Constrained Probability

Jan C. Willems
KU Leuven, Department of Electrical Engineering-ESAT, SCD-SISTA
B-3001 Leuven, Belgium
Jan.Willems@esat.kuleuven.be
http://www.esat.kuleuven.be/∼jwillems

***Abstract*—A stochastic system is defined as a probability triple. Models often require a coarse event sigma-algebra. A notion that emerges in this setting is constrained probability, a concept that is reminiscent but quite distinct from conditional probability. We end with applying these ideas to a binary channel.**

I. Introduction

By an 'open' system we mean a model that explicitly incorporates the influence of the environment, but as an unmodeled feature. Open systems can be interconnected with other systems. We view interconnection of systems as 'variable sharing': before interconnection, the variables pertaining to the subsystems are regarded as independent, while after interconnection some of these variables are required to be equal.

Our interest is mainly in systems with as outcome space a (finite or) countable set, or $\mathbb{R}^n$. If the event space consist of all subsets of the countable outcome space, or of the Borel sets in the case of $\mathbb{R}^n$, then we call the σ-algebra of events 'rich' or 'fine', in contrast to 'coarse' σ-algebras. Openness of systems requires a coarse event σ-algebra, in contrast to classical stochastic systems, where the event σ-algebra is assumed to be rich. The event σ-algebra is an intrinsic, not to be ignored, feature of the stochastic phenomenon that is modeled.

This is not a paper about the interpretation of probability. Neither is it a paper about the mathematical foundations of probability. The article functions completely within the orthodox measure theoretic setting of probability, with a σ-algebra of events, the mathematical framework of probability that is usually attributed to Kolmogorov. The main point of this article is basically pedagogical in nature, namely that the 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. Concepts, as linearity, interconnection, and constrained probability, function comfortably only within the context of coarse σ-algebras.

This conference paper is a summary of a full article [1] that has recently been submitted.

II. Stochastic systems

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 of $\mathbb{W}$ called *outcomes*,
- ▶ $\mathscr{E}$ a σ-algebra of subsets of $\mathbb{W}$, with elements of $\mathscr{E}$ are called *events*,
- ▶ P a *probability* measure on $\mathscr{E}$. ■

The construction of a stochastic model involves therefore three steps. In the first step, the phenomenon is formalized mathematically by determining the outcome space $\mathbb{W}$. For the purposes of the present paper, determining the outcome space is considered to be evident. In the second step, the set of events $\mathscr{E}$ to which we are willing to assign a probability is specified. We view the specification of the events as a crucial part of probabilistic modeling, contrary, as we shall see, to the classical practice of probabilistic modeling. As the third step, we need to quantify the probability of these events. The specification of P yields the numerical probabilistic features of the model numerically.

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 a Borel probability $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n),P)$, where $\mathscr{B}(\mathbb{R}^n)$ denotes the class of Borel subsets of $\mathbb{R}^n$. P can then be specified by a probability distribution, or, if the distribution is sufficiently smooth, by the probability density function $p:\mathbb{R}^n\to[0,\infty)$ 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 assigned a probability. Thus for classical stochastic systems, the events are obtained from the structure of the outcome space. No probabilistic modeling enters in the specification of the events.

We will illustrate the relevance of specifying $\mathscr{E}$ by a binary channel in Section VII.

Example: A noisy resistor. Consider a 2-terminal 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.

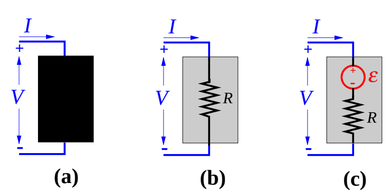

Fig. 1. **2-terminal electrical circuit**

As an example of a noisy circuit, consider a resistor with thermal noise, governed by the following relation between the current I through the resistor and the voltage V across it

$$V = RI + \varepsilon$$

with $R > 0$ and ε the voltage generated by the noisy voltage source, taken to be gaussian, zero mean, and with standard deviation $\sigma \sim \sqrt{RT}$ with T the temperature of the resistor. 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 are the sets of the form

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

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

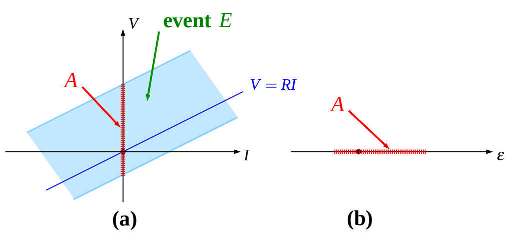

Fig. 2. **Events for the noisy resistor**

equals the probability that $\varepsilon \in A$ (see Figure 2(b)).

Hence, whereas ε is a classical random variable, $\left[\begin{smallmatrix} V \\ I \end{smallmatrix}\right]$ 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 V and I are not classical random variables. ■

III. Linearity

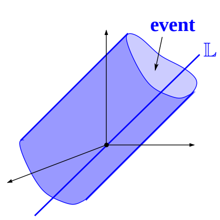

Fig. 3. **Events for a linear system**

Definition 2: Let $\mathbb{F}$ be a finite field, or $\mathbb{R}$. The stochastic system $(\mathbb{F}^{\mathtt{n}}, \mathscr{E}, P)$ is said to be *linear* if there exists a linear subspace $\mathbb{L}$ of $\mathbb{F}^{\mathtt{n}}$ such that the events consist of all subsets of the quotient space $\mathbb{F}^{\mathtt{n}}/\mathbb{L}$ when $\mathbb{F}$ if a finite field, or of the Borel subsets of $\mathbb{F}^{\mathtt{n}}/\mathbb{L}$ when $\mathbb{F} = \mathbb{R}$. $\mathbb{F}^{\mathtt{n}}/\mathbb{L}$ is a finite dimensional vector space over $\mathbb{F}$ of dimension = $\mathtt{n} - \mathtt{dimension}(\mathbb{L})$. $\mathbb{L}$ is called the *fiber* and $\mathtt{dimension}(\mathbb{L})$ *the number of degrees of freedom* of the linear stochastic system. ■

The idea behind Definition 2 is illustrated in Figure 3. The events are cylinders in $\mathbb{R}^{\mathtt{n}}$ with rays parallel to the fiber $\mathbb{L}$. A linear stochastic system is classical if and only if $\mathbb{L} = \{0\}$. Every classical random vector with $\mathbb{W} = \mathbb{F}^{\mathtt{n}}$ defines a linear stochastic system. At the other extreme, when $\mathbb{L} = \mathbb{F}^{\mathtt{n}}$, the event set $\mathscr{E}$ becomes the trivial σ-algebra $\{\emptyset, \mathbb{F}^{\mathtt{n}}\}$.

Observe that the definition of linearity involves only the event σ-algebra, but not the probability measure.

IV. Interconnection

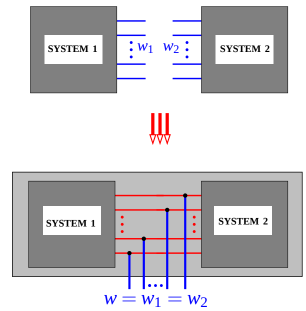

Fig. 4. **Interconnection**

In this section we discuss interconnection. We consider the situation illustrated in Figure 4 with the assumption that the two interconnected systems are stochastically independent. Note that interconnection comes down to imposing two distinct probabilistic laws on the same set of variables. *Is it possible to define one law which respects both laws?* As we shall see, this is indeed possible, provided some regularity condition, called 'complementarity', holds.

Definition 3: Two σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$ on $\mathbb{W}$ are said to be *complementary* if for all nonempty $E_1, E_1' \in \mathscr{E}_1, E_2, E_2' \in \mathscr{E}_2$

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

Two 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 $E_2, E_2' \in \mathscr{E}_2$

$$[E_1 \cap E_2 = E_1' \cap E_2'] \Rightarrow [P_1(E_1)P_2(E_2) = P_1(E_1')P_2(E_2')]. \quad ■$$

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 determines the intersecting sets uniquely.

It is readily proven that

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

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

and, furthermore, that complementarity of two stochastic systems is implied by complementarity of the associated σ-algebras. It is easy to construct examples that show that complementarity of two stochastic systems does not imply complementarity of the associated σ-algebras. The problem is that the stochastic systems may have too many zero probability events (the curse of probability theory). Complementarity of the event σ-algebras is a more primitive condition that is convenient for proving complementarity of stochastic systems.

Definition 4: Let $\Sigma_1 = (\mathbb{W}, \mathscr{E}_1, P_1)$ and $\Sigma_2 = (\mathbb{W}, \mathscr{E}_2, P_2)$ be stochastic systems and assume that they are complementary. The *interconnection* of Σ_1 and Σ_2, assumed stochastically independent, denoted by $\Sigma_1 \wedge \Sigma_2$, is defined as the system

$$\Sigma_1 \wedge \Sigma_2 := (\mathbb{W}, \mathscr{E}, P),$$
$$\mathscr{E} := \text{ the } \sigma\text{-algebra generated by } \mathscr{E}_1 \cup \mathscr{E}_2,$$

and P defined through 'rectangles' by

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

The definition of the probability P for rectangles $\{E_1 \cap E_2 \mid E_1 \in \mathscr{E}_1, E_2 \in \mathscr{E}_2\}$ uses complementarity in an essential way. $\mathscr{E}$ is 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 sets. The probability of rectangles defines the probability of the union of a finite number of disjoint rectangles. By the Hahn-Kolmogorov extension theorem, this leads to a unique probability measure P on $\mathscr{E}$.

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 $P(E_1) = P(E_1 \cap \mathbb{W}) = P_1(E_1)$ and $P(E_2) = P(\mathbb{W} \cap E_2) = P_2(E_2)$. Hence interconnection refines the event σ-algebras and the probabilities. 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 also that for $E_1 \in \mathscr{E}_1$ and $E_2 \in \mathscr{E}_2$, $P(E_1 \cap E_2) = P_1(E_1)P_2(E_2) = P(E_1)P(E_2)$. 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.

V. 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.

Consider for example 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. Stochastic systems with a coarse σ-algebra do allow to incorporate the unexplained environment.

Another way of looking at 'open' versus 'closed' systems 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 illustrate that coarseness of the σ-algebras is essential for complementarity in the case the $\mathbb{W}$ is countable. Assume that $\Sigma_1 = (\mathbb{W}, 2^{\mathbb{W}}, P)$, with $\mathbb{W}$ countable, is a classical stochastic system and that $\Sigma' = (\mathbb{W}, \mathscr{E}', P')$ is another stochastic system. Then the σ-algebras associated with Σ_1 and Σ_2 can only be complementary if $\mathscr{E}'$ is trivial, that is, $\mathscr{E}' = \{\emptyset, \mathbb{W}\}$. More generally, if the stochastic systems Σ and Σ' are complementary then for $E \in \mathscr{E}'$, we have $E \cap \mathbb{W} = E \cap E = \mathbb{W} \cap E$, 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 each singleton has positive P-measure, then $\mathscr{E}' = \{\emptyset, \mathbb{W}\}$.

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

VI. 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.

Definition 5: 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

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

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

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 $w \in \mathbb{S}$ with probability 1, that is, 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 (VI) holds. It follows that constraining is interesting when $\mathbb{S} \notin \mathscr{E}$.

Note that constraining essentially corresponds to interconnecting $(\mathbb{W}, \mathscr{E}, P)$ with the 'deterministic' system $(\mathbb{W}, \{\emptyset, \mathbb{S}, \mathbb{S}^{\text{complement}}, \mathbb{W}\}, P')$ with $P'(\mathbb{S}) = 1$. The regularity condition corresponds to complementarity.

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 $P(\mathbb{S})$.

VII. Binary channel

Open stochastic systems are often thought of as classical stochastic systems with 'input' parameters, that is, as a family of probability measures on the output space, parameterized by the 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 of systems has many drawbacks when modeling open physical systems, as argued for example in [2] for the deterministic case: a physical system is not a signal processor. 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 (interconnection, linearity, constraining) of the present article using probability kernels in their full generality lies beyond our scope. We now 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*.

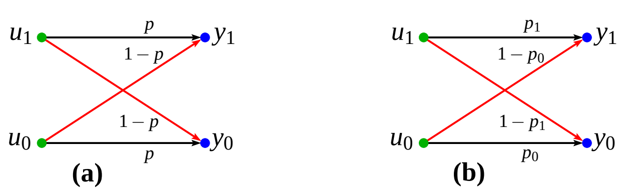

Fig. 5. **Binary channel**

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 y_1 with probability $1-p_0$, while the input u_1 leads to output y_1 with probability p_1 and to 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 5(a), while the asymmetric binary channel is shown in Figure 5(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 determined by

$$P_{u_0}(y_0) = p_0, P_{u_0}(y_1) = 1-p_0, P_{u_1}(y_0) = 1-p_1, P_{u_1}(y_1) = p_1.$$

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

A. The symmetric binary channel

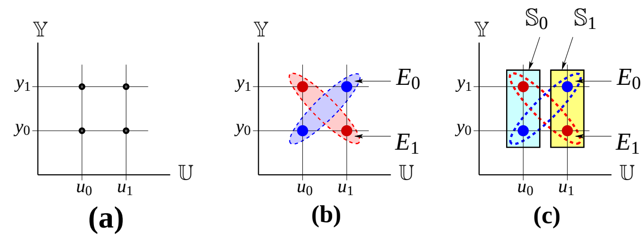

Fig. 6. **Events for the symmetric binary channel**

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

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

Thus the outcome space, shown in Figure 6(a), is $\mathbb{U} \times \mathbb{Y}$. The events $\mathscr{E}$ consist of the σ-algebra generated by

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

The generating set for $\mathscr{E}$ is shown in Figure 6(b). Note that the σ-algebra generated by $\{E_0, E_1\}$ is not equal to $2^{\mathbb{U}\times\mathbb{Y}}$. Σ_{sbc} is not a classical stochastic system. The probability P is determined by

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

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 6(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 $\Sigma_{\text{sbc}}|_{\mathbb{S}_0} = (\mathbb{Y}, 2^{\mathbb{Y}}, P|_{\mathbb{S}_0})$ with

$$P|_{\mathbb{S}_0}(y_0) = p, P|_{\mathbb{S}_0}(y_1) = 1-p,$$

and $\Sigma_{\text{sbc}}|_{\mathbb{S}_1} = (\mathbb{Y}, 2^{\mathbb{Y}}, P|_{\mathbb{S}_1})$ with

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

Observe that $\Sigma_{\text{sbc}}|_{\mathbb{S}_0}$ and $\Sigma_{\text{sbc}}|_{\mathbb{S}_1}$ yields *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 $P(\mathbb{L}) = p$ and $P((0,1)+\mathbb{L}) = 1-p$.

B. The asymmetric binary channel

We next show how to approach the asymmetric binary channel from our point of view. We start with the system

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

with $\mathbb{E} = \{e_1, e_2, e_3, e_4\}$. Thus the outcome space, shown in Figure 7(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 channel uncertainty. The events $\mathscr{E}$ consist of the σ-algebra generated by

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

The generating set for $\mathscr{E}$ is shown in Figure 7(b). Note that the σ-algebra generated by $\{E_1,E_2,E_3,E_4\}$ is not equal to $2^{\mathbb{U}\times\mathbb{Y}\times\mathbb{E}}$. Σ_{abc} is not a classical stochastic system. The probability P is determined by

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

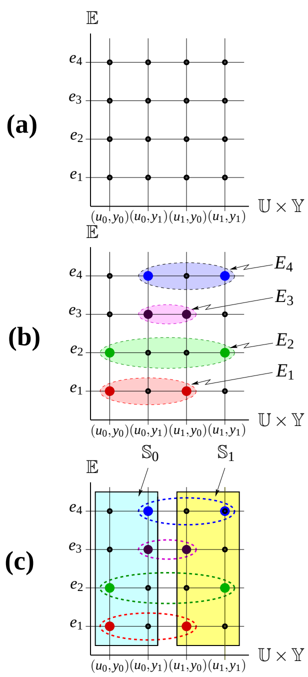

Fig. 7. **Events for the asymmetric binary channel**

Now consider the stochastic system Σ_{abc} with outcomes constrained to be in

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

respectively. The sets $\mathbb{S}_0$ and $\mathbb{S}_1$ are illustrated in Figure 7(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 $\Sigma_{\text{abc}}|_{\mathbb{S}_0} = (\mathbb{Y}\times\mathbb{E},\mathscr{E}|_{\mathbb{S}_0},P|_{\mathbb{S}_0})$ with $\mathscr{E}|_{\mathbb{S}_0}$ generated by

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

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

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

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

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

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

The introduction of $\mathbb{E}$ and Σ_{abc} shows that the specification of a channel as a probability kernel can be interpreted in a very natural way as constrained stochastic systems. The probability kernel can also be interpreted in terms of conditional probabilities by defining, for $\pi \in [0,1]$, $P_u(u_0) = \pi$ and $P_u(u_1) = 1-\pi$. We then obtain stochastic systems with Σ_{u_0} and Σ_{u_1} the conditional probabilities of y given u. Since the interpretation of the 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.

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 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 $e_1 \leftrightarrow (0,0), e_2 \leftrightarrow (0,1), e_3 \leftrightarrow (1,0), e_4 \leftrightarrow (1,1)$.

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] J.C. Willems, Open stochastic systems, submitted to the *IEEE Transactions on Automatic Control*.
[2] J.C. Willems, The behavioral approach to open and interconnected systems, *Control Systems Magazine*, volume 27, pages 46–99, 2007.
[3] J.C. Willems, Terminals and ports, *IEEE Circuits and Systems Magazine*, volume 10, pages 8–16, 2010.