An Epistemic Introduction to Sigma-Algebras and Conditional Probability

Andrew Clausen <clausen@econ.upenn.edu>

December 27, 2005

The use of σ-algebras in probability theory and measure theory is usually motivated in standard texts like Dudley [2002] by technical requirements, such as avoiding "division by zero" and various paradoxes related to combining an infinite collection of "almost empty" sets into a large set. However, an equally important role of σ-algebras is to partition spaces up, which is very important for conditional expectation and conditional probability. This role has an intuitive epistemic interpretation due to Aumann [1976] [1]. To exploit the intuition, this introduction to σ-algebras and conditional probability emphasizes the partitioning properties of σ-algebras, and their epistemic interpretations, rather than the technical paradoxes related to measure.

1 Partitions

A partition is a division of a set into separate pieces.

Definition 1.1 (Partition). *A set $\mathcal{P} \subset 2^\Omega$ is a partition of Ω if*

1. *$\emptyset \notin \mathcal{P}$.*
2. *Each $A, B \in \mathcal{P}$ with $A \neq B$ is disjoint. That is $A \cap B = \emptyset$.*
3. *$\mathcal{P}$ covers Ω. That is, $\cup \mathcal{P} = \Omega$.*

Partitions have a natural epistemic interpretation. Consider the following example, taken from Fagin et al. [1995]. Suppose there are four possible states of the weather in Philadelphia and Melbourne, namely

$$\Omega = \{(\text{sun}, \text{sun}), (\text{sun}, \text{rain}), (\text{rain}, \text{sun}), (\text{rain}, \text{rain})\}.$$

Alice, who lives in Philadelphia always knows what the weather is in Philadelphia but not Melbourne, and Bob who lives in Melbourne always knows the weather in Melbourne but Philadelphia. We can describe Alice's knowledge with the partition

$$\mathcal{P}_A = \{\{(\text{sun}, \text{sun}), (\text{sun}, \text{rain})\}, \{(\text{rain}, \text{sun}), (\text{rain}, \text{rain})\}\}.$$

Firstly, each set in $\mathcal{P}_A$ contains states that are indistinguishable to Alice. For example, $\{(\text{sun}, \text{sun}), (\text{sun}, \text{rain})\} \in \mathcal{P}_A$ means that Alice can not perceive any difference between sunny in both Philadelphia and Melbourne, and sunny in Philadelphia only (since she does not know anything about Melbourne).

Secondly, if two states $\omega_1, \omega_2 \in \Omega$ lie in two different sets in $\mathcal{P}_A$, then Alice can tell them apart. For example, (sun, sun) lies in the first set of $\mathcal{P}_A$ and $(\text{rain}, \text{rain})$ lies in the second set of $\mathcal{P}_A$. So, Alice can tell these two states apart, since in one state, it is sunny in Philadelphia, and in the other, it is raining.

[1] Halpern [2003] seems to backtrack on this on page 235. TODO: find out the history of the ideas.

Similarly, Bob's knowledge can be represented by a different partition,

$$\mathcal{P}_B = \{\{(\text{sun}, \text{sun}), (\text{rain}, \text{sun})\}, \{(\text{sun}, \text{rain}), (\text{rain}, \text{rain})\}\}.$$

Finally, Charlie, who lives in Philadelphia knows neither Philadelphia nor Melbourne's weather and would have the trivial partition $\mathcal{P}_C = \{\Omega\}$. Clearly, Charlie knows less than Alice and Bob. Section 3 will describe how to formally compare partitions; essentially, a partition represents less knowledge if it is "chunkier".

Summary: suppose $\omega \in \Omega$ is the real state of the world. Then $\omega \in P$, for some unique $P \in \mathcal{P}_A$. Alice knows that one the states in P is the real state, but not which one.

2 Sigma Algebras

The previous section showed how partitions can represent different people's knowledge of the states of the world. States that lie in the same set of Alice's partition $\mathcal{P}_A$ are indistinguishable to her, whereas she can tell two states apart if they lie in different sets. This section considers: What questions can Alice answer? In the example from the previous section, she can answer these questions:

- Is it raining in Philadelphia? (Or more precisely, is it either raining in both Philadelphia and Melbourne or raining in Philadelphia and sunny and Melbourne?) i.e. is $\omega \in \{(\text{rain}, \text{sun}), (\text{rain}, \text{rain})\}$?
- Is it sunny in Philadelphia? i.e. is $\omega \in \{(\text{sun}, \text{sun}), (\text{sun}, \text{rain})\}$?
- Is it neither raining or sunny in Philadelphia? i.e. is $\omega \in \emptyset$?
- Is it either raining or sunny in Philadelphia? i.e. is $\omega \in \Omega$?

Apparently, Alice's knowledge allows her to answer the following set of questions

$$\Sigma_A = \{\emptyset, \{(\text{rain}, \text{sun}), (\text{rain}, \text{rain})\}, \{(\text{sun}, \text{sun}), (\text{sun}, \text{rain})\}, \Omega\}.$$

Clearly, if Alice can answer A, and Alice can answer B, then she can answer $A \cup B$ and $A \cap B$. Moreover, if she can answer A, then she can also answer $\Omega \backslash A$. This motivates the following definition of σ-algebra:

Definition 2.1 (σ-algebra). $\Sigma \subseteq 2^\Omega$ *is a σ-algebra on Ω if*

1. $\emptyset \in \Sigma$ *an* $\Omega \in \Sigma$.
2. *If* $A \in \Sigma$ *then* $(\Omega \backslash A) \in \Sigma$.
3. *If A_n is a sequence of subsets of Ω, then* $(\cup_{i=1}^{\infty} A_i) \in \Sigma$.

The "σ" in σ-algebra refers to the countable union property. Firstly, anyone can answer the questions $\emptyset$ and Ω with *no* and *yes*, respectively. Secondly, if the answer to A is *yes*, then the answer to $\Omega \backslash A$ is *no* (Similarly, if the answer is *no*...) Finally, if the answers to A_i are $a_1, a_2, \ldots$, where $a_i \in \{\text{yes}, \text{no}\}$, then if one of $a_i = \text{yes}$, the answer to $(\cup_{i=1}^{\infty} A_i)$ is *yes*.

Alice's knowledge, represented by $\mathcal{P}_A$, determines which questions she can answer Σ_A in the following way.

Firstly, Alice can answer any question of the form "Is $\omega \in P$?", where $P \in \mathcal{P}$.

Definition 2.2 ($\mathcal{P}$-distinguishable σ-algebra). *A σ-algebra Σ on Ω is $\mathcal{P}$-distinguishable if $\mathcal{P} \subset \Sigma$.*

Secondly, Alice can not answer any question of the form "Is $\omega \in P$?" if P is not the union of some subset of $\mathcal{P}$.

Definition 2.3 ($\mathcal{P}$-indistinguishable σ-algebra). *A σ-algebra Σ on Ω is $\mathcal{P}$-indistinguishable if for every $A \in \Sigma$, and every $\omega \in A$, $\mathcal{P}(\omega) \subseteq A$.*

Definition 2.4. *If $\mathcal{P}$ is a partition of Ω, then the σ-algebra Σ represents $\mathcal{P}$ if Σ is both $\mathcal{P}$-distinguishable and $\mathcal{P}$-indistinguishable.*

3 Comparing Knowledge

What does it mean for Alice to know more than Bob? Is it possible for Charlie to know something iff Alice and Bob know something? Or iff either Alice and Bob know something?

Alice knows more than Bob if Alice can distinguish any world that Bob can. This means that Alice's partition $\mathcal{P}_A$ refines $\mathcal{P}_B$:

Definition 3.1 (Partition Refinement Relation). *A partition $\mathcal{P}_A$ refines a partition $\mathcal{P}_B$ if for all $B \in \mathcal{P}_B$, there exists some $A \in \mathcal{P}_A$ with $B \subseteq A$. This is written $\mathcal{P}_A \geq \mathcal{P}_B$.*

The analagous concept for σ-algebras is a little easier:

Definition 3.2 (σ-Algebra Refinement Relation). *A σ-algebra Σ_A refines Σ_B if $\Sigma_B \subseteq \Sigma_A$.*

This means that Alice knows more than Bob if Alice can answer all the questions that Bob can.

Now, suppose Charlie knows something iff Alice and Bob knows something. This means that Charlie can distinguish any two states that both Alice and Bob and distinguish, or equivalently can answer any question both Alice and Bob can answer. In this case, Alice's knowledge is the meet of Alice and Bob's knowledge[2].

Definition 3.3 (Meet of partitions). *Let $\Phi = \{\mathcal{P} : \text{both } \mathcal{P}_A \text{ and } \mathcal{P}_B \text{ refine } \mathcal{P}\}$. The meet of partitions, denoted $\mathcal{P}_A \wedge \mathcal{P}_B$ is the (unique) element of Φ that refines every element of Φ. That is meet $\mathcal{P}_A \wedge \mathcal{P}_B$ is the largest partition in that is refined by both $\mathcal{P}_A$ and $\mathcal{P}_B$.*

Definition 3.4 (Meet of σ-algebras). *The meet of σ-algebras, $\Sigma_A \wedge \Sigma_B$ is the largest σ-algebra that is refined by both Σ_A and Σ_B.*

Finally, suppose Charlie knows everything that either Alice or Bob knows. This means that Charlie can distinguish any two states that either Alice or Bob and distinguish, or equivalently can answer any question either Alice or Bob can answer. In this case, Charlie's knowledge is the join of Alice and Bob's knowledge:

Definition 3.5 (Join of partitions). *Let $\Phi = \{\mathcal{P} : \mathcal{P} \text{ refines both } \mathcal{P}_A \text{ and } \mathcal{P}_B\}$. The join of partitions, denoted $\mathcal{P}_A \vee \mathcal{P}_B$ is the (unique) element of Φ that is refined by every element of Φ. That is, $\mathcal{P}_A \vee \mathcal{P}_B$ is the smallest partition that refines both $\mathcal{P}_A$ and $\mathcal{P}_B$.*

Definition 3.6 (Join of σ-algebras). *The join of σ-algebras, $\Sigma_A \vee \Sigma_B$ is the smallest σ-algebra that refines both $\mathcal{P}_A$ and $\mathcal{P}_B$.*

4 Probability

The previous sections described σ-algebras as the set of questions a person can answer definitely with yes or no, based on their knowledge. This section gives a rather different interpretation: a σ-algebra here will respresent the set of questions that can be answered in proportions like "Q: Does it rain much in Melbourne? A: No, it only rains 40% of the time."

Definition 4.1 (Probability Measure). *Suppose Σ is a σ-algebra on Ω. $\mathbb{P} : \Sigma \to [0, 1]$ is a probability measure on Σ if*

1. *$\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega)$.*
2. *If $A \subset \Sigma$ is countable and disjoint, then $\mathbb{P}(\cup A) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_n)$.*

The triple $(\Omega, \Sigma, \mathbb{P})$ is called a probability space.

The motivation for using σ-algebras (rather than the powerset of Ω) here is purely technical. For example, if $\Omega = [0, 1]$, and we would like to have a uniform probability distribution such that any $\omega \in \Omega$ is "equally likely", then we clearly need $\mathbb{P}(\{\omega\}) = 0$ for all $\omega \in \Omega$. However, if $\mathbb{P}$ were additive (rather than just countably additive), then this would require $\mathbb{P}(\Omega) = 0$ rather than 1.

[2]This terminology is taken from lattice theory. Both of the refinement relations are lattices.

5 Random Variables

Definition 5.1 (Random Variable). *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability space.* X *is a random variable on* $(\Omega, \Sigma, \mathbb{P})$ *if* $X : \Omega \to \mathbb{R}^n$ *and* X *is measurable. That is if for all Borel sets* B*,* $X^{-1}(B) \in \Sigma$.

The measurability in the above definition of random variable is usually motivated by the ability to compute integrals. The second half of this section considers epistemic interpretations.

What does it mean for Alice to know X? The partition perspective is that if Alice can distinguish between any two states of the world in which X differs, then Alice knows X. The σ-algebra perspective is that if Alice can answer any question about X, then she knows X. Or, looking at this backwards, if Alice knows more than what she would know if she only knew X, then Alice knows X. I already defined "knows more than" earlier as the refinement of partitions and σ-algebras, so all that remains is a definition of "only knowing X":

Definition 5.2 (Generated partition). *If* X *is a random variable then its generated partition is* $\mathcal{P}_X = \{X^{-1}(c), c \in \mathbb{R}^n\}$.

Definition 5.3 (Generated σ-algebra). *If* X *is a random variable then its generated σ-algebra is* $\Sigma_X = \{X^{-1}(B), B \text{ is a Borel set}\}$.

A second equivalent but less obvious way to define knowing X is if X is Σ_A-measurable:

Definition 5.4 (Ξ-Measurable Random Variable). $X : \Omega \to \mathbb{R}^n$ *is* Ξ*-measurable if* X *is a random variable on the probability space* $(\Omega, \Xi, \mathbb{P})$.

The following proposition justifies measurability as a definition of knowledge. It says that if Alice's knowledge is represented by Ξ, and Ξ is $\mathcal{P}$-indistinguishable, then if some random variables X is Ξ-measurable, then Alice can deduce X with the information available to her. That is, if she is told that real state of the world ω lies in the set $A \in \Xi$, then a Ξ-measurable random variable X will only take on a single value in A, so Alice can compute $X(\omega) = X(A)$.

Proposition 5.1. *If* Ξ *is* $\mathcal{P}$*-indistinguishable and* X *is* Ξ*-measurable, then* $X(P)$ *is a constant for all* $P \in \mathcal{P}$.

Proof. Pick any $P \in \mathcal{P}$, and any $p, p' \in P$. Let $c = X(p)$. Now, observe that $\{c\}$ is a singleton, and hence a Borel set. Now let $A = X^{-1}(\{c\})$. Since X is Ξ-measurable, $A \in \Xi$. Clearly, $p \in A$. We claim also that $p' \in A$. Since Ξ is $\mathcal{P}$-indistinguishable, $P \subseteq A$. So $p' \in A$, and hence $X(p') = X(p) = c$. □

Corollary 5.2. *If* X *is* Ξ*-measurable, then* Ξ *refines* Σ_X.

Note the following trivial consequence of the definition of measurable:

Proposition 5.3. *If* X *is* Σ*-measurable and* Σ *refines* Ξ*, then* X *is* Ξ*-measurable.*

Finally, what does it mean to know X and Y? Clearly, we can treat (X, Y) as a single random variable, and the standard definitions apply. That is, if $X : \Omega \to \mathbb{R}^m$, $Y : \Omega \to \mathbb{R}^n$, then we can define $Z : \Omega \to \mathbb{R}^m \times \mathbb{R}^n$ as $Z(\omega) = (X(\omega), Y(\omega))$. Then knowing X and Y is the same as knowing Z. So, we can define $\Sigma_{X,Y} = \Sigma_Z$.

Equivalently, we can define $\Sigma_{X,Y} = \Sigma_X \vee \Sigma_Y$. That is, if $\Sigma_{X,Y}$ refines both Σ_X and Σ_Y.

6 Expectation

Definition 6.1 (Expectation). *Let* X *be a random variable on* $(\Omega, \Sigma, \mathbb{P})$*. Then the expectation of* X *is*

$$EX = \int_\Omega X(\omega) d\mathbb{P}(\omega).$$

7 Conditional Expectation

Dudley [2002] gives the following definition:

Definition 7.1 (Conditional Expectation). *Let $X : \Omega \to \mathbb{R}^k$ be a random variable on $(\Omega, \Sigma, \mathbb{P})$, and let $\Xi \subseteq \Sigma$ be a sub-σ-algebra of Σ. Then the conditional expectation $\mathbb{E}[X|\Xi]$ is the equivalence class of random variables $Y : \Omega \to \mathbb{R}^k$ satisfying*

- *Y is Ξ-measurable, and*
- *for all $A \in \Xi$,*

$$\int_A Y(\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega).$$

Before discussing what all this means, I will mention that conditional expectations neither exist (i.e. are non-empty) nor are unique in general. The following theorem gives sufficient conditions for this.[3]

Theorem 7.1. *If $X \in \mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$ (i.e. $\mathbb{E}X$ exists) and Ξ is a sub-σ-algebra of Σ, then $\mathbb{E}[X|\Xi]$ is non-empty. Moreover, if $Y, Y' \in \mathbb{E}[X|\Xi]$, then $Y = Y'$ (a.s.); that is, $\mathbb{P}(\{\omega : Y(\omega) = Y'(\omega)\}) = 1$. Hence $\mathbb{E}[X|\Xi] \in L^1(\Omega, \Sigma, \mathbb{P})$.*

Proof. See Dudley [2002], page 337. This theorem is an application of the Radon-Nikodym theorem. □

The first requirement that all $Y \in \mathbb{E}[X|\Xi]$ be Ξ-measurable says that Alice must know Y. (i.e. Alice's knowledge σ-algebra must refine Σ_Y.) This means that Alice must be able to distinguish between worlds in which Y takes different values.

The second requirement says that Alice must use all of the information contained in Ξ to compute a good estimate of X. That is, if Alice can answer a question $A \in \Xi$ about X, then the expectation of X on A must coincide with the expected conditional expectation.

Now, the first requirement that Y is Ξ-measurable means that Proposition 5.1 applies. That is, if Alice's knowledge is represented by Ξ, and Ξ is $\mathcal{P}$-indistinguishable, then Y is constant on all sets in $\mathcal{P}$. This is the only way Alice can possibly know $Y(\omega)$ without knowing exactly what ω is.

In particular, if Alice knows nothings, so $\mathcal{P} = \{\Omega\}$ giving $\Xi = \{\emptyset, \Omega\}$, and Y is Ξ-indistinguishable, then Y must be a constant, and the second requirement gives $Y(\omega) = EX$ for all $\omega \in \Omega$. That is, Alice's conditional expectation of X is unchanged if she learns nothing.

Theorem 7.2 (Know Nothing). *Suppose $X \in \mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$ and let $Y(\omega) = \mathbb{E}X$. Then, $Y \in \mathbb{E}[X|\{\emptyset, \Omega\}]$. With some abuse of notation, this is sometimes written $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}X$.*

On the other hand if Alice knows everything about X, then her conditional expectation coincides with X.

Theorem 7.3 (Know Everything). *If X is Ξ-measurable, then $X \in \mathbb{E}[X|\Xi]$.*

The previous result implies $X \in \mathbb{E}[X|\Sigma_X]$.

This is a (epistemically) boring algebraic property.

Theorem 7.4 (Linearity of $\mathbb{E}[\cdot|\Xi]$). *$\mathbb{E}[cX + Y|\Xi] = c\mathbb{E}[X|\Xi] + \mathbb{E}[Y|\Xi]$, for all $X, Y \in \mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$, all $c \in \mathbb{R}$ and any sub-σ-algebra Ξ of Σ.*

Note the slight abuse of notation in the next theorem, where the conditional expectation of an equivalence class is taken. (It is obviously well-defined.) It basically says that if Bob knows more than Alice, then what Alice expects is the same as what Alice expects that Bob will expect.

[3] This seems to contradict Bierens [2004], in which the writing seems to indicate an implicit assumption that $\mathbb{E}[X|\Xi]$ is unique. Is this a mistake?

Theorem 7.5 (Law of Iterated Expectations). *If Π is a sub-σ-algebra of Ξ, then for any $X \in \mathcal{L}^1$,*

$$\mathbb{E}[\mathbb{E}[X|\Xi]|\Pi] = \mathbb{E}[X|\Pi].$$

Proof. Suppose $A \in \Pi$. Since $\Pi \subseteq \Xi$, $A \in \Xi$ also. Then the definition of the conditional expectation requires

$$\int_A \mathbb{E}[X|\Pi] d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega),$$
$$\int_A \mathbb{E}[X|\Xi] d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega)$$

and

$$\int_A \mathbb{E}[\mathbb{E}[X|\Xi]|\Pi](\omega) d\mathbb{P}(\omega) = \int_A \mathbb{E}[X|\Xi](\omega) d\mathbb{P}(\omega).$$

Combining these equalities gives, for all $A \in \Pi$,

$$\int_A \mathbb{E}[\mathbb{E}[X|\Xi]|\Pi](\omega) d\mathbb{P}(\omega) = \int_A \mathbb{E}[X|\Pi](\omega) d\mathbb{P}(\omega).$$

The result follows directly from the definition of conditional expectation. □

Theorem 7.6. *If $X \in \mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$ and $X > 0$ (a.s.), then $\mathbb{E}[X|\Xi] > 0$ (a.s.).*

Theorem 7.7. *If X and $Z \in \mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$ where $Z(\omega) = X(\omega)Y(\omega)$, and Y is Ξ-measurable, then*

$$\mathbb{E}[Z|\Xi](\omega) = Y(\omega)\mathbb{E}[X|\Xi](\omega) \text{ (a.s.).}$$

Proof. See Dudley [2002] or Bierens [2004]. □

Example 1. *Now suppose X is uniformly distributed on $[0, 2)$. For example, $\Omega = [0, 2)$, $X(\omega) = \omega$, Σ is the Borel σ-algebra and $\mathbb{P}$ is the Lebesgue measure. Suppose that Alice knows the first digit of X. Then $\mathcal{P} = \{[0, 1)], [1, 2)\}$, and $\Xi = \{\emptyset, [0, 1], [1, 2), [0, 2)\}$. Now, the definition of conditional expectation requires all $Y \in E[X|\Xi]$ to have for all $A \in \Xi$ that*

$$\int_A Y(\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega).$$

This is satisfied trivially by $A = \emptyset$. Since Y must be Ξ-measurable, Y must be a constant on $[0, 1)$ and $[1, 2)$. Moreover,

$$\int_{[0,1)} Y(\omega) d\mathbb{P}(\omega) = \int_{[0,1)} X(\omega) d\mathbb{P}(\omega) = 1/4,$$
$$\int_{[1,2)} Y(\omega) d\mathbb{P}(\omega) = \int_{[1,2)} X(\omega) d\mathbb{P}(\omega) = 3/4.$$

So,

$$Y(\omega) = \begin{cases} 1/2 & \text{if } \omega \in [0, 1), \\ 3/2 & \text{if } \omega \in [1, 2). \end{cases}$$

8 Conditional Probability

Definition 8.1 (Indicator Function). *Let $X \subset \Omega$. Then $I_X : \Omega \to \mathbb{R}$ is the function*

$$I_X(\omega) = \begin{cases} 1 & \text{if } \omega \in X, \\ 0 & \text{if } \omega \notin X. \end{cases}$$

First notice that the inidicator function can be used to construct probability out of expectations, as $\int I_X d\mathbb{P} = \mathbb{P}(X)$. We can use this technique to construct conditional probability from conditional expectation.

Definition 8.2 (Conditional Probability). *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, and Ξ be a sub-σ-algebra of Σ. Then the conditional probability of X given Ξ is $\mathbb{P}(X|\Xi) = E[I_X|\Xi]$.*

9 Independence and Conditional Independence

This is the familiar definition of independence, from Dudley [2002] (page 452).

Definition 9.1 (Independent). *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Two sub-σ-algebras Δ and Ξ of Σ are independent if for all $U \in \Delta$ and all $V \in \Xi$,*

$$\mathbb{P}(U \cap V) = \mathbb{P}(U)\mathbb{P}(V).$$

Halpern [2003] (page 122) has the following intuitive alternate definition, which I have adapted to σ-algebras. It says that Alice and Bob's knowledge is independent if

- for any question Alice knows the answer to, Bob's guess is the same as if he knew nothing.
- for any question Bob knows the answer to, Alice's guess is the same as if she knew nothing.

Definition 9.2 (Independent (Conditional Version)). *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Two sub-σ-algebras Δ and Ξ of Σ are independent if both*

- *$\mathbb{P}(U|\Xi) = \mathbb{P}(U)$ for all $U \in \Delta$.*
- *$\mathbb{P}(V|\Delta) = \mathbb{P}(V)$ for all $V \in \Xi$.*

This definition means that if Alice can answer any question in Ξ, then her answer to any question $U \in \Delta$ is the same as if she did not know Ξ. That is, her (certain) knowledge of Ξ does not giver her any (probabilistic) knowledge of Δ. A symmetrical explanation applies to the second part of the definition.

Proposition 9.1. *The two definitions of independence are equivalent.*

Proof. See Proposition 4.1.2 in Halpern [2003] on page 122. □

Bierens [2004] (page 79) gives the following definition of conditional independence (which I have adapted to the independence of just two σ-algebras):

Definition 9.3 (Conditionally Independent). *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Two sub-σ-algebras Δ and Ξ of Σ are conditionally independent given the sub-σ-algebra Γ if for all $U \in \Delta$ and all $V \in \Xi$,*

$$\mathbb{P}(U \cap V|\Gamma) = \mathbb{P}(U|\Gamma)\mathbb{P}(V|\Gamma).$$

As usual, Halpern [2003] (page 125) has an intuitive definition, which I have adapted to σ-algebras. It says that if Alice and Bob's knowledge is conditionally independent given Charlies' knowledge if

- for any question Alice knows the answer to, Bob's guess (probability) is the same as Charlie's.

- for any question Bob knows the answer to, Alice's guess is the same as Charlie's.

Definition 9.4 (Conditionally Independent). *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability space. Two sub-*σ*-algebras* Δ *and* Ξ *of* Σ *are conditionally independent given the sub-*σ*-algebra* Γ *if both*

- $\mathbb{P}(U|\Xi \vee \Gamma) = \mathbb{P}(U|\Gamma)$ *for all* $U \in \Delta$*.*
- $\mathbb{P}(V|\Delta \vee \Gamma) = \mathbb{P}(V|\Gamma)$ *for all* $V \in \Xi$*.*

Note that neither conditional independence implies independence, nor independence implies conditional independence. Firstly, suppose Alice knows the weather in Philadelphia and London, Bob knows the weather in Melbourne and London, and Charlie knows the weather in London only. Then with uniform priors, Alice and Bob's knowledge would be conditionally independent given Charlie's, but not independent unconditionally.

Similarly, if Alice only knows Philadelphia's weather, Bob only knows Melbourne's weather and Charlie only knows if Alice and Bob are experiencing the same weather or not (i.e. Charlie's partition is $\{\{(\text{sun}, \text{sun}), (\text{rain}, \text{rain})\}, \{(\text{sun}, \text{rain}), (\text{rain}, \text{sun})\}\}$). Then with uniform priors, Alice and Bob's knowledge would be independent unconditionally, but dependent conditionally on Charlie's knowledge.

References

Robert Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236 – 1239, 1976.

Herman Bierens. *Introduction to the Mathematical and Statistical Foundations of Econometrics.* Cambridge University Press, 2004.

Richard Dudley. *Real Analysis and Probability.* Cambridge University Press, 2002.

Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi. *Reasoning About Knowledge.* The MIT Press, 1995.

Joseph Halpern. *Reasoning About Uncertainty.* The MIT Press, 2003.