It is easily seen that the diagram of any recursive path analysis model can be translated into a monotonically increasing family of sigma-fields so that the concepts of pre- and equiorderedness can be applied. For the path diagram of Figure 2, for example, we may define \(T := \{1, 2, 3\}, A_1 := A(Z_1, Z_2), A_2 := A(Z_1, Z_2, Z_3), \) and \(A_3 := A(Z_1, Z_2, Z_3, Z_4)\). Obviously, \(Z_1\) and \(Z_2\) are equiordered, \(Z_1\) and \(Z_2\) are both preordered to \(Z_3\) and \(Z_4\), and \(Z_3\) is reordered to \(Z_4\) with respect to \((A_t, t \in T)\) (see Definitions 3.2 and 3.3).

![Figure 2. Path diagram for a recursive model with four variables.](image)

We now treat some formal properties of the pre- and equiorderedness relations. The first one is that preorderedness of sigma-fields, as well as of stochastic variables, is a strict order relation. Note that all propositions in the following theorems are made with respect to a given family \((A_t, t \in T)\) of sigma-fields.

3.4. THEOREM

Let the presumptions and notations of 3.2 be valid, let \(F \subseteq A\) be a sigma-field and let \(X, Y, U\) be \((n, A')\)-, \((n', A')\)-, and \((n'', A'')\)-stochastic variables on the probability space \((n, A, P)\), respectively.

(i) If \(C\) is preordered to \(D\), then \(D\) is not preordered to \(C\) (asymmetry).

(ii) If \(C\) is preordered to \(D\) and \(D\) is preordered to \(F\), then \(C\) is also preordered to \(F\) (transitivity).

(iii) Preorderedness of sigma-fields is a strict order relation on the set of all sigma-fields being subsets of \(A\).

(iv) Propositions (i) to (iii) also hold for stochastic variables \(X, Y\) and \(U\) taking the role of \(C, D\), and \(F\), respectively.

Proof.

We only have to show that the properties (i) and (ii) hold for sigma-fields, because a strict order relation is defined by the asymmetry and transitivity properties, and preorderedness of stochastic variables is defined through their generated sigma-fields.

(i) If \(C\) is preordered to \(D\) with respect to \((A_t, t \in T)\), then there
is an element \( s \in T \) with \( C \subset A_s \), \( D \not\subset A_s \). As \( (A_t, t \in T) \) is monotonically increasing, there is no element \( t \in T \), \( t > s \), such that \( D \subset A_t \), \( C \not\subset A_t \), which implies that \( D \) is not preordered to \( C \).

(ii) If \( C \) is preordered to \( D \), then

(a) there is a \( s \in T \) with \( C \subset A_s \), \( D \not\subset A_s \).

If \( D \) is preordered to \( F \), then

(b) there is a \( t \in T \), \( t > s \), with \( D \subset A_t \), \( F \not\subset A_t \), and

(c) there is an element \( u \in T \), \( u > t \), with \( F \subset A_u \), \( F \not\subset A_u \), for all \( s \in T \), \( s < u \).

Hence, Conditions (a) to (c) imply that there is an element \( s \in T \) with \( C \subset A_s \), \( F \not\subset A_s \), and there is an element \( u \in T \), \( u > s \), with \( F \subset A_u \), \( F \not\subset A_u \), for all \( s \in T \), \( s < u \).

Theorem 3.4 does not imply that all sigma-fields being subsets of \( A \) are in this relation, but that the asymmetry and transitivity conditions are fulfilled.

3.5. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid. If \( C \) is preordered to \( D \), then \( C \) is also preordered to \( F := \Lambda(C \cup D) \).

Proof.

(i) Conditions 3.3a and 3.3b are obviously true for \( C = D \).

(ii) Is obvious.

(iii) If \( C \) and \( D \) are equiordered, then

(a) there is a smallest \( t \in T \) such that \( C \not\subset A_t \) and no \( s \in T \), \( s < t \), with \( C \subset A_s \) or \( D \subset A_s \).

If \( D \) and \( F \) are equiordered, this implies that also

(b) \( F \subset A_t \) and that

(c) there is no \( s \in T \), \( s < t \), with \( D \subset A_s \) or \( F \subset A_s \).

(a) to (c) imply that there is a \( t \in T \) with \( C \subset A_t \), whereas (a) and (c) imply that there is no \( s \in T \), \( s < t \), with \( C \subset A_s \) or \( F \subset A_s \).

Hence, \( C \) and \( F \) are equiordered, which implies the transitivity property.

(iv) An equivalence relation is defined by reflexivity, symmetry, and transitivity.

(v) This proposition is implied by 3.6i to 3.6iv, if we define \( C := X^{-1}(A^*), D := W^{-1}(A''), \) and \( F := U^{-1}(A''') \).

It should be noted that not every sigma-field \( C \subset A \) is equiordered to itself. It may happen that \( C \) is not in the equiorderedness relation, because there is no smallest \( t \in T \) with \( C \subset A_t \). If \( T = \emptyset \), \( C = (0,n) \), for example, there is no smallest \( t \in T \) with \( (0,n) \subset A_t \). The equivalence relation of equiorderedness is not defined on the set of all sigma-fields \( C \subset A \), but only on the set of all \( C_i \subset A, \) for which there is a smallest \( t \in T \) with \( C_i \subset A_t \).

3.7. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid.

If \( C \) and \( D \) are equiordered, then \( C \) and \( F := \Lambda(C \cup D) \), as well as \( D \) and
F are equiordered.

Proof. If C and D are equiordered, then there is a smallest $t \in T$ such that

(a) $C, D \subseteq A_t$, and
(b) there is no $s \in T, s < t$, with $C \subseteq A_s$ or $D \subseteq A_s$.

Condition (a) implies $F \subseteq A_t$, because F is the smallest sigma-field containing both $C$ and $D$ as subsets. Condition (b) implies that there is no $s \in T, s < t$, with $C \subseteq A_s$ or $F \subseteq A_s$. This implies that $C$ and $F$ as well as $D$ and $F$ are equiordered.

3.8. THEOREM

Let the assumptions and notations of 3.2 and 3.4 be valid.

(i) If $C$ is preordered to $D$, and $D$ and $F$ are equiordered to $C$, then $C$ is preordered to $F$.

(ii) If $C$, $D$ are equiordered and $D$ is preordered to $F$, then $C$ is preordered to $F$.

(iii) Propositions (i) and (ii) also hold for stochastic variables $X, W$ and $U$ taking the roles of $C$, $D$ and $F$, respectively.

Proof.

(i) In this case,

(a) there is an element $s \in T$ with $C \subseteq A_s, D \not\subseteq A_s$, and an element $t \in T, t > s$, with $D \subseteq A_t, D \not\subseteq A_t$, for all $s \in T, s < t$.

Furthermore,

(b) $F \subseteq A_t$ and there is no $s \in T, s < t$, with $D \subseteq A_s$ or $F \subseteq A_s$, for all $s \in T, s < t$.

(a) and (b) imply that there is a $s \in T$ with $C \subseteq A_s, D \not\subseteq A_s$, and an element $t \in T, t > s$, with $F \subseteq A_t, F \not\subseteq A_t$, for all $s \in T, s < t$, which implies that $C$ is preordered to $F$ (see 3.2).

(ii) Is obvious.

(iii) This proposition immediately follows from 3.8i and 3.8ii, if we define $C := U^{-1}(A^c)$, $D := U^{-1}(A)$, and $F := U^{-1}(A^c)$.

Another concept needed in the following sections is that of equi-or preorderedness, where 'or' denotes the logical disjunction. Equi- or preorderedness is a weak order relation. Again, the propositions are always made with respect to an underlying family $(A_t, t \in T)$ of sigma-fields, the explicit reference to which is omitted only for reasons of convenience.

3.9. DEFINITION

Let the assumptions and notations of 3.2 be valid.

(i) We say that $C$ is equi- or preordered to $D$, iff

(a) $C$ and $D$ are equiordered, or
(b) $C$ is preordered to $D$.

(ii) We say that $X$ is equi- or preordered to $W$, iff $C$ is equi- or preordered to $D$ and

(c) $C = X^{-1}(A')$ and $D = W^{-1}(A')$.

3.10. THEOREM

Let the assumptions and notations of 3.2 and 3.4 be valid.

(i) If there is a $t \in T$ with $C \subseteq A_t$ such that there is no $s \in T, s < t$, with $C \subseteq A_s$, then $C$ is equi- or preordered to itself (reflexivity).

(ii) If $C$ is equi- or preordered to $D$, and $D$ is equi- or preordered to $C$, then $C$ and $D$ are equiordered (antisymmetry).

(iii) If $C$ is equi- or preordered to $D$, and $D$ is equi- or preordered to $F$, then $C$ is equi- or preordered to $F$ (transitivity).

(iv) Equi- or preorderedness of sigma-fields is a weak order relation on the set of all sigma-fields $C \subseteq A_t$, for which there is a $t \in T$ with $C \subseteq A_t$ such that there is no $s \in T, s < t$, with $C \subseteq A_s$.

(v) Propositions 3.10i to 3.10iv also hold for equiorderedness of stochastic variables $X, W, U$ taking the roles of $C, D, F$, respectively.

Proof.

Property 3.10i follows from 3.6i. In order to prove 3.10ii, we consider four cases. First, both $C$ and $D$ are preordered to another, which can be excluded, because of 3.4i. Second, $C$ and $D$ are equiordered. In this case, the proposition 3.10ii is implied by 3.6ii. Third, $C$ is preordered to $D$ and $D$ is equiordered to $C$. This contradicts Condition (b) of Definition 3.3. Fourth, $D$ is preordered to $C$, and $C$ is equiordered to $D$. Again, this case cannot occur, because of 3.3b.

The proof of property (iii) follows the same line of arguments as that of property 3.10i. For the first case that $C, D$ and $F$ are equiordered, the property is implied by 3.6i. For the second case that $C$ is preordered to $D$ and $D$ is preordered to $F$, Property 3.10ii is
inplied by 3.4ii). For the third case that \( C \) is reordered to \( D \), and \( D \) and \( F \) are equiordered, it is implied by 3.8i, and for the last case that \( C \) and \( D \) are equiordered, whereas \( D \) is reordered to \( F \), it is implied by 3.8ii. (iv) A weak order relation is defined by reflexivity, antisymmetry, and transitivity. (v) This proposition follows from (i) to (iv), if we define \( C := X^{-1}(A') \), \( D := U^{-1}(A') \), and \( F := U^{-1}(A'^{\prime}) \).

3.11. THEOREM

Let the assumptions and notations of 3.2 and 3.4 be valid. If \( C \) is equi- or reordered to \( D \), then \( C \) is also equi- or reordered to \( F := A(C \cup D') \).

Proof. 
(a) If \( C \) is reordered to \( D \), then, according to 3.5i, \( C \) is also reordered to \( F \).
(b) If \( C \) and \( D \) are equiordered, then, according to 3.7i, \( C \) and \( F \) are also equiordered. (a) and (b) imply the theorem.

4. POTENTIAL DISTURBING SIGMA-FIELDS AND VARIABLES

By intuition, it is obvious that reorderedness of \( X \) to \( Y \) is a necessary but not sufficient condition for the stochastic dependence of \( Y \) on \( X \) to be causal. Two alternative second necessary conditions, the invariance and average conditions have been discussed by Steyer (1984). To state either of them precisely, we need a formal definition of a potential disturbing variable and a potential disturbing sigma-field. The aim is to define variables which are possibly confounded with \( X \) and thus make the dependence of \( Y \) on \( X \) a spurious one.

In the following definition we need the concept of probability measure that is nontrivial with respect to a sigma-field \( C \), which is defined to be a probability measure \( \Omega \) on \( C \) such that there is at least one \( C \in C \) with \( \Omega > \Omega(C) < 1 \).

4.1. DEFINITION

Let \((\Omega, A, P)\) be a probability space, let \((A_t, t \in T), T \subset \mathbb{R}\), be a monotonically increasing family of sigma-fields with \( A(\bigcup_{t \in T} A_t) \subset A \), let \( C, D, F \subset A \) denote sigma-fields, and finally let \( X, W \) be \((n', A')\)-, \((n'', A'')\)-stochastic variables on \((\Omega, A, P)\), respectively.

(i) \( D \) is called potential disturbing sigma-field with respect to \((A_t, t \in T)\) and \( C \) (or, if \( C = X^{-1}(A') \), with respect to \((A_t, t \in T)\) and \( X \)), iff the following two conditions hold:
(a) \( D \) is equi- or reordered to \( C \), and
(b) there is a probability measure \( \Omega \) on \( A \) that is nontrivial with respect to both \( C \) and \( D \) such that \( C \) and \( D \) are stochastically independent.

(ii) \( H \) is called a potential disturbing variable with respect to \((A_t, t \in T)\) and \( C \) (or, if \( C = X^{-1}(A') \), with respect to \((A_t, t \in T)\) and \( X \)), iff Conditions (a) and (b) hold, as well as
(c) \( D = U^{-1}(A'') \).

Condition 4.1a is chosen in order to exclude mediator variables from the set of potential disturbing variables. Mediator variables, say \( U \), are variables for which \( X \) is reordered to \( U \) and \( U \) is reordered to \( Y \). These variables do not threaten the validity of a causal proposition but may elaborate it. They are important to distinguish between direct and total effects. Condition 4.1b is added to exclude stochastic variables like \( X, X', Y \), etc. from the set of potential disturbing variables. It is not possible to construct a probability measure \( \Omega \) on \( A \) that is nontrivial with respect to \( A(X) \) and \( A(X') \), for example, such that \( X \) and \( X' \) are stochastically independent with respect to \( \Omega \). Furthermore, if \( U \) is a potential disturbing variable, then \( X+U, X-U \), for example, are not potential variables.

4.2. EXAMPLE

Consider a study on the effectiveness of a retraining program on the rate of recidivism (for a more detailed description of the example see Steyer (1984)). We might choose
\[
\Omega = \{(a_i, b_j, c_k): i, j, k = 0, 1\},
\]
where \((a_i, b_j, c_k)\) represents one of the eight possible elementary events that a parolee is male or female \((a_i)\), takes part at the retraining program or not \((b_j)\), and commits a crime or not after the retraining \((c_k)\). We choose the sigma-field \( A \) on \( \Omega \) to be the set of all subsets of \( \Omega \), which consists of \( \emptyset \), the eight events \((\{(a_i), b_j, c_k\})\) and all unions of these sets. Further, we might choose the stochastic variables...
If we follow this second path of interpreting the formal theory in the study on the effectiveness of the retraining with parolees, all properties of the parolee before the retraining are potential disturbing variables, e.g. the kind of the parolee's crime, the kind of experiences in jail, etc. Obviously, a causal proposition is in full accord with our intuition only for this second kind of interpretation of the sigma-fields $A_t$, allowing for all potential disturbing variables, disturbing variables which are there, even though they might not be observed.

However, so long as one mentions with respect to which family $(A_t, t \in T)$ a causal proposition is made, causal propositions as defined in the next section are meaningful and not trivial, so long as at least one potential disturbing variable is allowed for. The more potential disturbing variables are implicitly or explicitly allowed for by the interpretation of the sigma-fields $A_t, t \in T$, the more meaningful a causal proposition is. Only if the $A_t$ are interpreted to represent all events that may occur up to point $t$ of time inclusively, in the real process considered, will causal propositions fully coincide with the intuitive meaning of the term 'causal'. It should be noted, however, that the formal theory proposed is neutral with respect to these questions of adequate application.

5. CAUSAL LINEAR STOCHASTIC DEPENDENCE

Using the concepts of pre- and enuiergorssedness and of potential disturbing variables treated in previous sections, as well as that of a conditional expectation, we may now formally define causal linear stochastic dependence and study its formal properties. As not only causal linear stochastic dependence, but also independence is of interest, we first define causal linear stochastic dependence, which means dependence or independence. Independence additionally requires $E(Y|C) = E(Y)$, and dependence the negation of it.

5.1. DEFINITION

Let $Y$ be a real-valued stochastic variable on the probability space $(\Omega, A, P)$ with finite expectation $E(Y)$, let $X$ be a $(\alpha, A^*)$-stochastic variable and $H$ be a $(\eta, A')$-stochastic variable on $(\Omega, A, P)$, let $C \subset A$ be a sigma-field, let $(A_t, t \in T)$ be a monotonically increasing family of sigma-fields with $A(t) := \bigcup_{s \leq t} A_s \subset A$, and let $E(Y|C, P) := E(Y|A(C \cup D))$ denote the conditional expectation of $Y$ given $A(C \cup D)$. We say that $Y$ is causally linearly...
stochastically independent on \( C \) with respect to \( (A_t, t \in T) \), if all the following conditions hold:

(a) \( C \) is preordered to \( A(Y) \) with respect to \( (A_t, t \in T) \) (preorderedness).

(b) For all potential disturbing sigma-fields \( D \) with respect to \( (A_t, t \in T) \) and \( C \),

\[
E(Y|C,D) = E(Y|C) + H^C
\]

holds, where \( H^C \) denotes the composition of an \( A''\)-measurable real-valued function \( H \) with a \((P,A')\)-measurable stochastic variable \( U \) (invariance). If Conditions (a) and (b), as well as

(c) \( E(Y|C) = E(Y) \) (linear stochastic independence) hold, we say that \( Y \) is causally linearly stochastically independent from \( D \) with respect to \( (A_t, t \in T) \), and dependent on \( C \), otherwise.

If Conditions (a) and (b) hold, as well as (d) \( C = X^{-1}(A') \), we also say that \( Y \) is causally linearly stochastically independent on \( X \) with respect to \( (A_t, t \in T) \).

If no confusion is possible, we may omit the explicit reference to \( (A_t, t \in T) \).

There are a number of situations in which Equation (1) holds. Suppose for example, that the conditional expectation of the \( Y \) given \( C \) and \( D \) is the sum of a composition \( F(X) \) and a composition \( H^C \),

\[
E(Y|C,D) = F(X) + H^C
\]

where \( F \) is an \( A'\)-measurable and \( H \) an \( A''\)-measurable real-valued function, whereas \( X, W \) are \((C,A'),(D,A')\)-measurable, respectively. This will be referred to as the additivity condition, a special case of which is the multiple regression equation

\[
E(Y|X,W) = a_{01} + a_{02}X + b_{01}W.
\]

The additivity condition also holds, for example, if

\[
E(Y|X,W) = a_{01} + a_{02}X + b_{01}W + b_{02}X^2 + b_{03}W^2 + b_{04}XW^2,
\]

whereas it does not hold, for example, if

\[
E(Y|X,W) = a_{01} + a_{02}X + b_{01}W + b_{02}(X-W)X + b_{03}(X-W)W
\]

and \( b_{02}(X-W) \neq 0 \).

This last equation contradicts the additivity condition, because of the multiplicative term \( X \cdot W \), which is neither \( C \)- nor \( D \)-measurable.

5.2. Theorem

Let the presumptions and notations of 5.1 be valid. If (a) for all potential disturbing sigma-fields \( D \),

\[
E(Y|C,D) = F(X) + H^C,
\]

where \( F \) is an \( A'\)-measurable real-valued function, \( X \) is \((C,A')\)-measurable (additivity); if (b) \( X \) and \( W \) are stochastically independent, and if (c) \( C \) is preordered to \( A(Y) \) with respect to \( (A_t, t \in T) \), then \( Y \) is causally linearly stochastically independent on \( C \).

Proof.

In order to prove this proposition, we suppose that the expectation \( E(H^C) \) is equal to zero. Note that this does not restrict generality, because, by subtracting its expectation, any composition \( H^C \) with an expectation unequal zero can easily be transformed into a composition \( H^C \) with expectation zero. It is easily seen that the conditions (a) and (b) of theorem 5.2 imply Equation (1), because

\[
E(Y|X) = E(E(Y|C,D)|X)
\]

(see 4.17)

\[
= E(F(X,H^C)|X)
\]

(see 2)

\[
= E(F(X)|X) + E(H^C)|X)
\]

(see 4.15)

\[
= E(F(X)) + E(H^C) = F(X).
\]

(see 4.9)

where \( E(H^C) = 0 \) and the theorem is used that stochastic independence of \( Y \) and \( X \) implies \( E(H^C|X) = E(H^C) \). We may now insert \( F(X) \) as \( E(Y|X) \) into Equation (2), which yields Equation (1), which proves the theorem.

We now turn to another situation in which \( Y \) is causally linearly stochastically independent on \( C \). Suppose \( E(Y|X,W) = E(Y|X) \). In this situation, too, Equation (1) is fulfilled. \( Y \) being a constant is a special case in which \( E(Y|X,W) = E(Y|X) \) is true. Hence, if there is the possibility that a potential disturbing variable is an actual one, it might be held constant at one of its values. Thus, the experimental control technique of holding potential disturbing variables constant, can be based on the invariance condition. It should be noted, however, that the validity of a causal
proposition is then restricted to the case that \( W \) is constant at its value \( W \).

5.3. THEOREM

Let the presumptions and notations of 5.1 be valid. If

(a) \( C \) is preordered to \( A(Y) \) with respect to \( (A_t, t \in T) \), and if

(b) for all potential disturbing sigma-fields \( D \),

\[
E(Y|C,D) = E(Y|C),
\]

(C-conditional linear stochastic independence of \( Y \) from \( D \)), then \( Y \) is causally linearly stochastically independent on \( C \).

Proof.

Equation (3) implies Equation (1), because we may define \( H \) and \( W \) such that \( H W = 0 \).

5.4. COROLLARY

Let the presumptions and notations of 5.1 be valid. If \( C \) is preordered to \( A(Y) \) with respect to \( (A_t, t \in T) \) and

\[
E(Y|C) = Y
\]

(complete dependence of \( Y \) on \( C \)), then \( Y \) is causally linearly stochastically independent on \( C \).

Proof.

\[
E(Y|C) = Y \text{ implies } E(Y|C,D) = E(E(Y|C),D) = E(Y|C), \text{ for all sigma-fields } D \subseteq A,
\]

which implies 5.3b.

Hence, causal linear stochastic dependence does not require all of the variation of the dependent variable \( Y \) to be determined by \( C \) (see Equation (1)), nor is it necessary that all influencing variables that are equi- or preordered to \( C \) are known (see Equation (3)). It is only required that all potential disturbing sigma-fields \( D \) behave as described by Equation (1).

This equation is fulfilled already if \( Y \) is C-conditionally linearly stochastically independent from \( D \), i.e. if Equation (3) holds, or if \( X \) and \( W \) are stochastically independent and additivity can be assumed.

6. SIMPLE CAUSAL REG- LINEAR DEPENDENCE

In the case of a simple reg-linear dependence of \( Y \) on \( X \) which is characterized by the simple regression equation

\[
E(Y|X) = a_{y0} + a_{yx}X,
\]

the invariance condition implies that, for all potential disturbing variables \( W \),

\[
E(Y|X,W) = a_{y0} + a_{yx}X + H W.
\]

Whatever the type of the function \( H W \), the regression coefficient \( a_{yx} \) is left unchanged when turning from Equation (5) to (6), provided the invariance condition holds.

Another remarkable property of the simple regression coefficient \( a_{yx} \) is that it is equal to the corresponding regression coefficient \( a_{yx}\mid W \) of the conditional simple regression of \( Y \) on \( X \) given \( W \):

\[
E(Y|X,W) = a_{y0}\mid W + a_{yx}\mid W X,
\]

if the invariance condition holds. This is easily seen from Equation (6), because, for \( W \),

\[
E(Y|X,W) = a_{y0} + a_{yx}X + H(w),
\]

where \( H(w) \) is a constant so that we may define \( a_{y0}\mid W = a_{y0} + H(w) \) and \( a_{yx}\mid W = a_{yx} \). Thus, Equation (5) and the invariance condition imply, that \( a_{yx} \) is also the regression coefficient of the conditional simple regression of \( Y \) on \( X \) given \( W \), where \( W \) is any potential disturbing variable.

This motivates the following

6.1. DEFINITION

Let the presumptions and notations of 5.1 be valid. We say that \( Y \) is simply causally reg-linearly independent on \( X \) and that \( a_{yx} \) is the simple causal reg-linear effect of \( X \) on \( Y \), iff \( Y \) is causally linearly stochastically independent on \( X \) and Equation (5) holds.
7. WEAK CAUSAL LINEAR STOCHASTIC DEPENDENCE

It may be argued that the kind of causal linear stochastic dependence discussed above is unrealistically restrictive. Let us consider once again the invariance condition saying that, for all potential disturbing signalfields $\mathcal{O}$,

$$E(Y|\mathcal{O}) = E(Y|C) + H\mathcal{O},$$

where $H\mathcal{O}$ denotes the composition of an $\mathcal{A}^m$-measurable real-valued function $H$ with a $(\mathcal{D},\mathcal{A}^m)$-measurable stochastic variable $\mathcal{O}$. If, for example, $C$ is generated by the real-valued $N(J)$-dimensional stochastic variable $X = (X_1, \ldots, X_{N(J)})$, it is easily seen that there are no multipliative terms involving a $\mathcal{D}$-measurable function and one or more of the random variables $X_j$, $j \in J = \{1, \ldots, N(J)\}$. Therefore, the invariance condition excludes interactions in the analysis of variance sense of any order between the variables $X_j$ and any potential disturbing variable $\mathcal{O}$. This condition is a rather strict one that cannot be guaranteed to hold even in randomized experiments, where the term 'causal' seems to be appropriate, too, to characterize the dependence of $Y$ on the treatment variables $X_j$ indicating membership to the experimental groups.

Therefore, a weak type of causal linear stochastic dependence is now introduced. All that has to be done is to replace the invariance by the average condition, which can be formulated as follows: for all potential disturbing variables $\mathcal{O}$, the following equation holds for $P_{\mathcal{O}}$-almost all $E(Y|X=x)$:

$$E(Y|X=x) = \int E(Y|X=x, \mathcal{O}=w) \, P_{\mathcal{O}}(dw),$$

where $P_{\mathcal{O}}$ denotes the distribution of $\mathcal{O}$. If $\mathcal{O}$ is a discrete random variable, this equation can be written

$$E(Y|X=x) = \sum_{w} E(Y|X=x, \mathcal{O}=w) \cdot P(\mathcal{O}=w),$$

where the summation is over all values $w$ of $\mathcal{O}$. This average condition simply means that $P_{\mathcal{O}}$-almost all conditional expected values $E(Y|X=x)$ are the average of the conditional expectations $E(Y|X=x, \mathcal{O}=w)$ of $Y$ given $X=x$ and $\mathcal{O}=w$ across the values $w$ of $\mathcal{O}$. Together with preorderedness, the average condition defines weak causal linear stochastic dependence of $Y$ on $X$.

7.1. DEFINITION

Let the presumptions and notations of 5.1 be valid. We say that $Y$ is weakly causally linearly stochastically independent on $X$, i.e.

(a) $X$ is preordered to $Y$ with respect to $(A_+, t \in T)$ (preorderedness).

(b) For all potential disturbing variables $\mathcal{O}$ with respect to $X$ and $(A_+, t \in T)$

$$E(Y|X=x) = \int E(Y|X=x, \mathcal{O}=w) \, P_{\mathcal{O}}(dw),$$

for $P_{\mathcal{O}}$-almost all $x \in \mathcal{O}$ (average condition). If additionally

(c) $E(Y|X) = E(Y)$ (linear stochastic independence),

we say that $Y$ is weakly causally linearly stochastically independent from $X$ and, dependent on $X$, otherwise.

A special situation in which Equation (7) is fulfilled is characterized by $E(Y|X, \mathcal{O}) = E(Y|X)$, which defines $X$-conditional linear stochastic independence of $Y$ from $\mathcal{O}$.

7.2. THEOREM

Let the presumptions and notations of 5.1 be valid. If

(a) $X$ is preordered to $Y$ with respect to $(A_+, t \in T)$ and

(b) for all potential disturbing variables $\mathcal{O}$,

$$E(Y|X, \mathcal{O}) = E(Y|X),$$

then $Y$ is weakly causally linearly stochastically independent on $X$.

Proof.

Equation (8) implies $E(Y|X=x, \mathcal{O}) = E(Y|X=x)$, for $P_{\mathcal{O}}$-almost all $(x, w) \in \mathcal{O} \times \mathcal{O}$. Hence,

$$\int E(Y|X=x, \mathcal{O}=w) \, P_{\mathcal{O}}(dw) = E(Y|X) \cdot P_{\mathcal{O}}(dw) = E(Y|X).$$

The importance of the average condition, however, stems from the fact that it is implied by the stochastic independence of all potential disturbing variables $\mathcal{O}$ at one side and the $N(J)$-dimensional variable $X = (X_1, \ldots, X_{N(J)})$ on the other side.

7.3. THEOREM

Let the presumptions and notations of 5.1 be valid. If
(a) X is preordered to Y and
(b) X and all potential disturbing variables are stochastically independent, then Y is weakly causally linearly stochastically independent on X.

Proof.
A.17 yields $E[E(Y|X,W)|X]=E(Y|X)$, which implies
(a) $E(Y|X=x) = E(E(Y|X,W)|X=x)$, for $P_x$-almost all $x \in \Omega'$.

Stochastic independence of X and Y then implies
$E(Y|X=x) = E(E(Y|X=x,W)) = \int E(Y|X=x,W=w) P_w(dw)$
where the last line follows from (a) and 5.3.22 of Gänssler and Stute (1977), p. 149).

In a randomized experiment, $X = (X_1, \ldots, X_N)$ may represent the vector of treatment variables indicating the experimental group to which an experimental unit (subject) belongs. Randomization implies that all potential disturbing variables $W$ (representing any properties of the experimental units before or at the treatment) and the treatment variable $X = (X_1, \ldots, X_N)$ are stochastically independent. Therefore, the average condition holds in randomized experiments. Thus, weak causal linear stochastic dependence (or independence, if $E(Y|X) = E(Y)$) is guaranteed by randomization (for an example see Steyer (1984)). Matching, of course, serves the same purpose, but only for those potential disturbing variables $W$ with respect to which the experimental groups are matched. Thus, the control techniques of randomization and matching can be based on the theory proposed, too.

However, and maybe more importantly, the theory of causal linear stochastic dependence is also relevant for nonexper imental studies, because both the invariance and the average condition may be tested, once a potential disturbing variable $W$ is observed. If neither holds, any causal hypothesis should be rejected for the model considered. If one or both of them hold, one may maintain the causal hypothesis (in its weak or its strict form) as long as no potential disturbing variable is found for which the invariance and the average conditions do not hold. Testing if the average condition (see Equation (7)) holds is a test of weak causal linear stochastic dependence and testing if the invariance condition (see Equation (2)) holds is a test of causal linear stochastic dependence (in the stricter sense). Of course, such tests make sense only if X is assumed to be preordered to Y.

In the last theorem, it is shown that causal linear stochastic independence implies weak causal linear stochastic independence.

7.4. THEOREM

Let the presumptions and notations of 5.1 be valid. If Y is causally linearly stochastically dependent on X, then Y is also weakly causally linearly stochastically dependent on X.

Proof.
We have to show that Equation (1) implies Equation (7). Equation (1) implies:

\[ \begin{align*}
E(Y|X=x,W) &= E[E(Y|X,W)|X=x,W] \\
\end{align*} \]

Hence, $E(Y|X=x) = E(Y|X=x,W) - H(W)$ and

\[ \begin{align*}
E(Y|X=x) &= \int E(Y|X=x,W) P_w(dw) - \int H(W) P_w(dw) \\
&= \int E(Y|X=x,W) P_w(dw) - \int H(W) P_w(dw) = \int E(Y|X=x,W) P_w(dw),
\end{align*} \]

because $\int H(W) P_w(dw) = E(Ho|W) = 0$, which is easily seen from

\[ E[E(Y|X,W)] = E[E(Y|X)] + E(Ho|W), \]

which implies $E(Ho|W) = 0$, because $E[E(Y|X,W)] = E[E(Y|X)] = f(Y).

For examples illustrating the application of the theory proposed, as well as for a discussion relating this theory to simultaneous equation modeling, the reader is referred to Steyer (1984).

APPENDIX

In this appendix we presented the definition of, and some theorems pertinent to the conditional expectation $E(Y|C)$ of Y given C that are frequently cited in this chapter. For an introduction into the concept and its background see e.g., Bauer (1974), Breiman (1961), Gänssler and Stute (1977), or Loève (1977, 1978). A shorter and therefore more convenient name for $E(Y|C)$ is C-conditional expectation of Y. The C-conditional expectation of Y is a very general and useful concept. It is used, for example, to define the C-conditional probability $P(\\cdot|C)$ of an event $\\cdot$, as well as the C-conditional variance $V(Y|C)$ and covariance $C(Y,Z|C)$ of stochastic variables. Special C-conditional expectations are also obtained, if C is the sigma-field generated by a stochastic variable X or by a family $(X_j, j \in J)$ of stochastic variables. In these cases, the notations $E(Y|X), E(Y|X_j, j \in J)$, etc. may be used.
The mathematical definition given below does not appeal very much to intuitive insight. Therefore, it might be helpful to recall that \( E(Y|X) \) is a stochastic variable, the values of which are identical with the conditional expectations \( E(Y|X=x) \) given \( X=x \). Another way to think about \( E(Y|X) \) is that it is a stochastic variable consisting of the best predictions of \( Y \) given a value \( x \) of \( X \). In many textbooks on applied statistics, the variable consisting of the best predictions of \( Y \) is denoted by \( \bar{Y} \). \( E(Y|X) \) may be thought of as the mathematical equivalent of \( \bar{Y} \). Similarly, \( E(Y|C) \) consists of the best predictions of \( Y \) based on the sigma-field \( C \).

**Definition 1**

Let \( Y: \Omega \to \mathbb{R} \) be a real-valued stochastic variable on the probability space \((\Omega, \mathcal{A}, P)\) with finite expectation \( E(Y) \) and let \( C \subset A \) be a sigma-field. The stochastic variable \( E(Y|C): \Omega \to \mathbb{R} \) is called the \( C \)-conditional expectation of \( Y \), iff \( E(Y|C) \) is \( C \)-measurable and

\[
E[1_C E(Y|C)] = E[1_C Y], \quad \text{for all } C \subset C.
\]

(1)

Let \( A \in \mathcal{A} \) be an event and \( 1_A \) its indicator function. \( P(A|C): \Omega \to \mathbb{R} \) is called the \( C \)-conditional probability of \( A \), iff

\[
P(A|C) = E(1_A|C).
\]

(2)

Let \( X \) be a \((\Omega, A, \mathbb{P})\)-stochastic variable on \((\Omega, A, \mathbb{P})\). \( E(Y|X) \) is called the \( X \)-conditional expectation of \( Y \), iff \( C = A(X, A') \) is the sigma-field generated by \( X \) and \( A' \), and

\[
E(Y|X) = E(Y|C).
\]

(3)

Let \((X_j, j \in J)\) be a sequence of \((\Omega, A, \mathbb{P})\)-stochastic variables on \((\Omega, A, \mathbb{P})\). \( E(Y|X_j, j \in J) \) is called the \((X_j, j \in J)\)-conditional expectation of \( Y \), iff \( C = A(\bigcup_{j \in J} A(X_j, A_j)) \) is the sigma-field generated by \( \{X_j, A_j\}, j \in J \), and

\[
E(Y|X_j, j \in J) = E(Y|C).
\]

(4)

\(E(Y|X_1, \ldots, X_{N(J)})\) is called the \((X_1, \ldots, X_{N(J)})\)-conditional expectation of \( Y \), iff \( C = A(\bigcup_{j \in J} A(X_j, A_j)) \) is the sigma-field generated by \( \{X_j, A_j\}, j \in J \), where \( J = \{1, \ldots, N(J)\} \), \( N(J) \in \mathbb{N} = \{1, 2, \ldots\} \), and

\[
E(Y|X_1, \ldots, X_{N(J)}) = E(Y|C).
\]

(5)

Note that \( E(Y|C) \) is the general concept. Hence, if propositions are true for \( E(Y|X) \), then they also hold for \( P,A(C), E(Y|X), E(Y|X_j, j \in J) \), and \( E(Y|X_1, \ldots, X_{N(J)}) \). Through the definition above, the \( C \)-conditional expectation \( E(Y|C) \) is uniquely determined only with probability one. Therefore, there are generally different versions of \( E(Y|C) \), which are, however, equivalent, with probability 1. Equations on \( E(Y|C) \) are therefore true only almost surely in general, which will be abbreviated by the symbol 'as-'.

Notice that there is a difference between the \( C \)-conditional expectation \( E(Y|C) \) of \( Y \), i.e. the conditional expectation of \( Y \) with respect to the sigma-field \( C \) and the conditional expectation \( E(Y|C) \) of \( Y \) given the event \( C \), and correspondingly, a difference between the conditional probability \( P(A|C) \) of the event \( A \) with respect to the sigma-field \( C \) and the conditional probability \( P(A|C) \) of the event \( A \) given the event \( C \). \( E(Y|C) \) and \( P(A|C) \) are stochastic variables, whereas \( E(Y|C) \) and \( P(A|C) \) are real-valued constants.

In the following theorem, a number of propositions on \( E(Y|C) \) are gathered. Some of them are on conditions under which \( E(Y|C) \) is equal to a constant function. In these and related contexts, the symbol for the function and the constant (its values) will be the same. The symbol 'a', for example, does not only denote a real constant, but also a function \( a: \Omega \to \mathbb{R} \), taking the value \( a \) for all \( \omega \in \Omega \). The notation \( E(Y|C) = 0 \), almost sure, in Proposition (II) is an abbreviation for \( E(Y|C)(\omega) = 0(\omega) \), for almost all \( \omega \in \Omega \), where 0: \( \Omega \to \mathbb{R} \) is a function defined by \( 0(\omega) = 0 \), for all \( \omega \in \Omega \).

**Theorem 1**

Let \( Y \) be a real-valued stochastic variable on the probability space \((\Omega, A, \mathbb{P})\) with finite expectation \( E(Y) \). If \( E(Y|C) \) is the \( C \)-conditional expectation of \( Y \), then the following propositions are true:

\[
E[1_C E(Y|C)] = E(Y)
\]

(6)

\[
E(Y|C) = E(Y), \text{ if } C = (\Omega, \varnothing), \text{ where } E(Y): \Omega \to \mathbb{R} \text{ is defined by } E(Y)(\omega) = E(Y), \text{ for all } \omega \in \Omega.
\]

(7)

\[
E(Y|C) = E(Y), \text{ if } C = (\Omega, \varnothing), \text{ where } E(Y) = E(Y)|C.
\]

(8)

\[
E(Y|C) = a, \text{ if } Y \text{ is } C\text{-measurable}
\]

(9)

\[
E(Y|C) = a, \text{ if } Y \text{ is } C\text{-measurable}
\]

(10)

\[
E(Y|C) = 0, \text{ almost sure, if } Y \geq 0, \text{ almost sure, where 0: } \Omega \to \mathbb{R} \text{ is defined by } 0(\omega) = 0, \text{ for all } \omega \in \Omega.
\]

(11)
According to Equation (6), the expectation of the C-conditional expectation of Y is equal to the expectation of Y. Proposition (7) shows that the C-conditional expectation of Y is equal to the constant function E(Y): R → R, if C = ∅. According to Equation (8), the expectation of the difference Y - E(Y|C) of Y and its C-conditional expectation is zero. Note that the difference Y - E(Y|C), called the error or residual of Y with respect to E(Y|C), plays an important role in applications. Proposition (9) is often applied. Y is C-measurable, for example, if C is the sigma-field generated by the (Oi, ai) stochastic variables Xj, j ∈ J, and Y = Xj, j ∈ J, or Y = Xj + Xk, where j, k ∈ J. Hence, special cases of Proposition (9) are, for example,

E(Y|X) = Y,
E(Xj|Xj, j ∈ J) = Xj, if j ∈ J,
E(Xj + Xk|Xj, j ∈ J) = Xj + Xk, if j, k ∈ J.

The following theorem contains propositions on E(Y|C), where Y is the weighted sum or the product of other stochastic variables. Again, the notation Y ∈ Z, almost sure, which occurs in Proposition (16), is an abbreviation for Y(ω) ∈ Z(ω), for almost all ω ∈ ∅.

THEOREM 2

Let Y and Z be real-valued stochastic variables on the probability space (Ω, P) with finite expectations E(Y) and E(Z), respectively, and let C, C_0 ⊆ A be two sigma-fields. If E(Y|C) and E(Z|C) are the C-conditional expectations of Y and Z, respectively, then the following propositions are true:

E(Y|C) = E(Z|C), if Y = Z. (12)
E(Y - Z|C) = E(Y - Z|C), if Y is C-measurable and if E(Y - Z) is finite. (13)
E(Y|C_0 - Z|C) = E(Y|C_0) - E(Z|C), if C_0 ⊆ C. (14)
E(Y - Z|C) = E(Y - Z|C), if a, b ∈ R. (15)
E(Y|C) ∈ E(Z|C), almost sure, if Y ∈ Z, almost sure. (16)

THEOREM 3

Let Y be a real-valued stochastic variable on the probability space (Ω, P, M) with finite expectation E(Y). If E(Y|C) is the C-conditional expectation of Y and C_0 ⊆ C is a sigma-field, then the following equations are true:

E(E(Y|C_0)|C) = E(Y|C_0) = E(E(Y|C_0)|C), (17)
E(Y - E(Y|C_0)) = 0. (18)

The sigma-field C_0 will be a subset of the sigma-field C, for example, if C is generated by the (Oi, ai) stochastic variables Xj, j ∈ J, and C_0 by the (Oi, ai) stochastic variables X_k, k ∈ K, where K ⊆ J. Hence, special cases of the Equations (17) are

E(E(Y|X_1 X_2)|X_1) = E(Y|X_1) = E(E(Y|X_1|X_1 X_2),
E(E(Y|X_j, j ∈ J)|X_k, k ∈ K) as = E(Y|X_k, k ∈ K) as = E(E(Y|X_k, k ∈ K)|X_j, j ∈ J).

Equation (16) reveals that the C_0-conditional expectation of the error or residual F = Y - E(Y|C) is zero, if C_0 ⊆ C. This proposition on F is much stricter than that of Equation (8), according to which the (unconditional) expectation of F is zero.

REFERENCES

