# CONDITIONING AND INDEPENDENCE

Most of the material in this lecture is covered in [Bertsekas & Tsitsiklis] Sections 1.3-1.5 and Problem 48 (or problem 43, in the 1st edition), available at http://athenasc.com/Prob-2nd-Ch1.pdf. Solutions to the end of chapter problems are available at:http://athenasc.com/prob-solved_2ndedition.pdf. These lecture notes provide some additional details and twists.

**Contents**



## 1 CONDITIONAL PROBABILITY

**Definition 1.** *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and an event* $B \in \mathcal{F}$ *with* $\mathbb{P}(B) > 0$*. For every event* $A \in \mathcal{F}$*, the* conditional probability *that* $A$ *occurs given that* $B$ *occurs is denoted by* $\mathbb{P}(A \mid B)$ *and is defined by*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

**Theorem 1.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

*(a) If $B$ is an event with $\mathbb{P}(B) > 0$, then $\mathbb{P}(\Omega \mid B) = 1$, and for any sequence $\{A_i\}$ of disjoint events, we have*

$$\mathbb{P}\big( \cup_{i=1}^{\infty} A_i \mid B \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B).$$

*As a result, suppose $\mathbb{P}_B : \mathcal{F} \to [0, 1]$ is defined by $\mathbb{P}_B(A) = \mathbb{P}(A \mid B)$. Then, $\mathbb{P}_B$ is a probability measure on $(\Omega, \mathcal{F})$.*

*(b) Let $A$ be an event. If the events $B_i$, $i \in \mathbb{N}$, form a partition of $\Omega$, and $\mathbb{P}(B_i) > 0$ for every $i$, then*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

*In particular, if $B$ is an event with $\mathbb{P}(B) > 0$ and $\mathbb{P}(B^c) > 0$, then*

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c).$$

*(c)* **(Bayes' rule)** *Let $A$ be an event with $\mathbb{P}(A) > 0$. If the events $B_i$, $i \in \mathbb{N}$, form a partition of $\Omega$, and $\mathbb{P}(B_i) > 0$ for every $i$, then for every $i$*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A \mid B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_i)\mathbb{P}(A \mid B_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B_j)\mathbb{P}(A \mid B_j)}.$$

*(d) For any sequence $\{A_i\}$ of events, we have*

$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = \mathbb{P}(A_1) \prod_{i=2}^{\infty} \mathbb{P}(A_i | A_1 \cap \cdots \cap A_{i-1}),$$

*as long as all conditional probabilities are well defined.*

**Proof.**

(a) We have $\mathbb{P}(\Omega \mid B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$. Also

$$\mathbb{P}\big( \cup_{i=1}^{\infty} A_i \mid B \big) = \frac{\mathbb{P}\big(B \cap (\cup_{i=1}^{\infty} A_i)\big)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{i=1}^{\infty} (B \cap A_i))}{\mathbb{P}(B)}.$$

Since the sets $B \cap A_i$, $i \in \mathbb{N}$ are disjoint, countable additivity, applied to the right-hand side, yields

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \,|\, B\big) = \frac{\sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i \,|\, B),$$

as claimed.

(b) We have

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}\Big(A \cap (\cup_{i=1}^{\infty} B_i)\Big) = \mathbb{P}\big(\cup_{i=1}^{\infty} (A \cap B_i)\big) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i).\end{aligned}$$

In the second equality, we used the fact that the sets $B_i$ form a partition of $\Omega$. In the next to last equality, we used the fact that the sets $B_i$ are disjoint and the countable additivity property.

(c) This follows from the fact

$$\mathbb{P}(B_i \,|\, A) = \mathbb{P}(B_i \cap A)/\mathbb{P}(A) = \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i)/\mathbb{P}(A),$$

and the result from part (c).

(d) Note that the sequence of events $\cap_{i=1}^{n} A_i$ is decreasing and converges to $\cap_{i=1}^{\infty} A_i$. By the continuity property of probability measures, we have $\mathbb{P}\big(\cap_{i=1}^{\infty} A_i\big) = \lim_{n\to\infty} \mathbb{P}\big(\cap_{i=1}^{n} A_i\big)$. Note that

$$\begin{aligned}\mathbb{P}\big(\cap_{i=1}^{n} A_i\big) &= \mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \cdots \frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})} \\ &= \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i \,|\, A_1 \cap \cdots \cap A_{i-1}).\end{aligned}$$

Taking the limit, as $n \to \infty$, we obtain the claimed result. $\square$

## 2 INDEPENDENCE

Intuitively we call two events $A$, $B$ independent if the occurrence or nonoccurrence of one does not affect the probability assigned to the other. The following definition formalizes and generalizes the notion of independence.

**Definition 2.** *Let $(\Omega, \mathcal{F}.\mathbb{P})$ be a probability space.*

*(a) Two events, $A$ and $B$, are said to be* **independent** *if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Notation: $A \perp\!\!\!\perp B$. If $\mathbb{P}(B) > 0$, an equivalent condition is $\mathbb{P}(A) = \mathbb{P}(A \mid B)$.*

*(b) Let $S$ be an index set (possibly infinite, or even uncountable), and let $\{A_s \mid s \in S\}$ be a family (set) of events. The events in this family are said to be independent if for every* finite *subset $S_0$ of $S$, we have*

$$\mathbb{P}\big(\cap_{s \in S_0} A_s\big) = \prod_{s \in S_0} \mathbb{P}(A_s).$$

*(c) Let $\mathcal{F}_1 \subset \mathcal{F}$ and $\mathcal{F}_2 \subset \mathcal{F}$ be two $\sigma$-fields. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent (write $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2$) if any two events $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$ are independent.*

*(d) More generally, let $S$ be an index set, and for every $s \in S$, let $\mathcal{F}_s$ be a $\sigma$-field contained in $\mathcal{F}$. We say that the* **$\sigma$-fields $\mathcal{F}_s$ are independent** *if the following holds. If we pick one event $A_s$ from each $\mathcal{F}_s$, the events in the resulting family $\{A_s \mid s \in S\}$ are independent.*

**Example.** Consider an infinite sequence of fair coin tosses, under the model constructed in the Lecture 2 notes. The following statements are intuitively obvious (although a formal proof would require a few steps).

(a) Let $A_i$ be the event that the $i$th toss resulted in a "1". If $i \neq j$, the events $A_i$ and $A_j$ are independent.

(b) The events in the (infinite) family $\{A_i \mid i \in \mathbb{N}\}$ are independent. This statement captures the intuitive idea of "independent" coin tosses.

(c) Let $\mathcal{F}_n$ be the collection of all events whose occurrence can be decided by looking at the results of tosses $2n$ and $2n + 1$. (Note that each $\mathcal{F}_n$ is a $\sigma$-field comprised of finitely many events.) Then, the families $\mathcal{F}_n$, $n \in \mathbb{N}$, are independent.

(d) Let $\mathcal{F}_1$ (respectively, $\mathcal{F}_2$) be the collection of all events whose occurrence can be decided by looking at the results of the coin tosses at odd (respectively, even) times $n$. More formally, Let $H_i$ be the event that the $i$th toss resulted in a 1. Let $\mathcal{C}$ be the collection of events $\mathcal{C} = \{H_i \mid i \text{ is odd}\}$, and finally let $\mathcal{F}_1 = \sigma(\mathcal{C})$, so that $\mathcal{F}_1$ is the smallest $\sigma$-field that contains all the events $H_i$, for odd $i$. We define $\mathcal{F}_2$ similarly, using even times instead of odd times. Then, the two $\sigma$-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ turn out to be independent. This statement captures the intuitive idea that knowing the results of the tosses at odd times provides no information on the results of the tosses at even times.

### 2.1 How to check independence of $\sigma$-algebras? $p$-systems.

How can one establish that two complicated $\sigma$-fields (e.g., as in the last example above) are independent? It turns out that one only needs to check independence for smaller collections of sets – see the theorem below. This is similar to the question of uniqueness of extension that we discussed in previous Lecture. There we have seen an example of a collection of sets $\mathcal{C}$ and a pair of distinct probability measures that coincide on $\mathcal{C}$ but differ on $\sigma(\mathcal{C})$. At the same time we have shown that measures coinciding on an algebra $\mathcal{A}$ must necessarily coincide on $\sigma(\mathcal{A})$. Similarly, one can show that checking independence between $\sigma$-algebras can be reduced to checking independece between any two generating algebras. In fact, in both of these questions we can reduce to checking collections that are even smaller than algebras.

**Definition 3.** *A collection of sets $\Pi$ closed under finite intersections (that is, $A, B \in \Pi \Rightarrow A \cap B \in \Pi$) is called a $p$-system.*

Examples of $p$-systems are intervals $(a, b)$ on $\mathbb{R}$, rectangles $(a, b) \times (c, d)$ on $\mathbb{R}^2$, convex sets in $\mathbb{R}^d$, etc.

**Theorem 2.** *Let $\Pi_1$ and $\Pi_2$ be $p$-systems and $\mathcal{F}_i = \sigma(\Pi_i)$, $i = 1, 2$. If*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{1}$$

*for every $A \in \Pi_1$, $B \in \Pi_2$, then $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent.*

**Proof.** Fix an arbitrary $B \in \Pi_2$ and define a collection of sets

$$\mathcal{L}_B \triangleq \{E \in \mathcal{F}_1 : \mathbb{P}(E \cap B) = \mathbb{P}(E)\mathbb{P}(B)\}.$$

By assumption $\Pi_1 \subseteq \mathcal{L}_B$. We also have:

1. Clearly $\Omega \in \mathcal{L}_B$.

2. If $A_1 \subset A_2$ and both belong to $\mathcal{L}_B$ then from

$$\mathbb{P}((A_2 \setminus A_1) \cap B) + \mathbb{P}(A_1 \cap B) = \mathbb{P}(A_2 \cap B)$$

   we conclude that $A_2 \setminus A_1 \in \mathcal{L}_B$.

3. $\mathcal{L}_B$ is a monotone class. Indeed, if $A_n \nearrow A$ and $A_n \in \mathcal{L}_B$ then $A_n \cap B \nearrow A \cap B$ and by continuity of $\mathbb{P}$ we have

$$\mathbb{P}(A \cap B) = \lim_{n\to\infty} \mathbb{P}(A_n \cap B) = \mathbb{P}(B) \cdot \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(B)\mathbb{P}(A)\,,$$

   implying $A \in \mathcal{L}_B$. Similar argument holds for $A_n \searrow A$.

Recall that $\alpha(\mathcal{C})$ denotes the smallest algebra of sets containing collection $\mathcal{C}$. It turns out that 1 and 2 imply that $\mathcal{L}_B$ contains $\alpha(\Pi_1)$ (see Proposition to follow). Thus by the monotone class theorem $\mathcal{L}_B = \mathcal{F}_1$. Thus (1) holds for all $A \in \mathcal{F}_1$ and $B \in \Pi_2$. By symmetry it also holds for all $A \in \Pi_1$ and $B \in \mathcal{F}_2$. And applying the above argument again (with $\Pi_2$ replaced by $\mathcal{F}_2$) for all of $\mathcal{F}_1$ and $\mathcal{F}_2$. $\square$

**Proposition 1.** *Let $\Pi$ be a $p$-system on $\Omega$. Let $\mathcal{D}$ be a collection containing $\Pi$ satisfying the following:*

1. $\Omega \in \mathcal{D}$
2. *For all $A, B \in \mathcal{D}$ such that $A \subset B$ we have $B \setminus A \in \mathcal{D}$ (i.e. $\mathcal{D}$ closed under "punching holes").*

*Then $\mathcal{D} \supset \alpha(\Pi)$. Thus $\alpha(\Pi)$ is the smallest collection of sets containing $\Pi \cup \{\Omega\}$ and closed under punching holes.*

**Proof.** Let $\mathcal{D}_0$ be the smallest collection of sets containing $\Pi$ and satisfying conditions 1 and 2. We will show $\mathcal{D}_0 \supset \alpha(\Pi)$. Note that any $p$-system satisfying 1 and 2 is automatically an algebra. Thus it is sufficient to prove $\mathcal{D}_0$ is a $p$-system. Fix $C \in \Pi$ and let

$$\mathcal{L}_C = \{A \in \mathcal{D}_0 : A \cap C \in \mathcal{D}_0\}$$

On one hand, $\mathcal{L}_C$ contains $\Pi$ and $\Omega$. On the other hand, $\mathcal{L}_C$ is closed under punching holes: For $A \subset B$ we have $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$. Thus $\mathcal{L}_C = \mathcal{D}_0$ by minimality of $\mathcal{D}_0$. Hence $\mathcal{D}_0$ is closed under intersections with elements of $\Pi$.

Next take an arbitrary $D \in \mathcal{D}_0$. We have

$$\mathcal{L}_D = \{A \in \mathcal{D}_0 : A \cap D \in \mathcal{D}_0\}$$

containing $\Pi$ (and $\Omega$) by the previous argument and closed under punching holes (same reasoning). Thus $\mathcal{L}_D = \mathcal{D}_0$ and $\mathcal{D}_0$ is closed under intersections. $\square$

**Exercise 1.** *Let $\Pi$ be a $p$-system and $\mathcal{A} = \alpha(\Pi)$ the algebra generated by it. Suppose $P$ and $Q$ are two finitely additive non-negative set-functions with $P(\Omega) = Q(\Omega) = 1$. Show that if $P$ and $Q$ agree on $\Pi$ then they agree on $\mathcal{A}$. (Hint: As usual let $\mathcal{D} = \{E \in \mathcal{A} : P(E) = Q(E)\}$ and use the proposition above).*

**Exercise 2.** *Show that $\alpha(\mathcal{C})$ consists of Ø, $\Omega$ and all sets that can be written in the sum-of-products form (this should remind you of disjunctive normal form)*

$$(A_{1,1} \cap \ldots \cap A_{1,n_1}) \cup \ldots \cup (A_{m,1} \cap \ldots \cap A_{m,n_m}),$$

*with $A_{i,j} \in \mathcal{C}$ or $A^c_{i,j} \in \mathcal{C}$.*

## 3 THE BOREL-CANTELLI LEMMA

The Borel-Cantelli lemma is a tool that is often used to establish that a certain event has probability zero or one. Given a sequence of events $A_n$, $n \in \mathbb{N}$, the event $\{A_n \text{ i.o.}\}$ (read as "$A_n$ occurs infinitely often") is defined to be the event consisting of all $\omega \in \Omega$ that belong to infinitely many $A_n$. Show that equivalently

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

This event is also sometimes denoted by $\limsup_{n\to\infty} A_n$.

**Theorem 3.** (**Borel-Cantelli lemma**) *Let $\{A_n\}$ be a sequence of events and let $A = \{A_n \text{ i.o.}\}$.*

*(a) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A) = 0$.*

*(b) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and the events $A_n$, $n \in \mathbb{N}$, are independent, then $\mathbb{P}(A) = 1$.*

**Remark:** The result in part (b) is not true without the independence assumption. Indeed, consider an arbitrary event $C$ such that $0 < \mathbb{P}(C) < 1$ and let $A_n = C$ for all $n$. Then $\mathbb{P}(\{A_n \text{ i.o.}\}) = \mathbb{P}(C) < 1$, even though $\sum_n \mathbb{P}(A_n) = \infty$.

The following lemma is useful here and in may other contexts.

**Lemma 1.** *Suppose that $0 \le p_i \le 1$ for every $i \in \mathbb{N}$. Then:*

$$\sum_{i=1}^{\infty} p_i = \infty \quad \Rightarrow \quad \prod_{i=1}^{\infty} (1 - p_i) = 0 \tag{2}$$

$$\sum_{i=1}^{\infty} p_i = \infty \quad \Leftarrow \quad \prod_{i=1}^{\infty} (1 - p_i) = 0, p_i < 1 \tag{3}$$

**Proof.** Note that $\log(1-x)$ is a concave function of its argument, and its derivative at $x = 0$ is $-1$. It follows that $\log(1-x) \le -x$, for $x \in [0, 1]$. We then have

$$
\begin{aligned}
\log \prod_{i=1}^{\infty}(1-p_i) &= \log\Big(\lim_{k\to\infty}\prod_{i=1}^{k}(1-p_i)\Big) \\
&\le \log \prod_{i=1}^{k}(1-p_i) \\
&= \sum_{i=1}^{k}\log(1-p_i) \\
&\le \sum_{i=1}^{k}(-p_i).
\end{aligned}
$$

This is true for every $k$. By taking the limit as $k \to \infty$, we obtain $\log \prod_{i=1}^{\infty}(1-p_i) = -\infty$, and $\prod_{i=1}^{\infty}(1-p_i) = 0$.

For the converse statement, note that under $p_i < 1$ we have

$$
\prod_{i=1}^{\infty}(1-p_i) = 0 \qquad \Longleftrightarrow \qquad \forall n \prod_{i=n}^{\infty}(1-p_i) = 0
$$

If $p_i \not\to 0$ the result is automatic. Hence, we may also assume $p_i \to 0$. Then taking $n$ so large that $p_i \le 1 - e^{-1}$ for all $i \ge n$ we may apply the lower bound

$$
\log(1-x) \ge -\frac{e}{e-1}x \qquad \forall 0 \le x \le 1 - e^{-1}.
$$

Then, for arbitrary large $C$ we have for all sufficiently large $n$

$$
-C \ge \log \prod_{i=1}^{n}(1-p_i) \ge -\frac{e}{e-1}\sum_{i=1}^{n} p_i \,,
$$

implying $\sum_{i=1}^{\infty} p_i \ge C'$ for all $C' > 0$. □

**Proof of Theorem 3.**

(a) The assumption $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$ implies that $\lim_{n\to\infty}\sum_{i=n}^{\infty}\mathbb{P}(A_i) = 0$. Note that for every $n$, we have $A \subset \cup_{i=n}^{\infty} A_i$. Then, the union bound implies that

$$
\mathbb{P}(A) \le \mathbb{P}\big(\cup_{i=n}^{\infty} A_i\big) \le \sum_{i=n}^{\infty}\mathbb{P}(A_i).
$$

We take the limit of both sides as $n \to \infty$. Since the right-hand side converges to zero, $\mathbb{P}(A)$ must be equal to zero.

(b) Let $B_n = \cup_{i=n}^{\infty} A_i$, and note that $A = \cap_{n=1}^{\infty} B_n$. We claim that $\mathbb{P}(B_n^c) = 0$. This will imply the desired result because

$$\mathbb{P}(A^c) = \mathbb{P}\big(\cup_{n=1}^{\infty} B_n^c\big) \leq \sum_{n=1}^{\infty} \mathbb{P}(B_n^c) = 0.$$

Let us fix some $n$ and some $m \geq n$. We have, using independence (show that independence of $\{A_n\}$ implies independence of $\{A_n^c\}$)

$$\mathbb{P}(\cap_{i=n}^{m} A_i^c) = \prod_{i=n}^{m} \mathbb{P}(A_i^c) = \prod_{i=n}^{m} \big(1 - \mathbb{P}(A_i)\big).$$

The assumption $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ implies that $\sum_{i=n}^{\infty} \mathbb{P}(A_i) = \infty$. Using Lemma 1, with the sequence $\{\mathbb{P}(A_i) \mid i \geq n\}$ replacing the sequence $\{p_i\}$, we obtain

$$\mathbb{P}(B_n^c) = \mathbb{P}(\cap_{i=n}^{\infty} A_i^c) = \lim_{m\to\infty} \mathbb{P}(\cap_{i=n}^{m} A_i^c) = \lim_{m\to\infty} \prod_{i=n}^{m} \big(1 - \mathbb{P}(A_i)\big) = 0,$$

where the second equality made use of the continuity property of probability measures. $\square$