# Some Remarks on Sets of Lexicographic Probabilities and Sets of Desirable Gambles

**Fabio Gagliardi Cozman**
Escola Politécnica, Universidade de São Paulo, Brazil

## Abstract

Sets of lexicographic probabilities and sets of desirable gambles share several features, despite their apparent differences. In this paper we examine properties of marginalization, conditioning and independence for sets of lexicographic probabilities and sets of desirable gambles.

## 1 Introduction

The standard theory of probabilities is widely used to represent situations that display uncertainty. In that theory, events form a field, and probabilities are real-valued, non-negative, and countably additive. There are *many* variants of this Kolmogorovian theory of probabilities [8, 10, 21, 29, 32], including proposals that abandon the real scale and focus on infinitesimal probabilities [16] or on lexicographic probabilities [3]. Other departures from probability theory attempt to represent imprecision in numeric values [33, 34]. For instance, the theory of credal sets uses sets of probability measures as its basic entities [20]. Yet another departure from probability theory is the theory of sets of desirable gambles [35]. Matters become even more involved if one allows a language with negation and conjunction of desirability statements [27].

The purpose of this paper is to examine some properties of sets of lexicographic probabilities and of sets of desirable gambles. We present these formalisms through a hopefully illuminating analysis, emphasizing their close connection (Section 2). Because both lexicographic probabilities and sets of desirable gambles represent the same sort of preference orderings, by studying one of them, we obtain insights about the other; perhaps contentious concepts and drawbacks can be clarified by such a study.

Sections 3 to 6 examine features of marginalization and conditioning. We first compare lexicographic and full conditional probabilities, and show they are not as similar as usually suggested by the literature. We then examine convexity, non-uniqueness and independence, always together with marginalization and conditioning. Several of the properties discussed here are well-known, but still they may be somewhat surprising as a whole, and call for further study concerning these formalisms.

## 2 Lexicographic Probabilities and Sets of Desirable Gambles

In this paper we only deal with finite objects, so that complications arising from infinity are entirely ignored. We assume there is a finite set of states, denoted by $\Omega$, and that every subset of $\Omega$ is an event. A *gamble* is a function from states to real numbers. If $\Omega$ contains $n$ states, a gamble can be thought of as an $n$-dimensional point. Hence we will treat sets of gambles as subsets of $\Re^n$.

The plan for this section is to emphasize the relationship between sets of lexicographic probability measures and sets of desirable gambles. Previously, Couso and Moral have studied this relationship in some restricted cases [5], and Quaeghebeur has dealt with this relationship in considerable detail [24, 25]. Most of the following discussion touches on topics that may be familiar to readers with background on imprecise probabilities.

Probabilities are often justified and derived by assuming axioms about preferences [1, 11, 30]. To simplify matters, we always take preferences over the set of all gambles. A popular way to do so is to take a preference relation $\succ$ between gambles, such that $f \succ g$ is interpreted as "$f$ is preferred to $g$". Suppose $\succ$ is a (strict) partial order, meaning that it is irreflexive and transitive [11], and that $\succ$ satisfies a monotonicity condition: if $f(\omega) > g(\omega)$ for all $\omega$, then $f \succ g$. Suppose additionally that it satisfies an "independence condition" such as: for any $\alpha \in (0, 1]$ and any $f$, $g$, $h$, we have $f \succ g$ if and only if $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.

In this case the set of gambles that are preferred to the zero gamble is a cone that completely represents $\succ$. Suppose one assumes that this cone is open, an assumption that encodes an "Archimedean condition" on $\succ$ [31]. One then obtains the following representation: there is a unique maximal convex set $\mathbb{K}$ of probability measures such that $f \succ g$ if and only if $\forall \mathbb{P} \in \mathbb{K} : \mathbb{E}_{\mathbb{P}}[f] > \mathbb{E}_{\mathbb{P}}[g]$. Such a set of probability measures is called a *credal set*. Note that a preference profile may be completely characterized by more than one credal set, but there is a unique maximal credal set that offers such a representation, and this credal set is convex.

Suppose that $\succ$ is such that absence of preference is an equivalence (reflexive, transitive, symmetric); we then say that $\succ$ is a *strict weak order* [11]. If $\succ$ is a strict weak order, the credal set $\mathbb{K}$ is a singleton, so we obtain the usual representation by a single probability measure [11].

One might consider replacing the monotonicity condition by the following one: if $f(\omega) \geq g(\omega)$ for all $\omega$ and $f(\omega) > g(\omega)$ for some $\omega$, then $f \succ g$. Following Blume et al., we refer to this property as *admissibility* [3, Definition 4.1]. Note that a standard probability measure may fail to represent admissibility (if $\mathbb{P}(\omega) = 0$, differences on this $\omega$ do not matter).

### 2.1 Lexicographic Probabilities

Now suppose $\succ$ is a strict partial order that satisfies the "independence condition" and admissibility, but no Archimedean condition. We then obtain a representation using lexicographic probabilities. A lexicographic probability measure is simply a sequence of standard probability measures $[\mathbb{P}_0, \ldots, \mathbb{P}_K]$, and the representation is of the form: $f \succ g$ if and only if there is $[\mathbb{P}_0, \ldots, \mathbb{P}_K]$ such that

$$[\mathbb{E}_{\mathbb{P}_1}[f], \ldots, \mathbb{E}_{\mathbb{P}_K}[f]] >_{\mathrm{L}} [\mathbb{E}_{\mathbb{P}_1}[g], \ldots, \mathbb{E}_{\mathbb{P}_K}[g]],$$

where $>_{\mathrm{L}}$ denotes lexicographic comparison (for $a, b \in \Re^K$, $a >_{\mathrm{L}} b$ iff $a_j > b_j$ for some $j \leq K$ and $a_i = b_i$ for $1 < i < j$).

To produce the representation of $\succ$ in terms of lexicographic probability measures, note that $\succ$ can always be extended to a total order $\succ^*$ over gambles, such that $\succ^*$ satisfies the "independence condition" [31, Theorem 1] and admissibility. Every $\succ^*$ can be represented by a lexicographic linear utility [12, Chapter 4], and this lexicographic linear utility can be expressed as the expected value of a lexicographic probability measure (using the arguments by Blume et al. [3, Theorem 3.1]). Also the set of all extensions of $\succ$, and consequently $\succ$ itself, can be represented by a unique maximal convex set of lexicographic linear utilities [31, Theorem 2]. If $\succ$ is a strict weak order, the set of lexicographic linear utilities collapses to a single lexicographic probability measure [12, Chapter 4].

| | $H$ | $T$ |
|---|---|---|
| layer 0 | $\alpha$ | $(1-\alpha)$ |
| layer 1 | $\gamma$ | $(1-\gamma)$ |

Table 1: Lexicographic probabilities; $\alpha, \gamma \in (0, 1)$.

Consider a lexicographic probability measure $[\mathbb{P}_0, \ldots, \mathbb{P}_K]$. Then $\mathbb{P}_i$ is called the $i$th *layer* of the lexicographic probability measure. One important fact is that each $\mathbb{P}_i$ is only unique up to linear combinations of $\mathbb{P}_0, \ldots, \mathbb{P}_i$ that assign positive weight to $\mathbb{P}_i$ [3, Theorem 3.1]. So in fact there is no intrinsic uniqueness in the lexicographic representation, as emphasized in the following example.

**Example 1** Consider the lexicographic probability measure in Table 1, where each row contains a layer. The question is whether a gamble $f$ such that $f(H) = a$ and $f(T) = b$ is preferred to the zero gamble. Using the first layer, $\mathbb{E}[f] = 0$ only if $a\alpha = -b(1-\alpha)$, so we might focus on the question of whether the gamble $(1-\alpha, -\alpha)$ is desirable or not. As the next layer gives value $\gamma - \alpha$ to this gamble, we only need to determine whether $\gamma > \alpha$ or $\gamma < \alpha$ to fix all preferences (if $\gamma = \alpha$, the second layer can be discarded). □

Admissibility requires each event to have positive probability with respect to at least one layer. This follows from the fact that any indicator function is nonnegative and positive for at least one $\omega$; hence any indicator function if preferred to zero, and consequently for any event there is a probability measure that assigns it positive probability.

### 2.2 Sets of Desirable Gambles

For all preference orderings already discussed, axioms about preferences guarantee that $f \succ g$ if and only if $f - g \succ 0$. As noted already, we can then capture the preference relation by the set of gambles that are preferred to the zero gamble. This latter set is called the *set of desirable gambles* generated by the preference relation. But we can also start with sets of desirable gambles, properly axiomatized, and obtain preferences from them. For instance, here is a set of axioms that has been proposed for sets of desirable gambles [35]: a set of desirable gambles $\mathbb{D}$ is *coherent* if the zero gamble is not in $\mathbb{D}$; if all $f$ such that $f \geq 0$ and $f \neq 0$ are in $\mathbb{D}$; if for any $\lambda > 0$ and any $f \in \mathbb{D}$, we have $\lambda f \in \mathbb{D}$; and if for any $f, g \in \mathbb{D}$, we have $f + g \in \mathbb{D}$. To obtain preferences from a given set of desirable gambles, just say that $f \succ g$ if and only if $f - g \in \mathbb{D}$. By doing so, one notes that irreflexivity follows from the condition that the zero gamble is not in $\mathbb{D}$. Note also that admissibility follows from the second condition: if $f(\omega) \geq g(\omega)$ for all $\omega$ and $f(\omega) > g(\omega)$ for some $\omega$, then $f \succ g$. Finally, transitivity and the independence condition follow from the other axioms. Hence a coherent set of desirable gambles can be completely represented by a (unique maximal convex) set of lexicographic linear utilities. From now on, every set of desirable gambles is assumed coherent, so we drop the qualifier "coherent" whenever possible.

### 2.3 Marginalization and Conditioning

Now consider a pair of random variables $X$ and $Y$ defined over $\Omega$.

Marginalization of lexicographic probability measures is usually understood in a layer-wise fashion [14]. That is, if $[\mathbb{P}_0, \ldots, \mathbb{P}_K]$ are the layers of a lexicographic probability measure, then the marginal for $X$ is a lexicographic probability measure where each layer is a probability measure over $\Omega_X$ with value (for $X$ at $x$) $\sum_{\omega : X(\omega) = x} \mathbb{P}_i(\omega)$.

Given a set of desirable gambles $\mathbb{D}$, the *marginal* set of desirable gambles for $X$, denoted by $\mathbb{D}(X)$, is simply the set of all desirable gambles in $\mathbb{D}$ that are functions of $X$. For instance, if $\Omega$ is the Cartesian product of the set of values of $X$ and the set of values of $Y$, respectively $\Omega_X$ and $\Omega_Y$, then the $Y$-marginal $\mathbb{D}(Y)$ is $\{g : g \text{ is a function of } Y \text{ and } g \in \mathbb{D}\}$, with the understanding that $g \in \mathbb{D}$ means that the cylindrical extension of $g$ belongs to $\mathbb{D}$ [26].

It should be apparent that marginalization means the same thing both for sets of desirable gambles and sets of lexicographic probability measures, given appropriate interpretation. If one starts with a set of desirable gambles, generates a set of lexicographic probability measures, marginalizes the latter, and generates the corresponding set of desirable gambles, one reaches the marginal of the original set of desirable gambles.

Conditioning of lexicographic probability measures has also received a layer-wise definition by Blume et al. [3]. That is, if we again have the lexicographic probability measure $[\mathbb{P}_0, \ldots, \mathbb{P}_K]$, then conditioning on $A$ yields $[\mathbb{P}_0(\cdot|A), \ldots, \mathbb{P}_{K'}(\cdot|A)]$, where each layer that assigns positive probability to $A$ is processed through Bayes rule, and all other layers are discarded. This sort of layer-wise Bayes rule is actually derived from preferences, as follows. From a preference relation $\succ$, obtain conditional preference given $A$, denoted by $\succ_A$, by saying that $f \succ_A g$ if and only if $Af \succ Ag$ [3, Definition 2.1]. Then $\succ_A$ is represented by a conditional lexicographic probability measure as just defined [3, Theorem 4.3].[1]

Given a set of desirable gambles $\mathbb{D}$ and an event $A$, the conditional set of desirable gambles $\mathbb{D}|A$ is simply $\{f : I_A f \in \mathbb{D}\}$, where $I_A$ denotes the indicator function of $A$ [26]. In fact, by using de Finetti's convention where an event and its indicator function are denoted by the same symbol, we have [35]:

$$\mathbb{D}|A = \{f : Af \in \mathbb{D}\}.$$

But this is clearly equivalent to representing the preferences $Af \succ Ag$. That is, both conditional lexicographic probability measures and conditional sets of desirable gambles represent the same operation.

In short, sets of (admissible) lexicographic probability measures and (coherent) sets of desirable gambles are equivalent representations for preferences under uncertainty.

## 3 Full Conditional Probabilities: Not Really

One of the attractive features of lexicographic probabilities and sets of desirable gambles is the fact that conditioning is well defined for any nonempty conditioning event (because every event has positive probability in some layer). Thus it is not surprising that lexicographic probability measures have been connected with the theory of full conditional probabilities [8, 19], because the latter also offers conditioning on every nonempty event.

In fact, there are some recurring themes in the connection between lexicographic and full conditional probabilities [3, 15, 16]. On the one hand, the structure of full conditional probabilities can be understood through lexicographic probabilities, and full conditional probabilities can be justified using the axioms of lexicographic probabilities. On the other hand, full conditional probabilities can be treated as if they were a class of lexicographic probabilities that are easy to specify, interpret, and handle. We now examine to which extent these intuitions are valid.

### 3.1 A Brief Review

To recap, a full conditional probability $\mathbb{P} : \mathcal{B} \times (\mathcal{B} \backslash \emptyset) \to \Re$, where $\mathcal{B}$ is a Boolean algebra, is a two-place set-function such that for every event $H \neq \emptyset$ [9]:
(1) $\mathbb{P}(H|H) = 1$;
(2) $\mathbb{P}(G|H) \geq 0$ for all $G$;
(3) $\mathbb{P}(G_1 \cup G_2|H) = \mathbb{P}(G_1|H) + \mathbb{P}(G_2|H)$ whenever

[1] Note that Blume et al. actually assumes that preference relations are reflexive, but their analysis of conditional probability is not affected by that.

$G_1 \cap G_2 = \emptyset$;
(4) $\mathbb{P}(G_1 \cap G_2|H) = \mathbb{P}(G_1|G_2 \cap H) \times \mathbb{P}(G_2|H)$ whenever $G_2 \cap H \neq \emptyset$.
Whenever the conditioning event $H$ is equal to $\Omega$, we suppress it and write the "unconditional" probability $\mathbb{P}(G)$.

The theory of coherent probabilities advocated by de Finetti adopts full conditional probabilities, and offers a justification for them that is based on betting (in fact de Finetti's original arguments were later formalized by Holzer [17] and Regazzini [28]). It should be noted that similar (but more general) axioms have been proposed by Renyi [29] and Popper [23]; there are also variants of those theories that we do not discuss for the sake of space.

**Example 2** Take a coin with heads ($H$), tails ($T$), a sharp edge ($S$), and a blunt edge ($B$). We can have $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$, hence $\mathbb{P}(S) = \mathbb{P}(B) = 0$, but still $\mathbb{P}(B|S \cup B) = 2/3$. □

A full conditional probability can always be represented as a sequence of standard probability measures $\mathbb{P}_0, \ldots, \mathbb{P}_K$ [3, 4, 16, 19]. To obtain this representation, we must partition $\Omega$ into several events $L_0, \ldots, L_K$. Take $L_0$ to be the set of elements of $\Omega$ that have positive unconditional probability. Then take $L_1$ to be the set of elements of $\Omega$ that have positive probability conditional on $\Omega \backslash L_0$. And then take $L_i$, for $i \in \{2, \ldots, K\}$, to be the set of elements of $\Omega$ that have positive probability conditional on $\Omega \backslash \cup_{j=0}^{i-1} L_j$. The event $L_i$ denotes the support of the *layer* $\mathbb{P}_i$ of the full conditional probability. The *layer number* of layer $\mathbb{P}_i$ is $i$. For nonempty $G$, denote by $L_G$ the support of the first layer such that $\mathbb{P}(G|L_G) > 0$. We then have $\mathbb{P}(G|H) = \mathbb{P}(G|H \cap L_H)$ [2, Lemma 2.1a]. Note that some authors use a different terminology, using instead the sequence $\cup_{j=i}^{K} L_j$ rather than $L_i$ [4, 19].

**Example 3** In Example 2, we have two layers. The first consists of $H$ and $T$, with associated probabilities $\mathbb{P}_0(H) = \mathbb{P}_0(T) = 1/2$. The second layer consists of $S$ and $B$, with associated probabilities $2\mathbb{P}_1(S) = \mathbb{P}_1(B) = 2/3$. □

### 3.2 Admissibility and Marginalization

Given the results enumerated in the previous section, it is natural to think that full conditional probabilities are just instances of lexicographic probability measures. However, strictly speaking, full conditional probabilities cannot pose as admissible lexicographic probabilities, as the theory of full conditional probabilities does not satisfy admissibility. For instance, consider the gamble $f$ such that $f(H) = f(T) = f(S) = 0$, $f(B) = 1$ in Example 2. Computing expected value in the usual way with respect to this full conditional probability, we obtain zero; as far as preferences are to be extracted from expected values, this gamble is indistinguishable from the zero gamble. But if we were to interpret the layers of the full conditional probability as the layers of a lexicographic probability measure, then $f \succ 0$.

To obtain admissibility in actual decisions, one might use lexicographic expected values with respect to layers of a full conditional probability whenever necessary. For instance, in the previous paragraph one might say that $f \succ 0$ by looking at all layers of the full conditional probability. However, matters become even more delicate when we look at marginalization.

**Example 4** Consider two variables $X$ and $Y$, each with values $\{0, 1, 2\}$. Take a full conditional probability over $(X, Y)$ with two layers (layer numbers are indicated by subscripts):

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---|---|---|---|
| $X = 0$ | $(1/5)_0$ | $(1/5)_0$ | $(1/5)_0$ |
| $X = 1$ | $(1/5)_0$ | $(1/4)_1$ | $(1/4)_1$ |
| $X = 2$ | $(1/5)_0$ | $(1/4)_1$ | $(1/4)_1$ |

We have marginal probabilities for $X$: $\mathbb{P}(X = 0) = 3\mathbb{P}(X = 1) = 3\mathbb{P}(X = 2) = 3/5$. These marginal probabilities characterize a full conditional probability with a single layer. For this marginal full conditional probability, the gamble $f(X)$ such that $f(0) = -1$, $f(1) = 1$, $f(2) = 2$ has expected value equal to zero.

So, with respect to the marginal full conditional probabilities for $X$, there is not much to say about $f$; it is just indistinguishable from the zero gamble. There are no deeper layers to look at because the marginal full conditional probability does not assign zero probability to any event. So, there is no way to produce a lexicographic comparison if we first produce the full conditional probability that is the marginal of $X$.

However, as $f$ can be obviously viewed as a function of $X$ and $Y$, its expected value can be computed with respect to the joint full conditional probability. But now we see that we can have a lexicographic comparison: the expected value of $f$ with respect to the second layer is $3/2$, hence $f \succ 0$.

That is, marginalization of full conditional probabilities loses information concerning layers, information that is needed if we were to compute lexicographic expected values. If we were to treat full conditional probabilities as lexicographic probabilities, we would need to have marginal full conditional probabilities that carry some extra information.

Indeed, if we took the joint full conditional probability in our example as a lexicographic probability measure to begin with, and then marginalized it, we would obtain the following marginal lexicographic probabilities:

| $X = 0$ | $X = 1$ | $X = 2$ |
|---|---|---|
| $(3/5)_0$ | $(1/5)_0$, $(1/2)_1$ | $(1/5)_0$, $(1/2)_1$ |

With respect to this marginal lexicographic probability measure, we obtain $f \succ 0$, as we must. $\square$

So, if we wish to preserve admissibility by using lexicographic expectation with respect to full conditional probabilities, then the marginal of a full conditional probability must actually be represented as a lexicographic probability measure. The message is that lexicographic probabilities (and sets of desirable gambles) do offer conditioning on events of probability zero, but their solution is different from the one offered by full conditional probabilities. Lexicographic probabilities may be a representation for full conditional probabilities, but both behave differently.

## 4 Convexity?

When we adopt sets of lexicographic probability measures (or equivalently sets of desirable gambles), we seem to have convexity at hand. First, a set of desirable gambles is a convex object. Second, a strict ordering with independence and admissibility can be represented uniquely by a maximal convex set of lexicographic linear utilities.

However, convexity deserves further scrutiny. Again, it is useful to start this discussion with full conditional probabilities. Typically one assumes that, conditional on an event $A$, the set of probability measures $\mathbb{K}(\cdot|A)$ is convex [34]. But a set of full conditional probabilities cannot always be convex [13, 22], even if all probabilities are positive:

**Example 5** Suppose $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathbb{P}_1(\omega_1) = \mathbb{P}_1(\omega_2) = \mathbb{P}_1(\omega_3) = 1/3$ and $\mathbb{P}_2(\omega_1) = 2\mathbb{P}_2(\omega_2) = 2\mathbb{P}_2(\omega_3) = 1/2$. Build the convex combination $\mathbb{P}_\alpha = \alpha\mathbb{P}_1 + (1-\alpha)\mathbb{P}_2$. There is no $\alpha \in (0,1)$ such that $\mathbb{P}_\alpha(\omega_1|\omega_1 \cup \omega_3) = 2(\alpha-3)/(\alpha-9)$ is equal to $\alpha\mathbb{P}_1(\omega_1|\omega_1 \cup \omega_3) + (1-\alpha)\mathbb{P}_2(\omega_1|\omega_1 \cup \omega_3) = \alpha/2 + 2(1-\alpha)/3$. That is, $\mathbb{P}_\alpha$ cannot be a convex combination of the functions $\mathbb{P}_1$ and $\mathbb{P}_2$. $\square$

Consider a preference $\succ$ that can be extended to at least two orders, the former encoded by lexicographic linear utility $u_1$ and the latter by lexicographic linear utility $u_2$. On the one hand, any convex combination of these lexicographic linear utilities generates the same preference profile [31]. On the other hand, admissibility allows us to normalize each utility in the lexicographies, so as to obtain lexicographic probability measures [3]. However, suppose we wish to both normalize *and* do convex combinations. Apparently, matters are simple:

**Example 6** Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\alpha, \beta \in (0,1)$, and lexicographic probability measures $\mathbb{LP}_1$ and $\mathbb{LP}_2$:

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_1(\omega_i)$ | $(\alpha)_0$ | $(1-\alpha)_0$ | $1_1$ |

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_2(\omega_i)$ | $(1)_0$ | $(\beta)_1$ | $(1-\beta)_1$ |

Their half-half convex combination is:

| $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|
| $((1+\alpha)/2)_0$ | $((1-\alpha)/2)_0$, $(\beta/2)_1$ | $(1-\beta/2)_1$ |

As a digression: $\mathbb{LP}_1$ and $\mathbb{LP}_2$ have disjoint layers, so they *could* be representations for full conditional probabilities. But their convex combination is certainly not the representation of a full conditional probability as the supports of the layers are not disjoint. $\square$

The convex combination of lexicographic probability measures works perfectly if all lexicographic probability measures involved in the convex combination have the same number of layers. But suppose that modeling decisions have built two preference orderings with distinct depths; what to do?

**Example 7** Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\alpha, \beta, \gamma \in (0,1)$, all distinct, and lexicographic probability measures:

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_1(\omega_i)$ | $(\alpha)_0$, $(\gamma)_2$ | $(1-\alpha)_0$, $(1-\gamma)_2$ | $1_1$ |

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_2(\omega_i)$ | $(1)_0$ | $(\beta)_1$ | $(1-\beta)_1$ |

Note that $\mathbb{LP}_1$ reproduces Example 1, with one additional intervening layer. In fact $\mathbb{LP}_1$ defines a total order over gambles. And $\mathbb{LP}_2$ appeared in the previous example; for $\mathbb{LP}_2$ there are gambles that get zero expectation with respect to all layers (for instance, $h(\omega_1) = 0$, $f(\omega_2) = 1-\beta$, $f(\omega_3) = -\beta$).

Consider $\mathbb{LP}_{1/2}$, a half-half combination of $\mathbb{LP}_1$ and $\mathbb{LP}_2$. If we operate layer-wise,

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_{1/2}(\omega_i)$ | $(1+\alpha/2)_0$,<br><br>$(\gamma/2)_2$ | $((1-\alpha)/2)_0$,<br>$(\beta/2)_1$<br>$((1-\gamma)/2)_2$ | $(1-\beta/2)_1$ |

This is not a very satisfying result as probabilities in the last layer do not add up to one. □

One possible way to avoid the difficulties in this last example is always represent $\succ$ as a collection of total orders, all of which have the same depth. Indeed, the sets of lexicographic utilities by Seidenfeld et al. [31] are explicitly built from all such total orders, hence this sort of modeling decision makes sense conceptually.

However, there is a significant inconvenience. Suppose we wish to represent a set of preference orderings, some of which do display absence of preference. For instance, the ordering generated by $\mathbb{LP}_2$ does display absence of preferences (there are gambles that are not preferred nor dispreferred to the zero gamble). To represent such an ordering using total orders, we may need to introduce (possibly many) layers and measures that are apparently useless. To understand this, consider again $\mathbb{LP}_2$: to represent the strict weak order generated by $\mathbb{LP}_2$ using total orders, we might use a set consisting of two lexicographic probability measures:

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_3(\omega_i)$ | $(1)_0$ | <br>$(\beta)_1$<br>$(\delta_1)_2$ | <br>$(1-\beta)_1$<br>$(1-\delta_1)_2$ |

| | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| $\mathbb{LP}_4(\omega_i)$ | $(1)_0$ | <br>$(\beta)_1$<br>$(\delta_2)_2$ | <br>$(1-\beta)_1$<br>$(1-\delta_2)_2$ |

It is particularly annoying that there is great latitude in selecting the probability values: as long as $(\delta_1-\beta)(\delta_2-\beta)<0$, we have the desired strict weak order collectively represented by $\mathbb{LP}_3$ and $\mathbb{LP}_4$ (and their convex combinations, if desired). The lack of control over the representation, given the ordering, is apparent.

One might look for alternative approaches. For instance, we might define the convex combination of two lexicographic probability measures so that a final normalization step is applied to each layer. Another possibility: adopt lexicographic "probability" measures that are not normalized below the first layer, and allow convex combinations without further concern. Whatever the solution, it seems that convexity deserves further analysis when applied to sets of lexicographic probability measures.

To a great extent, this discussion does not affect the theory of sets of desirable gambles. However, in practice one may be interested in representations for sets of desirable gambles that are based on probability values. When such representations are needed, the challenges in mixing sets of lexicographic probabilities and convexity are bound to surface.

## 5 Non-Uniqueness and Weakness

Some of the discussion in Section 3 concentrated on the fact that, given joint probabilities, marginals may not carry all necessary information. Now consider the reverse situation; that is, we have marginal and conditional lexicographic probabilities, and we wish to construct a joint lexicographic probability measure out of them. We find this not to be an easy problem. In fact, matters are difficult already for full conditional probabilities [6], as the next example shows. (Again, we resort to subscripts to indicate layer numbers.)

**Example 8** Consider two binary variables $X$ and $Y$. Suppose $\mathbb{P}(X=0)=1$ and $\mathbb{P}(Y=0|X=0)=\mathbb{P}(Y=0|X=1)=1$ (that is, the conditional probability of $Y$ given $X$ is actually not affected by $X$). The following joint full conditional probabilities:

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $1_0$ | $1_1$ |
| $X=1$ | $1_2$ | $1_3$ |

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $1_0$ | $1_2$ |
| $X=1$ | $1_1$ | $1_3$ |

satisfy all marginal and conditional assessments. □

The fact that marginal and conditional assessments cannot always uniquely characterize a joint full conditional probability has been noted before [6, 18]. In fact, one should take this phenomenon to suggest that as long as statistical modeling employs full conditional probabilities, one should not abide by any axiom that enforces uniqueness of probability values.

Lexicographic probabilities suffer from the same lack of uniqueness, only they suffer more deeply down their layers. Consider the following example.

**Example 9** Suppose we have two variables $X$ and $Y$, each with values $\{0,1,2\}$. Consider the following marginal assessments

| $X=0$ | $X=1$ | $X=2$ |
|---|---|---|
| $(1/2)_0$ | $(1/2)_0$, $(1/2)_1$ | $(1/2)_1$ |

and the following conditional assessments (for $Y$ given $X$)

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---|---|---|---|
| $X = 0$ | $(1/2)_0$ | $(1/2)_0$, $(1/2)_1$ | $(1/2)_1$ |
| $X = 1$ | $(1/2)_1$ | $(1/2)_0$, $(1/2)_1$ | $(1/2)_0$ |
| $X = 2$ | $(1/2)_0$ | $1_1$ | $(1/2)_0$ |

There are *many* possible joint lexicographic probability measures that are compatible with these assessments. One possibility:

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---|---|---|---|
| $X = 0$ | $(1/4)_0$ | $(1/4)_0$, $(1/4)_1$ | $(1/4)_1$ |
| $X = 1$ | $(1/4)_1$, $(1/4)_3$ | $(1/4)_0$, $(1/4)_1$, $(1/4)_2$, $(1/4)_3$ | $(1/4)_0$, $(1/4)_2$ |
| $X = 2$ | $(1/4)_2$ | $(1/2)_3$ | $(1/4)_2$ |

Another possible joint lexicographic probability measure is obtained, for instance, by exchanging the second and third layers of this latter lexicographic probability measure. But we can be more creative still, by adding layers in various ways; for instance, consider the following joint lexicographic probability measure, with *eight* layers, that satisfies all assessments. Here we use the notation $(\alpha)_{i:j}$ to indicate that value $\alpha$ appears in all layers between layer $i$ (inclusive) and layer $j$ (inclusive).

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---|---|---|---|
| $X = 0$ | $(1/4)_{0:1}$ | $(1/4)_{0:3}$ | $(1/4)_{2:3}$ |
| $X = 1$ | $(1/4)_1$, $(1/4)_3$ | $(1/4)_{0:7}$ | $(1/4)_0$, $(1/4)_2$, $(1/4)_{4:7}$ |
| $X = 2$ | $(1/4)_4$, $(1/4)_6$ | $(1/2)_5$, $(1/2)_7$ | $(1/4)_4$, $(1/4)_6$ |

We can produce many more joint lexicographic probabilities by combining marginal and conditional layers in various ways. □

To emphasize how information is lost through marginalization, consider one more example.

**Example 10** Consider the following joint lexicographic probability measure.

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---|---|---|---|
| $X = 0$ | $1_0$ | $(1/3)_1$ | $(3/8)_2$ |
| $X = 1$ | $(1/6)_1$ | $(1/6)_1$ | $(1/8)_2$ |
| $X = 2$ | $(1/6)_1$ | $(1/2)_2$ | $(1/2)_3$ |
| $X = 3$ | $(1/6)_1$ | $(1/2)_3$ | $1_4$ |

To obtain the marginal lexicographic probability measure for $Y$, marginalize for each layer. We get $[1, 0, 0]$ for the first layer, $[1/2, 1/2, 0]$ for the second, $[0, 1/2, 1/2]$ for the third, $[0, 1/2, 1/2]$ for the fourth, and $[0, 0, 1]$ for the fifth. Note that the third and fourth layers collapse in the marginal; hence the "relative depth" of the fifth layer is lost.

One can interpret these facts as indicating that, once lexicographic probabilities are adopted, uniqueness of joint probabilities should be abandoned. So, one should be prepared to use sets of lexicographic probabilities (and the corresponding sets of desirable gambles) from the outset. This is a nice thought for anyone interested in imprecise and indeterminate probabilities; however, one can also interpret these examples as suggesting that marginalization and conditioning are quite weak when applied to lexicographic probabilities (and sets of desirable gambles). Consider this. If we start with a joint lexicographic probability measure, then its marginal and conditional probabilities contain some useful information, but not all the information needed to rebuild the joint. Specifically, we do not have information concerning which layers of marginal and conditional probabilities should be combined together to produce the joint. Similarly, if we start with marginal and conditional lexicographic probabilities, we do not have all the information to build a single joint. Should we really have all this indeterminacy?

## 6 Independence

In this section we briefly comment on the concept of independence in the context of lexicographic probabilities. To do so, first we must agree on what "independence" means here.

One might try to define independence by requiring the joint to be a product of the marginals. But a little reflection suggests this not to be easy: because a lexicographic probability does not fundamentally change if we transform linearly its layers, one can destroy an "independence" just by rewriting its terms through linear transformations. It seems wiser to define independence as a property of the preference orderings that are implied by conditional and marginal probabilities. This sort of definition is proposed by Blume et al. [3]. They use conditional preferences, denoted by $\succ_A$ (Section 2), as follows. Variables $X$ and $Y$ are independent when we have, first, $[f_1(X) \succ_{\{Y=y_1\}} f_2(X)] \Leftrightarrow [f_1(X) \succ_{\{Y=y_2\}} f_2(X)]$ for any $f_1$, $f_2$, $y_1$, $y_2$, and second, the same condition with $X$ and $Y$ exchanged.[2] A stronger condition is [7]: $X$ and $Y$ are independent when $[f_1(X) \succ_{\{Y=y\}} f_2(X)] \Leftrightarrow [f_1(X) \succ f_2(X)]$ for any $f_1$, $f_2$, $y$, and second, the same condition with $X$ and $Y$ exchanged. These concepts of independence fail

[2] The fact that $X$ and $Y$ are independent does not guarantee any factorization of lexicographic probabilities. Blume et al. show that even hyperreal representations of lexicographic probabilities fail to factorize under their definition [3].

| | $W=0,Y=0$ | $W=1,Y=0$ | $W=0,Y=1$ | $W=1,Y=1$ |
|---|---|---|---|---|
| $X=0$ | $(1/2)_0$ | $(1/2)_0$ | $(1)_2$ | $(1)_3$ |
| $X=1$ | $(1/2)_1$ | $(1/2)_1$ | $(1/2)_4$ | $(1/2)_4$ |

Table 2: Lexicographic probabilities in Example 11.

the Decomposition property [7]; that is, we may find that $X$ and $(W,Y)$ are independent but still $X$ and $W$ are *not* independent. An even stronger concept of independence has been proposed [7]: $X$ and $Y$ are independent when $[f_1(X) \succ_B f_2(X)] \Leftrightarrow [f_1(X) \succ f_2(X)]$ for any $f_1$, $f_2$, and any set $B$ of values of $Y$, and second, the same condition with $X$ and $Y$ exchanged. But this fails the Contraction property [7]: we may have $X$ and $Y$ independent, and $W$ and $X$ independent given any value of $Y$, and yet $X$ and $(W,Y)$ fail to be independent. Failure of these properties reveal weaknesses of existing concepts and deserve further debate. Moreover, such concepts of independence do not guarantee a unique joint lexicographic measure for given marginals (consider again Example 8; $X$ and $Y$ are independent and there is no uniqueness). However, the purpose of this section is not to insist on these facts, but rather to examine a point that seems particularly hard to handle.

**Example 11** Suppose we have three binary variables, $W$, $X$ and $Y$, and joint lexicographic probabilities in Table 2. If we look at the marginal probabilities for $(X,Y)$, we see that $X$ and $Y$ are independent according to all definitions above. Indeed, the preferences on $(X,Y)$ are represented by:

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $(1)_0$ | $(1)_2$ |
| $X=1$ | $(1)_1$ | $(1)_3$ |

However, there is something intuitively strange about this independence. If we observe $\{Y=0\}$, the difference between $\{X=0\}$ and $\{X=1\}$ is a single "jump" between layers. We might interpret that $\{X=1\}$ is infinitesimally smaller than $\{X=0\}$. But given $\{Y=1\}$, the jump between them is twice as big as we go down two layers of the joint distribution. The interpretation should be that, given $\{Y=1\}$, $\{X=1\}$ is infinitesimally smaller than some event that is infinitesimally smaller than $\{X=0\}$. In a sense, one feels that the marginal for $(X,Y)$ should be

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $(1)_0$ | $(1)_2$ |
| $X=1$ | $(1)_1$ | $(1)_4$ |

Now if all we had were these marginal lexicographic probabilities, it would be difficult to argue that $X$ and $Y$ should be considered independent, because there are different jumps given distinct conditioning events. But lexicographic probabilities do not let us keep the jumps between layers intact. In fact there seems to be no way to extract such differences in relative depth of layers by looking at preferences that only involve $X$ and $Y$; by the same token, there seems to be no way to extract such differences from the corresponding set of desirable gambles. $\square$

## 7 Discussion

This paper discussed properties of sets of lexicographic probability measures and sets of desirable gambles. Most of the discussion actually dealt with lexicographic probabilities and sets of them. However, any conclusions we reach for these objects should be easily transferred to the equivalent language of sets of desirable gambles. Even though sets of desirable gambles avoid some of the non-uniqueness inherent to lexicographic probabilities, most examples in this paper could also be expressed through sets of desirable gambles. Moreover, even if one wishes to focus on sets of desirable gambles, at some point their natural representation as lexicographic probabilities must be considered, and then the features of lexicographic probabilities must be properly understood.

In many ways, sets of desirable gambles offer an attractive formalism to handle uncertainty. We basically have to deal with cones of gambles; these are linear structures with clear geometric appeal. But this simplicity may be illusory; even though the geometry is simple, matters get complicated when we wish to represent in detail operations such as marginalization and conditioning. By playing with sets of desirable gambles and sets of lexicographic probabilities, we can better understand both operations.

To summarize, we have started by emphasizing the link between lexicographic probabilities and sets of desirable gambles. We have then examined the connection between lexicographic probabilities and full conditional probabilities; this connection seems to be weaker than sometimes assumed in the literature. We have emphasized the fact that modeling with full conditional probabilities and lexicographic probability measures leads one to deal with non-uniqueness of probability values. The move to non-uniqueness led us to consider differences between full conditional probabilities and lexicographic probabilities concerning convexity. And we have examined some challenges in interpreting independence for lexicographic probabilities (and consequently for sets of desirable gambles).

## Acknowledgements

The author is partially supported by CNPq (grant 305395/2010-6).

## References

[1] Francis J. Anscombe and Robert J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

[2] Pierpaolo Battigalli and Pietro Veronesi. A note on stochastic independence without Savage-null events. *Journal of Economic Theory*, 70(1):235–248, 1996.

[3] Lawrence Blume, Adam Brandenburger, and Eddie Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 58(1):61–79, January 1991.

[4] Giulianella Coletti and Romano Scozzafava. *Probabilistic Logic in a Coherent Setting*. Trends in logic, 15. Kluwer, Dordrecht, 2002.

[5] Inés Couso and Serafín Moral. Sets of desirable gambles: Conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52:1034–1055, 2011.

[6] Fabio G. Cozman. Independence for full conditional probabilities: Structure, factorization, non-uniqueness, and Bayesian networks. *International Journal of Approximate Reasoning*, 54:1261–1278, 2013.

[7] Fabio G. Cozman. Independence for sets of full conditional measures, sets of lexicographic probabilities, and sets of desirable gambles. In *International Symposium on Imprecise Probability: Theories and Applications*, pages 87–98, 2013.

[8] Bruno de Finetti. *Theory of Probability, vol. 1-2*. Wiley, New York, 1974.

[9] Lester E. Dubins. Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics*, 3(1):89–99, 1975.

[10] Terrence L. Fine. *Theories of Probability*. Academic Press, 1973.

[11] Peter C. Fishburn. *Utility Theory for Decision Making*. John Wiley and Sons, Inc., New York, 1970.

[12] Peter C. Fishburn. *The Foundations of Expected Utility*. D. Reidel Publishing Company, 1982.

[13] Angelo Gilio and Salvatore Ingrassia. Totally coherent set-valued probability assessments. *Kybernetica*, 34(1):3–15, 1998.

[14] Srihari Govindan and Tilman Klumpp. Perfect equilibrium and lexicographic beliefs. *International Journal of Game Theory*, 31:229–243, 2002.

[15] Joseph Y. Halpern. Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior*, 68:155–179, 2010.

[16] Peter J. Hammond. Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher; Volume 1*, pages 25–59. Kluwer, Dordrecht, The Netherlands, 1994.

[17] S. Holzer. On coherence and conditional prevision. *Bolletino Un. Mat. Ital. Serie VI, Analisi Funzionale e Applicazioni*, IV-C(1):441–460, 1985.

[18] Elon Kohlberg and Philip J. Reny. Independence on relative probability spaces and consistent assessments in game trees. *Journal of Economic Theory*, 75:280–313, 1997.

[19] Peter Krauss. Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae*, 19(3-4):229–241, 1968.

[20] Isaac Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

[21] Charles F. Manski. Learning and decision making when subjective probabilities have subjective domains. *Annals of Statistics*, 9(1):59–65, 1981.

[22] F. Matus. Conditional probabilities and permutahedron. *Annales de l'Institut H. Poincaré, Probabilités et Statistiques*, 39:687–701, 2003.

[23] Karl R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1975.

[24] Erik Quaeghebeur. The CONEstrip algorithm. In *Advances in Intelligent Systems and Computing*, volume 190, pages 45–54, 2013.

[25] Erik Quaeghebeur. A propositional CONEstrip algorithm. In *IPMU*, 2014.

[26] Erik Quaeghebeur. Desirability. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 1–27. Wiley, 2014.

[27] Erik Quaeghebeur, Gert de Cooman, and Filip Hermans. Accept & reject statement-based uncertainty models. *International Journal of Approximate Reasoning*, 57:69–102, 2015.

[28] Eugenio Regazzini. Finitely additive conditional probability. *Rend. Sem. Mat. Fis.*, 55:69–89, 1985.

[29] Alfred Renyi. On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungarian*, 6:285–335, 1955.

[30] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, Inc, New York, 1972.

[31] Teddy Seidenfeld, Mark J. Schervish, and Joseph B. Kadane. Decisions without ordering. In W. Sieg, editor, *Acting and Reflecting*, pages 143–170. Kluwer Academic Publishers, 1990.

[32] Teddy Seidenfeld. Remarks on the theory of conditional probability: some issues of finite versus countable additivity. In *Statistics – Philosophy, Recent History, and Relations to Science*, pages 167–177. Kluwer Academic, 2001.

[33] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[34] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[35] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.