Open Stochastic Systems

Jan C. Willems *Life Fellow, IEEE*

***Abstract*—The problem of providing an adequate definition of a stochastic system is addressed and motivated using examples. A stochastic system is defined as a probability triple. The specification of the set of events is an essential part of a stochastic model and it is argued that for phenomena with as outcome space a finite dimensional vector space, the framework of classical random vectors with the Borel sigma-algebra as events is inadequate even for elementary applications. Models very often require a coarse event sigma-algebra. A stochastic system is linear if the events are cylinders with fibers parallel to a linear subspace of a vector space. We address interconnection of stochastic systems. Two stochastic systems can be interconnected if they are complementary. We discuss aspects of the identification problem from this vantage point. A notion that emerges is constrained probability, a concept that is reminiscent but distinct from conditional probability. We end up with a comparison of open stochastic systems with probability kernels.**

***Index Terms*—Stochastic system, linearity, gaussian system, interconnection, system identification, constrained probability.**

I. INTRODUCTION

Open systems and their interconnection lie at the heart of system theory. By an 'open' system we mean a model that incorporates the influence of the environment explicitly, as an unmodeled feature. We view interconnection as 'variable sharing': before interconnection, the variables pertaining to the interconnected subsystems regarded as independent, while after interconnection some of subsystem variables are required to be equal.

The aim of the present paper is to present open systems in a stochastic setting. Our interest is mainly in systems with as outcome space $\mathbb{R}^{\mathtt{n}}$ or a subset of $\mathbb{R}^{\mathtt{n}}$. If the corresponding event space consists of the Borel sets, or of all subsets of the outcome space if this space is countable, then we call the σ-algebra of events 'rich' or 'fine', in contrast to 'coarse' σ-algebras. As we shall see, openness of systems requires probability spaces with a coarse σ-algebra of events, in contrast to what we call 'classical' stochastic systems, where the σ-algebra of events is assumed to be rich. The theme developed in this paper is that the σ-algebra of events should not be taken for granted, but is a not to be ignored feature of the stochastic phenomenon that is modeled.

This is not a paper about the interpretation or about the mathematical foundations of probability. The article functions completely in the orthodox measure theoretic setting of probability with a σ-algebra of events and countable additivity, the mathematical framework of probability theory usually attributed to Kolmogorov [1]. The main point of the paper is pedagogical in nature, namely that the usual emphasis in the teaching of probability on settings where *essentially* every subset of the outcome space is an event is unduly restrictive, even for elementary applications. We shall show that a coarse σ-algebra of events is needed in order to study open systems and their interconnection. Related concepts, such as linearity and constrained probability, also function comfortably only within the context of coarse σ-algebras.

The main original contributions of this paper are the notions of (i) interconnection of stochastic systems (Definition 6) and their complementarity (Definition 5), a condition required in order to be able to interconnect, (ii) linearity stochastic system (Definition 4), and (iii) constrained probability (Definition 7). We feel that these notions are worthwhile additions to the arsenal of elementary concepts of mathematical probability.

We now introduce some of the notation used in the paper. $\mathbb{N} = \{1, 2, \ldots, \mathtt{k}, \ldots\}$ denotes the set of natural numbers. $\mathbb{R}, \mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{n}\times\mathtt{m}}$ denote the sets of real numbers, $\mathtt{n}$-dimensional real vectors, and $\mathtt{n}\times\mathtt{m}$ real matrices. $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$ with $\mathbb{L}$ a linear subspace of $\mathbb{R}^{\mathtt{n}}$, denotes the quotient space $\mathbb{R}^{\mathtt{n}}$ *modulo* $\mathbb{L}$, that is, the class of subsets of $\mathbb{R}^{\mathtt{n}}$ consisting of the affine subspaces $\{a + \mathbb{L} \mid a \in \mathbb{R}^{\mathtt{n}}\}$. For $M = M^\top \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$, $M \succeq 0$ means that M is nonnegative definite, that is, $x^\top M x \geq 0$ for $x \in \mathbb{R}^{\mathtt{n}}$, while $M \succ 0$ means that M is positive definite, that is, $x^\top M x > 0$ for $0 \neq x \in \mathbb{R}^{\mathtt{n}}$. For $M \in \mathbb{R}^{\mathtt{n}\times\mathtt{m}}$, $\mathtt{kernel}(M)$ denotes the *kernel* of M, defined by $\mathtt{kernel}(M) := \{x \in \mathbb{R}^{\mathtt{m}} \mid Mx = 0\}$. For a set $\mathbb{S}, 2^{\mathbb{S}}$ denotes the *power set* of $\mathbb{S}$, that is, the class of all subsets of $\mathbb{S}$. For a map $f : \mathbb{X} \to \mathbb{Y}$ with domain $\mathbb{X}$ and codomain $\mathbb{Y}$, $\mathtt{image}(f)$ denotes the *image* of f, defined by $\mathtt{image}(f) := \{y \in \mathbb{Y} \mid \text{there exists } x \in \mathbb{X} \text{ such that } y = f(x)\}$, while $f^{-1} : 2^{\mathbb{Y}} \to 2^{\mathbb{X}}$ denotes the *set theoretic inverse* of f, that is, for $\mathbb{S} \subseteq \mathbb{Y}$, $f^{-1}(\mathbb{S}) := \{x \in \mathbb{X} \mid f(x) \in \mathbb{S}\}$. The set-to-set map f^{-1} is called the *pullback* of f.

The paper is organized as follows. In Section II the concept of a stochastic system is introduced. A stochastic system is simply a standard probability triple as used in mathematical probability theory. This concept is contrasted with what we call a 'classical' stochastic system when the event space consists of the Borel sets. Our interest, however, is primarily in systems with a coarse σ-algebra of events. We illustrate the notion of stochastic system in Section III by means of two examples, a noisy resistor, and the price/demand and price/supply characteristics of an economic good. In Section IV we formalize linearity of stochastic systems, and gaussian systems as a special class of linear stochastic systems. In Section V we discuss in an informal manner several ways of combining systems, while in Section VI we formalize interconnection of two stochastic systems. Interconnection requires that the interconnected systems are complementary. We illustrate interconnection by means of two examples, the noisy resistor connected with a voltage source, and the equilibrium price/demand/supply of an economic good. In Section VII we study interconnection and complementarity

Jan Willems is with the Department of Electrical Engineering, KU Leuven, B-3001 Leuven, Belgium. Email: `Jan.Willems@esat.kleuven.be`, URL `http://www.esat.kuleuven.be/~jwillems`

of linear stochastic systems. In Section VIII we argue that classical stochastic systems are basically closed systems, since complementarity of stochastic systems requires a coarse event σ-algebra. In Section IX we discuss some of the implications to system identification of the view of stochastic systems and their interconnection put forward in the previous sections. In Section X we introduce the notion of the stochastic system with outcomes constrained to be in a subset of the outcome space. Constraining, while reminiscent of conditioning, is quite different from it. In fact, conditioning requires that the conditioning set is an event, while constraining basically requires that the constraining set is not an event. Constraining is effective only in the context of coarse event σ-algebras. In Section XI we show how to construct the stochastic system induced by a map with the outcome space as domain. One way of obtaining open stochastic systems is as a family of classical probability measures on the output space, parameterized by an input. Such families of probability measures go under the name of 'probability kernels'. In Section XII we discuss, for a binary channel, the relation between probability kernels and our notion of stochastic system combined with constraining.

Finally, a brief reminder of the most important probabilistic concepts used in the paper. We add these well known notions here for easy reference and in order to introduce the nomenclature and notation. More details about these notions may be found in Wikipedia and in any book on mathematical probability theory.

A class of subsets $\mathscr{F}$ of a set $\mathbb{F}$ is said to be an *algebra* on $\mathbb{F}$ if (i) $\mathbb{F} \in \mathscr{F}$, (ii) $F \in \mathscr{F}$ implies $F^{\text{complement}} \in \mathscr{F}$ ($F^{\text{complement}}$ denotes the complement of F with respect to $\mathbb{F}$), and (iii) $F_1, F_2 \in \mathscr{F}$ implies $F_1 \cup F_2 \in \mathscr{F}$. If (iii) is strengthened to (iii)' $F_{\mathtt{k}} \in \mathscr{F}$ for $\mathtt{k} \in \mathbb{N}$ implies $\bigcup_{\mathtt{k}\in\mathbb{N}} F_{\mathtt{k}} \in \mathscr{F}$, then $\mathscr{F}$ is said to be a *σ-algebra* on $\mathbb{F}$. For any class of subsets $\mathscr{F}$ of $\mathbb{F}$, there is a smallest σ-algebra of subsets of $\mathbb{F}$ that contains $\mathscr{F}$. This σ-algebra is called the σ-algebra *generated by* $\mathscr{F}$. A *measurable space* is a pair $(\mathbb{F}, \mathscr{F})$ with $\mathscr{F}$ a σ-algebra on $\mathbb{F}$. Let $(\mathbb{F}, \mathscr{F})$ and $(\mathbb{D}, \mathscr{D})$ be measurable spaces. A map $f : \mathbb{F} \to \mathbb{D}$ is said to be *measurable* with respect to $(\mathbb{F}, \mathscr{F})$ and $(\mathbb{D}, \mathscr{D})$ if the inverse image under f of a measurable set is measurable, that is, if $D \in \mathscr{D}$ implies $f^{-1}(D) \in \mathscr{F}$.

A *probability space* is a triple $(\Omega, \mathscr{A}, P)$ consisting of a non-empty set Ω, called the *basic space*, a σ-algebra $\mathscr{A}$ of subsets of Ω (elements of $\mathscr{A}$ are called *measurable*), and a map $P : \mathscr{A} \to [0,1]$ called the *probability measure*) or simply the *probability*. A probability measure P must satisfy (i) $P(\Omega) = 1$ and (ii) $P(\bigcup_{\mathtt{k}\in\mathbb{N}} A_{\mathtt{k}}) = \sum_{\mathtt{k}\in\mathbb{N}} P(A_{\mathtt{k}})$ for sets $A_{\mathtt{k}} \in \mathscr{A}$, $\mathtt{k} \in \mathbb{N}$, that are disjoint (i.e., $A_{\mathtt{k}'} \cap A_{\mathtt{k}''} = \emptyset$ for $\mathtt{k}' \neq \mathtt{k}''$). Property (ii) is called *countable additivity* of P.

Let $\mathscr{A}'$ be an algebra of subsets of Ω and $P' : \mathscr{A}' \to [0,1]$ a map that satisfies (i) $P'(\Omega) = 1$ and (ii) finite additivity: $P'(A'_1 \cup A'_2) = P'(A'_1) + P'(A'_2)$ for disjoint sets $A'_1, A'_2 \in \mathscr{A}'$. The *Hahn-Kolmogorov extension theorem* states that there exists a unique probability measure P on $\mathscr{A}$, the σ-algebra generated by $\mathscr{A}'$, such that $P(A') = P'(A')$ for all $A' \in \mathscr{A}'$.

Let $(\Omega, \mathscr{A}, P)$ be a probability space. A subset of a set with probability zero is called a *null set*. $(\Omega, \mathscr{A}, P)$ is said to be *complete* if every null set is measurable. If $(\Omega, \mathscr{A}, P)$ is a probability space, then the class of subsets of Ω of the form $A \cup N$, with $A \in \mathscr{A}$ and N a null set, forms a σ-algebra $\mathscr{A}'$. Define $P' : \mathscr{A}' \to [0,1]$ by $P'(A \cup N) := P(A)$. Then $(\Omega, \mathscr{A}', P')$ is a complete probability space, called the *completion* of $(\Omega, \mathscr{A}, P)$. Two probability spaces $(\Omega, \mathscr{A}, P)$ and $(\Omega, \mathscr{A}', P')$ are said to be *equivalent* if they have the same completion. Equivalence means that the measurable sets in both spaces correspond up to sets of probability zero.

We denote the σ-algebra on $\mathbb{R}^{\mathtt{n}}$ generated by the open sets with respect to the norm topology by $\mathscr{B}(\mathbb{R}^{\mathtt{n}})$. Elements of $\mathscr{B}(\mathbb{R}^{\mathtt{n}})$ are called *Borel measurable* or simply *Borel*. A probability measure on $\mathscr{B}(\mathbb{R}^{\mathtt{n}})$ is called a *Borel probability* on $\mathbb{R}^{\mathtt{n}}$. The *support* of a Borel probability P, denoted by $\mathtt{support}(P)$, is the smallest closed set $\mathbb{S} \subseteq \mathbb{R}^{\mathtt{n}}$ such that $P(\mathbb{S}) = 1$.

II. Stochastic systems

In this section we introduce the central concept of a stochastic system, which is nothing else than a probability space as put forward in orthodox mathematical probability theory.

Definition 1: A *stochastic system* is a triple $(\mathbb{W}, \mathscr{E}, P)$ with
► $\mathbb{W}$ a non-empty set, the *outcome space* with elements called *outcomes*,
► $\mathscr{E}$ a σ-algebra of subsets of $\mathbb{W}$ with elements called *events*,
► $P : \mathscr{E} \to [0,1]$ a *probability* measure. ■

The intuitive background underlying this definition is as follows. Assume that we have a stochastic phenomenon that we wish to model. The phenomenon produces variables in the outcome space. The aim of the model is to specify (i) the subsets of the outcome space to which a probability is assigned and (ii) the numerical value of the probability (in the sense of relative frequency, degree of belief, or whatever interpretation of probability is relevant in the application at hand) that the outcomes belong to a particular subset. The set in which the outcomes take on their value is the outcome space $\mathbb{W}$. The set of events $\mathscr{E}$ consists of those subsets of $\mathbb{W}$ to which the model assigns a probability. The probability that the outcomes belong to the set $E \in \mathscr{E}$ is $P(E)$. $\mathscr{E}$ is required to be a σ-algebra, and P a probability measure.

Two important special cases are obtained as follows. We refer to these special cases as *classical* stochastic systems.
► The first special case is $(\mathbb{W}, 2^{\mathbb{W}}, P)$ with $\mathbb{W}$ a countable set. P can then be specified by giving the probability p of the individual outcomes, $p : \mathbb{W} \to [0,1]$, and defining P by $P(E) := \sum_{e\in E} p(e)$. In this case, every subset of $\mathbb{W}$ is assumed to be an event, and P is completely determined by the probability of the singletons.
► The second special case is $(\mathbb{R}^{\mathtt{n}}, \mathscr{B}(\mathbb{R}^{\mathtt{n}}), P)$, a Borel probability on $\mathbb{R}^{\mathtt{n}}$. P can then be specified by a probability distribution on $\mathbb{R}^{\mathtt{n}}$, or, if the distribution is sufficiently smooth, by the probability density function $p : \mathbb{R}^{\mathtt{n}} \to [0, \infty)$ leading to $P(E) = \int_E p(x)\, dx$.

For a classical stochastic system 'essentially every' subset of $\mathbb{W}$ is an event and is therefore assigned a probability. In the countable case this is completely correct, since then every subset of $\mathbb{W}$ is in $\mathscr{E}$, while in the case of $\mathbb{R}^{\mathtt{n}}$, this is a consequence of the fact that every 'reasonable' subset of $\mathbb{R}^{\mathtt{n}}$

is a Borel set. Thus for classical stochastic systems, the events are obtained from the structure (finiteness, or the topology) on the outcome space. No probabilistic modeling enters into the specification of the events. In Definition 1, on the other hand, the event space $\mathscr{E}$ is very much a part of the stochastic model.

We formalize the second special case as a definition.

Definition 2: The stochastic system $(\mathbb{R}^{\mathtt{n}}, \mathscr{B}(\mathbb{R}^{\mathtt{n}}), P)$ is called a *classical* n*-dimensional random vector*. ■

We use the term classical random vector also when the outcome space $\mathbb{W}$ is a Borel subset of $\mathbb{R}^{\mathtt{n}}$, and the events are the elements of $\mathscr{B}(\mathbb{R}^{\mathtt{n}})$ that are contained in $\mathbb{W}$.

Classical stochastic systems and classical random vectors dominate the development and the teaching of probabilistic modeling and analysis techniques, as witnessed by the emphasis on notions as mean and variance, in the classical definition of a random variable, a random vector, and a stochastic process, in notions as marginal and conditional probabilities, in the concept of Markov process, etc. The aim of the present paper is to show that this emphasis on classical stochastic systems with rich σ-algebras is unduly restrictive, even for elementary applications. In addition, we aim to demonstrate that notions as linearity and interconnection of stochastic systems require coarse σ-algebras and the full generality of Definition 1. In Section VIII, we argue that Borel σ-algebras aim solely at 'closed' systems, while coarse σ-algebras allow to consider 'open' systems.

Deterministic systems emerge as special cases of stochastic systems, as they should.

Definition 3: The stochastic system $(\mathbb{W}, \mathscr{E}, P)$ is said to be *deterministic* if $\mathscr{E} = \{\emptyset, \mathbb{B}, \mathbb{B}^{complement}, \mathbb{W}\}$ and $P(\mathbb{B}) = 1$. $\mathbb{B}$ is called the *behavior* of the deterministic system. ■

Deterministic and classical stochastic systems are extremes of a spectrum ranging from systems with very coarse to systems with very rich σ-algebras.

III. EXAMPLES

We illustrate the relevance of coarse σ-algebras by two examples.

Example 1: A noisy resistor. Consider a 2-terminal 1-port electrical circuit shown as a black box in Figure 1(a). The aim is to model the relation between the voltage V and the current I. The outcomes are voltage/current pairs $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$. Hence $\mathbb{W} = \mathbb{R}^2$.

An example is an Ohmic resistor, shown in Figure 1(b), described by $V = RI$ with R the resistance. An Ohmic resistor defines a deterministic system with behavior $\mathbb{B} = \left\{ \left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right] \in \mathbb{R}^2 \mid V = RI \right\}$.

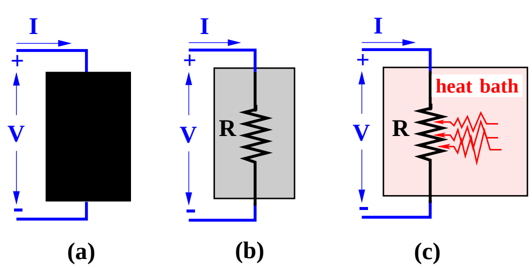


Fig. 1. **2-terminal electrical circuit**

As an example of a noisy circuit, consider a resistor with thermal noise. In 1928, John Bert Johnson, an engineer at Bell Labs, observed that a resistor in a heat bath (see 1(c)) produces current even when no voltage is applied. Harry Nyquist explained this phenomenon as resulting from thermal energy being transformed to electrical energy due to thermal agitation. This is a nice example of a physical phenomenon that can be described stochastically.

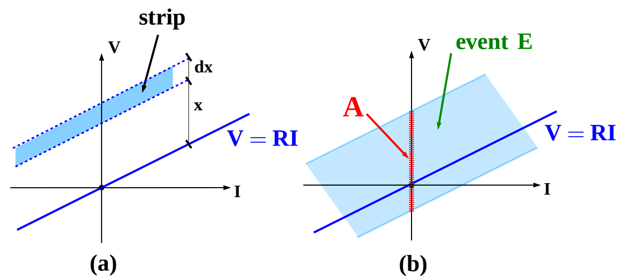


Fig. 2. **Events for the noisy resistor**

This noisy ('hot') resistor is modeled as follows. Without thermal agitation, the resistor is an Ohmic resistor, governed by the relation $V = RI$. With thermal agitation, the voltage/current pair (V, I) belongs to the incremental strip shown in 2(a) with a certain probability. Concretely, assume that the probability that $x \le V - RI \le x + dx$ is equal to $1/\sqrt{2\pi}\sigma\ e^{-x^2/2\sigma^2}\,dx$ with $\sigma \sim \sqrt{RT}$ and T the temperature of the resistor in the heat bath. More generally, the probability of $\left\{ \left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right] \in \mathbb{R}^2 \mid V - RI \in A \text{ with } A \text{ a Borel subset of } \mathbb{R} \right\}$ (the shaded region of 2(b)) is thus $1/\sqrt{2\pi}\sigma \int_A e^{-x^2/2\sigma^2}\,dx$.

The noisy resistor defines a stochastic system with outcome space $\mathbb{W} = \mathbb{R}^2$ and as outcomes voltage/current vectors $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$. The events $E \in \mathscr{E}$ are the sets of the form

$$E = \left\{ \left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right] \in \mathbb{R}^2 \mid V - RI \in A \text{ with } A \subseteq \mathbb{R} \text{ Borel} \right\}. \qquad (1)$$

The event E is illustrated in Figure 2(b). The probability of E is

$$P(E) = \frac{1}{\sqrt{2\pi}\sigma} \int_A e^{-\frac{x^2}{2\sigma^2}}\,dx. \qquad (2)$$

The parameters that specify the stochastic laws of the noisy resistor are R and σ.

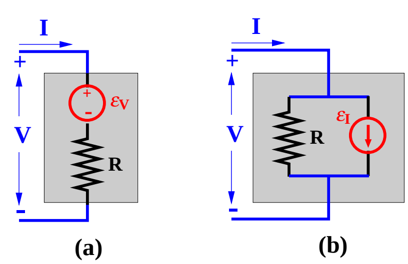


Fig. 3. **Equivalent circuits for a noisy resistor**

This noisy resistor can be represented by equivalent circuits. For example, as an Ohmic resistor in series with a random voltage source as shown in Figure 3(a). This leads to the following relation between the current I through the resistor and the voltage V across it

$$V = RI + \varepsilon_V \tag{3}$$

with $R > 0$ the value of the Ohmic resistor and ε_V the voltage generated by the noisy voltage source. In the standard Johnson-Nyquist model, the noise ε_V is taken to be gaussian, with zero mean and standard deviation σ. Alternatively, the noisy resistor can also be represented as an Ohmic resistor in parallel with a random current source as shown in Figure 3(b). This leads to the following relation between the current I through the resistor and the voltage V across it

$$I = V/R + \varepsilon_I$$

with $R > 0$ the value of the Ohmic resistor and ε_I the current generated by the noisy current source. In the standard Johnson-Nyquist model, the noise ε_I is taken to be gaussian, with zero mean and standard deviation σ/R. These equivalent circuits are merely representations of the noisy resistor. The basic physical phenomenon is best described by the events (1) and their probabilities (2).

Hence, whereas ε_V and ε_I are classical random variables, $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$ is not a classical random vector. Only cylinders with rays parallel to $V = RI$ (see Figure 2(a)) are events that are assigned a probability. In particular, V and I are not classical random variables. Indeed, the basic model of a noisy resistor does not imply a stochastic law for V or I, in the sense that (1, 2) does not model V and I individually as classical random variables. ■

Example 2: Price/demand and price/supply. Important characteristics of an economic good are the responsiveness of the demand and of the supply to the price. Typical deterministic price/demand and price/supply characteristics are shown in Figure 4(a) and Figure 4(b). These characteristics define deter-

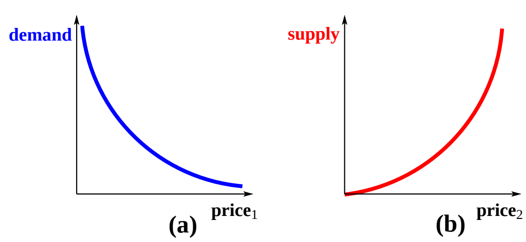


Fig. 4. **Deterministic price/demand and price/supply characteristics**

ministic systems with $\mathbb{W} = (0,\infty)^2$ and behavior given by the graph of respectively the price/demand and the price/supply characteristics.

In order to express that the demand is influenced by uncertain factors in addition to the price, randomness can be added to the price/demand characteristic. This leads to models that state that the price/demand vector lies in certain regions like those shown in Figure 5(a) with a certain probability. While it is viable to assign a probability to certain regions of the price/demand plane, it is not reasonable to assume that the price/demand is modeled as a classical 2-dimensional random vector. Indeed, the uncertainty of the price/demand phenomenon does not imply a probability distribution for the price itself. *No such probability distribution for the price is implied in the deterministic case, so why should it be implied in the stochastic case?* Similarly, for the price/supply, it is reasonable to assume for example that the price/supply vector lies in certain regions like those shown in Figure 5(b) with a certain probability.

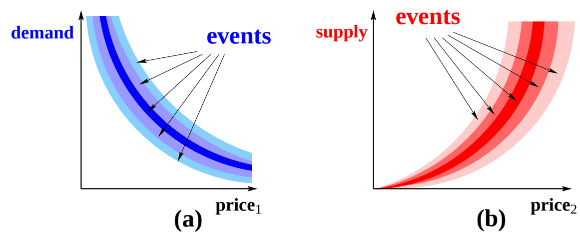


Fig. 5. **Stochastic price/demand and price/supply events**

One can make these examples more concrete by assuming that the price/demand relation is for instance given by $p_1 d = \varepsilon_1$ with ε_1 a classical positive real random variable. The events $\mathscr{E}$ of the stochastic price/demand system then consist of the sets

$$E = \left\{ \left[\begin{smallmatrix} p_1 \\ d \end{smallmatrix}\right] \in (0,\infty)^2 \mid \varepsilon_1 = p_1 d \in A \text{ with } A \subseteq (0,\infty) \text{ Borel} \right\}$$

and $P(E)$ equal to the probability that $\varepsilon_1 \in A$. Similarly, we could assume that the price/supply relation is for instance given by $s = \varepsilon_2 p_2^2$ with ε_2 a classical positive real random variable. The events $\mathscr{E}$ of the stochastic price/supply system then consist of the sets

$$E = \left\{ \left[\begin{smallmatrix} p_2 \\ s \end{smallmatrix}\right] \in (0,\infty)^2 \mid \varepsilon_2 = s/p_2^2 \in A \text{ with } A \subseteq (0,\infty) \text{ Borel} \right\}$$

and $P(E)$ equal to the probability that $\varepsilon_2 \in A$. ■

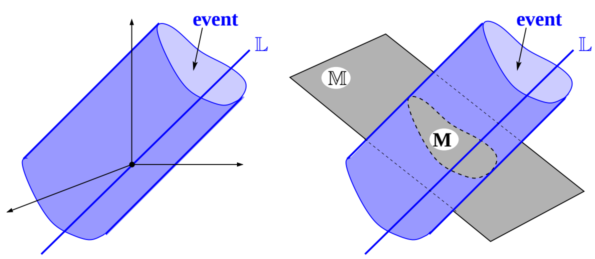
Many modeling problems studied in physics, economics, and statistics aim at the stochastic relation between two real variables (voltage versus current, price versus demand, price versus supply, weight versus size, intelligence versus scores on tests, age versus medical expenditures, and so forth). In many situations, it is reasonable to assume in the deterministic case that the relation between the two variables in question is given by a curve (such as Ohm's law $V = RI$ for a resistor, a price/demand characteristic, or a price/supply characteristic as those shown in Figure 4). Arguably, when we study a stochastic version of such a relation, we invariably end up with a stochastic system in the sense of Definition 1 with a coarse σ-algebra. It is unreasonable to expect that a classical stochastic system will emerge, since it is the characteristics, and not the variables, that become fuzzy by adding uncertainty. So, we need a coarse σ-algebra in order to obtain the deterministic case as a special situation. In order to model the relation between variables, classical random vectors should be more the exception than the rule.

IV. Linearity

Definition 4: The $\mathtt{n}$-dimensional stochastic system $(\mathbb{R}^{\mathtt{n}}, \mathscr{E}, P)$ is said to be *linear* if there exists a linear subspace $\mathbb{L}$ of $\mathbb{R}^{\mathtt{n}}$ such that the events are the Borel subsets of the quotient space $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$, and the probability is a Borel probability on $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$. Note that $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$ is a finite dimensional real vector space, with, therefore, well-defined Borel sets. The dimension of $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$ is equal to $\mathtt{n}-\mathtt{dimension}(\mathbb{L})$. $\mathbb{L}$ is called the *fiber* and $\mathtt{dimension}(\mathbb{L})$ *the number of degrees of freedom* of the linear stochastic system. The stochastic system $(\mathbb{R}^{\mathtt{n}},\mathscr{E},P)$ is said to be *gaussian* if it is linear and if the Borel probability on $\mathbb{R}^{\mathtt{n}}/\mathbb{L}$ has a gaussian distribution. ■

We consider a probability measure that is concentrated on a singleton to be gaussian. More generally, a gaussian probability measure may be concentrated on a linear variety.

The idea behind Definition 4 is illustrated in Figure 6(a). The events are cylinders in $\mathbb{R}^{\mathtt{n}}$ with rays parallel to the fiber



(a) (b)

Fig. 6. **Events for a linear system**

$\mathbb{L}$. A linear stochastic system is a classical random vector if and only if $\mathbb{L}=\{0\}$. Hence every classical random vector with $\mathbb{W}=\mathbb{R}^{\mathtt{n}}$ defines a linear stochastic system. At the other extreme, when $\mathbb{L}=\mathbb{R}^{\mathtt{n}}$, the event set $\mathscr{E}$ becomes the trivial σ-algebra $\{\emptyset,\mathbb{R}^{\mathtt{n}}\}$.

A concrete way of thinking about a linear $\mathtt{n}$-dimensional stochastic system is in terms of two linear subspaces $\mathbb{L},\mathbb{M}$ of $\mathbb{R}^{\mathtt{n}}$ that are complementary, $\mathbb{L}\oplus\mathbb{M}=\mathbb{R}^{\mathtt{n}}$, and a Borel probability $P_{\mathbb{M}}$ on $\mathbb{M}$. Take as events the sets of the form

$$E=\{\bigcup_{w\in M}(w+\mathbb{L})\mid M \text{ a Borel subset of } \mathbb{M}\}$$

(see Figure 6(b)) and $P(E)$ equal to $P_{\mathbb{M}}(M)$. A linear $\mathtt{n}$-dimensional stochastic system is thus parameterized by its linear fiber $\mathbb{L}$, a linear subspace $\mathbb{M}$ complementary to $\mathbb{L}$, and a Borel probability on $\mathbb{M}$.

Let $R\in\mathbb{R}^{\mathtt{p}\times\mathtt{n}}$ be a matrix of full row rank (that is, $\mathtt{rank}(R)=\mathtt{p}$) and ε a classical $\mathtt{p}$-dimensional random vector with Borel probability P_ε. Consider the equation

$$Rw=\varepsilon \tag{4}$$

describing the stochastic laws of the vector $w\in\mathbb{R}^{\mathtt{n}}$. This equation defines the linear stochastic system $\Sigma=(\mathbb{R}^{\mathtt{n}},\mathscr{E},P)$ with

$$[\![E\in\mathscr{E}]\!]:\Leftrightarrow[\![E=R^{-1}(A)\text{ for some Borel subset } A\subseteq\mathbb{R}^{\mathtt{p}}]\!]$$

and

$$P\left(R^{-1}(A)\right):=P_\varepsilon(A).$$

R^{-1} denoted the pullback of R. The fiber of this linear stochastic system is $\mathtt{kernel}(R)$. The number of degrees of freedom equals $\mathtt{n}-\mathtt{p}$. We call (4) a *kernel representation* of Σ. Every $\mathtt{n}$-dimensional linear stochastic system admits a kernel representation. Note that (4) defines a gaussian stochastic system if and only if ε is gaussian. An $\mathtt{n}$-dimensional gaussian system with $\mathtt{n}-\mathtt{p}$ degrees of freedom represented by (4) is hence parameterized by the triple (R,m,S) with $R\in\mathbb{R}^{\mathtt{p}\times\mathtt{n}}$ a matrix of full row rank, $m\in\mathbb{R}^{\mathtt{p}}$ the mean, and $S\in\mathbb{R}^{\mathtt{p}\times\mathtt{p}}, S=S^\top\succeq 0$, the covariance of ε. All triples (R,m,S) that define the same gaussian system are obtained by the transformation group

$$(R,m,S)\quad\underset{U\in\mathbb{R}^{\mathtt{p}\times\mathtt{p}}\text{ nonsingular}}{\longmapsto}\quad(UR,Um,USU^\top). \tag{5}$$

Observe that our definition of linearity involves only the event σ-algebra, but not the probability measure. This fact has subtle consequences when applied to deterministic systems. But it is easy to see that a deterministic system with behavior $\mathbb{B}$ is equivalent to a linear stochastic system if and only if $\mathbb{B}$ is an affine subspace of $\mathbb{R}^{\mathtt{n}}$. Hence, while a deterministic system with a linear behavior $\mathbb{B}$ does not define a linear stochastic system, it is equivalent to a linear stochastic system.

V. Combination of Stochastic Systems

One of the central aspects of systems thinking is the possibility of combining systems and viewing a complex system as an architecture of interconnected subsystems. This feature is important in all aspects of systems theory and control, in modeling, in analysis, and in synthesis. In [2] we have discussed 'tearing, zooming, and linking' modeling procedures for deterministic systems, while in [3] we applied these ideas to the modeling of *RLC* circuits. In the present section we deal with the composition of stochastic systems in an informal way. In Section VI we formalize interconnection in detail.

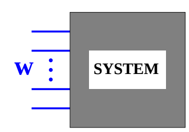


Fig. 7. **A system as a black box**

A convenient way to visualize systems is by block diagrams. Figure 7 shows a pictorial representation of a system as a black box with terminals. The variables w that are relevant in the model are shown as associated with terminals. In some applications (as electrical circuits, and some mechanical, thermal, and hydraulic systems) these terminals can be taken literally, while for other applications they should be thought of as virtual terminals. For example, if $w\in\mathbb{R}^{\mathtt{n}}$, we may think of each of the terminals as corresponding to one of the components of the vector $w=(w_1,w_2,\ldots,w_{\mathtt{n}})$. The black box indicates that the variables on the terminals are interrelated, for example through the laws of a stochastic system.

We consider several ways in which systems can be combined. The first way is juxtaposition. We start with two systems with variables w_1 and w_2 (Figure 8(a)) respectively, and obtain a new system with variables $w=(w_1,w_2)$ (Figure 8(b)).

A second way of combining systems is by interconnection. We start again with two systems with variables w_1 and w_2 (Figure 9(a)) respectively, and obtain a new system with variables w (Figure 9(b)). The interconnection imposes $w_1=w_2=w$. Interconnection can also be viewed as an operation on the terminal variables of a single system. We start with a system with variables w_1 and w_2 (Figure 10(a)), and obtain a new

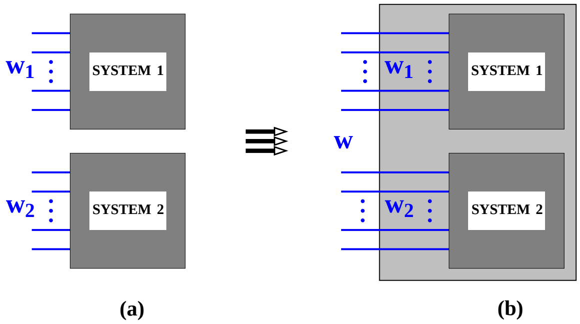


Fig. 8. **Juxtaposition of systems**

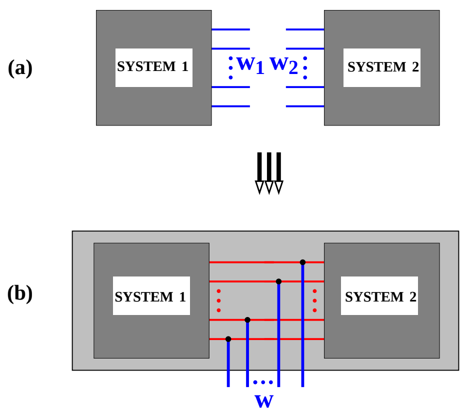


Fig. 9. **Interconnection of systems**

system with variables w by setting $w = w_1 = w_2$ (Figure 10(b)). The situations of Figures 9 and 10 are not really different,

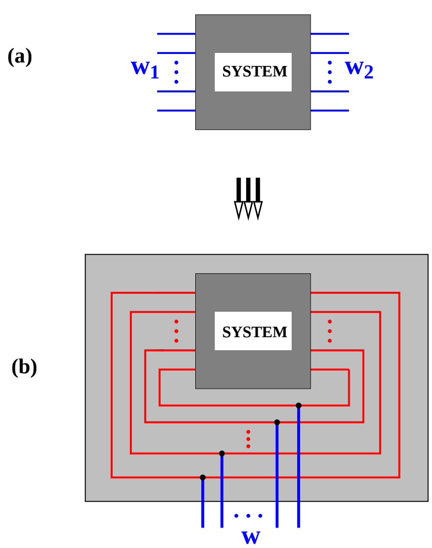


Fig. 10. **Interconnection of terminals**

since by combining juxtaposition of Figure 8 with terminal interconnection of Figures 10 applied to a single system, we obtain system interconnection as in Figure 9.

The basic idea of interconnection is *variable sharing*, in the sense explained in [2] and [3] for deterministic systems. Series, parallel, and feedback interconnections are readily seen to be special cases. We formalize interconnection of stochastic systems in Section VI.

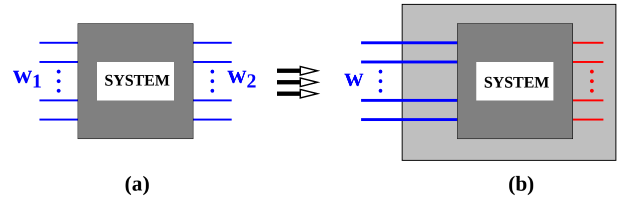


Fig. 11. **Elimination of variables**

A third way of obtaining a system from another one is by elimination of variables. We start with a system with variables w_1 and w_2 (Figure 11(a)) and obtain a new system with variables w by setting $w = w_1$ (Figure 11(b)). In other words, the variables w_2 are eliminated. Elimination of variables of stochastic systems is discussed briefly in Section XI.

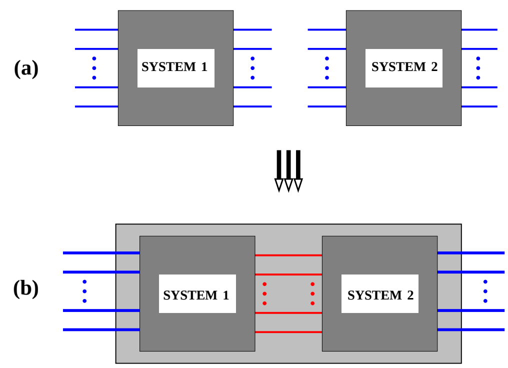


Fig. 12. **Interconnection and elimination**

By combining the operations explained above, it is possible to obtain complex interconnected systems from simpler subsystems. For example, by combining juxtaposition, interconnection, and elimination we can deal with the situation illustrated in Figure 12.

We have discussed so far the combination of two systems. These operations are of course readily extended sequentially to more than two systems, and therefore to complex architectures of interconnected systems.

VI. Interconnection

In this section we formalize interconnection. We start by considering the situation discussed in Figure 9 with the assumption that the two to-be-interconnected systems are stochastically independent. Note that interconnection comes down to imposing two distinct probabilistic laws on the same set of variables. The question is: *Is it possible to define one law which respects both laws?* As we shall see, this is indeed possible, provided a regularity condition, called 'complementarity', is satisfied.

Definition 5: Two σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$ on a set $\mathbb{W}$ are said to be *complementary* if for all nonempty sets $E_1, E_1' \in \mathscr{E}_1, E_2, E_2' \in \mathscr{E}_2$ there holds

$$[E_1 \cap E_2 = E_1' \cap E_2'] \Rightarrow [E_1 = E_1' \text{ and } E_2 = E_2'].$$

The stochastic systems $\Sigma_1 = (\mathbb{W}, \mathscr{E}_1, P_1)$ and $\Sigma_2 = (\mathbb{W}, \mathscr{E}_2, P_2)$ are said to be *complementary* if for all $E_1, E_1' \in \mathscr{E}_1$ and $E_2, E_2' \in$

$\mathscr{E}_2$ there holds

$$[\![E_1 \cap E_2 = E_1' \cap E_2']\!] \Rightarrow [\![P_1(E_1)P_2(E_2) = P_1(E_1')P_2(E_2')]\!]. \qquad \blacksquare$$

In words, complementarity of stochastic systems requires that the intersection of two events, one from each of the σ-algebras, determines the product of the probabilities of the intersecting events uniquely, while complementarity of the σ-algebras requires that the intersection of two sets, one from each of the σ-algebras, determines the intersecting sets uniquely.

Note that

$$[\![\mathscr{E}_1, \mathscr{E}_2 \text{ complementary}]\!] \Rightarrow [\![\mathscr{E}_1 \cap \mathscr{E}_2 = \{\emptyset, \mathbb{W}\}]\!]$$

and

$$[\![\mathscr{E}_1, \mathscr{E}_2 \text{ complementary}, E_1 \in \mathscr{E}_1, E_2 \in \mathscr{E}_2, \text{ and } E_1 \cap E_2 = \emptyset]\!] \Rightarrow [\![E_1 = \emptyset \text{ or } E_2 = \emptyset]\!].$$

Indeed, $S \in \mathscr{E}_1 \cap \mathscr{E}_2$ implies $S \cap S = S \cap \mathbb{W}$. Therefore, if $\mathscr{E}_1, \mathscr{E}_2$ are also complementary, then $S = \emptyset$ or $S = \mathbb{W}$. Further, $E_1 \cap E_2 = \emptyset$ implies $E_1 \cap E_2^{\text{complement}} = E_1 = E_1 \cap \mathbb{W}$. Hence, if $\mathscr{E}_1, \mathscr{E}_2$ are also complementary, then either $E_1 = \emptyset$, or $E_2^{\text{complement}} = \mathbb{W}$, that is, $E_2 = \emptyset$.

Complementarity of systems is a refinement of complementarity of σ-algebras in order to accommodate for zero probability events that may violate the complementarity of σ-algebras. It is easy to construct examples involving zero probability events that show that complementarity of two stochastic systems does not imply complementarity of the associated σ-algebras. Complementarity of the event σ-algebras is a more primitive condition that is convenient for proving complementarity of stochastic systems.

Complementarity of two stochastic systems is indeed implied by complementarity of the associated σ-algebras. In order to see this, let $E_1, E_1' \in \mathscr{E}_1, E_2, E_2' \in \mathscr{E}_2$. On the one hand, if the sets E_1, E_1', E_2, E_2' are all non-empty and $\mathscr{E}_1, \mathscr{E}_2$ are complementary, then $E_1 \cap E_2 = E_1' \cap E_2'$ implies $E_1 = E_1'$ and $E_2 = E_2'$, and, therefore, $P_1(E_1)P_2(E_2) = P_1(E_1')P_2(E_2')$. On the other hand, assume that at least one of the sets E_1, E_1', E_2, E_2', say E_1, is empty. Then $E_1 \cap E_2 = E_1' \cap E_2'$ implies $E_1' \cap E_2' = \emptyset$. By what we proved in the previous paragraph, complementarity of $\mathscr{E}_1, \mathscr{E}_2$ therefore implies that either $E_1' = \emptyset$, or $E_2' = \emptyset$. Consequently, also in this case $E_1 \cap E_2 = E_1' \cap E_2'$ implies $P_1(E_1)P_2(E_2) = 0 = P_1(E_1')P_2(E_2')$.

Definition 6: Let $\Sigma_1 = (\mathbb{W}, \mathscr{E}_1, P_1)$ and $\Sigma_2 = (\mathbb{W}, \mathscr{E}_2, P_2)$ be complementary stochastic systems. Then the *interconnection* of Σ_1 and Σ_2, assumed stochastically independent, denoted by $\Sigma_1 \wedge \Sigma_2$, is defined as the stochastic system

$$\Sigma_1 \wedge \Sigma_2 := (\mathbb{W}, \mathscr{E}, P).$$

The σ-algebra $\mathscr{E}$ and the probability P are defined as follows. $\mathscr{E}$ is the σ-algebra generated by $\mathscr{E}_1 \cup \mathscr{E}_2$, while P is defined on the 'rectangles' $\{E_1 \cap E_2 \mid E_1 \in \mathscr{E}_1, E_2 \in \mathscr{E}_2\}$ by

$$P(E_1 \cap E_2) := P_1(E_1)P_2(E_2) \text{ for } E_1 \in \mathscr{E}_1, E_2 \in \mathscr{E}_2,$$

and extended to all of $\mathscr{E}$ by the Hahn-Kolmogorov extension theorem. $\blacksquare$

Note that the definition of the probability P for rectangles uses complementarity in an essential way. $\mathscr{E}$, the σ-algebra generated by $\mathscr{E}_1 \cup \mathscr{E}_2$, is in fact also the σ-algebra generated by these rectangles. It is readily seen that the class of subsets of $\mathbb{W}$ that consist of the union of a finite number of disjoint rectangles forms an algebra of subsets of $\mathbb{W}$ (it is closed under taking the complement, intersection, and union). The probability of rectangles defines the probability on the subsets of $\mathbb{W}$ that consist of a union of a finite number of disjoint rectangles. By the Hahn-Kolmogorov extension theorem, this leads to a unique probability measure P on $\mathscr{E}$, the σ-algebra generated by the rectangles. This construction of the σ-algebra $\mathscr{E}$ and of the probability measure P is completely analogous to the construction of a product measure in measure theory.

The notions of interconnection of stochastic systems and of complementarity of stochastic systems and σ-algebras constitute the main original concepts of this paper, viewed as a contribution to mathematical probability theory.

Obviously, there holds $\mathscr{E}_1, \mathscr{E}_2 \subseteq \mathscr{E}$. Also, for $E_1 \in \mathscr{E}_1$ and $E_2 \in \mathscr{E}_2$, we have $P(E_1) = P_1(E_1)$ and $P(E_2) = P_2(E_2)$. Hence interconnection refines the event σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$ and the probabilities P_1 and P_2 . This implies in particular that Σ_1 and Σ_2 are unfalsified by $\Sigma_1 \wedge \Sigma_2$. The stochastic system $(\mathbb{W}, \mathscr{E}, P)$ is said to be *unfalsified* by $(\mathbb{W}, \mathscr{E}', P')$ if for all $E \in \mathscr{E} \cap \mathscr{E}'$ there holds $P(E) = P'(E)$. Note that for $E_1 \in \mathscr{E}_1$ and $E_2 \in \mathscr{E}_2$, $P(E_1 \cap E_2) = P_1(E_1)P_2(E_2) = P(E_1 \cap \mathbb{W})P(\mathbb{W} \cap E_2) = P(E_1)P(E_2)$. Hence $\mathscr{E}_1$ and $\mathscr{E}_2$ are stochastically independent sub-σ-algebras of $\mathscr{E}$. This expresses that Σ_1 and Σ_2 model phenomena that are stochastically independent.

The deterministic systems $(\mathbb{W}, \mathscr{E}_1, P_1)$ and $(\mathbb{W}, \mathscr{E}_2, P_2)$ with behavior $\mathbb{B}_1$ and $\mathbb{B}_2$ respectively, are complementary if either $\mathbb{B}_1 = \mathbb{W}$, or $\mathbb{B}_2 = \mathbb{W}$, or if $\mathbb{B}_1$ and $\mathbb{B}_2$ are both strict subsets of $\mathbb{W}$ and $\mathbb{B}_1 \cap \mathbb{B}_2 \neq \emptyset$. Their interconnection is equivalent to the deterministic system $(\mathbb{W}, \mathscr{E}, P)$ with behavior $\mathbb{B}_1 \cap \mathbb{B}_2$.

We illustrate interconnection by our two examples.

Example 3: The noisy resistor, interconnected. Consider the interconnection of a noisy resistor and a voltage source with an internal resistance and thermal noise. In terms of the equivalent circuits with a random voltage source, this leads to

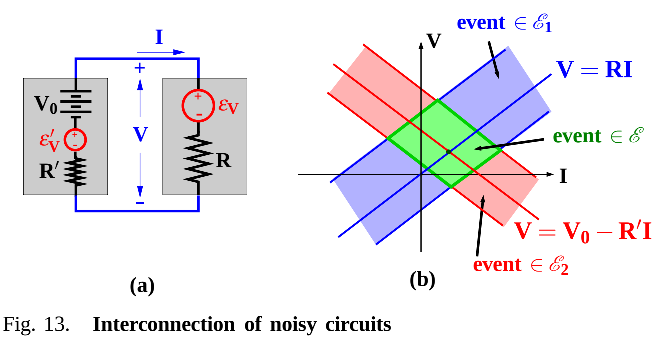


Fig. 13. **Interconnection of noisy circuits**

the configuration shown in Figure 13(a). System 1 corresponds to the noisy resistor described by (1, 2). System 2 correspond to the voltage source, and is described by equation $V = V_0 - R'I + \varepsilon_V'$ with V_0 a constant voltage, R' the internal resistance, and ε_V' a random variable independent of ε_V. Assume that ε_V' is gaussian, with zero mean and standard deviation σ'. A rectangular event of the interconnection is shown in Figure

13(b). It is easily seen that the corresponding σ-algebras are complementary if and only if $R+R' \neq 0$. The σ-algebra of the interconnected system is then the Borel σ-algebra on $\mathbb{R}^2$, and $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$ is the classical 2-dimensional random vector governed by

$$\begin{bmatrix} V \\ I \end{bmatrix} = \frac{1}{R+R'} \begin{bmatrix} R' & R \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_V \\ \varepsilon'_V + V_0 \end{bmatrix},$$

with ε_V as in equation (3).

■

Example 4: Equilibrium price/demand/supply. Consider first the deterministic price/demand and price/supply characteristics of an economic good shown in Figure 4. Assuming that these characteristics pertain to the same good is expressed by price$_1$=price$_2$, while equilibrium implies demand = supply. We

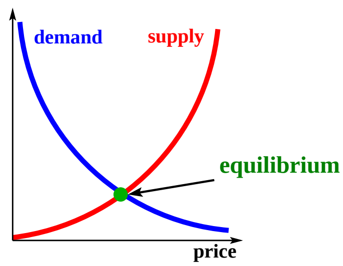


Fig. 14. **Deterministic equilibrium price/demand/supply**

view imposing the equilibrium conditions as interconnection. It is readily verified that the interconnection of the deterministic price/demand and price/supply systems yields the deterministic system with equilibrium behavior the intersection of the price/demand and price/supply characteristics as illustrated in Figure 14.

In the stochastic case, we start with the stochastic system $\Sigma_1 = ((0,\infty)^2, \mathscr{E}_1, P_1)$ that models the price/demand, and $\Sigma_2 = ((0,\infty)^2, \mathscr{E}_2, P_2)$ that models the price/supply. The elements of $\mathscr{E}_1$ and $\mathscr{E}_2$ are those to which a probability is assigned (see the discussion of Example 2 in Section II). Interconnection of Σ_1 and Σ_2 means $p_1 = p_2 = p$ (expressing that the prices pertain to the same good), and $d = s$ (expressing the equilibrium condition demand = supply).

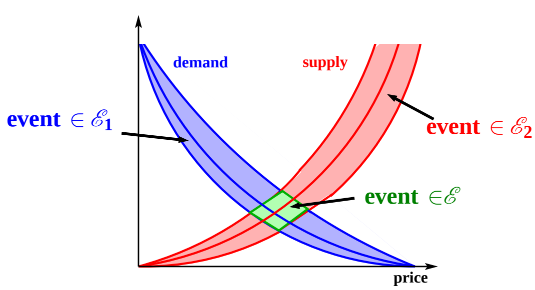


Fig. 15. **Price/demand/supply event**

Under reasonable conditions (related, for example, to the cardinality, shape, and monotonity of the price/demand and price/supply events) the associated σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$ are complementary, and the interconnection σ-algebra consists of the Borel subsets of $(0,\infty)^2$. A rectangular event for the interconnected stochastic system is shown in Figure 15. The probability for the interconnected stochastic system follows the construction of Definition 6. The specific case discussed in Example 2 leads to the following equations for the interconnected system

$$p = \sqrt[3]{\frac{\varepsilon_1}{\varepsilon_2}}, \qquad d = s = \sqrt[3]{\varepsilon_1^2 \varepsilon_2}.$$

■

For the interconnection of the stochastic system $\Sigma_1 = (\mathbb{W}, \mathscr{E}_1, P_1)$ with the deterministic system $\Sigma_2 = (\mathbb{W}, \mathscr{E}_2, P_2)$ with behavior $\mathbb{B}$, stochastic independence is trivially satisfied. Σ_1 and Σ_2 are then complementary if and only if

$$[E_1, E'_1 \in \mathscr{E}_1, \text{ and } E_1 \cap \mathbb{B} = E'_1 \cap \mathbb{B}] \Rightarrow [P_1(E_1) = P_1(E'_1)].$$

Assuming this complementarity, interconnection leads to the stochastic system that is equivalent to $(\mathbb{W}, \mathscr{E}, P)$ with $\mathscr{E} = \mathscr{E}_\mathbb{B} \cup \{\mathbb{B}^{\text{complement}}, \mathbb{W}\}$, where $\mathscr{E}_\mathbb{B} = \{E_1 \cap \mathbb{B} \mid E_1 \in \mathscr{E}_1\}$. The probability P of the interconnection is given by $P(E) = P_1(E_1)$ with E_1 any element of $\mathscr{E}_1$ such that $E = E_1 \cap \mathbb{B}$. This implies that $P(\mathbb{B}) = 1$, and the probability in the interconnected system is therefore concentrated on $\mathbb{B}$.

We now consider the interconnection of terminals as shown in Figure 10. Before interconnection, we have the stochastic system $\Sigma = (\mathbb{W} \times \mathbb{W}, \mathscr{E}, P)$ with variables (w_1, w_2). Both w_1 and w_2 have their outcomes in $\mathbb{W}$, and the laws governing these outcomes are coupled through $\mathscr{E}$ and P. The interconnection imposes $w_1 = w_2$ and we wish to consider the stochastic system that governs $w = w_1 = w_2$. This stochastic system a special case of the interconnection with a deterministic system discussed in the previous paragraph with the behavior given by $\mathbb{B} = \{(w_1, w_2) \in \mathbb{W} \times \mathbb{W} \mid w_1 = w_2\}$. Complementarity requires that

$$[E_1, E_2 \in \mathscr{E} \text{ and } E_1 \cap \mathbb{B} = E_2 \cap \mathbb{B}] \Rightarrow [P(E_1) = P(E_2)].$$

Assuming complementarity, interconnection yields the stochastic system $\Sigma' = (\mathbb{W}, \mathscr{E}', P')$ with

$$[E' \in \mathscr{E}'] \Leftrightarrow [\ \exists\ E \in \mathscr{E} \text{ such that } E' = \{(w,w) \mid (w,w) \in E\}]$$

and

$$P'(E') = P(E).$$

System interconnection (see Figure 9) and terminal interconnection (see Figure 10) are closely related. However, terminal interconnection is more general, since it also deals with interconnection of systems that are not stochastically independent.

VII. Interconnection of Linear Systems

Theorem 1: Consider the linear $\mathtt{n}$-dimensional stochastic systems $\Sigma_1 = (\mathbb{R}^\mathtt{n}, \mathscr{E}_1, P_1)$ and $\Sigma_2 = (\mathbb{R}^\mathtt{n}, \mathscr{E}_2, P_2)$ with associated fibers $\mathbb{L}_1$ and $\mathbb{L}_2$. The σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$ are complementary if and only if

$$\mathbb{L}_1 + \mathbb{L}_2 = \mathbb{R}^\mathtt{n}.$$

If this relation is satisfied, then the interconnected system $\Sigma_1 \wedge \Sigma_2$ is again a linear $\mathtt{n}$-dimensional stochastic system. Its fiber is $\mathbb{L}_1 \cap \mathbb{L}_2$. Hence $\Sigma_1 \wedge \Sigma_2$ is a classical $\mathtt{n}$-dimensional random vector if and only if

$$\mathbb{L}_1 \oplus \mathbb{L}_2 = \mathbb{R}^\mathtt{n}.$$

If Σ_1 and Σ_2 are gaussian, so is $\Sigma_1 \wedge \Sigma_2$. ■

The straightforward proof is omitted.

VIII. Open versus closed systems

As a general principle, it is best to aim for models that are *open* systems, and a mathematical theory of modeling should reflect this aspect from the very beginning. Models usually leave some of the individual variables free, unexplained, and merely express what one can conclude about a coupled set of variables. A model should incorporate the influence of the environment, but should leave the environment as unmodeled. The gas law does not explain the pressure all by itself, but it explains what one can conclude about the simultaneous occurrence of the pressure, the temperature, the volume, and the quantity of a gas. Which value of these variables actually occurs as the result of an experiment depends on both the system and the environment. A model of the environment is not part of the gas law. Newton's second law does not explain the position of a pointmass all by itself, but it explains it in combination with the force acting on it. The way the force is generated is not part of Newton's second law. Maxwell's equations do not explain the magnetic field, the current, and the charge density, but they explain them in combination with the electric field.

As an illustration of what we mean by 'open' versus 'closed' stochastic systems, consider the noisy resistor. Figure 1(c) and Equation (3) describe an open system, since knowledge of the system parameter R and the value of the internal noisy voltage source ε_V do not suffice to decide what V and I will be. The actual value of V and I depends in addition on some environmental conditions. Figure 13(a) is an example of a closed system, since knowledge of the system parameters R, R', V_0 and of the values of the internal noisy voltage sources ε_V and ε'_V suffices to decide what V and I are. Analogously, the price/demand characteristic of Figure 5(a) describes an open system. So does the price/supply characteristic of Figure 5(b). On the other hand, the equilibrium situation of Figure 15 describes a closed stochastic system, assuming that the interconnection σ-algebra consists of the Borel subsets of $(0,\infty)^2$. In the former case, the actual price, demand, and supply depend on environmental conditions in addition the characteristics and the randomness, while in the latter case, the characteristics and the randomness determine the price, demand, and supply.

Consider the classical notion of an $\mathtt{n}$-dimensional stochastic vector process as a family of measurable maps $f_t : \Omega \to \mathbb{R}^{\mathtt{n}}, t \in \mathbb{T}$ ($\mathbb{T}$ denotes the time-set), from a basic probability space Ω with σ-algebra $\mathscr{A}$, to $\mathbb{R}^{\mathtt{n}}$ with σ-algebra $\mathscr{B}(\mathbb{R}^{\mathtt{n}})$. This is very much a closed systems view, since once the uncertain parameter $\omega \in \Omega$ has been realized, the complete trajectory $t \in \mathbb{T} \mapsto f_t(\omega) \in \mathbb{R}^{\mathtt{n}}$ is determined. Such models leave no room for the influence of the environment and this is, in our view, a shortcoming. Stochastic systems with a coarse σ-algebra do allow to incorporate the unexplained environment. Definition 6 shows that if stochastic systems are complementary, it is possible to interconnect them. This feature shows the open nature of stochastic systems with coarse σ-algebras.

Another way of looking at 'open' versus 'closed' is by considering interconnection. An open stochastic system can be interconnected with other systems, a closed system cannot be interconnected (or, more accurately, it can only be interconnected with a trivial stochastic system). We now illustrate that coarseness of the σ-algebras is essential for complementarity. Assume that $\Sigma_1 = (\mathbb{R}^{\mathtt{n}}, \mathscr{B}(\mathbb{R}^{\mathtt{n}}), P)$ is a classical random vector and that $\Sigma' = (\mathbb{R}^{\mathtt{n}}, \mathscr{E}', P')$ is a stochastic system with $\mathscr{E}' \subseteq \mathscr{B}(\mathbb{R}^{\mathtt{n}})$. Then the σ-algebras associated with Σ_1 and Σ_2 can only be complementary if $\mathscr{E}'$ is trivial, that is, if $\mathscr{E}' = \{\emptyset, \mathbb{R}^{\mathtt{n}}\}$. More generally, if the stochastic systems Σ and Σ' are complementary then for $E \in \mathscr{E}'$, we have $E \cap \mathbb{R}^{\mathtt{n}} = E \cap E = \mathbb{R}^{\mathtt{n}} \cap E$, and hence $P(E) = P(E)P'(E) = P'(E)$. Therefore the following zero-one law must hold:

$$[\![E \in \mathscr{E}']\!] \Rightarrow [\![P(E) = P'(E) = 0 \text{ or } P(E) = P'(E) = 1]\!].$$

This is a very restrictive condition on Σ'. For example, if $\mathtt{support}(P) = \mathbb{R}^{\mathtt{n}}$, then $\mathscr{E}'$ cannot contain sets E such that both E and $E^{\text{complement}}$ have a non-empty interior.

We conclude that *classical random vectors are models of closed systems*. These systems cannot be interconnected with non-trivial systems. Open systems require a coarse σ-algebra. This shows a serious limitation of the classical stochastic framework, since interconnection ought to be one of the basic tenets of model building.

IX. Identification

In this section we discuss some implications to the problem of building models from data of the view of stochastic systems and their interconnection that emerges from the previous sections. The question we deal with is system identification: *how can we recover the laws that govern a stochastic system from measurements?*

Consider the stochastic system $\Sigma = (\mathbb{W}, \mathscr{E}, P)$. Assume that outcomes, realizations of the variables $w \in \mathbb{W}$, of the phenomenon that is modeled by Σ are observed. The aim is to identify the model, that is $\mathscr{E}$ and P, from these observations. In order to generate these observations, experimental conditions need to be set up during the data collection process. The data do not emerge from the stochastic system all by itself, but from observing Σ in interaction with an environment (see Figure 16). One of the questions that arises is whether it is

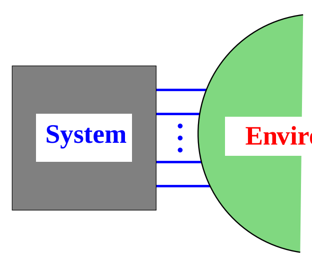


Fig. 16. **Data collection**

possible to disentangle from the data the laws of the stochastic system from the laws of the environment. There is a clear distinction between modeling a stochastic system from data and obtaining the statistical features of a random vector from samples. The latter problem consists of inferring the statistical laws by sampling a random vector in an experimental set-up, while the former problem requires in addition disentangling the laws of the system from the laws of the environment that was active while sampling.

Let us illustrate this issue by means of the noisy resistor of Figure 1(c). The variables $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$ are governed by Equations (1, 2) and the identification problem consists in deducing the parameters of the model, that is R and σ, from measured voltage/current pairs. These measurements may be generated in various ways. One possibility is to fix the current by driving the noisy resistor by a constant current source and measure various realizations of the voltage. Another possibility is to fix the voltage by putting a constant voltage source across the noisy resistor and measure various realizations of the current. A third possibility is to terminate the noisy resistor by a voltage source with internal resistance and thermal noise as shown in Figure 13, and measure various realizations of the voltage/current pair. These terminations of the noisy resistor give rise to three data clouds, each with completely different statistical features, and from each of these data clouds we may attempt to deduce the parameters R and σ.

For the noisy resistor it may be reasonable to assume that the experimenter can control the environmental conditions that are active during data collection. On the other hand, in many situations, for instance in economics, in the social sciences, or in biology, the data are collected in a passive way, *in vivo*, so to speak. The problem of disentangling the laws of the system from the laws of the environment then becomes imperative. As an example, assume that we wish to identify the stochastic system that governs the price/demand of an economic good. We could attempt to deduce the laws of this stochastic system from observing various realizations of the variables (p,d). If these measurements are obtained under the equilibrium condition demand = supply, then, as shown in Section VI, under reasonable conditions, the data are realizations of a classical 2-dimensional random vector, and then the probability distribution of the vector $\left[\begin{smallmatrix} p \\ d \end{smallmatrix}\right]$ depends not only on the stochastic price/demand system, but also on the stochastic price/supply system. The stochastic laws of the price/supply may also be unknown. *Is it nevertheless possible to identify the stochastic price/demand system from the data?*

In this paper we discuss only a very special case of the identification problem. We assume that the system to be identified is an $\mathtt{n}$-dimensional gaussian stochastic system. We further assume that the data are collected while the system is interconnected with another $\mathtt{n}$-dimensional gaussian stochastic system that is stochastically independent and complementary to the system to be identified, and such that the interconnected system is a classical random vector. As we have seen in Section VII, this classical random vector is also gaussian and we assume that from sampling, its mean and covariance matrix have been deduced. We assume therefore that the data consist of the mean and covariance of the probability distribution of the outcomes in the interconnected system.

Let $\mathbb{L} \subseteq \mathbb{R}^{\mathtt{n}}$ be the fiber of the gaussian system $\Sigma = (\mathbb{R}^{\mathtt{n}}, \mathscr{E}, P)$ to be identified and let $Rw = \varepsilon$ be a kernel representations of Σ. $R \in \mathbb{R}^{\mathtt{p}\times\mathtt{n}}$ is a matrix of full row rank and $\mathbb{L} = \mathtt{kernel}(R)$. Since Σ is assumed to be gaussian, ε is a classical gaussian $\mathtt{p}$-dimensional random vector. Let $m \in \mathbb{R}^{\mathtt{p}}$ be the mean and $S \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}, S = S^\top \succeq 0$, the covariance of ε. Let $\mathbb{L}' \subseteq \mathbb{R}^{\mathtt{n}}$ denote the fiber of the gaussian system $\Sigma' = (\mathbb{R}^{\mathtt{n}}, \mathscr{E}', P')$ that is interconnected with Σ during data collection. Assume that Σ and Σ' are stochastically independent and that $\mathbb{L} \oplus \mathbb{L}' = \mathbb{R}^{\mathtt{n}}$. Then, as shown in Section VII, the σ-algebras of Σ and Σ' are complementary and the interconnected system $\Sigma_{\text{observed}} = \Sigma \wedge \Sigma'$ is a classical $\mathtt{n}$-dimensional stochastic system $(\mathbb{R}^{\mathtt{n}}, \mathscr{B}(\mathbb{R}^{\mathtt{n}}), P_{\text{observed}})$ with P_{observed} a gaussian probability distribution on $\mathbb{R}^{\mathtt{n}}$. Let $\mu \in \mathbb{R}^{\mathtt{n}}$ be its the mean and $\Gamma \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}, \Gamma = \Gamma^\top \succeq 0$, its covariance.

Σ is unfalsified by Σ_{observed} if and only if

$$R\mu = m \quad \text{and} \quad R\Gamma R^\top = S. \tag{6}$$

The disentanglement question becomes: *Is it possible to deduce from equations (6) the stochastic system Σ, that is (R,m,S) up to the equivalence (5), from the observed system Σ_{observed}, that is from (μ,Γ)?*

Let $R'w = \varepsilon'$ be a kernel representation of Σ'. $R' \in \mathbb{R}^{(\mathtt{n}-\mathtt{p})\times\mathtt{n}}$ is a matrix of full row rank with $\mathtt{kernel}(R') = \mathbb{L}'$. Let $m' \in \mathbb{R}^{(\mathtt{n}-\mathtt{p})}$ be the mean and $S' \in \mathbb{R}^{(\mathtt{n}-\mathtt{p})\times(\mathtt{n}-\mathtt{p})}, S' = S'^\top \succeq 0$, the covariance of ε'. Since Σ and Σ' are assumed to be stochastically independent, ε and ε' are independent. $\mathbb{L} \oplus \mathbb{L}' = \mathtt{kernel}(R) \oplus \mathtt{kernel}(R') = \mathbb{R}^{\mathtt{n}}$ implies that the matrix $\left[\begin{smallmatrix} R \\ R' \end{smallmatrix}\right] \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$ is nonsingular. Hence

$$\begin{bmatrix} R \\ R' \end{bmatrix} w = \begin{bmatrix} \varepsilon \\ \varepsilon' \end{bmatrix}$$

is a kernel representation of $\Sigma_{\text{observed}} = \Sigma \wedge \Sigma'$. The mean μ and covariance Γ of $\Sigma_{\text{observed}} = \Sigma \wedge \Sigma'$ are related to the parameters R,m,S,R',m',S' of Σ and Σ' by

$$\begin{aligned} \begin{bmatrix} R \\ R' \end{bmatrix} \mu &= \begin{bmatrix} m \\ m' \end{bmatrix} \\ \begin{bmatrix} R \\ R' \end{bmatrix} \Gamma \begin{bmatrix} R \\ R' \end{bmatrix}^\top &= \begin{bmatrix} S & O_{\mathtt{p}\times(\mathtt{n}-\mathtt{p})} \\ O_{(\mathtt{n}-\mathtt{p})\times\mathtt{p}} & S' \end{bmatrix}. \end{aligned} \tag{7}$$

The following theorem shows the extent to which it is possible to deduce the parameters R,m,S,R',m',S' of Σ and Σ' from the parameters μ,Γ of $\Sigma_{\text{observed}} = \Sigma \wedge \Sigma'$.

Theorem 2: Let $\mu \in \mathbb{R}^{\mathtt{n}}$ and $\Gamma \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}, \Gamma = \Gamma^\top \succeq 0$, be given. For every $R' \in \mathbb{R}^{(\mathtt{n}-\mathtt{p})\times\mathtt{n}}$ of full row rank, there exist

1) $R \in \mathbb{R}^{\mathtt{p}\times\mathtt{n}}$ with $\left[\begin{smallmatrix} R \\ R' \end{smallmatrix}\right] \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$ nonsingular,
2) $m \in \mathbb{R}^{\mathtt{p}}$ and $m' \in \mathbb{R}^{\mathtt{n}-\mathtt{p}}$,
3) $S \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}, S = S^\top \succeq 0$, and $S' \in \mathbb{R}^{(\mathtt{n}-\mathtt{p})\times(\mathtt{n}-\mathtt{p})}, S' = S'^\top \succeq 0$,

such that (7) holds. If $R'\Gamma R'^\top \succ 0$, then R,m,S are uniquely determined by (7), up to the equivalence (5). ■

Proof: By choosing a suitable bases in the domain and co-domain of R', we can assume that

$$R' = \begin{bmatrix} O_{(\mathtt{n}-\mathtt{p})\times\mathtt{p}} & I_{(\mathtt{n}-\mathtt{p})\times(\mathtt{n}-\mathtt{p})} \end{bmatrix}.$$

Choose $R = \begin{bmatrix} I_{\mathtt{p}\times\mathtt{p}} & -L \end{bmatrix}$ with $L \in \mathbb{R}^{\mathtt{p}\times(\mathtt{n}-\mathtt{p})}$ to be determined. Clearly $\left[\begin{smallmatrix} R \\ R' \end{smallmatrix}\right] \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$ is nonsingular. Partition μ and Γ conformably to R', as

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} \Gamma_{1,1} & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_{2,2} \end{bmatrix}.$$

Equations (7) become

$$m = \mu_1 + L\mu_2, \quad m' = \mu_2,$$

$$S = \Gamma_{1,1} - \Gamma_{1,2}L^\top - L\Gamma_{2,1} + L\Gamma_{2,2}L^\top, \;\; S' = \Gamma_{2,2}, \;\; \Gamma_{1,2} = L\Gamma_{2,2}.$$

These equations define m, m', S, S', provided there exists L such that $\Gamma_{1,2} = L\Gamma_{2,2}$. $\Gamma \succeq 0$ implies that $\texttt{kernel}(\Gamma_{2,2}) \subseteq \texttt{kernel}(\Gamma_{1,2})$. Hence there indeed exists an L such that $\Gamma_{1,2} = L\Gamma_{2,2}$. Hence there exist then L, m, S, m', S' such that (7) holds.

Since $R'\Gamma R'^\top \succ 0$ corresponds to $\Gamma_{2,2} \succ 0$, this implies that the solution L is unique and given by $L = \Gamma_{1,2}\Gamma_{2,2}^{-1}$. Hence there then exist unique L, m, S, m', S' such that (7) holds. ■

The above theorem obviously also holds with the roles of Σ and Σ' reversed. The theorem shows that without further assumptions on Σ or Σ', it is not possible to deduce the laws of Σ from the laws of Σ_{observed}. In fact, Σ being unfalsified from Σ_{observed} leaves the fiber of Σ completely unspecified. So, not only is Σ unidentifiable from Σ_{observed}, but the deterministic part of Σ, governed by $Rw = 0$, is left completely arbitrary. Without further structural information on the system or on the environment, it is not possible to recover the parameters of Σ from sampling. The theorem also implies that the parameters μ, Γ of Σ_{observed} together with the fiber $\mathbb{L}'$ of Σ' specify Σ and Σ' uniquely, provided $R'\Gamma R'^\top \succ 0$. The condition $R'\Gamma R'^\top \succ 0$ is called *sufficiency of excitation*. It requires that there is an adequate variety of experiments generated by the environment.

For the economic example the full complexity of the identifiability question emerges. Sampling under equilibrium conditions does not lead to identification of the price/demand elasticity. A more elaborate controlled experiment is needed to entangle the price/demand and price/supply systems.

There are many applications in statistics in which one attempts to identify the stochastic laws governing a phenomenon involving two real variables. As we remarked, such a law invariably leads to a coarse σ-algebra. The important observation here is that data generation through sampling requires interconnection with another system, and therefore data collection involves *two* distinct random systems. One of these stochastic systems expresses the intrinsic random laws one is after, while the other expresses the features of the environment that happens to be acting during the data collection experiment. Disentangling these laws requires further structural assumptions on the experimental set-up.

X. Constrained Probability

Consider the stochastic system $(\mathbb{W}, \mathscr{E}, P)$. Let $\mathbb{S}$ be a nonempty subset of $\mathbb{W}$. In this section we discuss the meaning of *the stochastic system induced by* $(\mathbb{W}, \mathscr{E}, P)$ *with outcomes constrained to be in* $\mathbb{S}$. We shall see that this is indeed a sensible concept.

Before entering into the mathematical development, we illustrate the concept which we will introduce in this section by means of some examples.

Example 5: The noisy resistor with constraints. Consider the noisy resistor described by Equations (1, 2). Now impose the constraint $I = 1$ amp. *What is the probability distribution of the resulting voltage?* From an equation point of view, the answer is easy. The noisy resistor is described by equation (3). When $I = 1$ amp, then V is then equal to $R + \varepsilon_V$, a gaussian random variable with mean R and standard deviation σ. The problem is to deduce this result from the events and the probability associated with the noisy resistor. Note that $V = R + \varepsilon_V$ is not the result from conditioning the random vector $\begin{bmatrix} V \\ I \end{bmatrix}$ by $I = 1$ amp, since $\{\begin{bmatrix} V \\ I \end{bmatrix} \in \mathbb{R}^2 \mid I = 1 \text{ amp}\}$ is not an event. Similarly, we want to deduce that imposing $V = 10$ volt, leads to $I = 10/R - \varepsilon_V/R$, a gaussian random variable with mean $10/R$ and standard deviation σ/R.

Example 6: Price/demand and price/supply with constraints. Consider the stochastic price/demand system discussed in Example 2. Impose the condition price_1 = 1€ as illustrated in Figure 17(a). *What is the resulting probability distribution of the demand?* Similarly, consider the price/supply system and impose the condition price_2 = 1€ as illustrated in Figure 17(b). *What is the resulting probability distribution of the supply?* It is readily seen that these are sensible questions which for the specific cases discussed in Example 2 lead to $d = \varepsilon_1$ and $s = \varepsilon_2$.

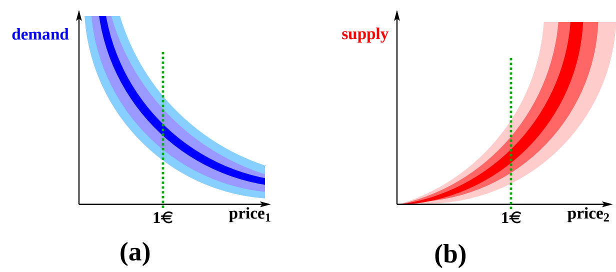


Fig. 17. **Stochastic price/demand and price/supply events**

The problem is to deduce these probability distributions from the σ-algebras and the probabilities associated with the stochastic systems that describe the noisy resistor and the price/demand and price/supply characterisitics.

Definition 7: Let $\Sigma = (\mathbb{W}, \mathscr{E}, P)$ be a stochastic system and $\mathbb{S} \subseteq \mathbb{W}$. Assume that the regularity condition

$$[\![E_1, E_2 \in \mathscr{E} \text{ and } E_1 \cap \mathbb{S} = E_2 \cap \mathbb{S}]\!] \Rightarrow [\![P(E_1) = P(E_2)]\!]$$

holds. Then the stochastic system

$$\Sigma|_{\mathbb{S}} := (\mathbb{S}, \mathscr{E}|_{\mathbb{S}}, P|_{\mathbb{S}})$$

with

$$\mathscr{E}|_{\mathbb{S}} := \{E' \in \mathscr{E} \mid E' = E \cap \mathbb{S} \text{ for some } E \in \mathscr{E}\},$$

and

$$P|_{\mathbb{S}}(E') := P(E) \text{ with } E \in \mathscr{E} \text{ such that } E' = E \cap \mathbb{S},$$

is called *the stochastic system* Σ *with outcomes constrained to be in* $\mathbb{S}$. ■

Constraining corresponds to interconnecting with the deterministic system $(\mathbb{W}, \{\emptyset, \mathbb{S}, \mathbb{S}^{\text{complement}}, \mathbb{W}\}, P)$ and regularity corresponds to complementarity. The regularity condition basically implies $\mathbb{S} \notin \mathscr{E}$. In fact, if $\mathbb{S} \in \mathscr{E}$, then regularity holds if and only if $P(\mathbb{S}) = 1$. In order to see this, observe first that $\mathbb{S} \cap \mathbb{S} = \mathbb{W} \cap \mathbb{S}$. Hence $\mathbb{S} \in \mathscr{E}$ and regularity yield $P(\mathbb{S}) = P(\mathbb{W}) = 1$. Conversely, assume that $\mathbb{S} \in \mathscr{E}$ and $P(\mathbb{S}) = 1$. Then $E \in \mathscr{E}$ implies $P(E) = P(E \cap \mathbb{W}) = P(E \cap \mathbb{S}) + P(E \cap \mathbb{S}^{\text{complement}}) = P(E \cap \mathbb{S})$. Therefore $E_1, E_2 \in \mathscr{E}$ and $E_1 \cap \mathbb{S} = E_2 \cap \mathbb{S}$ imply

$P(E_1) = P(E_1 \cap \mathbb{S}) = P(E_2 \cap \mathbb{S}) = P(E_2)$. Hence regularity holds. It follows that constraining is interesting when $\mathbb{S} \notin \mathscr{E}$.

The notion of *the stochastic system* Σ *with outcomes constrained to be in* $\mathbb{S}$, while reminiscent of the notion of *the stochastic system* Σ *conditioned on outcomes in* $\mathbb{S}$, is quite different from it. The former basically requires $\mathbb{S} \notin \mathscr{E}$, while the latter requires $\mathbb{S} \in \mathscr{E}$. Secondly, constraining associates with the event $E \in \mathscr{E}$ of Σ, the event $E \cap \mathbb{S}$ of $\Sigma|_{\mathbb{S}}$ with probability $P(E)$, while conditioning associates with the event $E \in \mathscr{E}$ of Σ the event $E \cap \mathbb{S}$, also in $\mathscr{E}$, with probability $P(E \cap \mathbb{S})/P(\mathbb{S})$. So, constraining pulls the probability of E 'globally' into $E \cap \mathbb{S}$, while conditioning associates with E 'locally' the probability of $E \cap \mathbb{S}$, renormalized by dividing by $P(\mathbb{S})$.

Constraining allows to deduce probability distributions on the outcome space beyond those that are obtained during the identification process. Assume that the stochastic system $\Sigma = (\mathbb{R}^{\mathtt{n}}, \mathscr{E}, P)$ is interconnected with $\Sigma' = (\mathbb{R}^{\mathtt{n}}, \mathscr{E}', P')$ during data collection and that $\Sigma \wedge \Sigma' = \Sigma_{\text{observed}} = (\mathbb{R}^{\mathtt{n}}, \mathscr{B}(\mathbb{R}^{\mathtt{n}}), P'')$ is a classical $\mathtt{n}$-dimensional system. Assume further that Σ is identified by sampling Σ_{observed} and disentangling Σ from Σ', in the manner discussed in Section IX. The question is to determine the probability distribution of the outcomes on a subset $\mathbb{S} \subseteq \mathbb{R}^{\mathtt{n}}$. In its present form, this question is ambiguous. Analyzing the outcomes in $\mathbb{S}$ in the experimental set-up to leads to the conditional probability P'' conditioned by $w \in \mathbb{S}$. A more relevant interpretation of the above question is that we want the probability associated with the stochastic system $(\mathbb{W}, \mathscr{E}, P)$ with outcomes constrained to be in $\mathbb{S}$. Both the conditional and constrained probability may be deduced from the identified system after sampling. The former requires deducing P'' from the samples and Σ_{observed}, while the latter requires first disentangling Σ from Σ', assuming that Σ is identifiable from Σ_{observed}, and subsequently constraining the outcomes to be in $\mathbb{S}$. In general, these two probabilities are quite different.

In order to make this difference more concrete, consider the noisy resistor (1, 2). We can deduce the parameters R and σ by sampling for example under the experimental conditions $I = 1$ amp. The conditional distribution of I conditioned by $V = 10$ volt in the interconnected system is the point measure concentrated at $I = 1$ amp. The distribution of I with outcomes constrained to satisfy $V = 10$ volts is gaussian with mean $10/R$ and standard deviation σ/R. Thus constraining allows to obtain probability distributions beyond the experimental set-up used to identify the model parameters. In this and similar examples, constraining appears a more relevant notion than conditioning, because of the prevalence of coarse σ-algebras.

The notion of the stochastic system Σ constrained by $w \in \mathbb{S}$ appears to be an interesting and useful addition to the list of elementary concepts in mathematical probability. It is a concept that is effective for stochastic systems with a coarse σ-algebra.

XI. Functions on the outcome space

In this section we discuss functions on the outcome space of a stochastic system. Consider the equation

$$w' = f(w) \tag{8}$$

with $w \in \mathbb{W}$ governed by the stochastic system $(\mathbb{W}, \mathscr{E}, P)$ and f a map from $\mathbb{W}$ into $\mathbb{W}'$. We want to construct the stochastic system $(\mathbb{W}', \mathscr{E}', P')$ that governs the outcomes of the variables $w' \in \mathbb{W}'$. A special case of (8) of particular interest is the projection $(w_1, w_2) \mapsto w_1$, which in Section V we have referred to as 'elimination'.

In classical probability theory with, for example, $\mathbb{W} = \mathbb{R}^{\mathtt{n}}$ and $\mathbb{W}' = \mathbb{R}^{\mathtt{n}'}$, the assumption is usually made that the σ-algebras $\mathscr{E}$ and $\mathscr{E}'$ are given, for example as the Borel σ-algebras, and that f is measurable, for example continuous, leading to the definition of P' as $P'(E') = P\left(f^{-1}(E')\right)$. In this case the events $\mathscr{E}$ and $\mathscr{E}'$ are obtained from the (topological) structure of the outcome spaces $\mathbb{W}$ and $\mathbb{W}'$ and therefore the construction of $\mathscr{E}$ and $\mathscr{E}'$ does not involve the probabilistic laws. The main theme of the present article is that the events are an essential part of a stochastic model and must therefore be constructed in accordance to the sets to which the model assigns a probability. When the variable w' is generated by (8), the question therefore emerges how to choose $\mathscr{E}'$ and P' from $\mathscr{E}, P$, and f, with $\mathscr{E}'$ the class of subsets of $\mathbb{W}'$ to which a probability can be assigned. This situation has already been set up by Kolmogorov in his original book on probability theory [1, III §1]

We start with some facts about σ-algebras and pullbacks of maps. Let $f : \mathbb{W} \to \mathbb{W}'$. The pullback f^{-1} satisfies

$$f^{-1}(E'^{\text{complement}}) = (f^{-1}(E'))^{\text{complement}} \text{ and}$$
$$f^{-1}\big(\underset{k\in\mathbb{N}}{\cup} E'_k\big) = \underset{k\in\mathbb{N}}{\cup}(f^{-1}(E'_k)).$$

These relations show that f^{-1} takes σ-algebras into σ-algebras, in both directions. More concretely, if $\mathscr{E}$ is a σ-algebra of subsets of $\mathbb{W}$, then the class of subsets $\mathscr{E}'$ of $\mathbb{W}'$ defined by

$$[\![E' \in \mathscr{E}']\!] :\Leftrightarrow [\![f^{-1}(E') \in \mathscr{E}]\!] \tag{9}$$

is also a σ-algebra of subsets of $\mathbb{W}'$. Conversely, if $\mathscr{E}'$ is a σ-algebra of subsets of $\mathbb{W}'$, then the class of subsets $\mathscr{E}$ of $\mathbb{W}$ defined by

$$[\![E \in \mathscr{E}]\!] :\Leftrightarrow [\![E = f^{-1}(E') \text{ for some } E' \in \mathscr{E}']\!]$$

is a sub-σ-algebra of $\mathscr{E}$.

Let $(\mathbb{W}, \mathscr{E}, P)$ be a stochastic system and $f : \mathbb{W} \to \mathbb{W}'$. Define $\mathscr{E}'$ by (9). Then $f : \mathbb{W} \to \mathbb{W}'$ is measurable with respect to the measurable spaces $(\mathbb{W}, \mathscr{E})$ and $(\mathbb{W}', \mathscr{E}')$, leading to the probability

$$P'(E') := P\left(f^{-1}(E')\right) \text{ for } E' \in \mathscr{E}'. \tag{10}$$

Definition 8: The stochastic system $(\mathbb{W}', \mathscr{E}', P')$ with $\mathscr{E}'$ defined by (9) and P' defined by (10) is called the *stochastic system on* $\mathbb{W}'$ *induced by* $(\mathbb{W}, \mathscr{E}, P)$ *and* $f : \mathbb{W} \to \mathbb{W}'$. ■

The construction of $\mathscr{E}'$ defined by (9) leads to the largest class of subsets of $\mathbb{W}'$ for which the probability can be defined from the probability of events in $\mathscr{E}$.

When $f : \mathbb{R}^{\mathtt{n}} \to \mathbb{R}^{\mathtt{m}}$ is linear, then Definition 8 associates with a linear stochastic system another linear stochastic system. The fiber $\mathbb{L}' \subseteq \mathbb{R}^{\mathtt{m}}$ in the co-domain is related to the fiber $\mathbb{L}$ in the domain by $\mathbb{L}' = f(\mathbb{L})$.

For the noisy resistor with $R \neq 0$, the maps $\begin{bmatrix} V \\ I \end{bmatrix} \mapsto V$ and $\begin{bmatrix} V \\ I \end{bmatrix} \mapsto I$ both generate the trivial stochastic system $(\mathbb{R}, \{\emptyset, \mathbb{R}\}, P')$. The variables V and I are hence not classical random variables. The only non-zero real linear functional on $\mathbb{R}^2$ that generates a non-trivial stochastic system is the map $\begin{bmatrix} V \\ I \end{bmatrix} \mapsto V - RI$ which generates a classical gaussian random variable with mean zero and standard deviation σ.

A common way in which probability enters into a system is that some of the variables w are modeled as random and influence other related variables w', for example by $f(w, w') = 0$, and the aim is to describe the stochastic behavior of the related variables w'. As a typical example think of modeling the terminal current/voltage behavior of an electrical circuit that contains stochastic sources. We explained how to construct the stochastic laws governing w' from the stochastic laws of w when w and w' are related by (8). The definition of the w'-events from the w-events is more involved and in general not easy to sort out when w and w' are related by a general implicit equation as $f(w, w') = 0$.

While the stochastic system $(\mathbb{W}', \mathscr{E}', P')$ on $\mathbb{W}'$ induced by $(\mathbb{W}, \mathscr{E}, P)$ and $f : \mathbb{W} \to \mathbb{W}'$ is a well-defined notion, a great deal of information may be lost when passing from $(\mathbb{W}, \mathscr{E}, P)$ and $f : \mathbb{W} \to \mathbb{W}'$ to $(\mathbb{W}', \mathscr{E}', P')$. The problem is that $\mathscr{E}'$ as constructed by (9) may contain very few events and certain properties and operations on $(\mathbb{W}, \mathscr{E}, P)$ may be lost when passing to $(\mathbb{W}', \mathscr{E}', P')$. We shall see an example of such a situation involving constraining in the next section.

XII. Probability Kernels

Open stochastic systems are often thought of as classical stochastic systems with 'inputs', that is, as a family of probability measures on an output space, parameterized by an input. Such families of probability measures go under the name of *probability kernels*. The main distinction between probability kernels and our approach consists in the input/output view of open systems that underlies probability kernels. While inputs and outputs definitely have their place in modeling, especially in signal processing and in feedback control, the input/output view has many drawbacks when modeling open physical systems, as argued for example in [2] for the deterministic case. With input/output thinking one cannot get off the ground when modeling, for example, simple electrical circuits [3], the paradigmatic examples of interconnected systems.

Developing the themes of the present article using probability kernels in their full generality lies beyond the scope of the present article. We merely explain some of the connections between our notion of stochastic system on the one hand, and probability kernels on the other hand, by means of an example that is important in applications, namely, the *binary channel*.

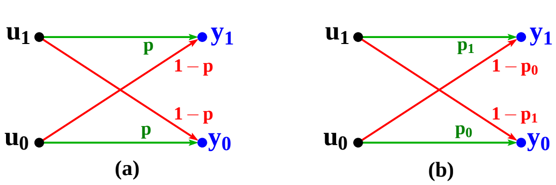


Fig. 18. **Binary channel**

The channel functions as follows. There are two possible inputs, u_0 and u_1. The channel transmits the input and produces two possible outputs, y_0 and y_1. The input u_0 leads to output y_0 with probability p_0 and to output y_1 with probability $1 - p_0$, while the input u_1 leads to output y_1 with probability p_1 and to output y_0 with probability $1 - p_1$. If $p_0 = p_1 = p$, then we call the channel *symmetric*, while if $p_0 \neq p_1$, then we call the channel *asymmetric*. The symmetric binary channel is shown in Figure 18(a), while the asymmetric binary channel is shown in Figure 18(b).

Formally, denote the input alphabet as $\mathbb{U} = \{u_0, u_1\}$ and the output alphabet as $\mathbb{Y} = \{y_0, y_1\}$. The channel is specified as two classical stochastic systems,

$$\Sigma_{u_0} = \left(\mathbb{Y}, 2^{\mathbb{Y}}, P_{u_0}\right) \quad \text{and} \quad \Sigma_{u_1} = \left(\mathbb{Y}, 2^{\mathbb{Y}}, P_{u_1}\right),$$

with the probabilities given by

$$\begin{aligned} P_{u_0}(\{y_0\}) &= p_0, \qquad P_{u_0}(\{y_1\}) = 1 - p_0, \\ P_{u_1}(\{y_0\}) &= 1 - p_1, \ P_{u_1}(\{y_1\}) = p_1. \end{aligned}$$

The pair of systems $\left(\Sigma_{u_0}, \Sigma_{u_1}\right)$ is an example of a probability kernel.

A. The symmetric binary channel

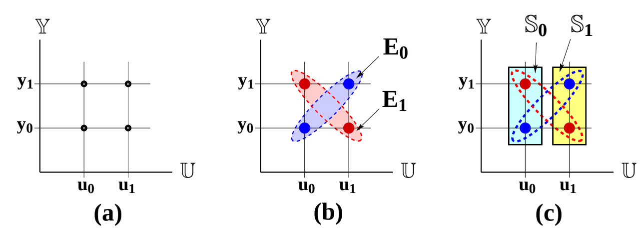


Fig. 19. **Events for the symmetric binary channel**

We now show how to approach the symmetric binary channel using constrained probability. Start with the system

$$\Sigma_{\text{sbc}} = (\mathbb{U} \times \mathbb{Y}, \mathscr{E}, P).$$

Thus the outcome space, shown in Figure 19(a), is $\mathbb{U} \times \mathbb{Y}$. The event set is $\mathscr{E} = \{\emptyset, E_0, E_1, \mathbb{U} \times \mathbb{Y}\}$. with

$$E_0 = \{(u_0, y_0), (u_1, y_1)\}, E_1 = \{(u_0, y_1), (u_1, y_0)\}$$

(see Figure 19(b)). Note that $\mathscr{E} \neq 2^{\mathbb{U} \times \mathbb{Y}}$. Hence Σ_{sbc} is not a classical stochastic system. The probability P is determined by

$$P(E_0) = p, \quad P(E_1) = 1 - p.$$

Now consider Σ_{sbc} with outcomes constrained to be in

$$\mathbb{S}_0 \ = \ \{(u, y) \mid u = u_0\} \text{ and } \mathbb{S}_1 = \{(u, y) \mid u = u_1\},$$

respectively. The sets $\mathbb{S}_0$ and $\mathbb{S}_1$ are illustrated in Figure 19(c). It is easily verified that the regularity condition of Definition 5 is satisfied for both $\mathbb{S}_0$ and $\mathbb{S}_1$. The resulting stochastic systems are $\Sigma_{\text{sbc}}|_{\mathbb{S}_0} = (\mathbb{Y}, 2^{\mathbb{Y}}, P|_{\mathbb{S}_0})$ and $\Sigma_{\text{sbc}}|_{\mathbb{S}_1} = (\mathbb{Y}, 2^{\mathbb{Y}}, P|_{\mathbb{S}_1})$ with

$$\begin{aligned} P|_{\mathbb{S}_0}(\{y_0\}) &= p, \qquad P|_{\mathbb{S}_0}(\{y_1\}) = 1 - p, \\ P|_{\mathbb{S}_1}(\{y_0\}) &= 1 - p, \ \ P|_{\mathbb{S}_1}(\{y_1\}) = p. \end{aligned}$$

Observe that $\Sigma_{\text{sbc}}|_{\mathbb{S}_0} = \Sigma_{u_0}$ and $\Sigma_{\text{sbc}}|_{\mathbb{S}_1} = \Sigma_{u_1}$ yield *precisely* the systems Σ_{u_0} and Σ_{u_1} that specify the channel as a probability kernel.

Note that the symmetric binary channel can be viewed as a linear stochastic system. Identify both $\mathbb{U}$ and $\mathbb{Y}$ with GF(2), the Galois field $\{0,1\}$. Set $\mathbb{W} = \mathbb{U} \times \mathbb{Y} = \text{GF}(2)^2$. Then Σ_{sbc} is a linear stochastic over the field GF(2) with fiber $\mathbb{L} = \{(0,0),(1,1)\}$ and probabilities $P(\{(0,0),(1,1)\}) = p$ and $P(\{(0,1),(1,0)\}) = 1-p$.

B. The asymmetric binary channel

We now show how to approach the asymmetric binary channel using probability kernels. Start with the stochastic system

$$\Sigma = (\mathbb{U} \times \mathbb{Y} \times \mathbb{E}, \mathscr{E}, P),$$

with $\mathbb{E} = \{e_1, e_2, e_3, e_4\}$. Thus the outcome space, shown in Figure 20(a), is the Cartesian product of $\mathbb{U} \times \mathbb{Y}$ and $\mathbb{E}$. The space $\mathbb{E}$ is introduced in order to generate the uncertainty in the channel. The events $\mathscr{E}$ consist of the σ-algebra generated by the pairs

$$\begin{aligned}
E_1 &= \{(u_0,y_0,e_1),(u_1,y_0,e_1)\},\\
E_2 &= \{(u_0,y_0,e_2),(u_1,y_1,e_2)\},\\
E_3 &= \{(u_0,y_1,e_3),(u_1,y_0,e_3)\},\\
E_4 &= \{(u_0,y_1,e_4),(u_1,y_1,e_4)\}.
\end{aligned}$$

Note that the σ-algebra generated by $\{E_1,E_2,E_3,E_4\}$ is not equal to $2^{\mathbb{U}\times\mathbb{Y}\times\mathbb{E}}$. So Σ is not a classical stochastic system. The probability P is determined by

$$\begin{aligned}
P(E_1) &= p_0(1-p_1), & P(E_2) &= p_0 p_1,\\
P(E_3) &= (1-p_0)(1-p_1), & P(E_4) &= (1-p_0)p_1.
\end{aligned}$$

The generating set for $\mathscr{E}$ is shown in Figure 20(b).

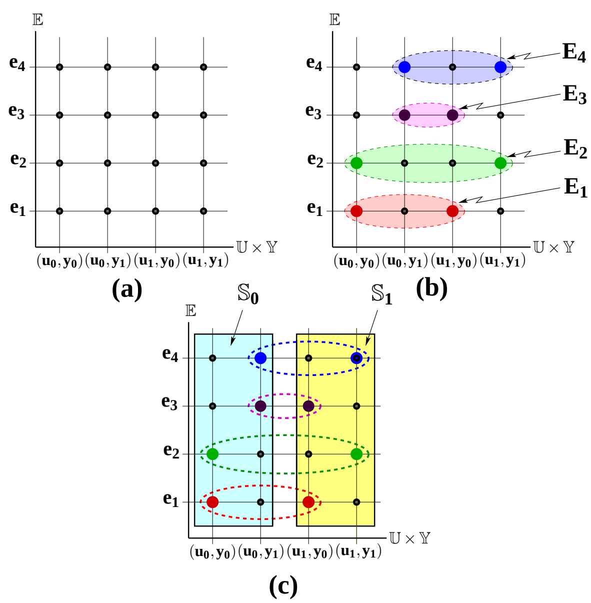


Fig. 20. **Events for the asymmetric binary channel**

Now consider the stochastic system Σ with outcomes constrained to be in $\mathbb{S}_0 = \{(u,y,e) \mid u = u_0\}$ and the stochastic system Σ with outcomes constrained to be in $\mathbb{S}_1 = \{(u,y,e) \mid u = u_1\}$. The sets $\mathbb{S}_0$ and $\mathbb{S}_1$ are illustrated in Figure 20(c). It is easily verified that the regularity condition of Definition 7 is satisfied for both $\mathbb{S}_0$ and $\mathbb{S}_1$. The constrained stochastic systems are denoted by $\Sigma|_{\mathbb{S}_0}$ and $\Sigma|_{\mathbb{S}_0}$ respectively and given by $\Sigma|_{\mathbb{S}_0} = (\mathbb{Y} \times \mathbb{E}, \mathscr{E}|_{\mathbb{S}_0}, P|_{\mathbb{S}_0})$ with $\mathscr{E}|_{\mathbb{S}_0}$ generated by

$$\begin{aligned}
E_1 &= \{(y_0,e_1)\}, & E_2 &= \{(y_0,e_2)\},\\
E_3 &= \{(y_1,e_3)\}, & E_4 &= \{(y_1,e_4)\},
\end{aligned}$$

and $P|_{\mathbb{S}_0}$ defined by

$$\begin{aligned}
P|_{\mathbb{S}_0}(E_1) &= p_0(1-p_1), & P|_{\mathbb{S}_0}(E_2) &= p_0 p_1,\\
P|_{\mathbb{S}_0}(E_3) &= (1-p_0)(1-p_1), & P|_{\mathbb{S}_0}(E_4) &= (1-p_0)p_1,
\end{aligned}$$

and $\Sigma|_{\mathbb{S}_1} = (\mathbb{Y} \times \mathbb{E}, \mathscr{E}|_{\mathbb{S}_1}), P|_{\mathbb{S}_1})$ with $\mathscr{E}|_{\mathbb{S}_1}$ generated by

$$\begin{aligned}
E_1 &= \{(y_0,e_1)\}, & E_2 &= \{(y_1,e_2)\},\\
E_3 &= \{(y_0,e_3)\}, & E_4 &= \{(y_1,e_4)\},
\end{aligned}$$

and $P|_{\mathbb{S}_1}$ defined by

$$\begin{aligned}
P|_{\mathbb{S}_1}(E_1) &= p_0(1-p_1), & P|_{\mathbb{S}_1}(E_2) &= p_0 p_1,\\
P|_{\mathbb{S}_1}(E_3) &= (1-p_0)(1-p_1), & P|_{\mathbb{S}_1}(E_4) &= (1-p_0)p_1.
\end{aligned}$$

Observe that after elimination of e, that is, the marginal probability for y, $\Sigma|_{\mathbb{S}_1}$ and $\Sigma|_{\mathbb{S}_2}$ yield *precisely* the systems Σ_{u_0} and Σ_{u_1} that specify the channel as a probability kernel.

The introduction of $\mathbb{E}$ and Σ shows that the asymmetric binary channel as a probability kernel can be interpreted in a as constrained stochastic systems. The probability kernel $(\Sigma_{u_0}, \Sigma_{u_1})$ can also be interpreted in terms of conditional probabilities. Define, for $\pi \in (0,1)$ the stochastic system $(\mathbb{U}, 2^{\mathbb{U}}, P_u)$ by $P_u(\{u_0\}) = \pi$ and $P_u(\{u_1\}) = 1-\pi$. We then obtain the family of stochastic systems

$$\Sigma_\pi = \left(\mathbb{U} \times \mathbb{Y}, 2^{\mathbb{U}\times\mathbb{Y}}, P_\pi\right)$$

with

$$\begin{aligned}
P_\pi(\{(u_0,y_0)\}) &= p_0 \pi,\\
P_\pi(\{(u_0,y_1)\}) &= (1-p_0)\pi,\\
P_\pi(\{(u_1,y_0)\}) &= (1-p_1)(1-\pi),\\
P_\pi(\{(u_1,y_1)\}) &= p_1(1-\pi).
\end{aligned}$$

For each of these systems, Σ_{u_0} and Σ_{u_1} are the conditional probabilities of y given u. Since the interpretation of a probability kernel as conditional probabilities requires modeling the environment, that is, interpreting the input u as a classical random variable, we feel that the interpretation in terms of constrained probability is a more satisfactory one conceptually.

Our treatment of the asymmetric binary channel is based on choosing the auxiliary outcomes in $\mathbb{E}$ and identifying the events E_1, E_2, E_3, E_4 with the associated probabilities such that constraining by $\mathbb{S}_0$ and $\mathbb{S}_1$ gives the channel probabilities after eliminating the variables e. Note that if we would have eliminated the e's immediately from the system Σ, equivalently, apply the projection $(u,y,e) \mapsto (u,y)$, then, as shown in Section XI, we obtain the stochastic system $\Sigma' = (\mathbb{U} \times \mathbb{Y}, \{\emptyset, \mathbb{U} \times \mathbb{Y}\}, P')$ with the trivial σ-algebra $\{\emptyset, \mathbb{U} \times \mathbb{Y}\}$. Constraining by $\mathbb{S}_0$ and $\mathbb{S}_1$ then becomes ineffective. This shows that in this example projection blends out a great deal of information. What is needed to correct this is allowing the projected atoms E_1, E_2, E_3, E_4 to

$$\tilde{E}_1 = \{(u_0,y_0),(u_1,y_0)\}, \quad \tilde{E}_2 = \{(u_0,y_0),(u_1,y_1)\},$$
$$\tilde{E}_3 = \{(u_0,y_1),(u_1,y_0)\}, \quad \tilde{E}_4 = \{(u_0,y_1),(u_1,y_1)\},$$

with the probabilities

$$\begin{aligned}
\tilde{P}(\tilde{E}_1) &= P(E_1) = p_0(1-p_1),\\
\tilde{P}(\tilde{E}_2) &= P(E_2) = p_0 p_1,\\
\tilde{P}(\tilde{E}_3) &= P(E_3) = (1-p_0)(1-p_1),\\
\tilde{P}(\tilde{E}_4) &= P(E_4) = (1-p_0)p_1
\end{aligned}$$

as the specification itself of a stochastic system on the outcome space $(\mathbb{U} \times \mathbb{Y})$. Unfortunately, the $\tilde{P}$'s do not define a probability on the σ-algebra generated by the $\tilde{E}$'s, and therefore the $(\tilde{E}, \tilde{P})$'s do not define a probability space in the orthodox sense of the term.

When ε a classical random vector, then $y = f(u, \varepsilon)$ can be dealt with by considering u as an input parameter which together with the random input ε generates the output y. For example, the symmetric binary channel can be realized this way by taking $\mathbb{U} = \mathbb{Y} = \{0,1\}$, ε a random variable taking values in $\{0,1\}$ with $P_\varepsilon(\{0\}) = p, P_\varepsilon(\{1\}) = 1-p$, and setting

$$u + y = \varepsilon$$

over GF(2). The asymmetric binary channel can be realized by setting $\mathbb{U} = \mathbb{Y} = \{0,1\}$, and

$$y = \varepsilon_0(1-u) + \varepsilon_1 u$$

with $\varepsilon_0, \varepsilon_1$ independent random variables both taking values in $\{0,1\}$ with $P(\{\varepsilon_0 = 0\}) = p_0$ and $P(\{\varepsilon_1 = 1\}) = p_1$. In terms of the e's discussed above, we have then $e_1 \leftrightarrow (0,0), e_2 \leftrightarrow (0,1), e_3 \leftrightarrow (1,0), e_4 \leftrightarrow (1,1)$.

Applying the thinking in terms of stochastic kernels to the noisy resistor, one could assume that I is an input which together with ε_V generates the random variable V through (3). This leads to an interpretation of the noisy resistor in terms of a probability kernel with $P_I(V)$ a gaussian real random variable with mean RI and standard deviation σ. There are several drawbacks of dealing with the noisy resistor in this way, the main one being that it does not put I and V a priori on equal footing. Our way of dealing with $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$ in terms of a coarse σ-algebra appears simpler, more general, and more satisfying conceptually. Modeling the noisy resistor with a random voltage source or a random current source is only an equivalent circuit view of a noisy resistor. Interpreting the events (1) and the probability (2) associated with the noisy resistor as the physics of a hot resistor is simpler, closer to reality, and generalizes to situations where the events are not cylindrical strips. Finally, it not evident how to deal with the stochastic price/demand and price/supply characteristic (see Figure 5) as stochastic kernels.

XIII. Conclusion

The main message of this paper is that a mathematical specification of a stochastic system should involve the events on an equal footing to the probability measure. The need to have not all Borel sets as events and to work with coarse σ-algebras is essential even for elementary applications.

Interconnection of stochastic systems can be defined effectively for stochastic systems with coarse event σ-algebras, but requires suitable properties of the event space, as complementarity of the stochastic systems or of the associated σ-algebras.

An interesting notion that emerges for systems with a coarse σ-algebra of events is constraining the outcomes to belong to a subset of the outcome space that is not an event.

One of the urgent directions of generalization of the notions of the present paper is to stochastic dynamical systems and stochastic processes. We have already pointed out that the classical notion of a stochastic process as a family of measurable maps $f_t : \Omega \to \mathbb{R}$, parametrized by the time parameter t, from a basic probability space Ω with σ-algebra $\mathscr{A}$ to $\mathbb{R}$ with the Borel σ-algebra is a closed systems view. It is quite reasonable to study stochastic processes in which the f_t's are not a classical random variables, even for elementary examples. For instance, Brownian motion is classically defined as a continuous process b on $[0,\infty)$ with (i) $b_0 = 0$, (ii) b has normally distributed increments with mean zero and variance proportional to the time elapsed between the increments, and (iii) the increments on non-overlapping time intervals are independent. Our view is that condition (i) is superfluous. This point of view implies in particular the b_t's need not a classical real random variables. A problem that is presently under investigation is to give a suitable definition of a Markov process f without assuming that the f_t's are classical random variables.

Ackowledgments

I thank Tzvetan Ivanov for many enlightening discussions regarding this paper.

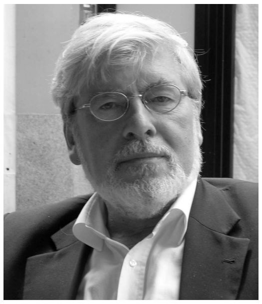
The SCD-SISTA research of the KU Leuven is supported by the projects. *Research Council KUL*: GOA/11/05 Ambiorics, GOA/10/09 MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC) en PFV/10/002 (OPTEC), IOF-SCORES4CHEM, several PhD, postdoc & fellowship grants. *Flemish Government: FWO:* PhD/postdoc grants, projects: G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine), Research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC); *IWT*: PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare. *Belgian Federal Science Policy Office*: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011), IBBT. *EU*: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC HIGHWIND (259 166). *Contract Research*: AMINAL, Helmholtz: viCERP, ACCM.

References

[1] A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin: Julius Springer, 1933.

[2] J.C. Willems, The behavioral approach to open and interconnected systems, *IEEE Control Syst. Mag.*, vol. 27, pp. 46–99, 2007.

[3] J.C. Willems, Terminals and ports, *IEEE Circuits Syst. Mag.*, vol. 10, pp. 8–16, 2010.



Jan C. Willems (S'66 - M'68 - SM'78 - F'80 - LF'05) was born in Bruges, Flanders, Belgium. He graduated from the University of Gent in 1963 in electro-mechanical engineering. He obtained an M.Sc. degree from the University of Rhode Island and the Ph.D. degree from MIT. He was on the electrical engineering faculty of MIT from 1968 to 1973, including a one-year postdoctoral stay at the Department of Applied Mathematics and Theoretical Physics of Cambridge University. From 1973 until 2003 he was professor of Systems and Control with the Mathematics Department of the University of Groningen in The Netherlands. In 2003, he became emeritus professor. He is presently Guest Professor at the Department of Electrical University of the K.U. Leuven. More details on his CV, a description of his research activities, and a list of publications can be found on his website `http://homes.esat.kuleuven.be/~jwillems`.